Journal of the Association for Information Systems

JAIS

The Evolving Nature of the Computer Self-Efficacy Construct: An Empirical Investigation of Measurement Construction, Validity, Reliability and Stability Over Time.

George M. Marakas
School of Business
University of Kansas
gmarakas@ku.edu

Richard D. Johnson
College of Business Administration
University of South Florida
drj_mis@bellsouth.net

Paul F. Clay
College of Business
Washington State University
pclay@wsu.edu

Abstract:

This paper reports an empirical study intended to provide detailed comparisons amongst and between the varieties of available measures of computer self-efficacy (CSE). Our purpose is to ascertain their relative abilities to isolate the CSE construct from other related constructs and to capture variance in performance attributed to changes in CSE level. In addition, we investigate the importance of ensuring the measure being used is sufficiently aligned with the task domain of interest. Finally, we explore the stability of CSE measures as they relate to the current state of evolution within the computing domain. Marakas, Yi, and Johnson (1998) proposed a framework for the construction of instruments intended to measure the CSE construct that we have adopted as a basis for this series of investigations.

To that end, we advance and test a set of hypotheses derived from the Marakas et al. (1998) framework. Results of the analyses support the need for adherence to the tenets of the proposed framework as well as provide evidence that CSE measures suffer from degradation of their explanatory power over time. Further, this study brings forth the importance of appropriately validating measures of CSE using approaches intended for a formative rather than a reflective construct. These results suggest that the common practices of instrument validation and reuse of long-standing instruments to measure CSE may not be the most effective approach to the study of the construct. Implications for future research are discussed.

Keywords: computer self-efficacy, social cognitive theory, measurement validation, computer performance, training, formative versus reflective constructs

Volume 8, Issue 1, Article 2, pp. 16–46, January 2007

Journal of the Association for Information Systems

Introduction

The computer self efficacy (CSE) construct, logically and theoretically derived from Bandura's (1977a, 1977b, 1986, 1997) broader concept of self-efficacy, is defined as "an individual's perception of efficacy in performing specific computer-related tasks within the domain of general computing" (Marakas et al. 1998, p. 127). More than simply an ability assessment, CSE reflects a dynamic composite of multiple factors, including not only perceived ability, but motivational and adaptation aspects as well (Gist and Mitchell, 1992; Wood and Bandura, 1989). The CSE construct is related to, but conceptually different from, other behavioral constructs commonly found in IS research such as ease of use, computer anxiety, and outcome expectancy. Originally conceptualized at the task-specific level, CSE has recently been hypothesized to be far more complex than previously suggested by earlier studies (cf. Compeau and Higgins, 1995a), and studies have established the construct at both the application-specific level (word processing, spreadsheet, etc.) and at a more general computing level (Bandura, 1997; Marakas et al., 1998).

Several researchers have developed and validated measures of CSE for use in studies that either focus on CSE as the primary construct of interest or as an ancillary construct related to the focus of the research. These measures have been used by a variety of disciplines including education (Brown, Lent, and Larkin, 1989; Delcourt and Kinzie, 1993), healthcare (Henderson, Deane, and Ward, 1995), computer training (Compeau and Higgins, 1995a; Johnson and Marakas, 2000), computer use (Burkhardt and Brass, 1990; Compeau and Higgins, 1995a), technology adoption (Hill, Smith, and Mann, 1987) and computer performance (Gist, Schwoerer, and Rosen, 1989; Webster and Martocchio, 1995), among many others. An even more varietal range of disciplines has investigated the broader, root construct self efficacy as originally conceptualized by Bandura (1977a, 1977b).

Adherence to Theory

One common denominator amongst the array of generally accepted measures of the broader construct has been adherence, either stated or identifiable, to the root theoretical concepts set forth by Bandura (1977a, 1977b) with regard to development of effective measures of the self-efficacy construct. Extending these original tenets, Marakas et al. (1998) proposed a framework for the development of computer self-efficacy measures intended to ensure such instruments effectively isolate the CSE construct while exhibiting sufficient discriminant validity from measures of other related constructs. An overarching principle in this framework is that an effective measure of task-specific CSE must be closely coupled to the task domain under study. In other words, a measure of CSE developed for testing a subject's perception of his or her ability to use a spreadsheet would not be suitable for testing a subject's perceived ability to perform statistical analysis using a spreadsheet. The former focuses on use of the tool and its functions, while the latter incorporates cross-domain skills requiring multiple domain estimations of perceived ability. Marakas et al. (1998) point out that both measures will, most likely, overlap in their measure of the domain of interest, but both the amount of variance explained in the dependent variable and the predictive power of the cross-domain measure will suffer significantly. In short, unless a task-specific measure of CSE is deployed in a task domain highly similar to that in which it was originally developed and validated, the results obtained from the measure will be subject to significant error and, thus, suspect.

