2007-07-18

Evidence Supporting the Validity of Inferences Required by the Intended Uses of the Technology Integration Confidence Scale

Jeremy Michael Browne
Brigham Young University - Provo

EVIDENCE SUPPORTING THE VALIDITY OF INFERENCES REQUIRED BY THE INTENDED USES OF THE TECHNOLOGY INTEGRATION CONFIDENCE SCALE

By

Jeremy M. Browne

A dissertation submitted to the faculty of

Brigham Young University

In partial fulfillment of the degree of

Doctor of Philosophy

Department of Instructional Psychology & Technology

Brigham Young University

July 2007

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a dissertation submitted by

Jeremy M. Browne

This dissertation has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date	Charles R. Graham, Chair
Date	C. Vic Bunderson
Date	Michael D. Bush
Date	Robert G. Erickson
Date	Richard R. Sudweeks

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the dissertation of Jeremy M. Browne in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date	Charles R. Graham Chair, Graduate Committee

Accepted for the Department

Andy S. Gibbons
Department Chair

Accepted for the College

K. Richard Young
Dean, David O. McKay School of Education

ABSTRACT

EVIDENCE SUPPORTING THE VALIDITY OF INFERENCES REQUIRED BY THE INTENDED USES OF THE TECHNOLOGY INTEGRETION CONFIDENCE SCALE

Jeremy M. Browne

Department of Instructional Psychology & Technology

Doctor of Philosophy

Many teacher preparation programs provide opportunities for their preservice educators to gain the requisite technology integration skills and knowledge. However, they often ignore the dispositions that affect whether a teacher will actually use technology in the classroom. In an effort to address this oversight, the McKay School of Education at Brigham Young University developed the Technology Integration Confidence Scale (TICS). It was hoped the TICS could be used to (a) establish a baseline preservice teacher profile, (b) monitor the effects of curricular adjustments, (c) identify preservice teachers in most need of intervention, and (d) predict in-practice behavior. Although a pilot test of the TICS revealed acceptable levels of reliability, the initial evidence gathered to support the validity of inferences to be drawn from TICS scores was based on underdeveloped, anachronistic views of validity.

The purpose of this dissertation was to gather evidence supporting the inferences required for each of the TICS' intended purposes, drawing on modern validity theory and codified testing standards, and employing state-of-the-art measurement methodology. Methods used to gather validity-supporting evidence included repeated measures ANOVA, regression analyses, and a synthesis of self-efficacy research. Evidence supported the use of the TICS to establish a baseline preservice teacher profile and to predict in-course preservice teacher performance, but only in the secondary education technology integration course. The evidence did not support using the TICS to monitor minor changes to the curriculum.

ACKNOWLEDGEMENTS

There are many to whom much credit is due for bringing these efforts to fruition: Dr. Charles Graham, whose faith in my abilities led me to this subject, Dr. Michael Bush who introduced me to Instructional Psychology & Technology and continued to mentor me even when my interests diverged from his, Dr. Richard Sudweeks whose enthusiasm for measurement theory and patient instruction inspired this work, and Dr. C. Vic Bunderson, who let me stand in his shadow and made me realize how important my ideas are. All of them deserve much gratitude. I could add to this list my loving parents for their never ending faith and my parents-in-law for keeping me humble.

But there is only one without whom this dissertation would never have seen the light of day. Sans her constant encouragement, relentless nagging, and loving correction, I wouldn't be where I am today. To my loving wife, who started a family the first week of my graduate studies, who put off the benefits of adulthood so that I could prolongate my childhood, and who feigned interest in chi-squares, the Rasch model, and self-efficacy, I dedicate this work.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Statement of the Problem

Despite the increased influence technology has in personal and business life, technology resources in public education have often failed to meet their potential (Brinkerhoff, 2006). Although teacher education programs are training preservice teachers in proper technology integration techniques, many of their efforts have failed to substantially increase in-practice technology integration (Kay, 2006) in part because they ignore the dispositional dimension of the issue (Butler & Sellbom, 2002).

Consider Figure 1. Many teacher education programs have oriented their technology integration training programs around national standards, which assure preservice teachers are able to effectively use technology in their classrooms. However, these programs have rarely addressed the dispositions that actually lead to such in-practice performance. Though such teacher preparation programs may have evidence that their graduating teachers can support student learning and professional productivity through technology integration, they cannot lack the evidence that their students will (Brzycki & Dudt, 2005; Swain, 2006).

Existing measures of preservice teacher competence in technology integration include exams (Charoula, 2005), performance assessments (Heide & Henderson, 2001; Mills, 2001; Persichitte, Caffarella, & Ferguson-Pabst, 2003), and self-assessments (Basham, Pallap, and Pianfetti, 2005; Foley et al., 2001; Shoffner & Dias, 2001; Kelley, Wetzel, Padgett, Williams, & Odom, 2005). Unfortunately, these measures almost exclusively assess respondents' level of knowledge and skill. Some researchers have employed measures of general dispositions concurrently with measures of ability

(Brinkerhoff, 2006; Levin, 1999), but these affective measures have not been aligned with the same standards as the measures of ability.

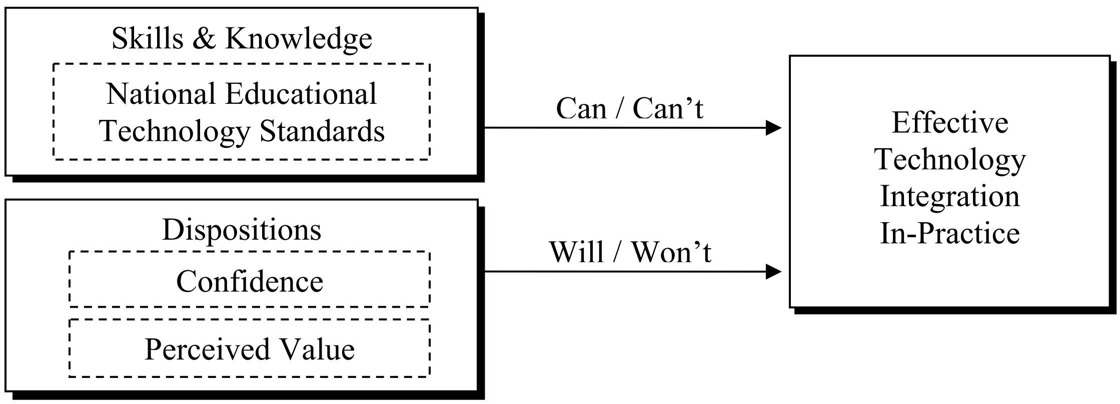


Figure 1. The relationship of skills and knowledge, dispositions, and in-practice technology integration.

Given that skills and knowledge are by themselves insufficient measures of propensity to integrate technology in education, there is a current need for measures of dispositions, such as perceptions of value and self-efficacy, and for these measures to be aligned with the same national standards as the current measures of ability. These new measures must also be comprehensively analyzed for their psychometric suitability, and the inferences users expect to make of the resulting scores must be supported with adequate evidence of validity. Indeed, the question of validity is "the most fundamental consideration in developing and evaluating tests" (American Educational Research Association [AERA], American Psychological Association [APA], National Council on Measurement in Education [NCME], 1999, p. 9).

Context

The McKay School of Education (MSE) at Brigham Young University understands the importance of preparing teachers to effectively integrate technology into their practice (Wentworth, Waddoups, & Earle, 2004). To address the significance of this process, the MSE's Instructional Psychology and Technology (IP&T) department, which is responsible for the technology integration training of preservice teachers, redesigned the required technology integration courses to better afford preservice teachers the opportunity to gain the necessary pedagogical skills (Graham, Culatta, Pratt, & West, 2004). However, given the above-cited issues of focusing solely on competence, there was concern regarding whether these courses affected preservice teacher dispositions as well.

In an effort to address this issue, the course coordinator directed a graduate student instructor to develop a measure of an affective trait that could (a) establish a baseline preservice teacher profile, (b) monitor the effects of curricular adjustments, (c) profile individual students to identify those in most need of intervention, and (d) could be expected to predict in-practice behavior.

The resulting measure, the Technology Integration Confidence Scale (TICS; Browne, in press), was intended to measure preservice teacher self-efficacy regarding tasks described in the National Educational Technology Standards for Teachers (NETS-T), the same influential standards to which the technology integration courses were designed (see Appendices A and B). Self-efficacy is a mature, well-defined psychometric construct (Pajares & Schunk, in press), with an established measurement methodology

(Bandura, in press) and has demonstrated power in predicting performance and persistence (Multon, Brown, & Lent 1991).

Despite the wherewithal of the TICS, its initial development only examined pilot test data using classical test theory and evaluated the meaning of its scores with an ill-defined concept of content validity. To assure that the inferences required for each of the TICS four purposes are sound, a more comprehensive analysis was necessary.

Statement of Purpose

The purpose of this project was to further the development of the TICS through (a) an item response theory (IRT) analysis of respondent data to ensure proper item and response category functioning and (b) an examination of the validity of specified inferences to be drawn from TICS scores. Although these two endeavors will serve the immediate needs of the technology integration courses in the MSE, they will also address the larger issue of preservice teacher disposition towards technology integration. Reliability analyses of most instruments used in technology integration teacher education research do not employ IRT, and even basic evidence of validity is "almost never noted" (Kay, 2006, p. 386). Performing these two steps would justify the use of the TICS in larger-scale studies.

Audience

This report is intended for teacher educators concerned with technology integration, administrators of teacher education programs, and current and potential users of the TICS. Psychometricians, evaluators, and others with an interest in systematic applications of validity theory may also find these research activities interesting.

Research Questions

This project attempted to answer the following research questions:

1. Given the stated intended uses for the TICS, what inferences are expected to be made from TICS scores (AERA et al., 1999)?
2. What evidence should and can be gathered to support the validity of each expected inference?
3. Given the appropriate evidence, which of the expected inferences are supported? Which are not?
4. What efforts should be undertaken to improve the validity of the expected inferences?

Scope

Validity can never be fully established for any inference. It can only be supported by an argument built on evidence and theory, and this support is measured in degrees (AERA et al., 1999; Cronbach & Meehl, 1955; Messick, 1995). Because of this challenge, the scope of this project had to be limited, or it would never have been completed. Therefore the following constraints were necessary:

1. Only inferences required for the stated intended uses of the TICS were considered.
2. Only evidence of validity required by those inferences were gathered (AERA et al., 1999).
3. Only definitions of validity and evidence gathering procedures that were compatible with the 1999 AERA, APA, NCME *Standards for Educational and Psychological Testing* were considered.

4. Due to the relatively homogeneous population of MSE students and the limited timeframe within which the TICS had been used, it would not have been effective to gather evidence supporting the generalizability of the intended inferences, or the consequences of test use.

Chapter 2: Review of Literature

Self-Efficacy

Self-efficacy grew out of the cognitive revolution, subsequent renewed interest in the self, and is a partial reaction against the pervasive concern of self-esteem (Bandura, in press). The theory holds that personal beliefs can predict behavior better than simple stimulus-response reactions, and such beliefs fit within "a theory of personal and collective agency" (Pajares & Schunk, in press, p. 18). It is important to note that, although self-efficacy is most often associated with measurement methodology, it is also a theory of behavioral change through "extraordinary personal feats [that] serve as transforming experiences" (Bandura, in press, p. 2), and formative feedback of each performance (Bandura, 1977).

"Self-efficacy is concerned with people's beliefs in their capabilities to produce given attainments" (Bandura, in press, p. 2). Such beliefs not only reflect a person's ability to perform a task, but also the likelihood that the performance will take place, thus increasing the predictive value of its measures. When combined with self-efficacy treatments, self-efficacy measures accurately predict outcomes of individual and group performances, in both pre- and post-treatment situations (Bandura, 1977).

In a massive meta-analytic investigation covering published reports from 1977 to 1988, Multon, Brown, & Lent (1991) found self-efficacy measures to account for 14% of variance in student performance and 12% of variance in student persistence. However, they also found evidence that "the relationship of self-efficacy to performance and persistence may vary across types of students, measures, and study characteristics" (p. 34).

It is important to draw a distinction between self-efficacy, particularly teacher self-efficacy, and general teacher efficacy. Heavily researched (Goddard, Hoy & Hoy, 2000, 2004; Henson, Kogan, Vacha-Haase, 2001; Hoy & Spero, 2005a, 2005b; Hoy & Woolfolk, 1993; Milner & Hoy 2003; Tschannen-Moran & Hoy, 2001; Tschannen-Moran, Hoy, & Hoy, 1998; Woolfolk & Hoy, 1990; Woolfolk, Rosoff, & Hoy, 1990), teacher efficacy grew out of Rotter's work on internal vs. external locus of control, which predated Bandura's self-efficacy by more than a decade (Tschannen-Moran et al., 1998). It differs from self-efficacy in that "self-efficacy concerns beliefs about whether one can produce certain actions (perceived self-efficacy) [which is] not the same as beliefs about whether actions affect outcomes (locus of control)" (p. 211). In fact, the work of Tschannen-Moran et al. (1998) was to unify these two efficacies, though what they created is obviously larger in scope than Bandura's self-efficacy.

The Technology Integration Confidence Scale

The TICS was developed for use in the MSE technology integration courses to track preservice teacher self-efficacy vis-à-vis tasks described in the National Educational Technology Standards for Teachers (NETS-T). The details of its development and pilot testing are described in the SITE 2007 proceedings (Browne, 2007) and in an upcoming volume of *Computers in the Schools* (Browne, in press), but they are summarized below to familiarize the reader with the instrument.

The developer of the TICS refused to apply his own definition of technology integration, and adopted instead the broadly-accepted NETS-T (International Society for Technology in Education [ISTE], n.d.; Waddoups, Wentworth, & Earle, 2004). Following Bandura's (in press) recommendations, the TICS comprised items that presented a task and asked the respondent to rate their confidence in accomplishing it. Although Bandura recommended an analog response scale ranging from *0-100*, with *0* being *not confident at all* and *100* being *absolutely confident,* recent research demonstrated that fewer response categories may function better than many (Reeve, Kitchen, Sudweeks, Bell, & Bradshaw, in preparation). Therefore the TICS presented six response categories: *not confident at all, slightly confident, somewhat confident, fairly confident, quite confident, completely confident*.

Although the TICS as a whole was reliable, the subscales aligned with individual NETS-T differed wildly in their reliability. This was due to some NETS-T being underrepresented, and a few malfunctioning items, identified by an item analysis. These items were revised for the present version. Still, the analysis of the pilot study data was based in classical test theory and did not investigate the functioning of the TICS response categories.

More importantly, the initial TICS development process had gathered only a few forms of evidence for an ill-defined concept of content validity. Though the items were relevant to the teaching domain, the developers did not determine how well the collection of items represented the domain of technology integration, in part because of the broadness of the NETS-T.

Item Response Theory and the Rating Scale Model

As in most self-efficacy scales, the TICS pilot test data were analyzed using classical test theory. While such analyses are informative, they yield a single estimate of item difficulty across the sample, and their results are sample-dependent (Hambleton, Swaminathan, & Rogers, 1997). In the case of the TICS, the item analysis completed during its initial development provided a mean score between 0 (*Not confident at all*) and 5 (*Absolutely confident*) for each self-efficacy item (Browne, in press). This item mean did not distinguish between respondents with generally high and generally low total scores. Also, if the TICS were administered to a similar sample (e.g. to preservice teachers in the same course, but a different semester), that estimate of item difficulty would not be identical to the pilot test estimate.

These classical test theory issues were overcome through the application of item response theory (IRT), which (a) discovered estimates of item difficulty that are conditional upon the respondents' level of self-efficacy and (b) established a scoring system that is independent of the sample. Further, using the rating scale model (RSM), the functioning of each response category (*Not confident at all, Slightly confident,* etc.) was analyzed for each subscale.

As an introductory example, imagine a self-efficacy scale with items presenting only two options: *Not confident at all*, and *Absolutely* c*onfident.* Such an item would not be effective, given Bandura's description of self-efficacy items (Bandura, in press), but is used here only to convey a simplest-case scenario. An IRT analysis would plot the probability of a respondent having confidence on any task in the scale against their overall level of self-efficacy. Consider Figure 2. The horizontal axis represents the self-efficacy of the respondents (with 0 representing average self-efficacy), and the vertical axis represents the probability that they would endorse the self-efficacy statement (ranging from 0 to 1.0). The *conditional probability curve* represents the probability that the statement would be endorsed by respondents given their level of self-efficacy. Respondents with an average level of self-efficacy have a .50 probability of endorsing the statement, whereas the endorsement probability for respondents with a self-efficacy of +18 is over .90. The self-efficacy level at which the probability curve crosses .50 (in this case, 0) is considered the item's *difficulty parameter*.

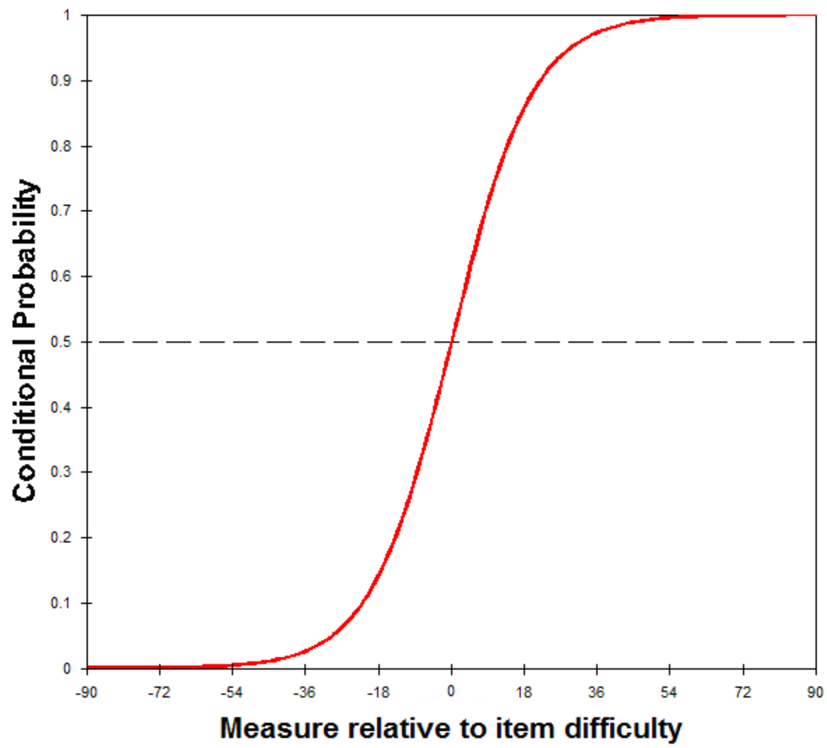


Figure 2. Conditional probability curve for a dichotomous item.

An RSM analysis extends the IRT model to investigate the probability that each response category would be chosen at any level of self-efficacy. Figure 3 displays the *category probability curves* for our example dichotomous self-efficacy scale. Notice that the probability that a respondent with low self-efficacy (-36) would select the second response category (*Absolutely confident*) is low, but increases with the respondents' self-efficacy.

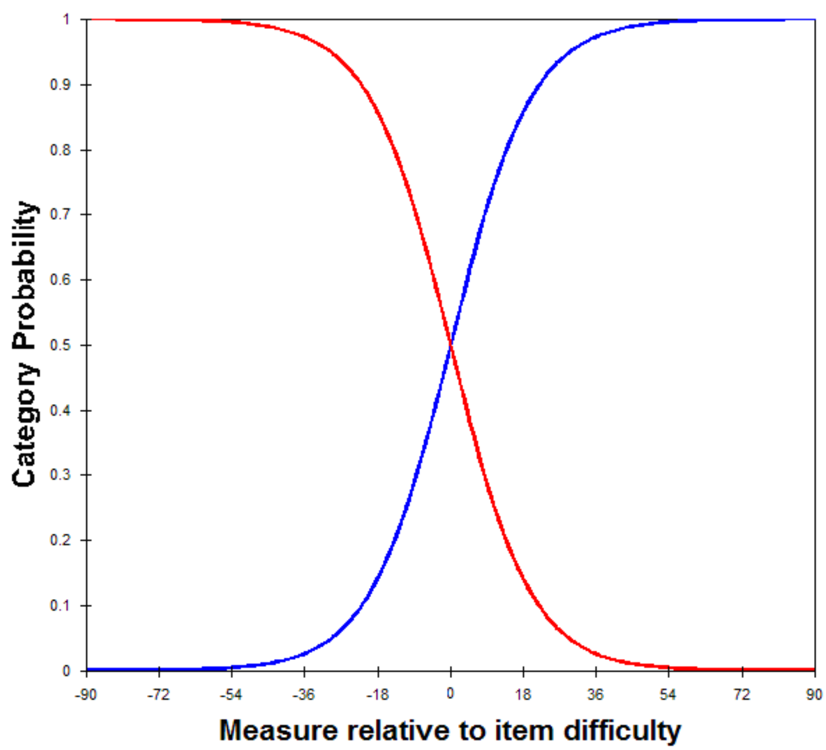


Figure 3. Conditional category curves for a dichotomous item.

Moving from the dichotomous simplest-case to a self-efficacy scale with three response categories (*Not confident at all*, *Somewhat confident*, and *Absolutely confident*), the conditional probability curves are more difficult to interpret. For example, in Figure 4, the curve on the left represents the probability a respondent with a given self-efficacy would select a response category higher than *Not confident at all*. For respondents with extremely low self-efficacy (-90) the probability of such a selection taking place is almost zero. The curve on the right represents the probability respondents would choose a response category higher than *Somewhat confident*.

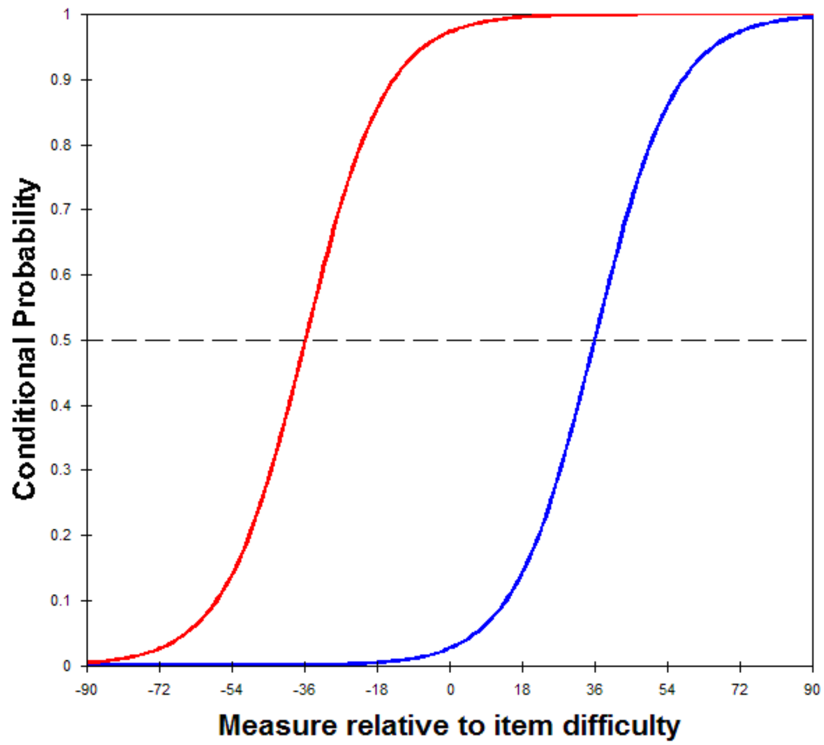


Figure 4. Conditional probability curve for three response categories.