This suggests even the most rigorously validated measure of task-specific CSE, when adopted and deployed by unrelated studies investigating the CSE construct across dissimilar domains, may suffer from limited generalizability and applicability. Further, implicit in this perspective, if the intention is to closely isolate the CSE construct for the purpose of explaining the maximum amount of variance in one or more dependent variables, it is likely that a new measure of CSE, constructed to be closely aligned to the task or application under study, may need to be developed from scratch rather than adopted for reuse from a previously published measure.

This treatise of redevelopment is not limited to CSE research nor is it new to self-efficacy research. Vispoel and Chen (1990) stated that no single standardized measure of self-efficacy is appropriate for all studies and advised researchers to develop new, or to significantly revise and revalidate, existing measures for each study. This perspective was again advanced by Marakas et al. (1998) and was clearly reiterated by Bandura (2001) in a paper directly addressing the construction of self-efficacy measures. Yet to date, we consistently see attempts to reuse available CSE instruments without sufficient regard to their applicability to the domain under study (Carlson and Grabowski, 1992; Harrison and Rainer, 1992; Taylor and Todd, 1995; Venkatesh and Davis, 1996).

Measurement of the CSE Construct

To effectively measure self-efficacy, one must understand its multi-dimensional nature (Marakas et al., 1998). The fundamental self-efficacy construct is described as having three dimensions. Magnitude refers to the level of task difficulty that individuals believe they can attain; strength indicates whether the conviction regarding magnitude is strong or weak; and generality describes the degree to which the expectation is, or can be, generalized across situations (Gist, 1987).

Typically, instruments reflecting the multi-dimensional nature of the CSE construct require a subject to respond dichotomously to whether he or she is capable of performing at one or more levels on a specific task. The sum of the positive, or "yes," responses is considered to represent the magnitude of that individual's specific self-efficacy. Each affirmative response collected during magnitude measurement is then rated by the subject on a scale that ranges from either 1 or 10 (quite uncertain) to 100 (quite certain) at intervals of either 1 or 10 points, respectively. The sum of these confidence ratings is used as a measure of CSE strength. The two scores are then correlated with performance measures across subjects. The determination of generality is a function of the design of the instrument itself. Instruments that focus more on the general level of a domain will have greater generality than those that focus on a specific task or application.

Estimations of self-efficacy are formed through a dynamic weighing, integration, and evaluation of complex cognitive, social, and personal performance experiences. Further, it is important to note that self-efficacy involves more than skill assessment. Self-efficacy reflects not only a perception of one's ability to perform a particular task based on past performance or experience but also forms a critical influence on future intentions (Bandura, 1997). Studies across a wide range of research domains have consistently found self-efficacy to be a strong predictor of subsequent task-specific performance.[1]

CSE versus GCSE

Marakas et al. (1998) theorized that perceptions of CSE exist at both the general computing behavior level and at the specific computer task or application level. In contrast to the definition of CSE above, general computer self-efficacy (GCSE) refers to an individual's judgment of his or her ability to perform across multiple computer application domains. GCSE, therefore, is more a product of a lifetime of related experiences and can be thought of as a weighted collection of all CSEs accumulated over time. While not yet formally and empirically tested, this conceptualization tends to more closely conform to the definition of computer self-efficacy that is often proffered and applied in the IS literature (cf. Carlson and Grabowski, 1992; Compeau and Higgins, 1995a; 1995b; Martocchio, 1994). It is important to note, while recent statements by Bandura (2001) and others acknowledge the possibility of a general level of self-efficacy in certain domains, the original SE construct was developed purely at the task-specific domain level, and no empirical evidence clearly establishing the true relationship between CSE and GCSE has yet appeared in the literature.