Resulting from the same transformation that yielded Figure 3, the category probability curves in Figure 5 are quite easy to interpret. The three curves represent the probability that each response category (*Not confident at all*, *Somewhat confident*, and *Absolutely confident*) would be select by respondents with the corresponding level of self-efficacy. Notice that respondents with average self-efficacy (0) have the greatest probability of selecting *Somewhat confident*, but that this probability is not 1.0. There is still a chance they would select either *Not confident at all* or *Absolutely confident*.

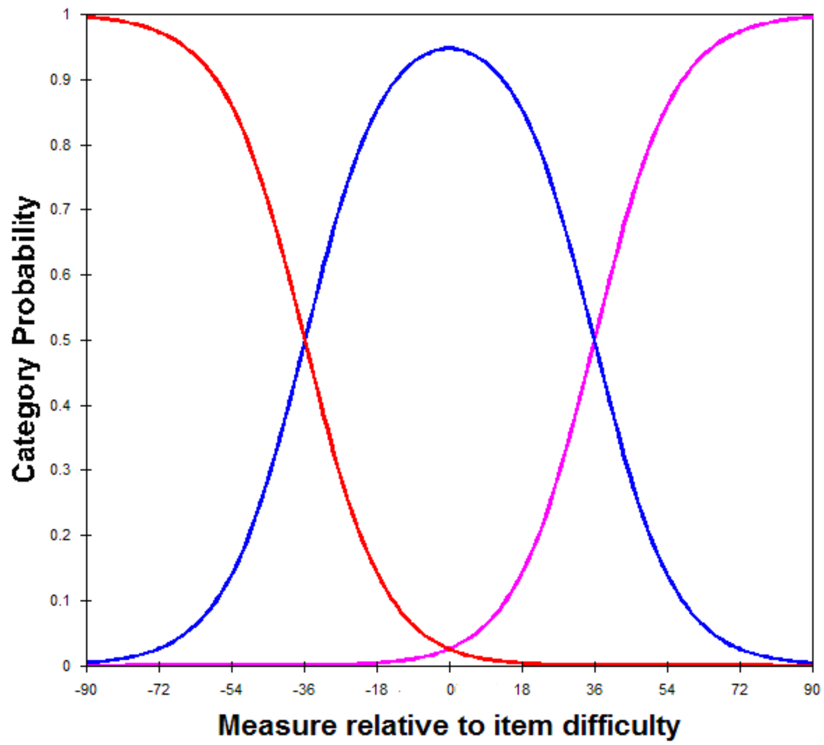


Figure 5. Conditional category curves for three response categories.

Besides the unconditional difficulty estimates, another weakness of the classical test theory analysis used in most self-efficacy scores is that it assumes the response categories are linear and equal-interval. For example, *Not confident at all, Somewhat confident,* and *Absolutely confident* are assumed to represent increasing levels of self-efficacy (linear), and the distance between them is expected to be constant (equal-interval). This must be the case if averaging across items is to be effective.

The rating scale model does not assume response categories to be linear or equal-interval. As shown in Figures 3 and 5, it plots the probability that preservice teachers will choose each response category against their overall level of self-efficacy. Figure 6 displays a properly functioning, five-category response set. Notice that the curves are well spaced (equal-interval), and each crosses the previous curve before crossing the next (linear). The curves displayed in Figure 6 support the assumptions of most self-efficacy scales.

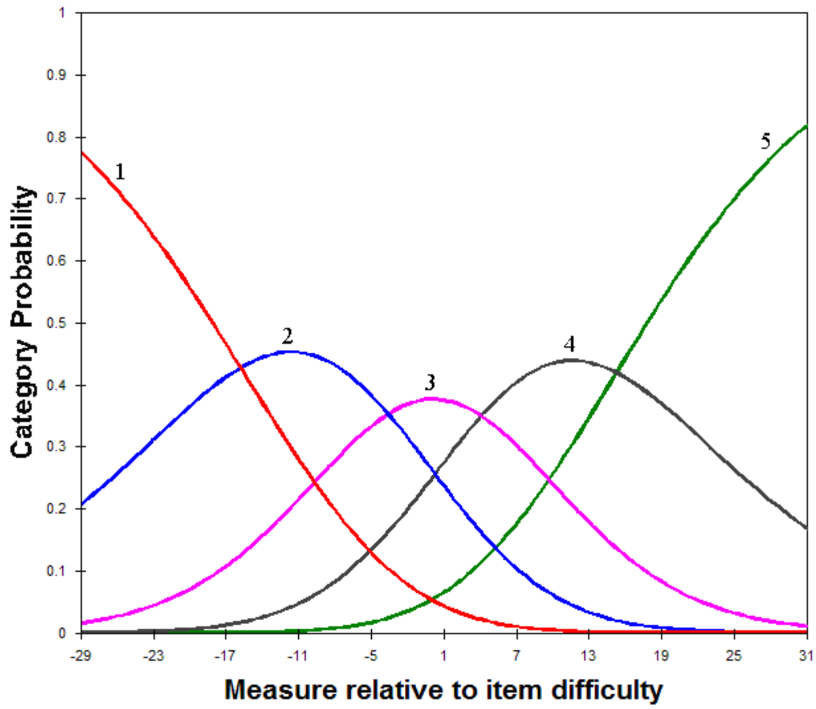


Figure 6. Conditional category curves for functioning response categories.

Figure 7 shows a malfunctioning response category set. Notice there is a higher probability that respondents with a self-efficacy level of +15 will choose the third or fifth response category than the fourth response category. Also, the curve of the fourth response category crosses the first, the fifth, the second, and the fourth categories – in that order. This represents response category disorder, or reversal, which does not support the scales' assumptions.

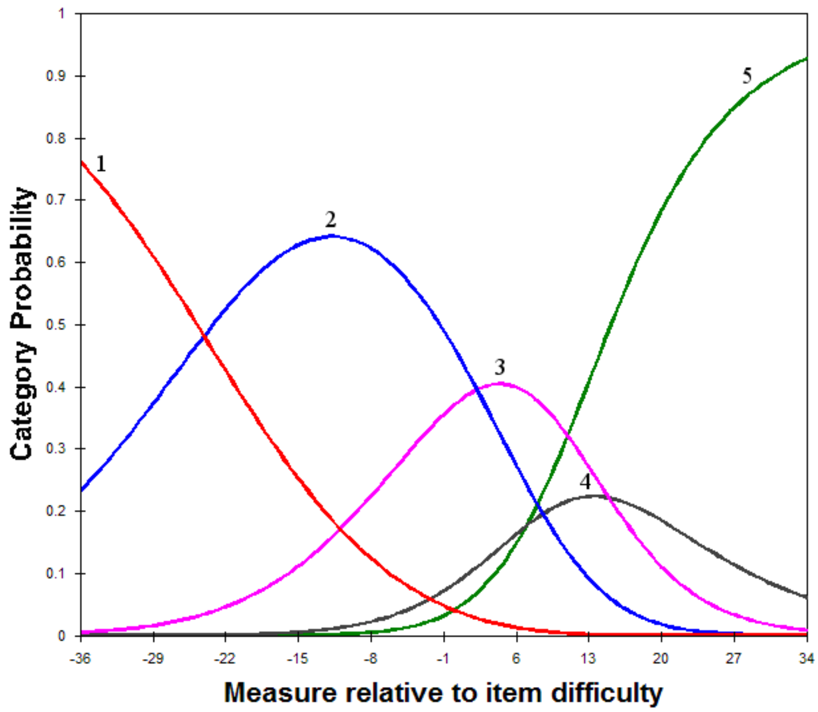


Figure 7. Conditional category curves for dysfunctional response categories.

RSM is not the only IRT approach to scales with multiple response categories, such as self-efficacy. Embretson and Reise (2000) describe five more in their chapter on polytomous IRT models. However, RSM was chosen for two reasons. First, it is easily interpretable by those with little knowledge of IRT, including this study's target audience. Second, it estimates fewer parameters than other IRT models, such as the partial credit model or modified graded response model, resulting in smaller errors from the same sample size.

Validity

In measurement, "Validity is the soundness of your interpretations and uses of students' assessment results" (Nitko & Brookhart, 2006, p. 38). Notice in this definition, and throughout this document, the measurement instrument is not valid or invalid, nor are the results obtained from the assessment valid or invalid. It is the inferences drawn from the scores, or how the scores are interpreted and used, that may be more or less valid. Validity is "an evaluative judgment," (Messick, 1995, p. 741) "integrating evidence from many different sources" (Cronbach & Meehl, 1955, p. 285). The 1974 *Standards for Educational and Psychological Tests and Manuals* (APA, AERA, NCME) declared that there is no single coefficient of validity. As Messick elegantly explains, "It should also be clear that test validity cannot ***rely*** on any one [form of evidence]. . . . What is ***required*** is a compelling argument that the available evidence justifies the test interpretation and use" (p. 744, original emphasis). Indeed, much of the decades-old debate surrounding validity concerns the development of a theoretically defensible and empirically grounded taxonomy of acceptable evidence.

Such a coherent list of acceptable validity-establishing evidence has been a moving target for more than a half century. Although social scientists were cognizant of several facets of validity before World War II, their methods were commonly restricted to correlations of test scores with some known criterion (Angoff, 1988). Under the direction of Lee Cronbach, the 1954 *Technical Recommendations for Psychological Tests and Diagnostic Techniques* compiled by the collaborative efforts of APA, AERA, and NCME, attempted to clarify and broaden the scope of validity by dividing it into four parts: (a) *concurrent* and (b) *predictive validity* (the traditional correlations between the test scores and known criteria), (c) *content validity* (considerations of how well items represent the targeted domain), and (d) *construct validity* (the alignment of the test results with the theoretical constructs they were meant to measure). Cronbach and Meehl's subsequent publication (1955) grouped predictive and concurrent validity into a *criterion-*

orientation, which eventually fossilized as *criterion validity* in the psychometric parlance and trimmed the number of validities to three.

This model of validity as three parts of a whole remained “something of a holy trinity” (Guion, 1980, cited in Angoff, 1988, p. 21), and many subsequent theories of validity developed in contrast to them. These newer views reflected more the Islamic profession of faith, لآ إِلَـهَ إِلاَّ أَللّٰه (“There is no god save God”; Pickthall, 555). Loevinger (1957), Messick (1980), Tenopyr (1977), and Guion (1977) seemed to declare there was no validity save construct validity (Angoff, 1988). Cronbach also acknowledged a need for a unified model of validity to place operationalism (construct validity), empirical validation (criterion validity), and judgments by educators (*consequential validity*), “under the same tent (though not in the same ring)” (Cronbach, 1969, p. 36).

This shift apparently encouraged Messick, who began to protest the disjointed structure of accepted validity models more than thirty years ago. In 1975, he moved away from criterion validity, and strongly questioned the legitimacy of content validity (Messick, 1975). He adopted Loevinger’s (1957) *substantive*, *structural*, and *external aspects* of construct validity as he put up the “tent” to which Cronbach had alluded. At the same time, he advocated ideals regarding consequential validity, or appraising the test, “in light of the probable future consequences of the testing” (Messick, 1975 p. 962). Through two subsequent articles (Messick, 1980, 1995), he developed this model, concluding that construct validity was the only validity, and all validity evidence could be organized as aspects of construct validity. He wrote, “Different kinds of inferences require different kinds of evidence, not different kinds of validity” (Messick 1980, p. 1014).

The 1999 *Standards for Educational and Psychological Testing* (AERA et al.) demonstrate Messick's influence. They describe five types of validity-supporting evidence that incorporate each of Messick's aspects. Table 1 harmonizes the 1999 *Standards* with Messick's model. Messick's admission that the "distinctions [between aspects] may seem fuzzy because they are not only interlinked but overlapping" (1995, p. 747), applies as well to the *Standards'* sources of evidence.

Table 1
Comparison of Messick's Aspects with the 1999 Standards' Sources of Validity-Supporting Evidence

Source of invalidity	Messick's aspect	1999 *Standards'* source of evidence	Suggested evidence-gathering methods
Internal	Content	Test Content	Domain analysis/specification
			Expert ratings of relevance & representativeness
	Substantive	Response Processes	Response pattern analyses
			Think-aloud protocols
	Structural	Internal Structure	Factor analysis
External	External	Relations to Other Variables	Criterion correlations (convergent/discriminant, predictive/concurrent)
	Generalizable		Meta-analysis of test-criterion correlations across testing circumstances
			G-Studies
	Consequential	Consequences of Testing	Empirical evidence of the consequences of test use
			Arguments of potential consequences of test use

Chapter 3: Method

Instrument

This study targeted scores resulting from administrations of an improved version of the TICS. This version included the alterations recommended in Browne's (in press) report, with input from a smaller pilot test ($N = 32$) of an intermediate version. It consisted of 33 self-efficacy items, covering all six NETS-T. (See Appendix A for the complete instrument and Appendix B for its alignment with the NETS-T.)

The TICS was administered during the first and last weeks of the Fall, 2006, semester to all preservice teachers enrolled in IP&T 286 and 287, the MSE technology integration courses. Although participation in the survey was required for class credit, students had the option to either endorse the use of their responses, or request their data not be used in research. More than 90% of respondents endorsed the use of their data. The course instructors removed any identification information from the data before the analysis began.

Participants

The participants were preservice teachers enrolled in IP&T 286 and 287. As in past semesters, enrollment in these courses was predominantly upperclassmen and female. The 287 class was restricted to Elementary, Early Childhood, and Special Education majors, while 286 was a mixture of most Secondary Education majors.

Evidence of Construct Validity

The proposed methods for gathering validity-supporting evidence were organized by the purposes for which the TICS was intended, and the inferences that were necessary for each (see Table 2). However, the classification of one type of evidence under a certain

Table 2
Summary of TICS Purposes, Inferences, and Methods for Gathering Validity Evidence

Intended purposes	Inferences	Basis of evidence	Evidence gathering methods
Establish a baseline preservice teacher profile	Scores accurately reflect self-efficacy vis-à-vis the NETS-T	Internal structure	RSM analysis of response categories Factor analysis
		Test content	Evaluation of the NETS-T functioning as a domain analysis Expert ratings of item relevance and domain representativeness
		Relation to other variables	Correlation of TICS scores with in-class performance Lack of strong correlation with NGSE scores
Monitor the effects of curricular adjustments	Scores are sensitive to resulting changes in self-efficacy	Response processes	Comparison of pre-/post-course TICS scores between elementary and secondary education courses
Identify preservice teachers in most need of intervention	Scores predict in-class difficulties	Relation to other variables	Regression analysis of pre-course TICS scores (and other factors) with course performance
Predict in-practice technology integration	Scores correlate with in-practice behavior	Relation to other variables	A research synthesis of the predictive power of self-efficacy measures

purpose did not preclude its support of another purpose. For example, the content-based evidence listed under Purpose 1 was also important to every other purpose.

Purpose 1: Establish a Baseline Preservice Teacher Profile

This purpose required the inference that the TICS (a) was a reliable measure, and (b) that it measured the construct it purported to measure. While the first inference was addressed by the RSM and factor analyses, the second inference required evidence of a substantive relationship between self-efficacy theory, the NETS-T, and the TICS scores.

Evidence of proper scale functioning. Because validity is "the extent to which a measure reflects only the desired construct without contamination from other systematically varying constructs" (Judd, Smith, & Kidder, 1991, p. 51), a lack of proper item and scale functioning would preclude any argument of validity. Therefore, this validation study began with an RSM analysis (as described in the previous chapter), which verified the functioning of each subscale, and investigated how well each response category functioned. Principle component factor analyses and reliabilities analyses were then carried out to verify that each TICS subscale was reliable and measured one and only one trait.

Evaluation of the NETS-T as a domain model. The TICS' items contained tasks described by the NETS-T, which were accepted as "a specification of the content domain" (AERA et al., p. 11), but also needed to be evaluated for its functioning as a domain analysis, such as recommended by Messick (1995). Bunderson (2003) described such an analysis as a descriptive "theory of progressive attainments" (p. 1), whose development is inseparable from the development of their associated measures and instruments. As described by Bunderson, and modeled by Bond & Fox (2001), the RSM analysis of the pre-course TICS scores ascertained the difficulty of the items and, therefore, helped evaluate the NETS-T as a domain model.

Evidence of the TICS items' relevance and representativeness. The initial development of the TICS included expert ratings of how relevant each item was to "the teaching field" (Browne, 2007, p. 5). Although most items were assigned acceptable relevance ratings, the survey suffered from two flaws that were corrected in the current study. First, the construct against which the raters were to judge each item's relevance, "the teaching field," was not well-defined. Second, the method for selecting expert raters was not established, and the resulting sample included only representatives of secondary education.

To better define the construct to which the items were meant to be relevant, the current expert survey presented each NETS-T individually, with its associated TICS items. This narrowed the raters' task from judging how relevant each item was to "the teaching field" (Browne, 2007, p. 5), to judging how relevant each item was to the standard it was designed to assess. Before moving to the next set of items, the raters assigned a summary rating of how well that particular set of items represented all possible tasks described by the corresponding NETS-T.

To remedy the issue of expert rater selection, a larger sample of preservice teacher educators, in-practice teachers, and public school administrators was drawn to represent the four major groups of preservice teachers in the MSE: (a) early childhood education, (b) elementary education, (c) secondary education, and (d) special education. The participants were not randomly chosen, but the researcher intentionally included individuals who were experienced in technology integration.

Evidence of TICS relation to self-efficacy. Self-efficacy is task-specific, and measures of self-efficacy should reflect that. For example, the TICS should measure self-efficacy regarding technology integration and not some other, more general trait such as confidence or self-esteem.

The New General Self-efficacy Scale (NGSE; Chen, Gully, & Eden, 2001) is a short, task-independent measure of self-efficacy that was administered concurrently to the TICS. Because the construct it purports to measure is associated with self-efficacy, one may expect its results will correlate with TICS subscale scores. However, those correlations should be practically insignificant, and variance in NGSE scores should not explain more than 5% of the variance in TICS subscale scores. Such a low amount of shared variance would mean the self-efficacy construct measured by the TICS is distinct from the construct of general self-efficacy.

Purpose 2: Monitor the Effects of Curricular Adjustments

Expecting the TICS to provide feedback on curricular adjustments implies the TICS would be sensitive to the effects of such adjustments. Although there were no major adjustments to the technology integration curriculum during the time the TICS was employed, there were obvious differences between the curriculum of the secondary education course (IP&T 286) and elementary education course (IP&T 287). While the secondary education course met only one hour each week, the elementary education course met two hours each week. During the semester from which data for this study was collected, the elementary education course introduced a unit addressing NETS-T IV, which was entirely neglected by the secondary education course. If the TICS did not detect this difference between the courses, it may not be sensitive enough to fulfill this purpose.

Purpose 3: Identify Preservice Teachers in Most Need of Intervention

If the TICS is to be used to profile individuals, it must have predictive power. That is, it must be able to predict performance in the technology integration courses. If such profiling is to be a valid use of the TICS, TICS scores should explain enough variance beyond that explained by demographic and other easily gathered data to justify the added time of the TICS administration. If basic demographic information explained enough variance without the TICS, it would be more efficient to identify preservice teachers' needs based on demographic information than on TICS scores. A comparison of three regression models with independent variables of (a) demographic data, (b) the pre-course TICS scores, and (c) both demographic data and TICS scores, and in-class performance as the dependent variable determined if this was the case.

Purpose 4: Predict In-practice Behavior

Effectively determining how well the TICS predicts in-practice use of technology would have required more time than was feasible for the current project. To do so would have required a longitudinal study to track preservice teachers through teacher preparation programs, student teaching, and several years of in-practice teaching. While possible designs for such a study are discussed in the Limitations section of Chapter 5, this project could only support the TICS' power to predict in-practice behavior through an appeal to the self-efficacy research literature.

Appeals to published research literature are accepted methods for gathering validity evidence (Gadbury-Amyot et al., 2003). In fact, "Use of existing data from similar tests and contexts can enhance the quality of the validity argument, especially when current data are limited" (AERA et al., 1999, p. 10-11). Therefore, if published studies had found that self-efficacy predicted behavior, and if the targets of those studies, their contexts, instruments, etc., are sufficiently similar to the TICS, it would be preliminary evidence that the inference required for this use was valid.

Summative Judgment

The 1999 *Standards* (AERA et al.) and others (Cronbach, 1988; Messick ,1995) described validity as an evaluative judgment. Messick used the term *empirical evaluation* to "convey that the validation process is scientific as well as rhetorical and requires both evidence and argument" (p. 747). Indeed, any validity study would be incomplete simply listing the gathered evidence. A summative conclusion, complete with recommendations, which drew upon the evidence and arguments laid out during the study, was required to complete this validation study. Because "validation [is] a continuing process" (Messick, 1995, p. 741), the argument had to be convincing, but wasn't required to establish certainty of the inferences' validity. To compare validation to the American court system, validity is not a criminal case, wherein allegations must be substantiated beyond a reasonable doubt, but a civil dispute where the burden of proof is simply a preponderance of evidence.

Chapter 4: Results

When the initial version of the TICS was pilot tested on preservice teachers enrolled in IP&T 286 and 287, it was only administered at the end of the semester (Browne, in press). Because these courses could be considered a treatment, the respondents' completion of the courses may have reduced the variance in the responses. In other words, the research posited that preservice teachers' self-efficacy as measured by the TICS varied more before the course than after. The version of the TICS investigated in this study was administered pre- and post-course, but, in order to avoid the lack of variance that may stem from using post-course responses[1], only the pre-course responses were used in the following analyses, unless otherwise stated.

Respondent Demographics

All six sections of IP&T 286 and 287 (three sections each) offered in the MSE participated in the study with 212 students completing the pre-course survey. Tables 3 and 4 summarize the distribution of genders, class standing, computer ownership and self-rated computer experience by course. Keep in mind that preservice teachers in 287 must have passed a basic technology skills assessment before enrolling, and those in 286 must have passed the same assessment before the end of the course.

---

[1]The pre-course TICS scores provided, on average, 24% more variance than the post-test scores. This supports the assertion that participation in the courses had a normalizing effect on the respondent sample.

Table 3
Number of Participants in Each Course by Gender and Class Standing

Class standing	IP&T 286 Male	IP&T 286 Female	IP&T 286 Total	IP&T 287 Male	IP&T 287 Female	IP&T 287 Total	Combined total
Sophomore	0	2	2	0	1	1	3
Junior	1	21	22	0	40	40	62
Senior	12	53	65	1	80	81	146
Combined Total	13	76	89	1	121	122	211

Table 4
Number of Participants in Each Course by Computer Expertise and Ownership

	IP&T 286						IP&T 287					
Computer	Computer expertise						Computer expertise					
ownership	1	2	3	4	5	Total	1	2	3	4	5	Total
Desktop	1	0	11	4	2	18	1	3	15	8	1	28
Laptop	3	12	26	7	1	49	2	6	45	11	2	66
Both	0	0	12	1	1	14	0	1	9	4	0	14
Total	4	12	49	12	4	81	3	10	69	23	3	108

Note. Computer expertise was self-rated on a 5-point scale ranging from 1 (novice) to 5 (expert).

Rating Scale Model Analysis

Although the rating scale model was the most complex analysis performed for this study, it is reported first because the resulting recommendations concerning the structure of the TICS response categories affected many other analyses. All RSM analyses were completed using Winsteps software.

Category Probability Curves with Original Data

Figures 8-14 each compare two methods of analysis conducted on the pre-course TICS responses. In Method A, all 33 TICS items were entered into the analysis simultaneously with each item designated as a member of one of the seven subscales. In Method B, each subscale was analyzed separately, with only the items in that subscale entered into the analyses. The major difference between these methods was that the step parameters and category probability curves were estimated across subscales in Method A, while in Method B the step parameters and category probability curves were estimated uniquely for each subscale.

As shown in these comparisons, the estimates of response category functioning changed based on the method of analysis. Method B resulted in better-functioning curves that were easier to interpret. This finding supports the assertion that the TICS consists of distinct subscales, despite their collective unidimensionality (Browne, in press). In other words, although previous analyses have found that the seven TICS subscales measure a common trait, theoretical differences have been postulated. The fact that RSM plots were more interpretable when each subscale was analyzed individually supports this premise. Therefore, the remaining RSM analyses were carried out with Method B.

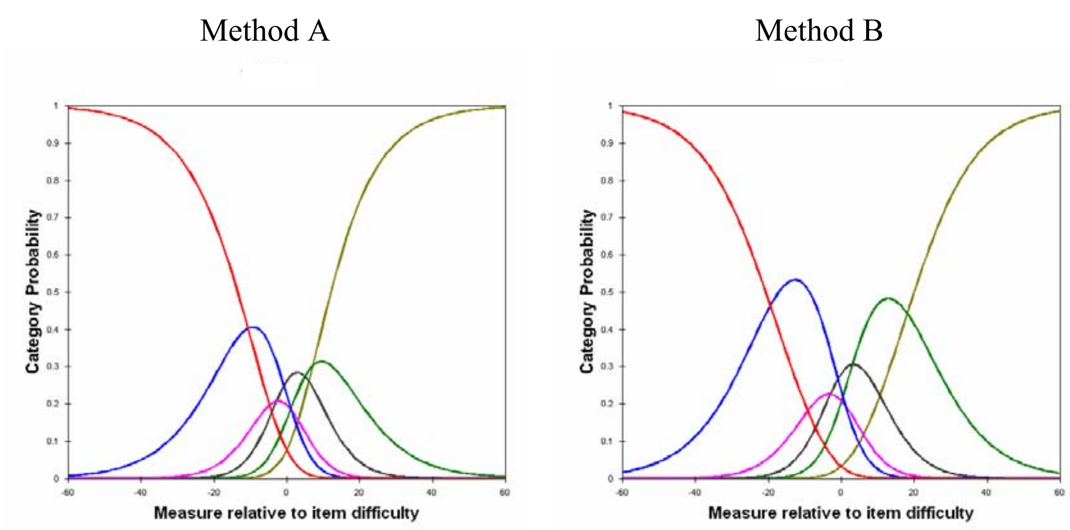


Figure 8. Comparison of RSM analysis methods for TICS Subscale IA.

Method A Method B

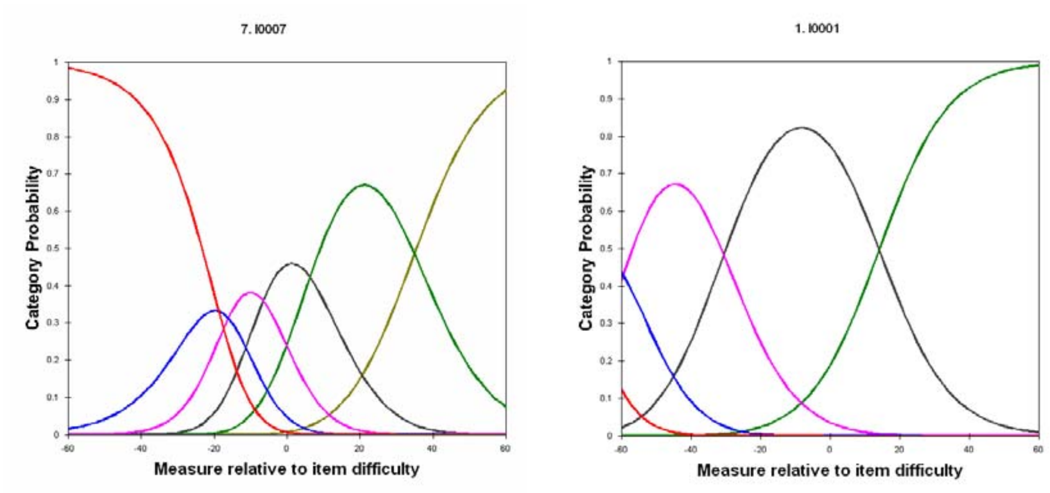


Figure 9. Comparison of RSM analysis methods for TICS Subscale IB.

Method A Method B

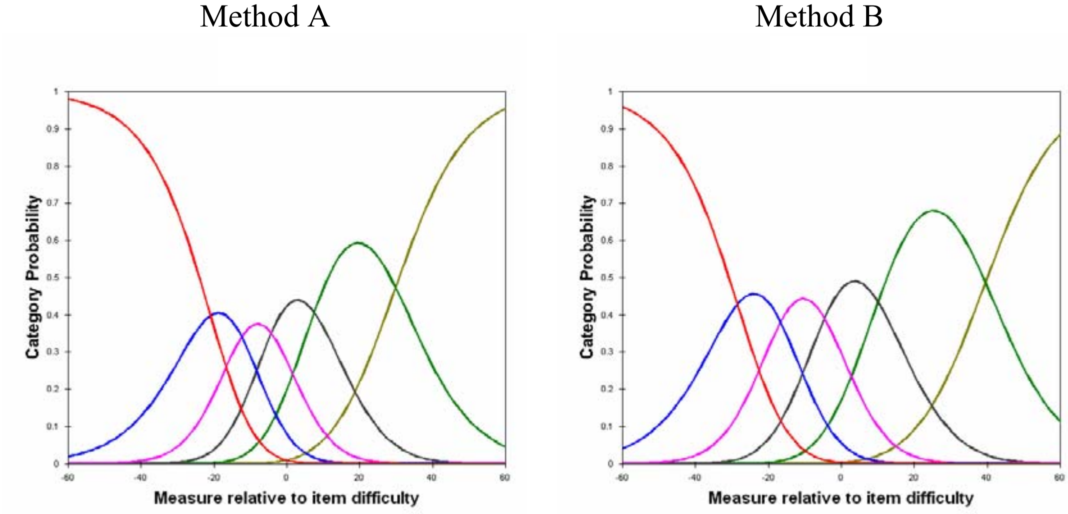


Figure 10. Comparison of RSM analysis methods for TICS Subscale II.

Method A Method B

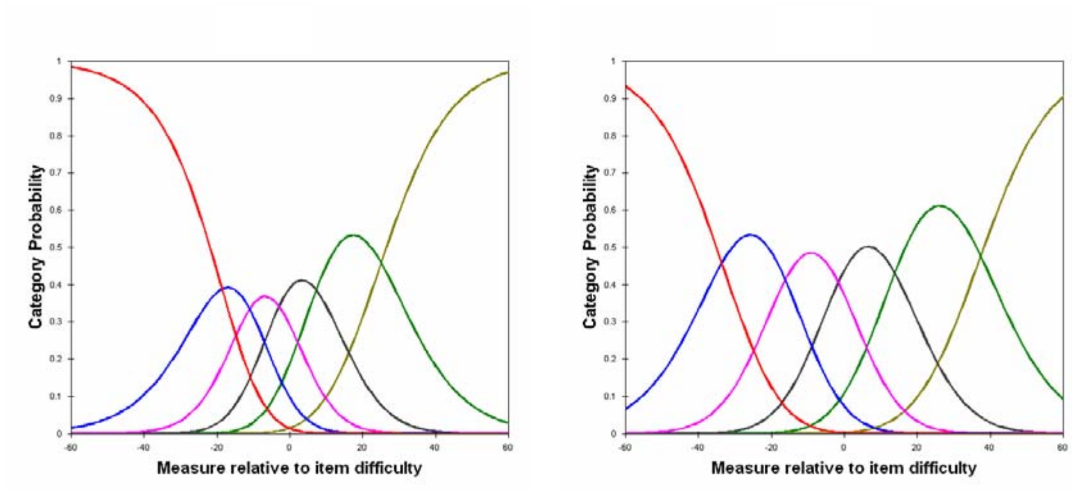


Figure 11. Comparison of RSM analysis methods for TICS Subscale III.

Method A Method B

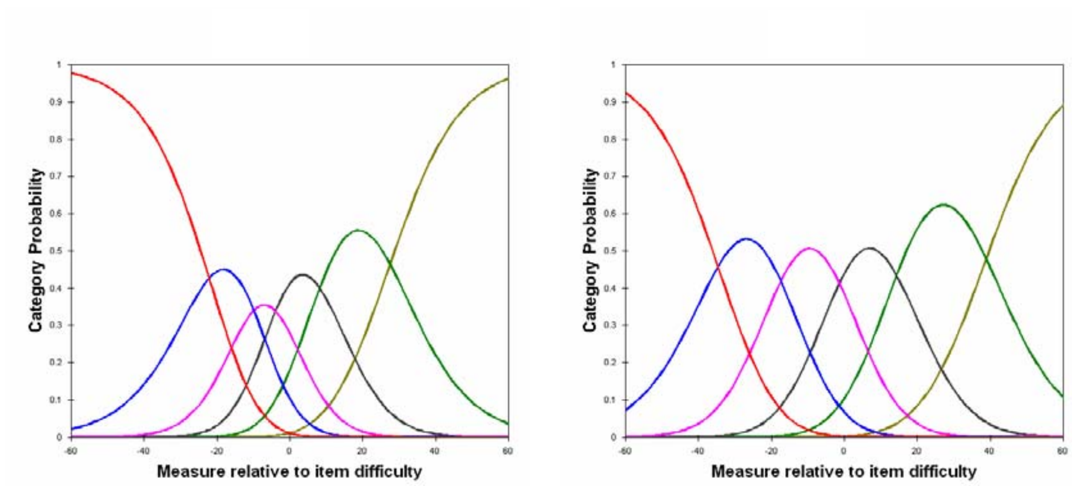


Figure 12. Comparison of RSM analysis methods for TICS Subscale IV.

Method A Method B

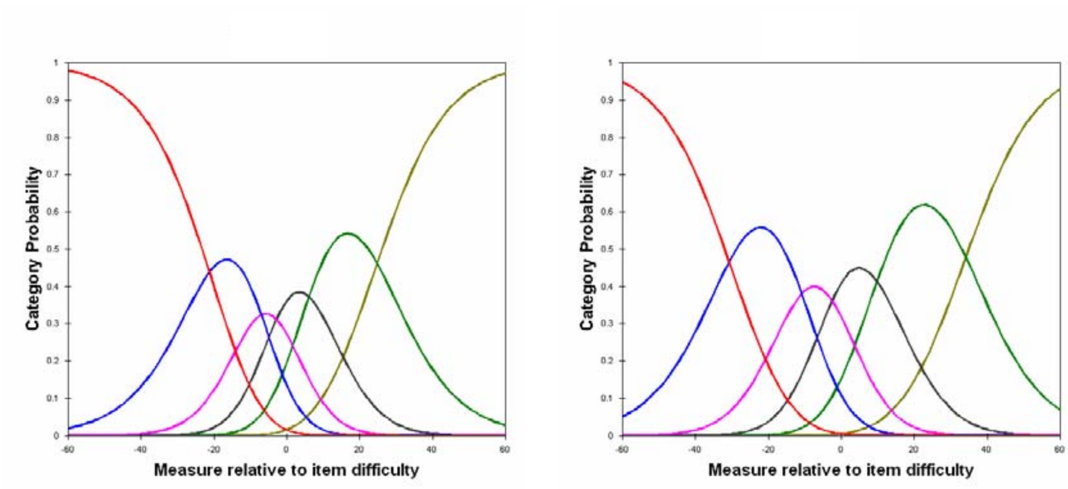


Figure 13. Comparison of RSM analysis methods for TICS Subscale V.

Method A Method B

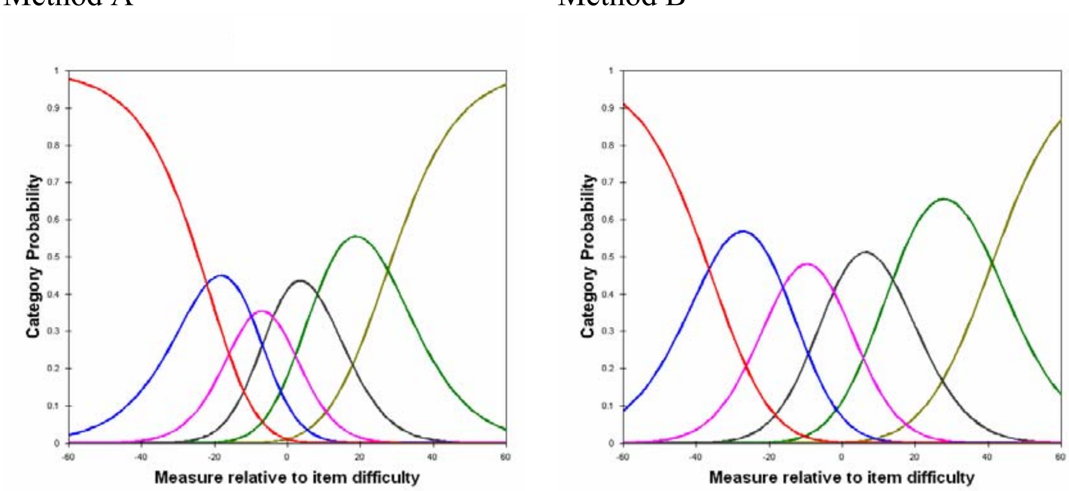


Figure 14. Comparison of RSM analysis methods for TICS Subscale VI.

Category Diagnostics

Table 5 displays the *Step Measures* for each subscale. These numbers represent the self-efficacy level whereat the probability of choosing one category or the next is equal. For example, in Subscale IA, the probability that respondent would choose *Not confident at all* (Category 0) is equal to the probability that they would choose *Slightly* *confident* (Category 1) for respondents with a self-efficacy level of -19.42. Remember that the mean self-efficacy level is 0.

Table 5
Step Measures for Each TICS Subscale

	Step				
Subscale	0-1	1-2	2-3	3-4	4-5
IA	-19.42	1.15	-2.93	3.09	18.12
IB	-72.61	-59.50	-30.59	14.11	148.58
II	-28.47	-16.62	-4.58	10.12	39.55
III	-33.56	-16.13	-1.60	13.86	37.43
IV	-34.59	-17.37	-1.33	14.36	38.93
V	-30.73	-10.80	-2.47	9.82	34.18
VI	-36.46	-16.23	-2.28	13.88	41.09

The primary concern in Table 5 is the disorder in Subscale IA's steps. Notice that the threshold between Categories 1 and 2 for Subscale IA is 1.15, while the threshold between Categories 2 and 3 for Subscale IA is -2.93. This may be caused by the low number of respondents who chose those categories (see Table 7 for details), or it may be due to deeper issues in the items. It should be noted that this is the only case of disorder in the TICS.

Secondary to the issue of disorder is the spacing of the step measures in Table 5. Notice that in Subscales II, III, IV, and VI, the distance between each step is fairly constant across the first four steps. On the other hand, Subscale IB exhibits severe attenuation in that the first step occurs at an impossibly low level of self-efficacy, and the last step at an impossibly high level. Both issues (category disorder in Subscale IA and unequal intervals in other subscales) may be mediated through category recoding, as is addressed in the following section.

Analysis of Item Fit and Recoding

The use of the RSM assumes that the observed data fit the model. This assumption was checked with principle component analyses (reported below) and item infit and outfit statistics. Fit statistics reflect the degree to which responses appeared where the model predicted they would appear. Theoretically, the infit and outfit statistics range from 0 to infinity, but the ideal range is from 0.5 to 1.5. Fit from 0 to 0.5 indicates that the model is not as productive for scale development, but does not degrade the results of the analysis, while 1.5 to 2.0 is unproductive, but also not degrading. Any fit statistic beyond 2.0 represents a distortion of the obtained measurements (Linacre, 1994). As shown in the left half of Table 6, only two items' infit (Items 20 and 29) and one item's outfit (Item 4) were beyond these acceptable limits.

The functioning of the response categories and the overall model fit were improved through minor adjustments in the response categories. Looking at the Method B category probability curves for Subscale IA (Figure 8), it was obvious that collapsing the two middle categories into one would result in a better-ordered and more interpretable plot. Additionally, the category representing the lowest level of self-efficacy (Category 0), was very rarely selected for any item. Therefore, Categories 0 and 1 were collapsed for subscales IB-VI. Figures 15-21 display the response category probability curves for the recoded responses. In other words, responses of *Not confident at all* and *Slightly confident* were both coded as 0, and the other response categories were coded as 1 through 4. As shown on the right side of Table 6, this recoding improved the infit and outfit statistics in two of the three instances of misfit. The distribution of response category selection, for each subscale and for each coding method, is displayed in Table 7.

Table 6
RSM Item Statistics by Coding Scheme

		Original data			Recoded data		
		Mean square		Point-	Mean square		Point-
Subscale	Item	Infit	Outfit	measure *r*	Infit	Outfit	measure *r*
IA	1	.98	.82	.66	.97	.86	.70
	2	.85	.83	.79	.87	.86	.82
	3	1.03	1.07	.68	1.08	1.09	.70
	4	1.25	1.54*	.61	1.24	1.47	.65
	5	.92	.74	.67	.83	.65	.71
	6	1.12	1.02	.72	1.06	1.02	.76
IB	7	1.03	.81	.84	1.00	.80	.85
	8	.87	.70	.86	.93	.72	.88
II	9	1.06	1.10	.71	1.11	1.13	.72
	10	1.02	1.10	.70	1.03	1.11	.72
	11	1.35	1.39	.69	1.31	1.32	.70
	12	.78	.80	.78	.80	.82	.79
	13	.76	.81	.83	.75	.80	.83
	14	.83	.81	.83	.88	.82	.83
	15	.97	.98	.81	.94	.99	.81
III	16	.79	.77	.86	.84	.82	.84
	17	.78	.79	.85	.82	.88	.83
	18	1.08	1.04	.76	1.07	1.04	.78
	19	.88	.90	.84	.82	.79	.83
	20	1.50*	1.44	.76	1.45	1.40	.77
IV	21	1.29	1.25	.75	1.27	1.25	.75
	22	1.37	1.22	.74	1.32	1.19	.76
	23	.66	.65	.85	.67	.66	.86
	24	.81	.81	.82	.83	.85	.82
V	25	.92	.91	.71	.94	.92	.72
	26	.83	.83	.82	.82	.84	.82
	27	1.02	.99	.72	1.00	.99	.73
	28	.79	.80	.77	.79	.80	.78
	29	1.55*	1.34	.68	1.54*	1.34	.69
VI	30	1.05	1.07	.75	1.08	1.09	.76
	31	.88	.88	.82	.90	.95	.82
	32	1.06	1.04	.83	1.02	1.01	.83
	33	.95	.94	.83	.94	.93	.83

Note. * indicates fit statistics that fall outside the boundaries of productivity.

Table 7

Percent of Responses by Subscale, Coding Scheme, and Response Category

	Response category					
Subscale	0	1	2	3	4	5
IA original	1	2	2	5	17	73
IA recode	1	2	7	17	73	
IB original	1	3	12	32	40	12
IB recode	4	12	32	40	12	
II original	4	10	19	28	30	9
II recode	14	19	28	30	9	
III original	4	10	18	26	28	14
III recode	14	18	26	28	14	
IV original	1	5	13	26	34	20
IV recode	6	13	26	34	20	
V original	1	4	7	18	37	34
V recode	4	7	18	37	34	
VI original	2	9	17	28	31	12
VI recode	12	17	28	31	12	
Original average	2	6	13	23	31	25
Recoded average	8	13	24	31	25	

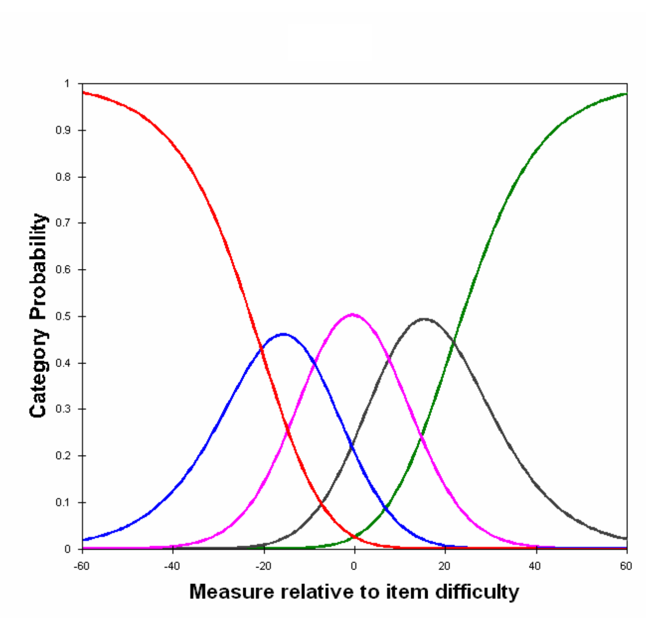


Figure 15. Category probability curves for recoded TICS Subscale IA.

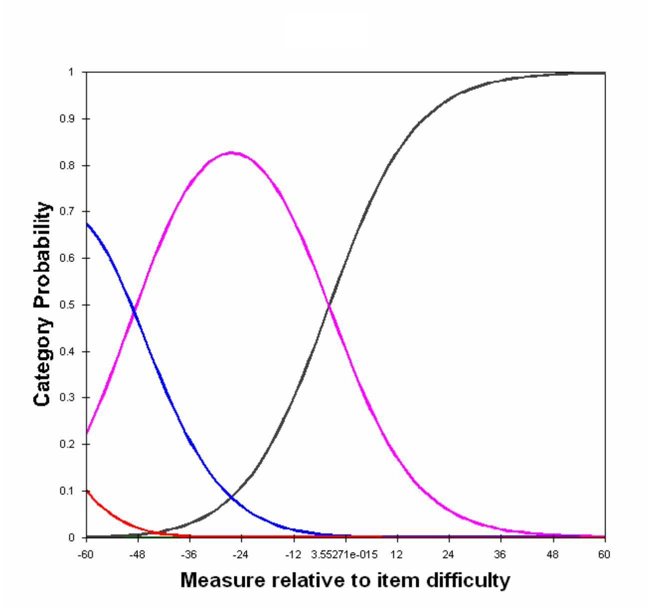


Figure 16. Category probability curves for recoded TICS Subscale IB.

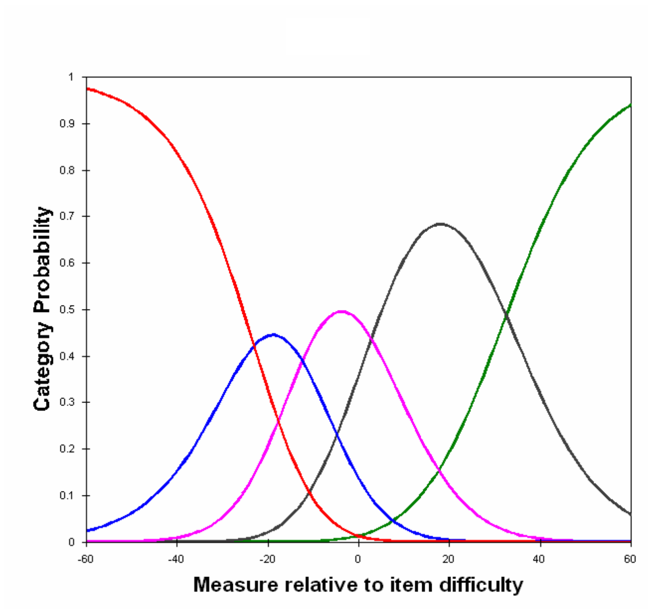


Figure 17. Category probability curves for recoded TICS Subscale II.