It has been theorized that, over time and multiple experiences within the general computing domain, a measure of GCSE will become an equally effective, or possibly superior, predictor of future performance within the domain as any appropriately designed task-specific measure of CSE. If found to be tenable, this suggests the ability to design and use a more general measure of CSE in situations where the domain experience level of the subject is considered to be beyond the novice level. Implicit in this is the elimination of the need to develop a new measure of CSE when faced with a new task or application domain, thus easing the burden on the researcher and the potential cost of assessment in an applied context.

While the results of an empirical test of this facet of the theory have not yet appeared in academic literature, the premise has been echoed and supported both conceptually and logically by a wide variety of researchers, including Bandura (2001). It is important to note, however, adoption of a general CSE measure does not supplant the need to measure CSE at the application or task level. If the desired efficacy estimation is focused at the task or application level, and the subject's expected CSE perception and computing experience is not extremely high, it is likely that adoption of a GCSE measure will result in a lower explained variance with regard to predicting task performance or variance in a dependent variable than that of a more targeted task or application-specific measure.

Despite examples of task-specific CSE measures developed with this framework appearing in recent literature (cf. Agarwal, Sambamurthy, and Stair, 2000; Johnson and Marakas, 2000), the majority of studies employing the CSE construct have traditionally relied on the adoption of measures culled from the extant literature and deployed in a variety of research settings and task domains. This practice of employing previously validated and published measures of a construct has been commonly used within the IS research community since its earliest days and, for the most part, has been widely accepted as

[1] Readers are referred to Bandura, A. (1997) *Self-efficacy: The exercise of control*. New York: W. H. Freeman. and Gist, M. E. and T. R. Mitchell (1992) "Self-efficacy: A theoretical analysis of its determinants and malleability," *Academy of Management Review* (17), pp. 183-211., for a more thorough review of the self-efficacy literature.

holding our various streams of research in good stead (cf. Baroudi and Orlikowski, 1988; Davis, 1989; Moore and Benbasat, 1991). The practice has also been identified as an important path toward building a more rigorous research tradition (Keen, 1980).

Boudreau, Gefen, and Straub (2001) point out, however, this treatise of reuse has been interpreted by some to mean that "use of previously validated instruments is a superior practice to revalidating and/or creating new measures for constructs," but they are quick to put forward that "[n]othing could be further from the truth" (pg.12). The reasoning behind this position is that a new and properly validated instrument exposes the construct of interest to a more robust test of the nomological validity of both the old and the new scales. As Campbell (1960) points out, validation works in both directions; it is both "symmetrical and egalitarian" (pg. 548). In the case of CSE estimations, we believe the unique dynamics of the computing domain can significantly affect the nomological validity of any extant instrument. Thus, an existing measure of CSE must be given substantial and careful consideration when weighing the costs and benefits of adopting the existing measure against the development of a measure more targeted to the task or application under study.

CSE/GCSE as Formative Indicator

Recently, much discussion has appeared in the management and marketing literature with regard to model misspecification and measurement validation associated with the differences between formative and reflective constructs (Jarvis, MacKenzie, and Podsakoff, 2003; Diamantopoulos and Winklhofer, 2001; Podsakoff, MacKenzie, Podsakoff, and Lee, 2003). This discussion, and an understanding of its importance to IS measurement development, has been brought to the attention of the IS research community via the work of Straub, Boudreau, and Gefen (2004) and Loch, Straub, and Kamel (2003). Through this literature, we learn that the common practice of using validation techniques appropriate for reflective constructs (those where the latent variable causes the observed variables or item responses) is not appropriate for the validation of measures of formative constructs (those where the observed variables or item responses are the causal formative of a latent variable).

Within the CSE literature, we can easily see that virtually all of the progress made in the development of extant measures of CSE and/or GCSE has been based on classical test theory and the assumption that the variation in the scores on a measure of CSE is a function of the true score of the latent construct — CSE or GCSE. This approach positions the CSE construct as a reflective indicator. Following the work of Diamantopoulos and Winklhofer (2001) and others, we argue the true nature of the CSE construct is that of a formative indicator, and the common reliability and validity practices associated with measurement development and validation are not appropriate. As such, instead of the more traditional approaches to measurement validation such as confirmatory factor analysis and the calculation of reliability indices, we posit a more appropriate validation is achieved via the application of principal components analysis (PCA), structural equation modeling (SEM), and modified multi-trait multi-method (MTMM) approaches developed specifically for the validation of measures of formative constructs.

Goals of the Study

In summary, the goal of this research is to provide detailed comparisons amongst and between the variety of available and commonly adopted measures of CSE to ascertain their relative abilities to isolate the CSE construct and to predict future performance. Additionally, we hope to effectively illustrate both the importance of ensuring the measure being used is sufficiently aligned with the task domain of interest and that the methods used to validate the measure are appropriate for a formative, rather than reflective, construct. Finally, we will explore the degree to which a valid measure of CSE can sustain its predictive abilities over time.

To that end, we advance a set of hypotheses (derived from the Marakas et al. (1998) proposed framework for the construction of CSE measures), review the extant literature regarding formative versus reflective constructs, discuss the time-related stability arguments presented above, and present the results of a series of empirical tests of those hypotheses. Finally, as a function of the analyses conducted, we offer a discussion related to the relative merits of several of the existing measures of CSE at both the task-specific and general domain levels. It is hoped that through such comparative tests, the community of CSE researchers can become better informed about the appropriateness of adopting an existing measure versus developing one specifically for the task or application domain under study.

Challenges in Measuring the CSE Construct

Equivocality of CSE/GCSE Findings

Notwithstanding its rich conceptual foundation, some of the published results obtained in CSE research to date have been equivocal despite the extensive reuse of previously developed measures. A meta-analysis of self-efficacy studies (Multon, Brown, and Lent, 1991) revealed effect sizes of self-efficacy on performance outcomes (the strongest and most commonly found empirical relationship) depend on specific characteristics of the studies, *most notably on the construction of the self-efficacy and performance measures*. The strongest effects were obtained by researchers who compared specific efficacy judgments with basic cognitive skills measures of performance, developed highly concordant self-efficacy/performance indices, and administered them proximate to each other. Of note, researchers obtained significant relationships even with generalized self-efficacy indices, albeit with much less variance explained in the performance variable. The results of the meta-analysis, however, indicated numerous confounded and misleading results when using a generalized measure to isolate changes in self-efficacy levels due to manipulation or when predicting levels of future performance. The authors concluded, however, that if generalized self-efficacy assessments are able to predict performances not tied closely to the particular self-efficacy construct under study, then the relationship between a properly assessed self-efficacy perception and subsequent performance should be even stronger.

Using the Compeau and Higgins (1995a; 1995b) measure of CSE (arguably the single most adopted and reused measure of the construct), Bolt, Killough, and Kuo (2001) found no significant relationship between CSE and performance (a counter-intuitive and counter-theoretic result) but did find a strong relationship between prior performance and CSE (a commonly found relationship with widespread theoretical and empirical support). The authors concluded that their findings may have resulted from the varying task complexity present in their treatments.

In direct contrast to the Bolt et al. (2001) findings, Venkatesh and Davis (1996) (also using the Compeau and Higgins measure) found no change in CSE with experience over time. Here again, we find evidence of both counter-theoretic and equivocal results using the same widely adopted measure.

In a more recent study of CSE, Shapka and Farrari (2003) distinguished between distal and proximal self-efficacy assessments, with the difference between the ability perceptions based upon the timing between the measure and the task. They defined distal self-efficacy assessment as a person's perception of his or her ability to perform a specific task at an unspecified point in time. In contrast, proximal self-efficacy assessments measure a person's perception of his or her ability to perform a specific task immediately prior to performing the task.

The Shapka and Farrari (2003) study, using 56 Canadian pre-service public school teachers and administering the CSE portion of the Computer Attitude Scale from Lloyd and Gressard (1984), produced counter-theoretic results. The findings indicated distal computer self-efficacy was related neither to performance success nor to prior experience — findings that run counter to established self-efficacy theory and numerous empirical studies employing the CSE construct.

In their seminal investigation into the construct, Compeau and Higgins (1995a) trained managers over a two-day period to use word-processing and spreadsheet packages. Contrary to their expectations (and in some cases, counter to theory), their hypotheses did not receive consistent support. For day one spreadsheet training, only five out of the nine hypothesized relations were significant in the hypothesized direction, and the remainder showed either no support or significance in the opposite direction. For day two training, however, nine of 13 hypotheses were significant in spreadsheet training subjects. For word-processing, a slightly different set of five of nine hypothesized relations were supported for day one, and only four of 13 hypotheses were supported on day two. Further, the overall results obtained for the word-processing component of the study did not receive as strong a support as the spreadsheet portion despite the fact that the same measure of CSE was used throughout.