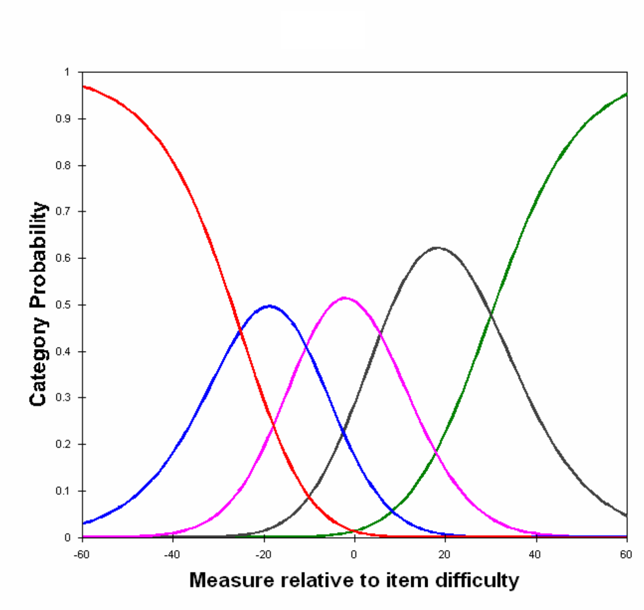


Figure 18. Category probability curves for recoded TICS Subscale III.

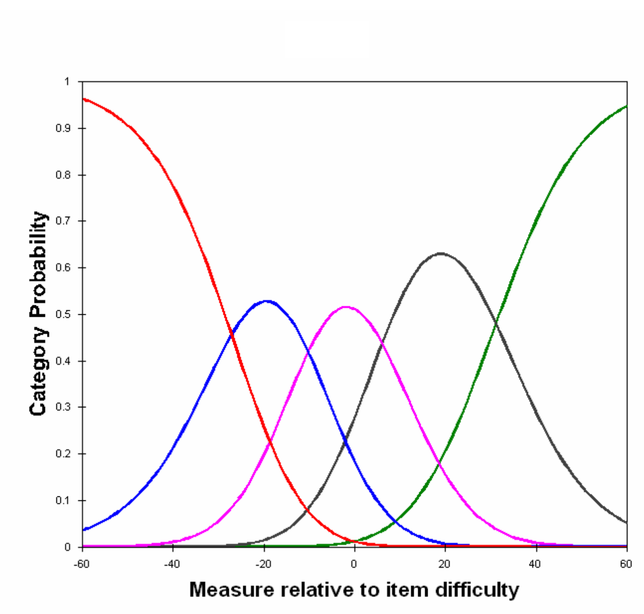


Figure 19. Category probability curves for recoded TICS Subscale IV.

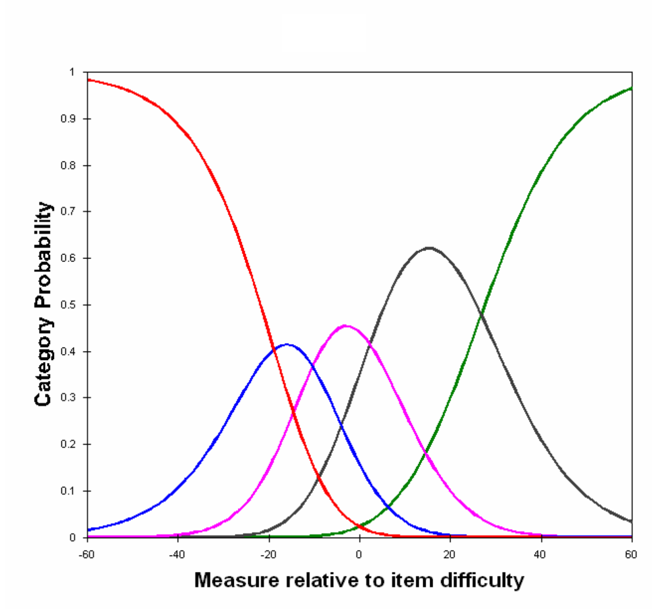


Figure 20. Category probability curves for recoded TICS Subscale V.

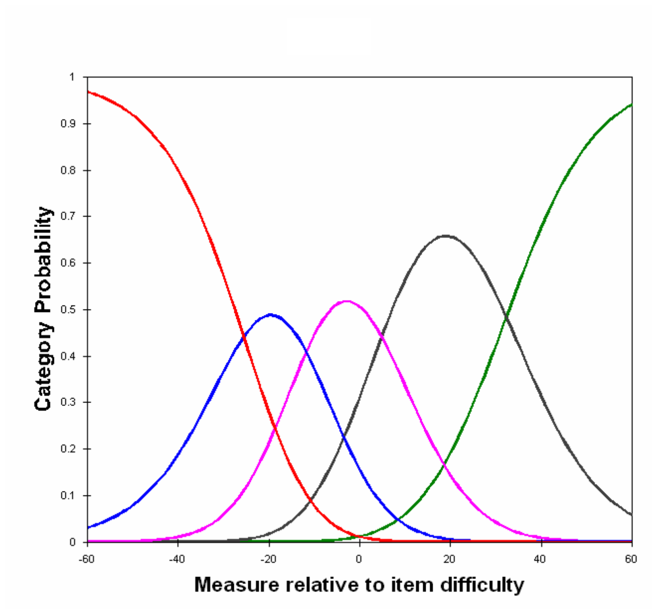


Figure 21. Category probability curves for recoded TICS Subscale VI.

As shown in Table 8, the recoding ameliorated the step measure disorder in Subscale IA, but did not affect the lack of equal distance between each step in the other subscales. In particular, the distance between Steps 2-3 and 3-4 are still much larger than between the other steps.

Table 8
Step Measures for Each TICS Subscale (Recoded Data)

Subscale	Step 0-1	Step 1-2	Step 2-3	Step 3-4
IA	-20.58	-8.82	7.57	21.83
IB	-78.92	-48.95	-3.83	131.70
II	-22.99	-12.37	2.83	32.53
III	-15.12	-10.65	5.69	30.80
IV	-27.07	-10.19	6.09	31.16
V	-18.83	-10.30	2.29	26.85
VI	-25.49	-11.73	4.88	32.34

Comparison of Original and Recoded Model Fit

Embertson & Reise (2000) recommend comparing the model fit statistics (χ^2) for competing models to determine whether the difference is statistically significant. As shown in Table 9, restructuring the response categories significantly improved the fit of the model to the data. Because these analyses demonstrated the superiority of the recoded response categories, the remaining analyses were carried out on the recoded data, unless explicitly stated.

The RSM analysis described above was performed to assess to what degree the response categories function as intended, and to recommend recoding if necessary. RSM also provides scores for each respondent, but because these scores were highly correlated with the recoded raw person mean scores (r was between .94 and .99), using RSM to score TICS responses is deemed unnecessary. In other words, though an RSM analysis was necessary to verify the response category functioning, averaging TICS scores is a sufficient method for scoring individual and group administrations of the TICS. Therefore, the remaining analyses will not consider the RSM item difficulties or person scores. They will present results from both the original and the recoded data.

Table 9

Comparison of Fit Statistics by Coding Scheme

	Chi-square fit statistic			
Subscale	Original data	Recoded data	Difference	*p*
IA	1260.43	1161.95	98.48	<.001
IB	424.55	395.47	29.08	<.001
II	3139.18	2916.14	223.04	<.001
III	2130.91	1953.73	177.18	<.001
IV	1600.80	1525.13	75.67	<.001
V	1867.06	1817.81	49.25	<.001
VI	1700.04	1599.29	100.75	<.001

Note. *df* in all comparisons was 1.

Item Analysis

Table 10 displays the item statistics for each TICS item. It is important to note the low means and relatively high standard deviations for items in Subscales IB through VI. When the initial version of the TICS was pilot tested, it was only administered post-course, and the resulting means were high and the variances were low. It had been hypothesized that this would not be the case in a pre-course administration. The lower means and higher variances displayed in Table 10 support this hypothesis.

In the current version of the TICS, the items associated with Subscale IA had relatively high means and low standard deviations, though they were slightly lower than in the pilot test. Still, with the exception of Item 4, item-corrected-total-subscale correlations were high.

Table 10

Item Statistics by Coding Scheme

		Original data			Recoded data		
Subscale	Item	*M**	*SD*	Item-corrected-total-subscale *r*	*M***	*SD*	Item-corrected-total-subscale
IA	1	4.67	.80	.73	3.70	.68	.69
	2	4.32	1.10	.71	3.39	.90	.71
	3	4.58	.84	.66	3.62	.71	.63
	4	4.64	.79	.57	3.66	.68	.56
	5	4.69	.75	.71	3.72	.64	.72
	6	4.47	.98	.63	3.52	.81	.64
IB	7	3.53	.94	.66	2.54	.92	.67
	8	3.18	1.18	.66	2.22	1.10	.67
II	9	3.34	1.12	.66	2.35	1.10	.64
	10	3.26	1.02	.63	2.27	.99	.63
	11	3.23	1.24	.62	2.26	1.16	.61
	12	3.29	1.06	.73	2.29	1.05	.74
	13	2.56	1.29	.78	1.64	1.17	.78
	14	2.65	1.40	.80	1.72	1.29	.78
	15	2.43	1.38	.73	1.52	1.24	.73
III	16	2.84	1.29	.78	1.89	1.21	.76
	17	2.73	1.25	.76	1.77	1.18	.75
	18	3.82	1.11	.68	2.82	1.10	.67
	19	2.53	1.31	.75	1.62	1.16	.75
	20	3.38	1.35	.63	2.41	1.27	.63
IV	21	3.34	1.18	.55	2.37	1.12	.55
	22	3.96	1.16	.60	2.97	1.13	.61
	23	3.42	1.09	.74	2.42	1.08	.75
	24	3.15	1.14	.68	2.16	1.10	.68
V	25	4.24	.89	.65	3.24	.89	.65
	26	3.28	1.26	.68	2.30	1.21	.68
	27	3.96	1.01	.62	2.97	.99	.63
	28	3.83	.96	.66	2.83	.94	.66
	29	4.11	1.12	.56	3.11	1.10	.56
VI	30	3.43	1.06	.61	2.43	1.05	.60
	31	2.73	1.20	.68	1.76	1.14	.67
	32	3.03	1.34	.69	2.07	1.27	.70
	33	3.29	1.26	.72	2.31	1.21	.72

* The maximum value possible in the original data was 5.0.

** The maximum value possible in the recoded data was 4.0.

Reliability Analysis

Each TICS subscale was found to be acceptably reliable both with the original and the recoded data. However, the recoded data did slightly increase the reliability estimates. This increase was within rounding error, except for Subscale IB, whose reliability coefficient increased from $\alpha = .7862$ to $\alpha = .7973$ (see Table 11).

With both the original and the recoded data, an acceptable reliability coefficient of $\alpha \geq .80$ may have been achieved with fewer items than the TICS constituted. However, removing items from any of the subscales would also remove some representation of the NETS-T and lower the validity of inferences drawn from TICS scores.

Table 11
Reliability Analyses by Coding Scheme

		Original data			Recoded data		
	Number		Number of Items for			Number of Items for	
Subscale	of items	α	α = 0.8	α = 0.9	α	α = 0.8	α = 0.9
IA	6	.86	4	9	.86	4	9
IB	2	.79	3	5	.80	2	5
II	7	.90	4	8	.90	4	8
III	5	.88	3	7	.88	3	7
IV	4	.82	4	8	.82	4	8
V	5	.83	5	10	.83	5	10
VI	4	.84	4	7	.84	4	7
Total	33		27	54		26	54

Factor Analysis

Items associated with each TICS subscale were analyzed independently of items in other subscales. It is desirable that self-efficacy scores be unidimensional, in other words, that the items vary along a single dimension (Bandura, in press). In factor analysis, a factor is considered a dimension when it has a unique value, or *eigenvalue* greater than 1.0. As shown in Table 12, each subscale's solution revealed a single factor with an eigenvalue greater than 1.0. Because of this, no rotation method was applicable.

The percentage of variance accounted for by the factor varied, with Subscales IA and IB exhibiting anomalously low and high results respectively.

It should be noted that the effect of the RSM-recommended recoding was minimal and mixed (it increased some percentages of explained variance and decreased others). It should also be noted that Subscale IB's *KMO* statistic, which assesses the data's factorability, or the appropriateness of factor analyses to the data, was technically *miserable* (Kaiser, 1974). Though this result was just above the cutoff for *unacceptable*, the high percentage of variance explained by the dominating factor in Subscale IB may simply reflect its low numbers of items.

Table 12
Principle Component Analyses by Coding Scheme

	Original data		Recoded data	
Subscale	*KMO*	% Variance explained	*KMO*	% Variance explained
IA	.84	60.78	.83	59.70
IB	.50	83.25	.50	83.67
II	.90	62.63	.91	62.22
III	.85	68.36	.85	67.73
IV	.79	65.29	.79	65.65
V	.84	60.26	.84	60.53
VI	.79	67.70	.79	67.41

Note. Each solution contained a single factor with an eigenvalue greater than 1.0.

Purpose 1: Establish a Baseline Preservice Teacher Profile

Evaluation of the NETS-T as a Domain Theory

Bunderson (2003) defines a domain theory as follows:

> A domain theory is a descriptive theory of the contents, substantive processes, and boundaries of a domain of human learning and growth that gives an account of construct-relevant sources of task difficulty; and conjointly, an account of the

> substantive processes operative at different levels of growth along the scale(s) that span the domain. (p. 5)

A cursory reading of the NETS-T was all that was needed to establish its failure to fulfill Bunderson's definition. The six NETS-T cover the areas of basic computing skill, technology assisted planning, teaching, assessing, and productivity, as well as the ethical issues surrounding educational technology. While this may have marked the borders of the domain, it lacked the "substantive processes" (p. 5) Bunderson required. The defined scope was also so broad that the universe of potential activities described by the NETS-T included every possible in-practice technology use.

The NETS-T and their indicators were not ordered by difficulty, and thus contained no "levels of growth" (p. 5). They did not establish a descriptive "theory of progressive attainments" (p. 1) and they were not developed "using measurement instruments linked to the constructs in the domain theory" (p. 5). Although ISTE published a book of NETS-T assessments, those assessments would not stand up to the AERA, APA, NCME *Standards* (1999), let alone Bunderson's more rigorous requirements.

One should not interpret the preceding paragraphs as a critique of the ISTE's efforts to create the NETS-T. Standards to guide the technology training of preservice teachers were sorely needed, and the NETS-T have filled that need. The purpose of evaluating the NETS-T vis-à-vis domain theory was an attempt to assess the validity of certain inferences, as both Messick (1995) and Bandura (in press) recommended domain analysis in their respective theories of validity and self-efficacy.

A deeper deconstruction of the NETS-T in light of Bunderson's path of domain theory development moved beyond the NETS-T's failings to investigate how they might become a domain theory and where the TICS fit into the process. Beyond his description of an ideal domain theory, Bunderson (2003) also provides a development model: "Observe/Compare → Measure → Interpret → Take Action (Use)" (p. 10). The NETS-T effectively accomplished the first step, "Observe/Compare," which is to delineate the "contents, problems, questions, etc., objectives, tasks, and models of what a proficient person does when performing the work" (p. 5). However, "the process is not complete until a set of essentially unidimensional measurement scales is developed, which together define the scope [of the domain]" (p. 5). The TICS represents part of the next step in Bunderson's process ("Measure") and the current validity study is part of the third step ("Interpret"). Therefore, a consideration of the difficulty of the TICS subscales was completed to inform the progression of the NETS-T towards a functioning domain theory.

Table 13 displays the TICS items in decreasing order of average pre-course response. Notice that preservice teachers in our sample reported higher self-efficacy for the tasks aligned with NETS-T IA (Technology Operations and Concepts) and V (Productivity and Professional Practice). The high average responses to items in Subscale IA may be due to the prerequisite Technology Skills Assessment that preservice teachers had to pass before or during IP&T 286 and 287, and may not reflect preservice teachers in general. NETS-T V addresses the use of technology to increase productivity and "to communicate and collaborate with peers, parents, and the larger community" (ISTE, n.d., Section V). Because many college-age preservice teachers are comfortable with

Table 13

TICS Items Ordered by Mean Response

Subscale	Item	*M*	*SD*
IA	5	3.72	.64
IA	1	3.70	.68
IA	4	3.66	.68
IA	3	3.62	.71
IA	6	3.52	.81
IA	2	3.39	.90
V	25	3.24	.89
V	29	3.11	1.10
IV	22	2.97	1.13
V	27	2.97	.99
V	28	2.83	.94
III	18	2.82	1.10
IB	7	2.54	.92
VI	30	2.43	1.05
IV	23	2.42	1.08
III	20	2.41	1.27
IV	21	2.37	1.12
II	9	2.35	1.10
VI	33	2.31	1.21
V	26	2.30	1.21
II	12	2.29	1.05
II	10	2.27	.99
II	11	2.26	1.16
IB	8	2.22	1.10
IV	24	2.16	1.10
VI	32	2.07	1.27
III	16	1.89	1.21
III	17	1.77	1.18
VI	31	1.76	1.14
II	14	1.72	1.29
II	13	1.64	1.17
III	19	1.62	1.16
II	15	1.52	1.24

Note. The maximum mean value for all items was 4.0.

technology-assisted communications (email, blogs, social networks, etc.), transferring those skills to their professional life may seem natural.

Items aligned with NETS-T II and III tended to elicit lower levels of self-efficacy. In contrast to the tasks with high average responses, tasks in Subscales II and III were taken from in-practice teaching experiences and are not aligned with typical pre-training computer use. The degree to which computer experiences transfer from the pre-training life to NETS-T tasks may be a "construct-relevant [source] of task difficulty" (Bunderson, 2000, p. 5). Therefore, a natural path of NETS-T progressive attainment would move the preservice teacher from tasks with which they are familiar (NETS-T IA and V) to activities that are sequentially less like non-teaching technology use and more like in-practice use (NETS-T II and III).

This does not explain the seemingly random difficulty distribution of the TICS items from Subscales IB, IV, and VI. A detailed analysis of the tasks contained in those items' would be beneficial, but is beyond the scope of the current research.

Evidence of the TICS Items' Relevance and Representativeness

Local teachers and teacher educators were asked to rate the relevance of each TICS item to its associated NETS-T. No random selection was employed because raters were expected to be well-versed in technology integration issues, which would exclude most teachers and teacher educators. Instead, a non-random stratified sample was asked to participate with representatives from early childhood, elementary, secondary, and special education. The respondents were evenly distributed between in-practice teachers and teacher educators.

Relevancy can be evaluated by using Aiken's (1980, 1996) *V* index, which is a statistic that can be tested for significance. As shown in Table 14, no items were rated *Irrelevant* by any rater and the majority of the tasks have high *V* indices. However, there

Table 14
Distribution of Relevancy Ratings of TICS Items

		Distribution of ratings across categories				
Subscale	Item	Relevant	Somewhat relevant	Somewhat irrelevant	Irrelevant	Aiken's *V*
IA	1	6	1	2	0	.81*
	2	6	1	2	0	.81*
	3	7	1	1	0	.89*
	4	8	0	1	0	.93*
	5	8	0	1	0	.93*
	6	7	1	1	0	.89*
	7	8	1	0	0	.96*
IB	8	6	3	0	0	.89*
	9	6	3	0	0	.89*
II	10	9	0	0	0	1.00*
	11	6	3	0	0	.89*
	12	5	4	0	0	.85*
	13	9	0	0	0	1.00*
	14	3	6	0	0	.78*
	15	8	1	0	0	.96*
III	16	8	1	0	0	.96*
	17	9	0	0	0	1.00*
	18	8	1	0	0	.96*
	19	2	4	3	0	.63
	20	8	1	0	0	.96*
IV	21	7	2	0	0	.93*
	22	8	1	0	0	.96*
	23	7	2	0	0	.93*
	24	6	2	1	0	.85*
V	25	9	0	0	0	1.00*
	26	9	0	0	0	1.00*
	27	9	0	0	0	1.00*
	28	9	0	0	0	1.00*
	29	9	0	0	0	1.00*
VI	30	9	0	0	0	1.00*
	31	7	2	0	0	.93*
	32	8	1	0	0	.96*
	33	9	0	0	0	1.00*

Note: * indicates a significant *V* index at $p = .05$.

was considerable disagreement on the relevance of tasks in Subscale IA, and Item 19's *V* index was not significant.

The same raters also judged how well each TICS subscale represented all possible tasks that were described by their NETS-T. Four response categories were presented: *Very well*, *Somewhat well*, *Somewhat poorly*, and *Very poorly*. Although the average ratings were 3.0 (*Somewhat well*) or higher, the high standard deviations revealed considerable disagreement among raters (see Table 15). It should be noted that most raters did agree, but a single rater declared *Very poor* representativeness of every NETS-T but NETS-T I.

Table 15
Representativeness Ratings for Each TICS Subscale

NETS-T	*M* rating	*SD*
I	3.00	.76
II	3.22	.97
III	3.11	1.05
IV	3.33	1.00
V	3.67	1.00
VI	3.44	1.01

Evidence of the TICS' Relation to Self-efficacy

Following Bandura's (in press) recommendations, the TICS are a domain-specific measure of self-efficacy. As such, little variance in TICS scores should be explained by variance in other measures of self-efficacy. The NGSE (Chen, et al. 2001) was administered in the same survey as the TICS to verify this assumption. Because the NGSE is a measure of *general* self-efficacy, it was expected to correlate with TICS scores, even at a statistically significant level, but the correlation should be $r \leq .20$. Such a low correlation would not be practically significant, meaning the NGSE scores would explain 5% or less of the variance in TICS scores.

As shown in Table 16, the NGSE explained less than 5% of the variance in every TICS subscale, compared to the 15% variance that is explained between TICS Subscales IA and IV, and the 70% explained between Subscales II and III. Clearly the TICS subscales and the NGSE were measures of distinct, though related, traits.

Table 16
Percentage of Variance Explained by TICS Subscales and NGSE Scores

	NGSE	IA	IB	II	III	IV	V
IA	3						
IB	0	20					
II	4	20	54				
III	3	16	36	70			
IV	3	22	34	54	58		
V	4	18	39	52	57	59	
VI	4	15	18	43	46	49	47

Purpose 2: Monitor the Effects of Curricular Adjustments

In order to use the TICS to monitor the positive or negative changes in preservice teacher self-efficacy that result from changes in the curriculum, one must assume that the TICS is sensitive to those changes. Rather than purposefully alter the curriculum in randomly selected sections, it was proposed to compare post-course TICS scores between preservice teachers enrolled in IP&T 286 and IP&T 287. These courses differed in several important aspects. First, 287 was a two-credit course, while 286 was only one credit. Second, the extra class time in 287 allows for more step-by-step tutorials, in-class practice, and additional instructional units. For example, while preservice teachers in 286 completed one capstone technology integration project, those in 287 completed two. Additionally, the IP&T 287 instructors introduced a unit on technology-assisted assessment and evaluation (NETS-T IV) the semester the TICS was administered. Finally, although most of the demographic data showed little difference between preservice teachers enrolled in the two courses, 287, which was designed for Elementary, Early Childhood, and Special Education majors, had significantly fewer males enrolled.

Repeated measure analyses with pre- and post-course TICS scores as the within-subjects factor and course enrollment as the between-subjects factor demonstrated that the difference between pre-course and post-course TICS scores was due to the treatment (the courses), and not to specific course enrollment (see Tables 17-23). In other words, on each TICS subscale, the difference between preservice teachers enrolled in one course and another, did not approach significance.

Table 17
Repeated Measures ANOVA for TICS Subscale IA

Source	*SS*	*df*	*MS*	*F*	*p*
Pre/Post	3.67	1	3.67	15.70	<.01
Course	.02	1	.02	.06	.80
Pre/Post * Course	.01	1	.01	.03	.86
Error	90.05	385	.23		
Total	93.74	388			

Table 18
Repeated Measures ANOVA for TICS Subscale IB

Source	*SS*	*df*	*MS*	*F*	*p*
Pre/Post	1.46	1	1.46	1.72	.19
Course	.81	1	.81	.95	.33
Pre/Post * Course	.11	1	.11	.13	.71
Error	331.77	392	.85		
Total	334.15	395			

Table 19

Repeated Measures ANOVA for TICS Subscale II

Source	*SS*	*df*	*MS*	*F*	*p*
Pre/Post	31.16	1	31.16	43.36	<.01
Course	.78	1	.78	1.08	.30
Pre/Post * Course	.22	1	.22	.31	.58
Error	278.07	387	.72		
Total	310.22	390			

Table 20

Repeated Measures ANOVA for TICS Subscale III

Source	*SS*	*df*	*MS*	*F*	*p*
Pre/Post	42.36	1	42.36	55.12	<.01
Course	.00	1	.00	.00	.96
Pre/Post * Course	.52	1	.52	.68	.41
Error	295.08	384	.77		
Total	337.97	387			

Table 21

Repeated Measures ANOVA for TICS Subscale IV

Source	*SS*	*df*	*MS*	*F*	*p*
Pre/Post	26.28	1	26.28	39.59	<.01
Course	.40	1	.40	.60	.44
Pre/Post * Course	.09	1	.09	.13	.71
Error	255.53	385	.66		
Total	282.30	388			

Table 22

Repeated Measures ANOVA for TICS Subscale V

Source	*SS*	*df*	*MS*	*F*	*p*
Pre/Post	10.81	1	10.81	20.69	<.01
Course	.90	1	.90	1.72	.19
Pre/Post * Course	.40	1	.40	.77	.38
Error	200.12	383	.52		
Total	212.23	386			

Table 23
Repeated Measures ANOVA for TICS Subscale VI

Source	*SS*	*df*	*MS*	*F*	*p*
Pre/Post	44.96	1	44.96	62.84	<.01
Course	.08	1	.08	.12	.73
Pre/Post * Course	.18	1	.18	.25	.62
Error	276.18	386	.72		
Total	321.39	389			

Despite specific course enrollment not being a contributor of variance to post-course TICS scores, TICS scores revealed an important difference between the courses. As shown in Table 24, both courses made significant (p = .05) gains pre-post-course on all TICS subscales except on Subscale IB in IP&T 287.

Table 24
Paired-samples t-test of Pre-/Post-course TICS Scores by Subscale and Course

	IP&T 286					IP&T 287				
Subscale	*M* diff.	*SD*	*t*	*df*	*p*	*M* diff.	*SD*	*t*	*df*	*p*
IA	-.19	.50	-3.45	77	<.01	-.24	.40	-5.86	92	<.01
IB	-.15	.63	-2.20	81	.03	-.05	.65	-.79	95	.43
II	-.54	.67	-7.21	79	<.01	-.57	.66	-8.36	92	<.01
III	-.65	.69	-8.26	78	<.01	-.72	.86	-8.11	92	<.01
IV	-.53	.68	-6.88	78	<.01	-.54	.83	-6.30	93	<.01
V	-.30	.76	-3.47	77	<.01	-.42	.72	-5.61	92	<.01
VI	-.67	.68	-8.81	80	<.01	-.74	.93	-7.66	91	<.01

Purpose 3: Identify Preservice Teachers in Most Need of Intervention

To justify the use of TICS scores to profile incoming preservice teachers, the TICS must demonstrate predictive ability vis-à-vis performance on course assignments. Regression analyses, with performance indicators as dependent variables and TICS scores, demographic information, and other easily gathered data as independent variables uncovered whether TICS scores contributed significant predictive power.

Because IP&T 286 and 287 are different courses, assignments in each course differ. Therefore, data from each course were analyzed individually. Only assignments with adequate score variance were considered in the analysis. For example, class readings, and pass/fail assignments on which the vast majority of the class received full credit were not considered.

As shown in Table 25, gender, ownership of a desktop computer, ownership of a laptop computer, and ownership of both, were dummy coded 0 or 1. The other items retained their polytomous codings from the pre-course survey. None of the models considered in this analysis exhibited colinearity issues in that tolerance statistics were greater than .20.

Table 25
Demographics Considered as Independent Variables in the Regression Analyses

Independent variable		Regression coding
Gender		0 = Male, 1 = Female
Computer ownership	Desktop only	0 = No, 1 = Yes
	Laptop only	0 = No, 1 = Yes
	Both	0 = No, 1 = Yes
Computer experience		1 = Novice, 2, 3, 4, 5 = Expert
Frustration frequency		1 = Frequently, 2, 3, 4, 5 = Rarely
Does technology improve education?		1 = Significantly, 2, 3, 4, 5 = Not really
Will technology be a part of your teaching?		1 = It won't be. 2 = It will be a minor part. 3 = It will be a major part.

Predicting In-course Performance in IP&T 286

As shown in Table 26, TICS scores explained more variance in assignment scores than did the demographic data, but neither the demographic variables nor the TICS scores explained more than 18% of the variance in assignment scores. Considering both sets of data together explained more than 20% of the variance in three of the assignments, including the capstone technology integration project. Figure 22 clearly shows the predictive power TICS scores contributed to the regression model. Notice that this contribution does not seem related to the degree of variance in each assignment score (see Table 27).

Table 26
Percent of Variance in IP&T 286 Assignment Scores Explained by Three Regression Models

	Independent variables		
Assignment	Demographic	TICS scores	Both
Attendance	6.7	17.2	22.1
iSafe	5.5	10.4	18.6
Video Modeling	3.6	6.1	14.3
Project Proposal	4.7	9.7	18.9
Concept Map	9.9	15.1	28.3
Project	7.4	12.8	23.2
Final Exam	5.1	8.6	16.8
Final Grade	4.9	9.0	16.3
Average	6.0	11.1	19.8

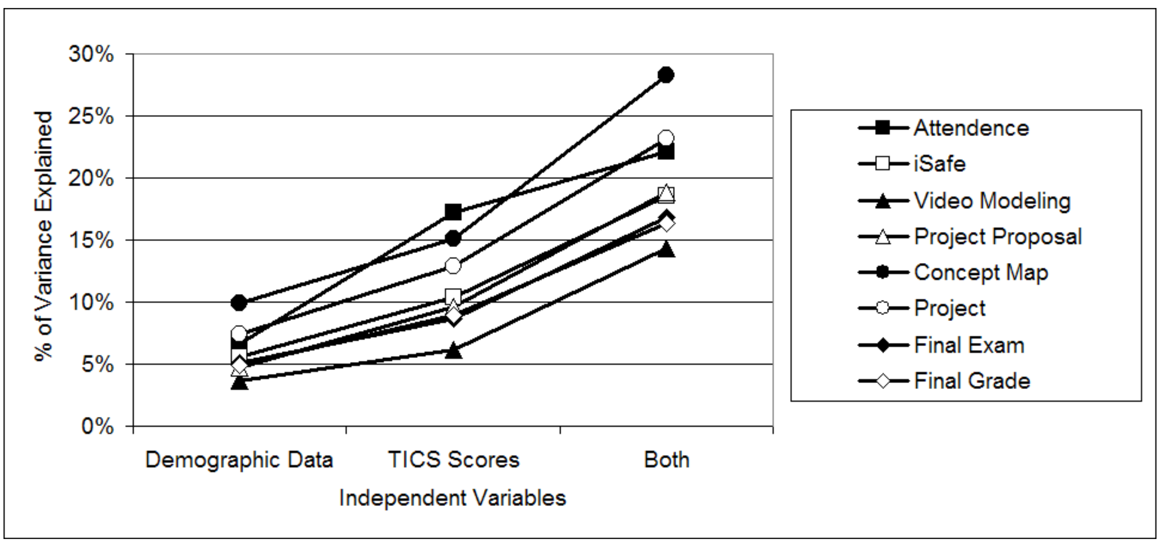


Figure 22. Variance in IP&T 286 assignment scores explained by three regression models.

Table 27
Descriptive Statistics for Assignment Scores in IP&T 286

Assignment	*N*	Mean	*SD*
Attendance	97	88.30	21.32
iSafe	97	97.94	14.28
Video modeling	97	92.54	18.48
Project proposal	97	92.99	11.59
Concept map	54	95.67	13.75
Project	97	91.54	14.06
Final exam	97	88.68	14.45
Final grade	97	90.09	11.39

Predicting In-course Performance in IP&T 287

Table 28 summarizes the percentage of variance in each IP&T 287 assignment that was explained by each model. Note two key differences between these results and those from 286: First, much less assignment score variance was explained in every model. Second, the demographic data were generally more effective predictors of assignment scores than were the TICS scores.

As shown in Figure 23, the TICS scores generally explained little variance in the assignment scores and, as in IP&T 286, the percentage of variance explained appeared unrelated to the degree to which assignment scores varied (see Table 29). In 287, the regression models that included both demographic data and TICS scores explained more than 15% of the variance in only three assignments. In 286, those same independent variables explained more than 15% of the variance in all but one assignment score. Clearly, there are some key differences between these courses that affect the predictive ability of TICS scores. One explanation may be that the increased duration of IP&T 287, and its more hands-on approach, may mitigate the effect of low self-efficacy in preservice teachers. IP&T 286 requires students to complete most computer-based assignments out of class, which may increase the importance of preservice teacher self-efficacy.

Table 28

Percent of Variance in IP&T 287 Assignment Scores Explained by Three Regression Models

Assignment	Independent variables: Demographic	TICS scores	Both
Attendance	6.4	5.2	9.6
iSafe	8.2	4.8	10.6
Copyright Quiz	15.9	4.1	20.2
Performance Assessment	5.3	6.0	12.6
Video Modeling Essay	9.1	13.7	19.5
Assessment Quiz	8.7	8.8	18.4
Project Proposal	7.9	4.7	10.2
Math/Science Project	12.5	5.2	14.8
Soc. Sci./Lang. Arts Project	7.5	3.1	8.7
Final Exam	3.2	7.0	10.7
Final Grade	8.0	5.5	10.7
Average	8.4	6.2	13.3

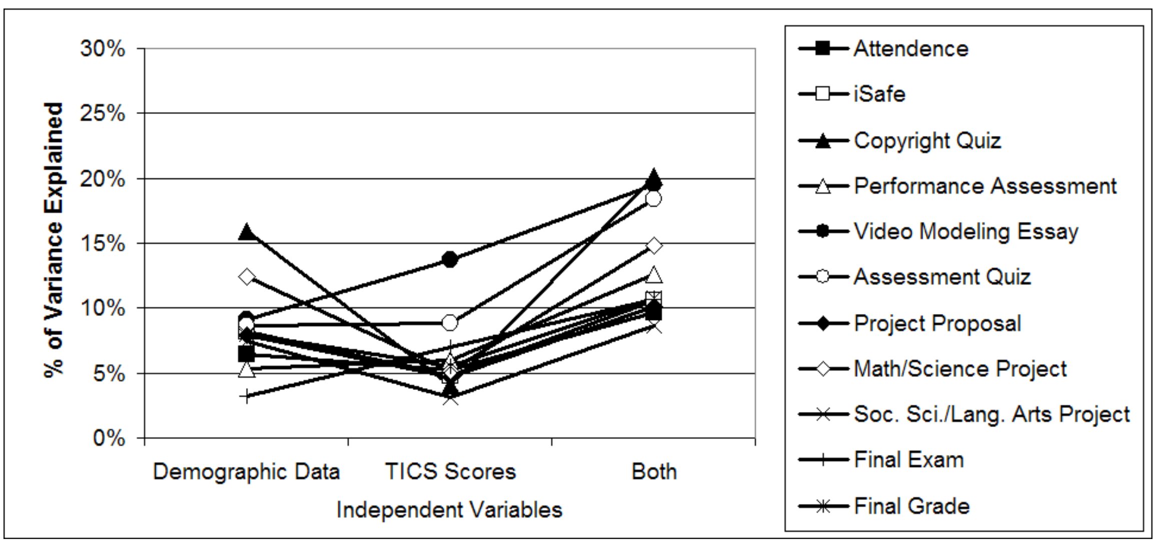


Figure 23. Variance in 287 assignment scores explained by three regression models.

Table 29
Descriptive Statistics for Assignment Scores in IP&T 287

Assignment	*N*	Mean	*SD*
Attendance	119	96.55	17.44
iSafe	119	96.92	16.80
Copyright quiz	119	87.76	13.48
Performance assessment	119	85.64	19.50
Video modeling essay	119	88.29	18.44
Assessment quiz	119	91.71	15.13
Project proposal	119	93.78	27.02
Math/Science project	119	91.24	22.61
Soc. Sci/Lang. arts project	119	90.35	22.27
Final exam	119	91.89	13.27
Final grade	119	90.78	12.53

Another possible explanation for the little variance explained in IP&T 287 assignment scores lies in the fact that the average scores for many of these assignments were quite high, which means the variances will be quite low. In IP&T 287, preservice teachers were allowed to redo many assignments until a satisfactory grade was achieved. Because of this, there was very little variance for the regression models to explain. Given a more selective grading method, TICS scores may prove more predictive.

Purpose 4: Predict In-practice Behavior

The ultimate goal of the TICS was to predict in-practice technology integration while teachers are still preservice. While a longitudinal study would best support the validity of the inferences required for this purpose, such an endeavor was well beyond the scope of the present research. Instead, a synthesis of predictive self-efficacy studies sufficed as validity-supporting evidence.

Five online databases of research publications were searched for the term *Self-efficacy* in article abstracts. As shown in Table 30, this initial search was too broadly defined; adding the term *Predicts* to the search also returned far too many results to be analyzed. Therefore, the search was narrowed to the phrases, *Self-efficacy predicts*, *Self-efficacy can predict*, *Self-efficacy does predict*, and the negative forms of those phrases.

These five queries resulted in 129 references with 85 representing unique articles. Following diligent effort, 36 abstracts and 25 full text articles were obtained and coded based on their target population, dependent variable, methods, findings, and the relationship between self-efficacy and the dependent variable. The coding resulted in 366 data points.

Table 30
Search Terms and Number of Articles Returned from Five Research Databases

	Research database					
Search term	Academic Search Premier	ComDis Dome	Education Full Text	ERIC	PsycInfo	Total
"Self-efficacy"	4,039	66	1,105	2,248	10,178	17,636
"Self-efficacy" and "predict"	227	1	188	68	714	1,198
"Self-efficacy predicts"	7	1	0	2	12	22
"Self-efficacy does predict"	1	0	0	4	31	36
"Self-efficacy can predict"	2	0	1	4	31	38
"Self-efficacy does not predict"	3	0	0	4	26	33
"Self-efficacy cannot predict"	0	0	0	0	0	0

Note: The search phrases "Self-efficacy doesn't predict" and "Self-efficacy can't predict" returned no results.

Frequently Researched Populations

Population was recorded in an open-ended field on the coding form. The most popular populations sampled in the included studies were medical patients ($n = 12$), students and teachers (*n* = 12), or substance abusers (*n* = 11). Nine studies focused on minorities and economically underprivileged groups or women. This diversity speaks well of the broad applicability of self-efficacy theory.

Dependent Variables

The specific trait or behavior of interest was not recorded for each study. Instead, it was coded as *A psychological trait*, *Educational performance*, *Professional performance*, or *Other Behavior*. Because a few studies used self-efficacy as a dependent variable, the additional category of *Self-efficacy predicted by another trait* was added to the analysis, though those articles are of little interest to this research.

Training and education were not considered separately. That is, studies of school counselors-in-training (Ridgway & Sharpley, 1990) and math-related college majors (Hackett, 1985) would both be considered *Educational performance*. If completion of a course or training program, such as an addiction treatment program (Steinhoff-Thorton, 1995), was the dependent variable, it was considered *Educational performance* as well. *Professional performance* was only indicated if the dependent variable was data collected from actual or simulated activities that represented the participants' employment. Table 31 lists the common dependent variables and Table 32 contains a summary of dependent variable codings and their frequency.

Table 31
Common Dependent Variables in Self-efficacy Research

Dependent variable	References
Abstinence from drug & alcohol use	Fiorentine & Hillhouse, 2003a, 2003b; Ilgen, McKellar, & Tiet, 2005; Vielva & Iraurgi, 2001
Anxiety	Nicastro, 1996
Bulimic symptoms	Bardone-Cone, 2002
Depression	Shnek, 1996; Simons, 2002
Disability following joint replacement surgery	Orbell, Johnston, Rowley, Davey, & Espley, 2001
Exercising/activity	Jitramontree, 2003; Buckelew; Luszczynska & Sutton, 2006; Luszczynska, Mazurkiewicz, Ziegelmann, & Schwarzer, 2007
Health status	Riazi, Thompson, & Hobart, 2004
Hormonal reaction to forgiveness training	Standard, 2004
Intent to quit smoking	Yzer, 2006
Intent to teach physically active PE classes	Martin & Kulinna, 2004.
Intent to use search engines as a learning assisted tool	Liaw, Chang, Hung, & Huang, 2006
Pain, stress, anxiety	Hunter, 1995
Parenting beliefs and parent-child relationships	Turner & Johnson, 2003
Perception of spousal abuse	Kugler, 2005
Psychosocial outcomes	Caprara, 2004
Quality of life	Joekes, Elderen, & Schreurs, 2007
Safe needle practices in injection drug users	Falck, 1995
Satisfaction	Seilheimer, 1995

Analysis Methods

Each study's methods were recorded in an open-ended field. If a single study employed multiple methods, each was recorded. Because these studies were predictive in nature, the most popular analyses were various flavors of regression. Twenty-two of the 61 studies employed linear, logistic, multiple, or hierarchal regression. Repeated measures and other analysis of variance approaches were the second-most-used methods (n = 15). Correlation analysis (n = 7), structural equation modeling (n = 6) and path analysis (n = 4) were the other common methods. One study used signal detection.

Because this synthesis focused on research into the predictive power of self-efficacy, qualitative techniques were not coded. However, there were several mixed-methods approaches, typically involving interviews with the participant and/or their parents or teachers.

When regression was used, multiple predictors were entered, including other psychometric scores such as locus of control, learned helplessness, etc. This was highly desirable for this study in that the influence of non-self-efficacy measures would be reported separately.

Research Findings

Each study's findings were coded as *Self-efficacy predicts*, *Self-efficacy doesn't predict*, *Unclear*, or *Not applicable (NA)*. The not applicable designation was only assigned when the study did not report any conclusion regarding predictive power, or when the study sought to predict self-efficacy from other traits.

In some cases of hierarchal regression, self-efficacy's effect on the dependent variable was mediated by other independent variables. For example, Campbell (1995) found that the influence of self-efficacy on hemodialysis patients' dietary compliance was filtered through the effect of the patients' families, background, and gender. In other cases, self-efficacy mediated the effect of other dependent variables. Bardone-Cone (2002) concluded that the effect of female college students' perfectionism influenced their exhibition of bulimic symptoms, but that influence was mediated by their self-efficacy. Hackett (1985) found self-efficacy to mediate the effect of gender and other variables in math-oriented career choices. Ilgen, Tiet, Finney, & Moos (2006) discovered that recovering alcoholics with high self-efficacy did not require as strong patient-therapist relationships to successfully abstain from drinking for one year.

When paths of mediation were considered and self-efficacy was considered a major enough component to include in the final predictive model, it was coded as predicting the dependent variable. In other words, whether contributing independently, as a mediator, or mediated by other variables, if researchers declared self-efficacy a predictor, it was coded as such. If self-efficacy's influence was explained by other dependent variables, it was not coded as a predictor (see Shnek, Foley, LaRocca, Smith, & Halper, 1995).

As shown in Table 32, most of the reviewed studies found that self-efficacy did predict the dependent variable. However, most of the research was not attempting to predict professional or educational performance, which was the focus of this study.

Additional Observations

The construct of self-efficacy. Several of the reviewed studies supported tenets of self-efficacy theory as described by Bandura (in press). For example, self-efficacy is domain and context specific. Joekes, Elderen, and Schreurs (2007) found their congestive heart failure (CHF) recovery self-efficacy scale predicted overall wellness in both CHF and myocardial infarction patients. However, it only predicted quality of life in CHF patients. In other words, their scale's predictive power decreased when they administered it to a different, albeit similar, population.

Table 32
Number of Research Articles by Dependent Variable and Findings

	Did self-efficacy predict *Y*?				
Dependent variable (*Y*)	Yes	No	Unclear	NA	Total
Professional performance	2	1	1	0	4
Educational performance	6	0	0	0	6
Psychological trait	14	4	1	0	19
Other behavior	20	4	1	0	25
Other traits predict self-efficacy	0	0	1	4	5
NA	0	0	0	2	2
Total	42	9	4	6	61

While Joekes et al. (2007) demonstrated how a self-efficacy instrument may not predict traits equally across different populations, Luszczynska, Mazurkiewicz, Ziegelmann, and Schwarzer (2007) found that different instruments did not function equally in a single population. Their study of recovering heart patients showed that recovery self-efficacy was a significant predictor of physical activity level, but maintenance self-efficacy was not.

This specificity may also depend on traits that are irrelevant to the construct of interest. Self-efficacy has been shown to vary across race, gender (Steinhoff-Thorton, 1995), and age group. Simons (2002) found it lacked power to predict life satisfaction in young adults, but it was a major predictor in the elderly (Simons, 2002). The length of time between when self-efficacy is measured and the desired outcome may also affect its predictive functioning (Gore, 2006).

On the other hand, the predictive power of self-efficacy appears culturally independent. Peetsma, Hascher, Van Der Veen, and Roede (2005) found that it predicted adolescent academic achievement in four Western European and former Soviet countries. Vielva & Iraurgi (2001) conducted their research entirely in Spain and found self-efficacy to independently predict alcohol abstinence. As mentioned above, minority populations are often targeted for self-efficacy research. While this does not mean self-efficacy functions similarly in more disparate cultures, it does appear fairly functional within European and American cultures.

The research has produced mixed conclusions regarding whether successful experiences lead to increased self-efficacy. While Waldman (1995) showed that success on the Wisconsin Card Sorting Task did not correlate with an increase in self-efficacy, Britner & Pajares (2006) demonstrated self-efficacy does improve with "mastery experiences" (p. 485). Britner & Pajares' work may carry more weight because a) the distinction between *success* and *mastery experience* is important in the self-efficacy literature, and b) Waldman measured self-efficacy with the General Self-Efficacy Scale Questionnaire, which ignores self-efficacy's domain dependence (Shelton, 1990).

Failures to predict outcomes. Still, there are notable instances where self-efficacy failed to predict, or negatively predicted outcomes. A year-long study concluded that dating violence was not predicted by self-efficacy (Wolfe, Wekerle, Scott, Straatman, & Grasley, 2004), neither were childbirth pain ratings (Hunter, 1995), increases in anxiety about public speaking (Nicastro, 1996), or hormonal changes in salivary cortisol (Standard, 2004),

However, it is difficult to disambiguate the construct's lack of predictive power from the quality of the instrument, its alignment with formal self-efficacy theory, and properties of the dependent variable. Martin and Kulinna (2004) found that self-efficacy's influence on physical education teachers' intent to hold rigorously active class periods was minimal. Perception of control and attitude towards physical activity were far better predictors of that dependent variable. Motl, Dishman, Ward, Saunders, Dowda, Felton, and Pate (2005) likewise found that self-efficacy failed to predict level of activity in adolescents.

While physical education may have presented a self-efficacy construct different from other domains, these negative findings more likely originated with the specific self-efficacy instrument the researchers employed. Both studies used the "Barrier Self-efficacy" scale (BSE; Martin & Kulinna, 2003), which did not measure "the level of difficulty individuals believe they can surmount" (Bandura, in press, p. 4), but what complications the participant would overcome to achieve a given outcome. In other words, the impediments in the BSE were external to the task, while the "level of difficulty" (p. 4), to which Bandura referred, was integral to the outcome in question. This, combined with the dependent variables' (intent) conceptual distance from actual performance, may have made the results difficult to interpret. The findings may not have indicated that self-efficacy, properly defined, does not predict active physical education classes, but that intent to perform a task is independent of difficulties that are exogenous to that task.

On the other hand, Stockman and Altmaier (2001) found that self-efficacy significantly predicted pain and medication use during childbirth even when controlled for other variables. Most apropos, they reported that items reflecting "barriers self-efficacy" were the strongest predictors. Therefore, the predictive impotence of the BSE may be an instrumental issue rather than one stemming from its construct.

Self-efficacy negatively predicted a desirable outcome. Similar reasoning helps explain the single case where self-efficacy negatively predicted a desirable dependent variable. Florentine and Hillhouse (2003b) found that their self-efficacy scale predicted which recovering addicts were amenable to abstinence from drugs and alcohol. Oddly, the self-efficacy scores were negatively correlated with measures of abstinence. However, when one considers their scale's self-efficacy domain, "controlled use" (p. 349), the reason for correlation becomes obvious. Of course recovering addicts who felt they would lose control if they used drugs or alcohol (measured as *low* self-efficacy) would be more willing to abstain from those substances. While this study may have informed theories of rehabilitation psychology, it seemed to have little applicability to self-efficacy theory in general. Contrarily, Florentine and Hillhouse's thinking (their paper was entitled: When low self-efficacy is efficacious) may help explain the negative correlations discovered between some TICS subscales and IP&T 286/287 final exam scores.

Synthesis of Research Literature

Two conclusions may be drawn from the research summarized above. First, self-efficacy theory has been applied in a wide variety of contexts to predict a vast array of traits and behaviors. Second, despite the broad spectrum in which self-efficacy has been applied, it has usually been effective at predicting dependent variables of interest. Infrequently, complex predictor-predicted variable relationships forced some theoretical gymnastics to properly account for the observations, but those situations were not inexplicable. Further, the likelihood that self-efficacy would predict the desired outcome appeared to depend on how closely the researchers and the measures they employed held to Bandura's (in press) description of self-efficacy and his counsel on developing self-efficacy scales.

It may be concluded, therefore, that the TICS would predict in-practice technology integration only to the extent that it a) followed Bandura's (in press) guidelines and b) the relationship between the TICS scores and actual technology integration was not confounded by other psychological and social factors. While the first requirement appears to have been satisfied, the second is largely unknown and should be the subject of further research.

Publication Bias in Predictive Self-efficacy Research

The threat of publication bias is well-accepted in social science meta-research. That is, it is assumed that researchers are more motivated to publish findings that are significant than those which fail to reject the null hypothesis. It is also suspected that peer reviewed journals are more apt to publish articles with significant findings than those without. In the context of this synthesis, this means there are many unpublished studies that did not find self-efficacy to predict the dependent variable. If they were not published, they could not have been considered. Therefore, the conclusions may be spurious. Fortunately, due to three fundamental differences between this research synthesis and the meta-analyses which usually consider publication bias, we can control for the effect of this threat.

First, this research synthesis was not a meta-analysis and flattened the findings to raw numbers for comparison. Rosenthal's (1979) Fail-Safe File Drawer calculation may have been refuted (Scargle, 2000), but his reasoning was astute and applicable to this research. Rosenthal asserted that if the number of unpublished non-significant findings required to invalidate the meta-research's conclusions was beyond reason, the effect of publication bias could be discounted. In the case of this research synthesis, for the number of studies which did not find self-efficacy to predict the dependent variable to equal the number that did, there must be 33 additional studies of publishable quality that went unpublished due to their non-significant findings. This would represent 50% more articles and dissertations than were included in the entire sample, an increase that is most likely not reasonable.

Second, this research synthesis considered both refereed journal articles and doctoral dissertation abstracts. It may be assumed that doctoral dissertations suffer almost no publication bias because their abstracts are published regardless of findings and the probability that a doctoral study would be "shelved" or redone because it resulted in insignificant findings is almost nil. Thus, the proportion of significant to non-significant findings in dissertations may establish a baseline to which the refereed article findings may be compared.

Of the 22 dissertations considered in this research synthesis whose findings were clear, 5, or 23%, reported that self-efficacy was not a predictor of the dependent variable. Only 4 of the 29 refereed journal articles, or 14%, reported the same findings. Therefore, publication bias may be assumed to operate in refereed journal articles, but its impact on the conclusions of this synthesis is dubious. Were the same findings ratio (predictive:non-predictive) to generalize from the doctoral research to the refereed journal research, there would only be three additional predictive self-efficacy studies of publishable quality that were not published because of their non-predictive findings. This would adjust the overall ratio from 42:9 to 42:12.

Third, the detailed breakdown of non-predictive findings in this synthesis showed that some may be dismissed as a misalignment of the researchers' definition of self-efficacy with the formal self-efficacy construct. This analysis, which moves beyond questions of *what* to investigate *why*, is not typically employed in meta-analysis. Because there were acceptable theoretical explanations for many of the non-predictive findings, it may be assumed that some of the unpublished non-predictive studies could be similarly explained.

Publication bias is real, and there is evidence of its effect in this research synthesis. However, publication bias would have to operate on an enormous scale to effect the conclusions of this synthesis. Conversely, were it to operate on the scale at which it was observed, its effect would be minimal. Therefore, it should not be considered a significant threat to this research synthesis' conclusions.

Summative Judgment of Validity

Because validity is a property of the interpretation or inferences drawn from data, and not a property of a test, judgments of validity are not necessarily generalizable between purposes and uses of the same instrument or data. Therefore, not one, but four summative judgments of validity will be rendered based on the results presented above.

Purpose 1: Establish a Baseline Preservice Teacher Profile

Unfortunately, the NETS-T, the standards to which the TICS was designed, do not appear to have been developed as a domain theory. Both Bandura (in press) and Messick (1995) recommend deconstructing the domain to discover the construct-relevant sources of variance and difficulty. Those recommendations are perfectly aligned with Bunderson's (2003) description of domain theory. The NETS-T lack several fundamental components of a domain theory, including the description of task difficulty required by Bandura. Therefore, whether valid inferences can be drawn from any data generated by any NETS-T-aligned measures may be in doubt. Simply put, the NETS-T a) define such a large universe of possible tasks that it would be difficult to sufficiently sample them, and b) do not present a detailed description of successful technology integration, thus the construct-relevant sources of variance in individual performance are unknown. Because two fundamental questions of validity regard the representativeness of the test and how influenced test scores are by construct-irrelevant sources of variance, it may be impossible to build a measure that provides data that is validly interpretable, without making key assumptions.

Fortunately, it seems the assumptions made in the TICS were shared by at least the panel of expert raters who reviewed the TICS items. All but one of its items were judged to be relevant to their specific NETS-T by a panel of experts, and the tasks in each subscale represented *Somewhat well* the universe of possible tasks described by their associated NETS-T. Given the above-described breadth of the NETS-T, this representativeness should be considered a great accomplishment.

It was also important that the TICS measure the trait it was designed to measure (self-efficacy regarding technology integration tasks) and not some other trait. The fact that less than 5% of the variance in the TICS scores was explained by NGSE scores supports this assumption. Paradoxically, the fact that TICS scores did not correlate with in-course performance indicators also bolstered this claim. This lack of correlation showed that the TICS measured something distinct from technology integration skills and knowledge.

With minor adjustments to the response categories, the TICS responses functioned well and demonstrated acceptable levels of reliability, especially for such a short instrument with so many subscales. Each TICS subscale was found to be unidimensional, which was an assumption of the RSM analysis that established the response category functioning, and a requirement of scales associated with domain theories (Bunderson, 2003).

It is, therefore, judged that the interpretations and inferences required to use TICS scores to create a baseline preservice teacher profile are valid. The issues inherent in the NETS-T were largely overcome (as evidenced by the expert ratings), the scale did not measure at least two heavily-related constructs (general self-efficacy and course performance), and it was psychometrically functional.

Purpose 2: Monitor the Effects of Curricular Adjustments

The change in TICS scores from pre-course to post-course was significant, showing that self-efficacy, as measured by the TICS, does change predictably through treatment. Though the expected differences between courses were not observed, the two courses did differ on their change in self-efficacy regarding NETS-T IB. This may be theoretically justified in that IB concerns the confidence to learn new technologies with varying levels of support. The fact that IP&T 286, the course with less time for in-class tutorials and computer work, resulted in a significant increase in Subscale IB scores, while 287, the course with more step-by-step instruction, did not, may be due to the requirement that preservice teachers in 286 work through the new technologies on their own. In other words, 286 was more conducive to the mastery experiences that build self-efficacy in this subscale than was 287.

Still the fact that there was no difference between changes in scores on Subscale IV, which is addressed in 287 and not in 286, did not support the validity of the inferences required to use the TICS to monitor curricular adjustments. It may have been that the adjustments investigated in this study were not designed to influence self-efficacy, but skills and knowledge, and, therefore, attempting to perceive a change in self-efficacy was naïve. No matter the case, there is no evidence to support using the TICS to track minor curricular adjustments, such as adding a unit.

Purpose 3: Identify Preservice Teachers in Most Need of Intervention

Though built on the same standards, and with highly similar syllabi, IP&T 286 and 287 did differ in some ways, and those differences affected the ability of TICS scores to predict in-course performance. In 286, there was evidence supporting the validity of the inference that TICS scores predict in-course performance on most assignments, so long as the scores were combined with demographic data. In 287, the evidence was insufficient to support this inference because the TICS scores explained little variance in assignment scores, and combining the TICS scores with demographic data typically did not explain more than 15% of the assignment score variance.

Purpose 4: Predict In-practice Behavior

The majority of studies that attempted to use self-efficacy to predict some behavior or psychological trait were able to do so. Despite a wide variety of populations and dependent variables, self-efficacy has proven to be a consistent predictor. However, few of these studies addressed in-practice teaching or other professional behaviors, and none concerned technology integration. The connection between self-efficacy as measured by the TICS and in-class use of technology has not been established, but there is evidence supporting self-efficacy's use to predict behaviors in general, inasmuch as those behaviors are aligned with the measured construct and that construct is aligned with formal self-efficacy theory. Thus, using TICS scores to predict in-practice behavior requires inferences which are likely sound, but currently tenuous. More research is required on this topic.

Chapter 5: Discussion

Research Questions

The first two questions of this research were addressed in Chapter 3: Methods. Specifically, Table 2 outlines the inferences that were required to be drawn from TICS data to meet each of its stated purposes, as well as the evidence that was to be gathered to support each of those inferences. Therefore, this section will address the final two questions.

Research Question 3: Given the Appropriate Evidence, Which of the Expected Inferences Are Supported? Which Are Not?

To use the TICS to develop a baseline teacher profile (Purpose 1) implies that TICS scores reflect the psychological construct they were intended to measure – self-efficacy regarding technology integration. This inference is also required by the other purposes. Although not perfect, the seven subscales that make up the TICS function appropriately, represent their target domain, and their items are relevant. Therefore the validity of this inference is supported.

Using the TICS to monitor how changes in the curriculum affect preservice teacher self-efficacy (Purpose 2) implies the TICS is sensitive to the resulting fluctuations in self-efficacy. There is no evidence to support this inference when the curricular changes are minor. Therefore, this inference is unsupported, though the adjustment investigated in this research may not have been effective enough or adequately aligned with self-efficacy to produce the desired evidence.

Inferring that TICS scores predict in-course performance is necessary if those scores are to identify preservice teachers in most need of intervention (Purpose 3). There is evidence to support the validity of this inference. However, the TICS' predictive power was limited to IP&T 286 and only when combined with demographic data. In 287, the evidence was insufficient to support the validity of this inference.

To interpret TICS scores as a prediction of in-practice technology integration (Purpose 4), one must infer that self-efficacy predicts in-practice behavior. This inference was supported as much as possible by a synthesis of research, but this support may be weak. Few of the dependent variables in the reviewed studies approached the context of in-practice teacher performance, or even professional behavior in general.

Research Question 4: What Efforts Should Be Undertaken to Improve the Validity of the Expected Inferences?

Beginning with methods to improve the psychometric functioning of the TICS, the reliability, factor, and RSM analyses all uncovered issues with Subscale IA, which should be significantly revised. In pilot testing, this subscale resulted in outrageously high means and low variances (Browne, in press), but it was unknown whether this was due to the pilot test's administration at the end of a semester, or to issues inherent in those items. Because this pre-course administration also resulted in high means and low variances, we may conclude that the items themselves need to be significantly revised. Additionally, Subscale IB proved to be informative and useful in many of the analyses, but it provided a reliability coefficient that bordered on unacceptability and a factor solution that may not be reliably interpreted due to its low *KMO* statistic. Additionally, the RSM step measures for Subscale IB were extremely large. Though Subscale IB was useful, despite it comprising only two items, it would be more psychometrically sound with one or two additional items.

Moving to more salient issues of validity, most TICS items were relevant and most subscales represented their domain relatively well. However, Item 19 (Subscale III) received by far the lowest and the only non-significant relevance score from expert raters. This low relevance rating should be investigated, expert raters who rated it low should be interviewed for their rationale, and the item should be rewritten.

It should not be expected that the TICS will ever fulfill Purpose 2. The increase in self-efficacy observed in preservice teachers taking IP&T 286/287 did not result from any specific focus on self-efficacy in the syllabus, but was likely a byproduct of the project-based curriculum. The course was designed to instill skills and knowledge, so using the TICS to monitor self-efficacy during minor curricular adjustments is useless because the TICS does not measure skills or knowledge. However, some differences between courses were revealed by TICS scores, but they were not the differences that had been expected. The TICS may provide useful feedback on how major course redesigns affect preservice teachers' self-efficacy, but there is no evidence to support the validity of this inference when the curricular changes are minor.

Limitations

This research was constrained to the sample of preservice teachers to which the researcher had access, and was limited in time and other necessary resources. Therefore four key limitations emerged. First, due to the small sample, the RSM analysis estimated response category functioning across all items of each subscale. In other words, response category diagnostics were only produced at the subscale level. A larger sample would permit the application of the partial credit model (PCM), which would estimate response category functioning for each item.

Second, because only preservice teachers at a single teacher preparation program participated in this study, and because some results were found not to hold across courses within that program, it may not be assumed that the results of this study would be the same at another institution. Were the TICS to be administered in other teacher education programs, and the results analyzed, patterns may emerge that allow for more generalization than the current research.

Third, because the scope of the TICS initial development was confined within a teacher preparation program, no efforts to include in-practice teachers were made, except as expert raters. It is unknown how the profiles of the preservice teachers who participated would compare to those of actual teachers.

Fourth, the research synthesis that was carried out to investigate whether TICS scores could predict in-practice behavior should not replace an eventual longitudinal study of preservice teacher attributes. Such a project would follow teachers from preservice through several years of in-practice experience, and measure multiple psychological and behavioral traits at multiple time intervals. This worthy endeavor was simply beyond the realm of feasability for this study.

Implications and Future Considerations

Implications for Technology Integration Teacher Preparation

This research began with the model that technology integration teacher education programs ignore some preservice teacher traits that may foster in-practice technology use. Courses in these programs focus on the skills and knowledge necessary to use technology effectively in the classroom, but little is known about how these courses affect the preservice teachers' self-efficacy or how they value technology in the classroom.

The creation of the TICS was an initial effort to address this concern, but, without evidence supporting the validity of inferences required by its intended uses, the soundness of any interpretation of TICS scores would be unknown. This study provides some empirical evidence, both quantitative and qualitative, to support the use of the TICS for specific purposes.

Additionally, the fledgling domain analysis of NETS-T tasks reported in this research may lead to a more comprehensive domain theory of technology integration teacher education. The prospect that difficulty in teaching tasks may in part be dependent on how similar those tasks are to pre-training life, as revealed in the modest domain analysis, may be novel and should be researched in more detail.

Implications for Self-efficacy Research

The RSM analysis that investigated how well each response category operated is above and beyond what Bandura (in press) recommended for self-efficacy scale development. However, given an adequate sample size, such analyses are not difficult and provide important psychometric feedback. Of course, Bandrua's suggested *0-100* response scale is not conducive to such analyses. Following the example of Reeve et al. (2006) an administration of the TICS with the 101-point scale, followed by an RSM analysis, may reveal which response scale functions better.

Additionally, provided a larger sample size, a two-parameter IRT model, such as the modified graded response model (M-GRM; Muraki, 1990), could be applied to TICS responses. The RSM analysis in this research assumed the slopes of the conditional probability curves (at the inflexion point) were 1.0. The M-GRM does not make this assumption, but estimates the slope from the observed responses, which provides an estimate of each item's effectiveness at discriminating between respondents with high and low self-efficacy. Winsteps estimations of TICS items' discrimination parameters ranged from .57 to 1.38, with 21 items lower than .90 or greater than 1.10. This indicates a two-parameter analysis would be worthwhile.

Unlike many self-efficacy instruments, the TICS relied wholly upon established standards to define the domain in question. This method may have increased the usability and appeal of the TICS by anchoring it in a well-known description of technology integration, but it also endangered the validity of certain inferences meant to be drawn from TICS scores. Specifically, Bandura (in press) calls for domains to be analyzed to discover what constitutes quality in a given performance, and what makes a task difficult. The NETS-T are not ordered by difficulty, nor provide descriptions of good versus great technology integration. Therefore, the tasks presented in TICS items represented a best-guess at what may be representative of these aspects of the NETS-T domain. These guesses seem to have been very close to the mark because the TICS items received high marks for relevancy and the TICS subscales were rated highly for representativeness.

Similarly, the complex relationship between self-efficacy and task performance is in need of further investigation. As shown in the research synthesis, self-efficacy may be mediated by other variables and may correlate negatively with desirable dependent variables. This complex relationship was evident in this research's attempt to predict assignment performance from demographics and pre-course TICS scores. It was hypothesized that self-efficacy's ability to predict in-course performance may decrease as instructional time increases. In other words, increased time on task may mediate the effect of low self-efficacy. Clearly, more research is needed in this area and IP&T 286 and 287, with their similar instructional objectives and different course structure, offer a unique environment in which one may investigate such questions.

Implications for Measurement Theory

This research did not invent any measurement methodologies, nor did it analyze responses in any novel manner. The types of validity-supporting evidence gathered were not revolutionary. Indeed, all of those methods are described in the 1999 *Standards* (AREA et al.). What was original to this work, and what it can contribute to the measurement field in general, is its systematic approach to gathering validity-supporting evidence. That process included the following steps:

1. Establish the intended purposes for the measure.
2. Determine the inferences required by each purpose.
3. Select sources of evidence that may support each inference.
4. Gather and analyze data for each source of evidence.
5. Form a judgment of validity for each inference based on the gathered evidence.
6. Determine steps to improve the validity of any inferences if needed.

Interestingly, this approach follows closely Bunderson's (2003) domain theory development path, with Steps 1 and 2 aligning with "Observe/Compare," 3 and 4 with "Measure," and 5 and 6 with "Interpret, and Take Action" (p. 10), respectively.

Conclusions

This project to gather evidence supporting the validity of inferences required by the TICS' intended uses was largely successful. Even when the validity of the inferences was not supported by the observed data, the effort was successful because it brought into question or narrowed that particular application of the TICS. However, there were instances where the evidence was less convincing due to issues in the sampled subgroups, or because the evidence that would best support or refute validity was beyond the reach of the research. These instances of failure should be considered in future research.

Chapter 6: Summary Article

This chapter is a compressed version of some of the research included in this dissertation. It focuses on the synthesis of research compiled in Chapter 4 to support the inference that TICS scores predict in-practice technology integration, but includes portions of Chapters 2 and 5 as well. The intent of this chapter was to have something immediately publishable in a refereed journal.

Abstract

Some researchers and educators may be wary of self-efficacy measures due to its self-report format and the potential of respondents to exaggerate their confidence ratings. However, the research literature is rife with examples of self-efficacy scales contributing considerable predicative power to pre-treatment measures. This research synthesis reviewed 61 predictive self-efficacy studies and found that the vast majority (82%) concluded that self-efficacy was a predictor of their dependent variable. However, there are few such studies in professional and educational contexts. Anomalous findings are investigated as is the effect of publication bias on the reviewed sample.

Introduction

Self-efficacy grew out of the cognitive revolution, subsequent renewed interest in the self, and is a partial reaction against self-esteem (Bandura, in press). The theory holds that personal beliefs can predict behavior better than simple stimulus-response reactions, and such beliefs fit within “a theory of personal and collective agency” (Pajares & Schunk, in press, p. 18). It is important to note that, although self-efficacy is most often associated with measurement, it is also a theory of behavioral change through “extraordinary personal feats [that] serve as transforming experiences” (Bandura, in press, p. 2), and formative feedback of each performance (Bandura, 1977).

“Self-efficacy is concerned with people’s beliefs in their capabilities to produce given attainments” (Bandura, in press, p. 2). Such beliefs not only reflect a person’s ability to perform a task, but also the likelihood that the performance will take place, thus increasing the predictive value of its results. When combined with self-efficacy treatments, self-efficacy measures accurately predict outcomes of both individual and group performances, in both pre- and post-treatment situations (Bandura, 1977).
A massive meta-analytic investigation covering published reports from 1977 to 1988, Multon, Brown, and Lent (1991) found self-efficacy measures to account for 14% of variance in student performance and 12% of variance in student persistence. However, they also found evidence that “the relationship of self-efficacy to performance and persistence may vary across types of students, measures, and study characteristics” (p. 34).

The efforts of Multon et al. (1991) were impressive for their scope, but their meta-analysis looked beyond the question of whether self-efficacy measures predict associated behaviors. In order to justify or controvert a future longitudinal study on the topic, a research synthesis was performed to ascertain whether the published research represented a consensus on self-efficacy’s predictive power.

Methodology

Five of the most prominent online databases of educational and psychological research publications were searched for the term *Self-efficacy* in article abstracts. As shown in Table 33, this initial search was too broadly defined; adding the term *Predicts* to the search also returned far too many results to be analyzed. Therefore, the search was narrowed to the phrases *Self-efficacy predicts*, *Self-efficacy can predict*, *Self-efficacy does predict*, and the negative forms of those phrases. These search phrases were specifically selected to return articles that made clear statements regarding the predictive power of self-efficacy.

These five queries resulted in 129 references with 85 representing unique articles. Following diligent effort, 36 abstracts and 25 full text articles were obtained and coded based on their target population, dependent variable, methods, findings, and the relationship between self-efficacy and the dependent variable. The coding resulted in 366 data points.

Table 33
Search Terms and Number of Articles Returned from Five Research Databases

	Research database					
Search term	Academic Search Premier	ComDis Dome	Education Full Text	ERIC	PsycInfo	Total
"Self-efficacy"	4,039	66	1,105	2,248	10,178	17,636
"Self-efficacy" and "predict"	227	1	188	68	714	1,198
"Self-efficacy predicts"	7	1	0	2	12	22
"Self-efficacy does predict"	1	0	0	4	31	36
"Self-efficacy can predict"	2	0	1	4	31	38
"Self-efficacy does not predict"	3	0	0	4	26	33
"Self-efficacy cannot predict"	0	0	0	0	0	0

Note: The search phrases "Self-efficacy doesn't predict" and "Self-efficacy can't predict" returned no results.

Frequently Researched Populations

The target populations of each study were recorded in an open-ended field on the coding form. The most popular populations sampled in the included studies were medical patients ($n = 12$), students and teachers ($n = 12$), or substance abusers ($n = 11$). Nine studies focused on minorities and economically underprivileged groups or women. This diversity speaks well of the broad applicability of self-efficacy theory.

Dependent Variables

The specific trait or behavior of interest was not recorded for each study. Instead, it was coded as *A psychological trait*, *Educational performance*, *Professional performance*, or *Other Behavior*. Because a few studies used self-efficacy as a dependent variable, the additional category of *Self-efficacy predicted by another trait* was added to the analysis, though those articles are of little interest to this research.

Training and education were not considered separately. That is, studies of school counselors-in-training (Ridgway & Sharpley, 1990) and math-related college majors (Hackett, 1985) would both be considered *Educational performance*. If completion of a course or training program, such as an addiction treatment program (Steinhoff-Thorton, 1995), was the dependent variable, it was considered *Educational performance* as well. *Professional performance* was only indicated if the dependent variable was data collected from actual or simulated activities that represented the participants' employment. Table 34 lists the common dependent variables and Table 35 contains a summary of dependent variable codings and their frequency.

Table 34
Common Dependent Variables in Self-efficacy Research

Dependent variable	References
Abstinence from drug and alcohol use	Fiorentine & Hillhouse, 2003a, 2003b; Ilgen, McKellar, & Tiet, 2005; Vielva & Iraurgi, 2001
Anxiety	Nicastro, 1996
Bulimic symptoms	Bardone-Cone, 2002
Depression	Shnek, 1996; Simons, 2002
Disability following joint replacement surgery	Orbell, Johnston, Rowley, Davey, & Espley, 2001
Exercising/activity	Jitramontree, 2003; Buckelew; Luszczynska & Sutton, 2006; Luszczynska, Mazurkiewicz, Ziegelmann, & Schwarzer, 2007
Health status	Riazi, Thompson, & Hobart, 2004
Hormonal reaction to forgiveness training	Standard, 2004
Intent to quit smoking	Yzer, 2006
Intent to teach physically active PE classes	Martin & Kulinna, 2004.
Intent to use search engines as a learning assisted tool	Liaw, Chang, Hung, & Huang, 2006
Pain, stress, anxiety	Hunter, 1995
Parenting beliefs and parent-child relationships	Turner & Johnson, 2003
Perception of spousal abuse	Kugler, 2005
Psychosocial outcomes	Caprara, 2004
Quality of life	Joekes, Elderen, & Schreurs, 2007
Safe needle practices in injection drug users	Falck, 1995
Satisfaction	Seilheimer, 1995

Analysis Methods

Each study's methods were recorded in an open-ended field. If a single study employed multiple methods, each was recorded. Because these studies were predictive in nature, the most popular analyses were various flavors of regression. Twenty-two of the 61 studies employed linear, logistic, multiple, or hierarchal regression. Repeated measures and other analysis of variance approaches were the second-most-used methods (n = 15). Correlation analysis (n = 7), structural equation modeling (n = 6) and path analysis (n = 4) were the other common methods. One study used signal detection.

Because this synthesis focused on research into the predictive power of self-efficacy, qualitative techniques were not coded. However, there were several mixed-methods approaches, typically involving interviews with the participant and/or their parents or teachers.

When regression was used, multiple predictors were entered, including other psychometric scores such as locus of control, learned helplessness, etc. This was highly desirable for this study in that the influence of non-self-efficacy measures would be reported separately.

Research Findings

Each study's findings were coded as *Self-efficacy predicts*, *Self-efficacy doesn't predict*, *Unclear*, or *Not applicable (NA)*. The not applicable designation was only assigned when the study did not report any conclusion regarding predictive power, or when the study sought to predict self-efficacy from other traits.

In some cases of hierarchal regression, self-efficacy's effect on the dependent variable was mediated by other independent variables. For example, Campbell (1995) found that the influence of self-efficacy on hemodialysis patients' dietary compliance was filtered through the effect of the patients' families, background, and gender. In other cases, self-efficacy mediated the effect of other dependent variables. Bardone-Cone (2002) concluded that the effect of female college students' perfectionism influenced their exhibition of bulimic symptoms, but that influence was mediated by their self-efficacy. Hackett (1985) found self-efficacy to mediate the effect of gender and other variables in math-oriented career choices. Ilgen, Tiet, Finney, and Moos (2006) discovered that recovering alcoholics with high self-efficacy did not require as strong patient-therapist relationships to successfully abstain from drinking for one year.

When paths of mediation were considered and self-efficacy was considered a major enough component to include in the final predictive model, it was coded as predicting the dependent variable. In other words, whether contributing independently, as a mediator, or mediated by other variables, if researchers declared self-efficacy a predictor, it was coded as such. If self-efficacy's influence was explained by other dependent variables, it was not coded as a predictor (see Shnek, Foley, LaRocca, Smith, and Halper, 1995).

Results

As shown in Table 35, most of the reviewed studies found that self-efficacy did predict the dependent variable. However, most of the research was not attempting to predict professional or educational performance.

Table 35
Number of Research Articles by Dependent Variable and Findings

	Did self-efficacy predict *Y*?				
Dependent variable (*Y*)	Yes	No	Unclear	NA	Total
Professional performance	2	1	1	0	4
Educational performance	6	0	0	0	6
Psychological trait	14	4	1	0	19
Other behavior	20	4	1	0	25
Other traits predict self-efficacy	0	0	1	4	5
NA	0	0	0	2	2
Total	42	9	4	6	61

The Construct of Self-efficacy

Several of the reviewed studies supported tenets of self-efficacy theory as described by Bandura (in press). For example, self-efficacy is domain and context specific. Joekes, Elderen, and Schreurs (2007) found their congestive heart failure (CHF) recovery self-efficacy scale predicted overall wellness in both CHF and myocardial infarction patients. However, it only predicted quality of life in CHF patients. In other words, their scale's predictive power decreased when they administered it to a different, albeit similar, population.

While Joekes et al. (2007) demonstrated how a self-efficacy instrument may not predict traits equally across different populations, Luszczynska, Mazurkiewicz, Ziegelmann, and Schwarzer (2007) found that different instruments did not function equally in a single population. Their study of recovering heart patients showed that recovery self-efficacy was a significant predictor of physical activity level, but maintenance self-efficacy was not.

This specificity may also depend on traits that are irrelevant to the construct of interest. Self-efficacy has been shown to vary across race, gender (Steinhoff-Thorton, 1995), and age group. Simons (2002) found it lacked power to predict life satisfaction in young adults, but it was a major predictor in the elderly (Simons, 2002). The length of time between when self-efficacy is measured and the desired outcome may also affect its predictive functioning (Gore, 2006).

On the other hand, the predictive power of self-efficacy appears culturally independent. Peetsma, Hascher, Van Der Veen, and Roede (2005) found that it predicted adolescent academic achievement in four Western European and former Soviet countries.

Vielva and Iraurgi (2001) conducted their research entirely in Spain and found self-efficacy to independently predict alcohol abstinence. As mentioned above, minority populations are often targeted for self-efficacy research. While this does not mean self-efficacy functions similarly in more disparate cultures, it does appear fairly functional within European and American cultures.

The research has produced mixed conclusions regarding whether successful experiences lead to increased self-efficacy. While Waldman (1995) showed that success on the Wisconsin Card Sorting Task did not correlate with an increase in self-efficacy, Britner and Pajares (2006) demonstrated self-efficacy does improve with "mastery experiences" (p. 485). Britner and Pajares' work may carry more weight because a) the distinction between *success* and *mastery experience* is important in the self-efficacy literature and b) Waldman measured self-efficacy with the General Self-Efficacy Scale Questionnaire, which ignores self-efficacy's domain dependence (Shelton, 1990).

Failures to Predict Outcomes

Still, there are notable instances where self-efficacy failed to predict or negatively predicted outcomes. A year-long study concluded that dating violence was not predicted by self-efficacy (Wolfe, Wekerle, Scott, Straatman, & Grasley, 2004), neither were childbirth pain ratings (Hunter, 1995), increases in anxiety about public speaking (Nicastro, 1996), or hormonal changes in salivary cortisol (Standard, 2004),

However, it is difficult to disambiguate the construct's lack of predictive power from the quality of the instrument, its alignment with formal self-efficacy theory, and properties of the dependent variable. Martin and Kulinna (2004) found that self-efficacy's influence on physical education teachers' intent to hold rigorously active class periods was minimal. Perception of control and attitude towards physical activity were far better predictors of that dependent variable. Motl, Dishman, Ward, Saunders, Dowda, Felton, and Pate (2005) likewise found that self-efficacy failed to predict level of activity in adolescents.

While physical education may have presented a self-efficacy construct different from other domains, these negative findings more likely originated with the specific self-efficacy instrument the researchers employed. Both studies used the "Barrier Self-efficacy" scale (BSE; Martin & Kulinna, 2003), which did not measure “the level of difficulty individuals believe they can surmount” (Bandura, in press, p. 4), but what complications the participant would overcome to achieve a given outcome. In other words, the impediments in the BSE were external to the task, while the “level of difficulty” (p. 4), to which Bandura referred, was integral to the outcome in question. This, combined with the dependent variables’ (intent) conceptual distance from actual performance, may have made the results difficult to interpret. The findings may not have indicated that self-efficacy, properly defined, does not predict active physical education classes, but that intent to perform a task is independent of difficulties that are exogenous to that task.

On the other hand, Stockman and Altmaier (2001) found that self-efficacy significantly predicted pain and medication use during childbirth even when controlled for other variables. Most apropos, they reported that items reflecting “barriers self-efficacy” were the strongest predictors. Therefore, the predictive impotence of the BSE may be an instrumental issue rather than one stemming from its construct.

Self-efficacy Negatively Predicted a Desirable Outcome

Similar reasoning helps explain the single case where self-efficacy negatively predicted a desirable dependent variable. Florentine and Hillhouse (2003b) found that their self-efficacy scale predicted which recovering addicts were amenable to abstinence from drugs and alcohol. Oddly, the self-efficacy scores were negatively correlated with measures of abstinence. However, when one considers their scale's self-efficacy domain, "controlled use" (p. 349), the reason for correlation becomes obvious. Of course recovering addicts who felt they would lose control if they used drugs or alcohol (measured as *low* self-efficacy) would be more willing to abstain from those substances. While this study may have informed theories of rehabilitation psychology, it seemed to have little applicability to self-efficacy theory in general. Contrarily, Florentine and Hillhouse's thinking (their paper was entitled: When low self-efficacy is efficacious) may help explain the negative correlations discovered between some TICS subscales and IP&T 286/287 final exam scores.

Conclusions

Two conclusions may be drawn from the research summarized above. First, self-efficacy theory has been applied in a wide variety of contexts to predict a vast array of traits and behaviors. Second, despite the broad spectrum in which self-efficacy has been applied, it has usually been effective at predicting dependent variables of interest. Infrequently, complex predictor-predicted variable relationships forced some theoretical gymnastics to properly account for the observations, but those situations were not inexplicable. Further, the likelihood that self-efficacy would predict the desired outcome appeared to depend on how closely the researchers and the measures they employed held to Bandura's (in press) description of self-efficacy and his counsel on developing self-efficacy scales.

It may be concluded, therefore, that the TICS would predict in-practice technology integration only to the extent that it a) followed Bandura's (in press) guidelines and b) the relationship between the TICS scores and actual technology integration was not confounded by other psychological and social factors. While the first requirement appears to have been satisfied, the second is largely unknown and should be the subject of further research.

Publication Bias in Predictive Self-efficacy Research

The threat of publication bias is well-accepted in social science meta-research. That is, it is assumed that researchers are more motivated to publish findings that are significant than those which fail to reject the null hypothesis. It is also suspected that peer reviewed journals are more apt to publish articles with significant findings than those without. In the context of this synthesis, this means there are many unpublished studies that did not find self-efficacy to predict the dependent variable. If they were not published, they could not have been considered. Therefore, the conclusions may be spurious. Fortunately, due to three fundamental differences between this research synthesis and the meta-analyses which usually consider publication bias, we can control for the effect of this threat.

First, this research synthesis was not a meta-analysis and flattened the findings to raw numbers for comparison. Rosenthal's (1979) Fail-Safe File Drawer calculation may have been refuted (Scargle, 2000), but his reasoning was astute and applicable to this research. Rosenthal asserted that if the number of unpublished non-significant findings required to invalidate the meta-research's conclusions was beyond reason, the effect of publication bias could be discounted. In the case of this research synthesis, for the number of studies which did not find self-efficacy to predict the dependent variable to equal the number that did, there must be 33 additional studies of publishable quality that went unpublished due to their non-significant findings. This would represent 50% more articles and dissertations than were included in the entire sample, an increase that is most likely not reasonable.

Second, this research synthesis considered both refereed journal articles and doctoral dissertation abstracts. It may be assumed that doctoral dissertations suffer almost no publication bias because their abstracts are published regardless of findings and the probability that a doctoral study would be "shelved" or redone because it resulted in insignificant findings is almost nil. Thus, the proportion of significant to non-significant findings in dissertations may establish a baseline to which the refereed article findings may be compared.

Of the 22 dissertations considered in this research synthesis whose findings were clear, 5, or 23%, reported that self-efficacy was not a predictor of the dependent variable. Only 4 of the 29 refereed journal articles, or 14%, reported the same findings. Therefore, publication bias may be assumed to operate in the refereed journal articles, but its impact on the conclusions of this synthesis is dubious. Were the same findings ratio (predictive:non-predictive) to generalize from the doctoral research to the refereed journal research, there would only be three additional predictive self-efficacy studies of publishable quality that were not published because of their non-predictive findings. This would adjust the overall ratio from 42:9 to 42:12.

Third, the detailed breakdown of non-predictive findings in this synthesis showed that some may be dismissed as a misalignment of the researchers' definition of self-efficacy with the formal self-efficacy construct. This analysis, which moves beyond questions of *what* to investigate *why*, is not typically employed in meta-analysis. Because there were acceptable theoretical explanations for many of the non-predictive findings, it may be assumed that some of the unpublished non-predictive studies could be similarly explained.

Publication bias is real, and there is evidence of its effect in this research synthesis. However, publication bias would have to operate on an enormous scale to affect the conclusions of this synthesis. Conversely, were it to operate on the scale at which it was observed, its effect would be minimal. Therefore, it should not be considered a significant threat to this research synthesis' conclusions.

References

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement 40*, 955-959.

Aiken, L. R. (1996). *Rating scales and checklists: Evaluating behavior, personality, and attitudes*. New York, NY: John Wiley.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1954). *Technical recommendations for psychological tests and diagnostic techniques.* Washington, DC: The Association.

American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1974). *Standards for educational & psychological tests.* Washington, DC: The Association.

Angoff, W. H. (1988). Validity: An evolving concept. In H. Wainer & H. Braun (Eds.), *Test Validity* (pp. 19-32). Hillsdale, NJ: Lawrence Erlbaum.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavior change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (in press). Guide for constructing self-efficacy scales. In Pajares, F., & Urdan, T. (Eds.). *Adolescence and education, Vol. 4: Self-efficacy beliefs of adolescents*. Greenwich, CT: Information Age Publishing.

Bardone-Cone, A. M. (2002). An interactive model of perfectionism, perceived weight status, and self-efficacy in the prediction of bulimic symptoms. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *62*(7), 3371.

Basham, J., Palla, A., & Pianfetti, E. (2005). An integrated framework used to increase preservice teacher NETS-T ability. *Journal of Technology and Teacher Education, 13*, 257-276.

Berkson, D. (1998). Family functioning and self-efficacy in heart-healthy dietary change. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *58*(7) 3916.

Blue, C. L. (1997). Theory of planned behavior and self-efficacy and exercise behavior in blue-collar workers. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *57*(11), 6848.

Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model: Fundamental measurement in the human sciences.* Hillsdale, NJ: Erlbaum.

Bouffard-Bouchard, T. (1990). Influence of self-efficacy on performance in a cognitive task. *Journal of Social Psychology*, *130*(3), 353-364.

Brinkerhoff, J. (2006). Effects of a long-duration, professional development academy on technology skills, computer self-efficacy, and technology integration beliefs and practices. *Journal of Research on Technology in Education, 39*, 22-43.

Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching*, *43*(5), 485-499.

Browne, J. M. (in press). Assessing preservice teacher attitudes and skills with the Technology Integration Confidence Scale. *Computers in the schools.*

Browne, J. M. (2007). Measuring preservice teacher self-efficacy of technology integration. *Society for Information Technology & Teacher Education International Conference: Proceedings of SITE 2007, 18*, San Antonio, TX.

Brzycki, D., & Dudt, K. (2005). Overcoming barriers to technology use in teacher preparation programs. *Journal of Technology and Teacher Education, 13*, 619-641.

Buckelew, S. P., Huyser, B., Hewett, J. E., & Parker, J. C. (1996). Self-efficacy predicting outcome among fibromyalgia subjects. *Arthritis Care & Research*, *9*(2) 97-104.

Bui, U. (2005). The career development of Vietnamese Americans: Factors in career self-efficacy and traditionality of career choice. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *66*(5), 2874.

Bunderson, C. V. (2003). How to build a domain theory: On the validity centered design of construct-linked scales of learning and growth. In M. Wilson (Ed.). *Objective Measurement: Theory into Practice.* Stamford, CT: Ablex Publishing Co.

Butler, D. L., Sellbom, M. (2002). Barriers to adopting technology for teaching and learning. *Educause Quarterly, 25,* 22-28.

Campbell, K. A. (1995). Compliance with chronic dietary regimens: Measurement, prediction, and outcome. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *56*(1), 0503.

Caprara, G. V., Barbaranelli, C., Pastorelli, C., & Cervone, D. (2004). The contribution of self-efficacy beliefs to psychosocial outcomes in adolescence: Predicting beyond global dispositional tendencies. *Personality and Individual Differences*, *37*(4), 751-763.

Charoula, A. (2005). Transforming a teacher education method course through technology: effects on preservice teachers' technology competency. *Computers & Education, 45*, 383-398.

Charrow, C. B. (2006). Self-efficacy as a predictor of life satisfaction in older adults. *Dissertation Abstracts International: Section A: Humanities and Social Sciences*, *67*(1), 292.

Chen, G., Gully, S. M., & Eden, D. (2001). Validation of a new general self-efficacy scale. *Organizational Research Methods*, *4*, 62-83.

Condly, S. J. (2000). Motivation to learn and to succeed: A path analysis of the cane model of cognitive motivation. *Dissertation Abstracts International: Section A: Humanities and Social Sciences*, *60*(12), 4315.

Cormier, R. M. A. (2002). Predicting treatment outcome in chemically dependent women: A test of Marlatt and Gordon's relapse model. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *62*(10), 4777.

Cronbach, L. J. (1969). Validation of educational measures. *Proceedings of the 1969 Invitational Conference on Testing Problems*. Princeton, NJ: Educational Testing Service, 35-52.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test Validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions* (2nd ed.). Urbana: University of Illinois Press.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: LEA, Inc.

Falck, R. S., Siegal, H. A., Wang, J., & Carlson, R. G. (1995). Usefulness of the health belief model in predicting HIV needle risk practices among injection drug users. *AIDS Education and Prevention*, *7*(6), 523-533.

Fiorentine, R., & Hillhouse, M. P. (2001). The addicted-self model: An explanation of "natural" recovery? *Journal of Drug Issues*, *31*(2), 395-424.

Fiorentine, R., & Hillhouse, M. P. (2003). Replicating the addicted-self model of recovery. *Addictive Behaviors*, *28*(6), 1063-1081.

Fiorentine, R., & Hillhouse, M. P. (2003). When low self-efficacy is efficacious: Toward an addicted-self model of cessation of alcohol-and drug-dependent behavior. *American Journal on Addictions*, *12*(4), 346-365.

Foley, S., Remley, C., Morales, D. R., Grega, L. M., Lantz, N., Haughton, N. A., et al. (2001). Designing and developing rubrics as an instrument to assess ISTE's (International Society for Technology in Education) National Educational Technology Standards for Teachers (NETS-T). *Annual proceedings of selected research and development [and] practical papers presented at the national convention of the Association for Educational Communications and Technology, 24*, Atlanta, GA.

Gadbury-Amyot, C. C., Kim, J., Palm, R. L., Mills, G. E., Noble, E., & Overman, P. R. (2003). Validity and reliability of portfolio assessment of competency in a baccalaureate dental hygiene program. *Journal of Dental Education*, *67*, 991-1002.

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2004). Collective efficacy beliefs: Theoretical developments, empirical evidence, and future directions. *Educational Researcher, 33*(3), 3-13.

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal, 37*, 479-507.

Gore, P. A. Jr. (2006). Academic self-efficacy as a predictor of college outcomes: Two incremental validity studies *Journal of Career Assessment*, *14*(1), 92-116.

Graham, C., Culatta, R., Pratt, M., & West, R. (2004). Redesigning the teacher education technology course to emphasize integration. *Computers in the Schools, 21*(1-2), 127-148.

Guion, R. M. (1977). Content validity–The source of my discontent. *Applied Psychological Measurement*, *1*, 1-10.

Guion, R. M. (1980). On trinitarian doctrines of validity. *Professional Psychology, 11,* 385-398.

Hackett, G. (1985). Role of mathematics self-efficacy in the choice of math-related majors of college women and men: A path analysis. *Journal of Counseling Psychology*, *32*(1) 47.

Hambleton, R. K., Swaminathan, H., & Rogers, J. R. (1997). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Harper, M. C. (2006). Retirement modeling: An exploration of the effects of retirement role model characteristics on retirement self-efficacy and life satisfaction in midlife workers. *Dissertation Abstracts International: Section A: Humanities and Social Sciences*, *66*(8), 2842.

Heide, A. & Henderson, D. (2001). *Active Learning in the Digital Age Classroom*. Portsmouth, NH: Heinemann.

Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A reliability generalization study of the Teacher efficacy scale and related instruments. *Educational and Psychological Measurement, 61*, 404-420.

Hernandez, C. A. (1998). Prediction of metabolic control in individuals with insulin dependent diabetes mellitus. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *58*(7), 3556.

Holden, G., Meenaghan, T., Anastas, J., & Metrey, G. (2002). Outcomes of social work education: The case for social work self-efficacy. *Journal of Social Work Education*, *38*(1), 115-134.

Hoy, A. W., & Spero, R. B. (2005a). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education, 21*, 343-356.

Hoy, A. W., & Spero, R. B. (2005b). Erratum to ‘Changes in teacher efficacy during the early years of teaching: A comparison of four measures’ [Teaching and Teachers Education 21(4) 343-356]. *Teaching and Teacher Education*, *21*, 743.

Hoy, W. K., & Woolfolk, A. E. (1993). Teachers' sense of efficacy and the organizational health of schools. *The Elementary School Journal, 93*, 355-372.

Hunter, W. J. (1995). Predicting labor outcomes in low-risk primiparas using cognitive-behavioral variables. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *56*(4), 1944.

Ilgen, M., McKellar, J., & Tiet, Q. (2005). Abstinence self-efficacy and abstinence 1 year after substance use disorder treatment. *Journal of Consulting and Clinical Psychology*, *73*(6), 1175.

Ilgen, M., Tiet, Q., Finney, J., & Moos, R. H. (2006). Self-efficacy, therapeutic alliance, and alcohol-use disorder treatment outcomes. *Journal of Studies on Alcohol*, *67*(3), 465-472.

International Society for Technology in Education. (n.d.). *Educational technology standards and performance indicators for all teachers.* Retrieved September 20, 2006, from http://cnets.iste.org/teachers/t_stands.html

Jitramontree, N. (2003). Predicting exercise behavior among Thai elders: Testing the theory of planned behavior. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *64*(4), 1685.

Joekes, K., Van Elderen, T., & Schreurs, K. (2007). Self-efficacy and overprotection are related to quality of life, psychological well-being and self-management in cardiac patients. *Journal of Health Psychology*, *12*(1), 4-16.

Judd, C. M., Smith, E. R., & Kidder, L. H. (1991). *Research methods in social relations* (6th ed.). Fort, Worth, TX: Holt, Rinehart and Winston.

Kaiser, H. F. (1974). Analysis of factorial simplicity. *Psychometrika, 39*, 31–36.

Kay, R. H. (2006). Evaluating strategies used to incorporate technology into preservice education: A review of the literature. *Journal of Research on Technology in Education, 38*, 383-408.

Kelley, M., Wetzel, K., Padgett, H., Williams, M. K., & Odom, M. (2005). Early childhood teacher preparation and technology integration: The Arizona State University West experience. *Information Technology in Childhood Education, 18*, 245-260.

King, J. D. (1996). Investigating the general and mathematics efficacy and attributions of African American students, as predictors of pursuit of mathematics-related careers. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, *56*(8), 3041.

Kugler, R. (2005). Predicting children's internalizing and externalizing behaviors: The moderating effect of battered mothers' self-efficacy on their perceptions of lethality. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *66*(3), 1723.

Levin, B. B. (1999). Is the class of 1998 ready for the 21st century school? Longitudinal study of computer-using teacher candidates. East Lansing, MI: National Center for Research on Teacher Learning. (ERIC Document Reproduction Service No. ED432556)

Liaw, S., Chang, W., Hung, W., & Huang, H. (2006). Attitudes toward search engines as a learning assisted tool: Approach of Liaw and Huang's research model. *Computers in Human Behavior*, *22*(2), 177-191.

Linacre, J. M., Wright, B., D. (1994) Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3) 370.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 634-694.

Luszczynska, A., Mazurkiewicz, M., Ziegelmann, J. P., & Schwarzer, R. (2007). Recovery self-efficacy and intention as predictors of running or jogging behavior: A cross-lagged panel analysis over a two-year period. *Psychology of Sport and Exercise*, *8*(2), 247-260.

Luszczynska, A., & Sutton, S. (2006). Physical activity after cardiac rehabilitation: Evidence that different types of self-efficacy are important in maintainers and relapsers. *Rehabilitation Psychology*, *51*(4), 314-321.

Madden-Szeszko, G. M. (2000). Variables contributing to teacher efficacy: An examination of burnout, affect, demographic variables, and general self-efficacy. *Dissertation Abstracts International: Section A: Humanities and Social Sciences*, *61*(3), 881.

Martin, J. J., & Kulinna, P. H. (2004). Self-efficacy theory and the theory of planned behavior: Teaching physically active physical education classes. *Research Quarterly for Exercise and Sport*, *75*(3), 288-298.

Martin, J. J., Kulinna, P. H. (2003). The Development of a Physical Education Teachers' Physical Activity Self-Efficacy Instrument. *Journal of Teaching in Physical Education, 22*, 219-232.

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*, *30*, 955-966.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, *35*, 1012-1027.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741-749.

Mills, S. C. (2001). Analyzing teacher preparation to integrate technology in classroom instruciton. Annual Proceedings of Selected Research and Development [and] Practice Papers Presented at the National Convention of the Association for Educational Communcations and Technology, 24, Atlanta, GA, November 8-12,2001). Volumes 1-2.

Milner, H. R., & Woolfolk Hoy, A. (2003). Teacher self-efficacy and retaining talented teachers: A case study of an African American teacher. *Teaching and Teacher Education, 19*, 263–276.

Motl, R. W., Dishman, R. K., Ward, D. S., Saunders, R. P., Dowda, M., Felton, G., & Pate, R. R. (2005). Comparison of barriers self-efficacy and perceived behavioral control for explaining physical activity across 1 year among adolescent girls. *Health Psychology*, *24*(1), 106-111.

Multon, K., Brown, S., & Lent R. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30-38.

Nicastro, R. S. (1996). Cognitive processes in social anxiety: The role of imperatives and self-efficacy. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *56*(11), 6467.

Nitko, A. J., & Brookhart, S. M. (2006). *Educational assessment of students* (5^{th} ed.). Upper Saddle River, NJ: Prentice Hall.

Orbell, S., Johnston, M., Rowley, D., Davey, P., & Espley, A. (2001). Self-efficacy and goal importance in the prediction of physical disability in people following hospitalization: A prospective study. *British Journal of Health Psychology*, *6*(1), 25-40.

Pajares, F., & Schunk, D. H. (in press) Self and self-belief in psychology and education: An historical perspective. In J. Aronson & D. Cordova (Eds.), *Psychology of education: Personal and interpersonal forces*. New York: Academic Press.

Peetsma, T., Hascher, T., van der Veen, I., & Roede, E. (2005). Relations between adolescents' self-evaluations, time perspectives, motivation for school and their achievement in different countries and at different ages. *European Journal of Psychology of Education*, *20*(3), 209-225.

Persichitte, K. A., Caffarella, E. P., & Ferguson-Pabst, D. (2003). A continuing journey toward technology infusion within teacher preparation. *TechTrends, 47*(2), 12-14.

Pickthall, M. M. (1977). *The glorious qur'an.* New York, NY: Muslim World League.

Rabinowitz, Y. G. (2005). The relationship between caregiving self-efficacy and cumulative health risk as determined by health behavior patterns in female caregivers of elderly relatives with Alzheimer's dementia. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *66*(6), 3423.

Reeve, S., Kitchen, E., Sudweeks, R. R., Bell, J. D., & Bradshaw, W. S. (2006). *Self-efficacy in cell biology*. Manuscript submitted for publication.

Riazi, A., Thompson, A. J., & Hobart, J. C. (2004). Self-efficacy predicts self-reported health status in multiple sclerosis. *Multiple Sclerosis (13524585)*, *10*(1), 61-67.

Ridgway, I. R., & Sharpley, C. F. (1990). Multiple measures for the prediction of counsellor trainee effectiveness. *Canadian Journal of Counselling*, *24*(3), 165-177.

Rivera, L. M. (2002). Relations of perceived barriers, acculturation, and role models to the career self-efficacy and career considerations of Hispanic women. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *63*(6), 3021.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86*, 638–641.

Scargle, J. D. (2000). Publication Bias: The "File-Drawer" Problem in Scientific Inference. *Journal of Scientific Exploration, 14*(1), pp. 91–106.

Schock, M. D. (1995). Adolescents, condoms and self-efficacy: A social cognitive study of condom use among adolescents in juvenile detention. *Dissertation Abstracts International: Section A: Humanities and Social Sciences*, *56*(4), 1529.

Seilheimer, T. A. (1995). Self-efficacy, quality of life domains, and residential satisfaction: A multivariate analysis of mental patients. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *56*(2), 1121.

Shadel, W. G., & Cervone, D. (2006). Evaluating social-cognitive mechanisms that regulate self-efficacy in response to provocative smoking cues: An experimental investigation. *Psychology of Addictive Behaviors*, *20*(1), 91-96.

Shannon, J., Kirkley, B., Ammerman, A., Keyserling, T., Kelsey, K. DeVellis, R., & Simpson, RJ Jr. (1997). Self-efficacy as a predictor of dietary change in a low-socioeconomic-status Southern adult population. *Health Education & Behavior*, *24*(3), 357-368.

Shelton, Suzanne H. (1990). Developing the construct of general self-efficacy. *Psychological Reports, 66*, 987-994.

Shnek, Z. M. (1996). Predictors of depression in multiple sclerosis and spinal cord injury. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *57*(3), 2210.

Shnek, Z. M., Foley, F. W., LaRocca, N. G., & Smith, C. R. (1995). Psychological predictors of depression in multiple sclerosis. *Journal of Neurologic Rehabilitation*, *9*(1), 15-23.

Shoffner, M., & Dias, L. (2001). On-line support and portfolio assessment for NETS-T standards in preservice program at a large Southeastern university. *Annual proceedings of selected research and development [and] practical papers presented at the national convention of the Association for Educational Communications and Technology, 24*, Atlanta, GA.

Simons, C. M. (2002). Proactive coping, perceived self-efficacy, and locus of control as predictors of life satisfaction in young, middle-aged, and older adults. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *63*(3), 1590.

Soet, J. E., DiLorio, C., & Dudley, W. N. (1998). Women's self-reported condom use: Intra and interpersonal factors. *Women & Health*, *27*(4), 19-32.

Standard, S. (2004). Effects of a forgiveness intervention on salivary cortisol, DHEA, and psychological variables. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *65*(4), 2114.

Steinhoff-Thornton, R. (1995). Self-efficacy as a predictor of attrition for African American and white client populations in treatment for poly-drug abuse. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *55*(12), 5281.

Stevens, T., Olivarez, A. Jr., Lan, W. Y., & Tallent-Runnels, M. K. (2004). Role of mathematics self-efficacy and motivation in mathematics performance across ethnicity. *Journal of Educational Research*, *97*(4), 208-221.

Stockman, A. F., & Altmaier, E. M. (2001). Relation of self-efficacy to reported pain and pain medication usage during labor. *Journal of Clinical Psychology in Medical Settings*, *8*(3), 161-166.

Swain, C. (2006). Preservice teachers self-assessment using technology: Determining what is worthwhile and looking for changes in daily teaching and learning practices. *Journal of Technology and Teacher Education, 14*, 29-59.

Tenopyr, M. L. (1977). Content-construct confusion. *Personnel Psychology*, *30*, 47-54.

Todd, S. Y. (2004). A causal model depicting the influence of selected task and employee variables on organizational citizenship behavior. *Dissertation Abstracts International: Section A: Humanities and Social Sciences*, *64*(10), 3757.

Tschannen-Moran, M., Hoy, A. E., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*, 202-248.

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783-805.

Turner, L. A., & Johnson, B. (2003). A model of mastery motivation for at-risk preschoolers. *Journal of Educational Psychology*, *95*(3), 495-505.

Vielva, I., & Iraugi, I. (2001). Cognitive and behavioural factors as predictors of abstinence following treatment for alcohol dependence. *Addiction*, *96*(2), 297-304.

Waddoups, G. L., Wentworth, N., & Earle, R. (2004). Principles of technology integration and curriculum development: A faculty design team approach. *Computers in the Schools, 21*(1-2)*,* 15-23.

Waldman, J. M. (1995). Can task difficulty alter one's perceived level of self-efficacy? *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *56*(1), 0547.

Wentworth, N., Waddoups, G. L., & Earle, R. (2004). Technology integration into a teacher education program. *Computers in the Schools, 21*(1-2)*,* 1-14.

White, C. M. (2004). Coping resources, coping responses, and employees' contextual performance in a merger context. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *65*(1), 471.

Wolfe, D. A., Wekerle, C., Scott, K., Straatman, A., & Grasley, C. (2004). Predicting abuse in adolescent dating relationships over 1 year: The role of child maltreatment and trauma. *Journal of Abnormal Psychology*, *113*(3), 406-415.

Woolfolk, A. E., & Hoy, W. K. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology, 82*, 81-91.

Woolfolk, A. E., Rosoff, B., & Hoy, W. K. (1990). Teacher's sense of efficacy and their beliefs about managing students. *Teaching and Teacher Education, 6*, 137-148.

Young, R. M., Connor, J. P., Ricciardelli, L. A., & Saunders, J. B. (2006). The role of alcohol expectancy and drinking refusal self-efficacy beliefs in University student drinking. *Alcohol and Alcoholism*, *41*(1), 70-75.

Yzer, M. C., & van den Putte, B. (2006). Understanding smoking cessation: The role of smokers' quit history. *Psychology of Addictive Behaviors*, *20*, 356-361.

Appendix A: The Technology Integration Confidence Scale

The Technology Integration Confidence Scale (Version 2)

Instructions: For this survey, you will be asked to rate how confident you are that you can complete certain technology integration tasks on the following scale:

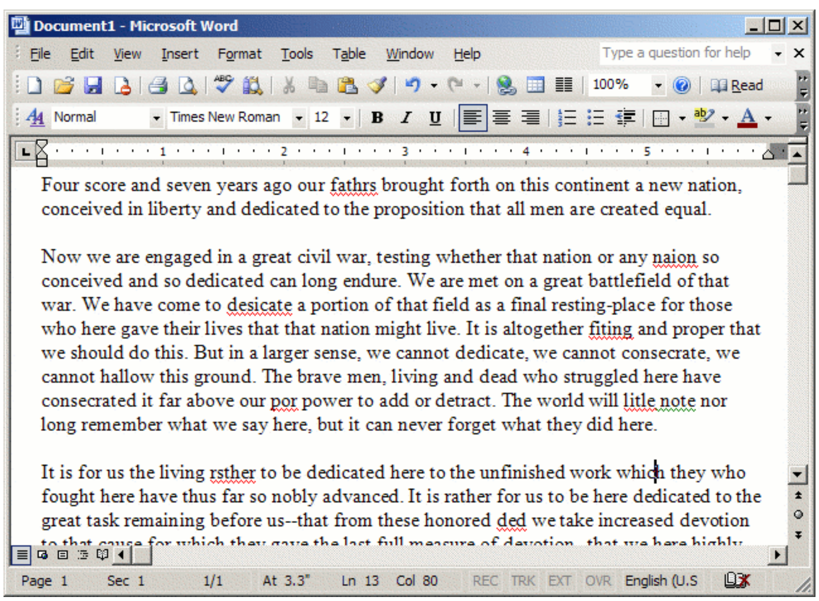
0 - Not confident at all
1 - Slightly confident
2 - Somewhat confident
3 - Fairly confident
4 - Quite confident
5 - Completely confident

Although these items are worded as if you were already teaching, rate your confidence as it is at this moment.

The items are presented in one of two formats. The first format presents an image and an associated task. For example:

Example Item #1:

In the document pictured below, how confident are you that you can find the misspelled words?



__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

Example Item #2:

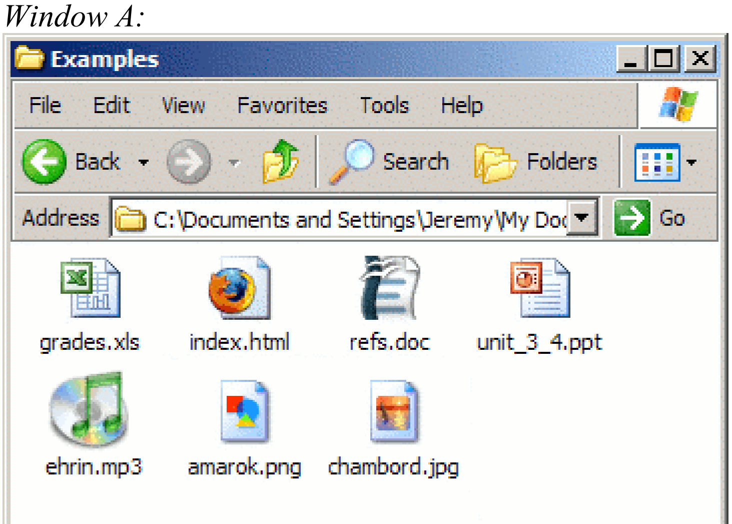
The club you sponsor will be giving a presentation to detail their activities at the next assembly. The assembly hall is equipped with a computer and an LCD projector. How confident are you that you can help the students create an effective presentation using PowerPoint, or another slide show program?

__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

Technology Integration Confidence Scale

Items 1 through 4 refer to this image (Window A). Rate how confident you are at this moment and without any further instruction or practice to accomplish the tasks listed.

Window A:



1. Identify the sound file in Window A
   __ Not confident at all
   __ Slightly confident
   __ Somewhat confident
   __ Fairly confident
   __ Quite confident
   __ Completely confident

2. Identify the graphic/image files in Window A
   __ Not confident at all
   __ Slightly confident
   __ Somewhat confident
   __ Fairly confident
   __ Quite confident
   __ Completely confident

3. Identify the word processing document in Window A
   __ Not confident at all
   __ Slightly confident
   __ Somewhat confident
   __ Fairly confident
   __ Quite confident
   __ Completely confident

4. Open, edit, and save the file named "grades.xls" in Window A
   __ Not confident at all
   __ Slightly confident
   __ Somewhat confident
   __ Fairly confident
   __ Quite confident
   __ Completely confident

5. Delete the file named "refs.doc" in Window A
   __ Not confident at all
   __ Slightly confident
   __ Somewhat confident
   __ Fairly confident
   __ Quite confident
   __ Completely confident

6. Rename the document "index.html" in Window A
   __ Not confident at all
   __ Slightly confident
   __ Somewhat confident
   __ Fairly confident
   __ Quite confident
   __ Completely confident

Read the following situations and rate how confident you are at this moment and without any further instruction or practice to accomplish the tasks they propose.

7. Your district is rolling out a new technology at each school. They invite representatives from each department to an inservice demonstration. How confident are you that you can learn this new technology during the inservice?
   __ Not confident at all
   __ Slightly confident
   __ Somewhat confident
   __ Fairly confident
   __ Quite confident
   __ Completely confident

8. The news has recently featured a new online program that you think may be helpful in your classes. How confident are you that you can learn this new program on your own?
   __ Not confident at all
   __ Slightly confident
   __ Somewhat confident
   __ Fairly confident
   __ Quite confident
   __ Completely confident

9. Unfortunately, your school will not be able to afford a computer lab attendant this year. Instead, each teacher will be assigned two lab hours per week. How confident are you that you can manage your students' time and activities during these lab sessions?
   __ Not confident at all
   __ Slightly confident
   __ Somewhat confident
   __ Fairly confident
   __ Quite confident
   __ Completely confident

10. A member of the PTA feels there is too much technology in the school and states that not all technologies are equally applicable to your classroom, and not all student learning goals are well suited for technology. How confident are you that to you can effectively judge when and how to use technology to support your students' learning?
    __ Not confident at all
    __ Slightly confident
    __ Somewhat confident
    __ Fairly confident
    __ Quite confident
    __ Completely confident

11. Your school assigns each class one computer lab period every two weeks. How confident are you that you can create lesson plans that effectively use the lab time for student learning?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

12. A teacher in another subject has found an article that claims students learn more when they use a certain computer program. How confident are you that you can identify the information in the article that applies to your classes?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

13. An educational software vendor gives a sales pitch to your department. How confident are you that you can evaluate their products for their suitability to your teaching environment?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

14. A vice principle is upset that the new equipment that was donated to the school is not being used. She asks if you can demonstrate how to use it at the next inservice meeting. How confident are you that you can accomplish this task?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

15. Your district has allocated money to purchase educational technology products for your subject/grade. The board has asked for input to help them decide between two competing products. How confident are you that you can advise them on this purchase by evaluating the products for their suitability to your teaching situation?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

16. Your principal promises full support for any technology that can be linked to the state's core curriculum standards. How confident are you that you can find technologies that will help you meet these standards in your subject?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

17. Current educational practice stresses 'higher order' thinking skills such as analysis, synthesis, and evaluation. How confident are you that you can use technology to improve these skills in your students?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

18. Thanks to a grant from the state, your classroom now has three computers, a video camera, and a digital camera. How confident are you that you can integrate some or all of these technologies into your teaching?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

19. Due to a personal emergency, a fellow teacher asks you to teach his computer lab period during your preparation time this afternoon. How confident are you that you can make good use of the class time without the opportunity to plan?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

20. Your students are using the Internet to research a topic. How confident are you that you can provide them with a list of high quality, trustworthy websites to get them started?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

21. The state has created a website where teachers can download test questions that have been written to the state's core curriculum in every subject. How confident are you that you can use these questions to track your students' learning?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

22. Mrs. Jones, an administrator, is acting as a mediator between you and Mr. Smith, a parent who feels a test you gave was unfair. Mrs. Jones asks you to email her evidence supporting your test, which she will review before her meeting with Mr. Smith. How confident are you that you can summarize the necessary information in an electronic format (document, spreadsheet, etc.)?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

23. Your students use computers to complete several assignments during the year. How confident are you that you can grade both the final product of these assignments and the students' use of the technology?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

24. In preparation for a performance review with an administrator, you are asked to critically evaluate several aspects of your teaching, including your use of technology in class. How confident are you that you can accurately do so?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

25. A speaker from the state Department of Education declares that effective teachers are also life-long learners, and that the Internet is a great source of information. How confident are you that you can use the Internet and other technology resources as part of your own lifelong learning?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

26. A group of teachers has lunch together once a month to share lesson plans to use in the computer lab. How confident are you that you can contribute to these discussions, including critiquing the other teachers' ideas?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

27. A resourceful teacher in your area has created a website where teachers can exchange ideas, resources, and lesson plans. How confident are you that you can use this site to improve your teaching?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

28. A senior teacher, known for creatively integrating technology into her teaching, allows you to observe her class. How confident are you that you can judge which of her techniques will be useful in your own class?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

29. The parents of several students have asked to be kept informed of class assignments and activities via regular emails or a class website. How confident are you that you can accommodate this request?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

30. Not all of your students will have equal access to technology at home. How confident are you that you can identify situations where access to technology might be an issue for one or more of your students?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

31. When some of your students do not have access to technology outside the classroom, how confident are you that you can appropriately, legally, and ethically lessen the effects of such unequal access?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

32. Because students are using the Internet and other technologies in school, they must be instructed how to stay safe while getting the most from these resources. How confident are you that you can model and teach safe usage of technology, including Internet safety?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

33. Technology can help students accomplish tasks, good or ill. For example, students can find images of rare historical artifacts, but they can also illegally obtain copyrighted materials online (such as music). Telecommunications technology can bring the world into your classroom, and allows students to text one another exam answers via cell phones. How confident are you that you can model and teach ethical and legal use of technology?
__ Not confident at all
__ Slightly confident
__ Somewhat confident
__ Fairly confident
__ Quite confident
__ Completely confident

Appendix B: TICS Subscales and Their Relation to the NETS-T

TICS subscale	TICS items	NETS-T
		I. Technology Operations and Concepts: Teachers demonstrate a sound understanding of technology operations and concepts.
IA	1-6	A. Teachers demonstrate introductory knowledge, skills, and understanding of concepts related to technology (as described in the ISTE National Education Technology Standards for Students)
IB	7-8	B. Teachers demonstrate continual growth in technology knowledge and skills to stay abreast of current and emerging technologies.
II	9-15	II. Planning and Designing Learning Environments and Experiences: Teachers plan and design effective learning environments and experiences supported by technology.
III	16-20	III. Teaching, Learning, and the Curriculum: Teachers implement curriculum plans, that include methods and strategies for applying technology to maximize student learning.
IV	21-24	IV. Assessment and Evaluation: Teachers apply technology to facilitate a variety of effective assessment and evaluation strategies.
V	25-29	V. Productivity and Professional Practice: Teachers use technology to enhance their productivity and professional practice.
VI	30-33	VI. Social, Ethical, Legal, and Human Issues: Teachers understand the social, ethical, legal, and human issues surrounding the use of technology in PK-12 schools and apply those principles in practice.