From a logical perspective, several sources for these examples of equivocal or unexpected results can be suggested: 1) the model under investigation may be theoretically misspecified, 2) the subjects used may have a high level of baseline CSE prior to manipulation, 3) the methodology employed may not completely capture the robustness of the phenomena under investigation, and/or 4) the constructs contained within the model may not be adequately isolated from other related constructs in their measurement.

Johnson and Marakas (2000) conducted a study that both replicated and extended the original Compeau and Higgins (1995a) work in an attempt to account for the unexpected results obtained. In this study, we compared the original Compeau and Higgins measure and one developed using the Marakas et al. (1998) CSE instrument development framework. Both measures were able to capture changes in levels of CSE and generally performed well. The two measures, however, displayed differential levels of explained variance in the dependent variable, with the framework developed measure explaining significantly more variance in the performance variable than the original Compeau and Higgins measure. In addition, the newly developed measure was more effective at capturing incremental changes in CSE throughout the experiment and was able to capture the relationship between CSE and enactive mastery (prior experience) to a much greater extent.

The results of this study demonstrated the difference in effectiveness between using a more general CSE measure (i.e. the Compeau and Higgins' measure) and using an application-specific CSE measure with regard to their relative abilities to explain variance in the dependent variable and to capture finer granularity changes in CSE perceptions. It is important to note that both instruments performed acceptably, and the Compeau and Higgins instrument has been shown to be a highly regarded measure of the construct. In this case, the issue lies not with the quality of either instrument, but rather with the appropriate application of an instrument designed to conform to the task context and level of analysis under study. While both measures attempt to capture the same construct, visual inspection of the respective measures suggests basic differences in their approach to CSE assessment. The application-specific measure developed using the Marakas et al. (1998) framework measures CSE as an individual's perceptions of his or her ability to accomplish the tasks and activities associated with the specific application domain under study. In contrast, the Compeau and Higgins instrument assesses CSE as an individual's perceptions of his or her ability to use the specific computer application in the accomplishment of an unspecified task and under varying sets of conditions. The former is clearly focused at the task/application level of analysis, and the latter is focused at a more general level with little or no task or application alignment or specificity.

The important issue in these examples is the value of employing a measure of CSE that is both closely aligned with the task or application under study and absent of any reference or inference to cross-domain skills necessary to complete the task.

CSE as a Formative versus a Reflective Construct

During the course of investigating the CSE construct and, for that matter, other IS-related constructs of interest, a great deal of effort and attention has been devoted to developing multi-item measures displaying sound psychometric properties through appropriate validation approaches. Historically, it has been observed, across multiple domains of interest, that conventional wisdom based on classical test theory is the basis for instrument validation. This approach assumes the items in a scale are perceived as reflective (effect) indicators of the underlying, or latent, construct. In other words, the true nature of the construct, whether it be personality, attitude, or CSE, is what gives rise to that which can be observed — the items contained within the scale or measure. While many constructs within the social sciences can be argued to be reflective in nature, others are, in fact, formative rather than reflective. This distinction necessitates significant alterations to classical test theory when validating measures and requires an assessment of the construct using techniques appropriate for formative constructs (Straub et al., 2004).

To better understand the differences between formative and reflective indicators, Diamantopoulos and Winklhofer (2001) proffer a comparison of the properties of the two types of indicators, which we summarize in Table 1.

Table 1: Comparison of Properties Associated With Formative and Reflective Constructs.

Property	Formative Construct	Reflective Construct
Direction of Causality	Direction of causality is from items to construct Indicators are defining characteristics of the construct Changes in indicators should cause changes in the construct Changes in the construct do not cause changes in the indicators	Direction of causality is from construct to items Indicators are manifestations of the construct Changes in indicators should not cause changes in the construct Changes in the construct do cause changes in the indicators
Interchangeability of indicators/items	Indicators need not be interchangeable Indicators need not have the same or similar content and need not share a common theme Dropping an indicator may alter the conceptual domain of the construct	Indicators should be interchangeable Indicators should have the same or similar content and should share a common theme Dropping an indicator should not alter the conceptual domain of the construct
Covariation Amongst the Indicators	Not necessary for indicators to covary A change in one indicator is not	Indicators are expected to covary A change in one indicator is expected to be associated with changes in the other

	necessarily associated with changes in the other indicators	indicators
Nomological Net	Nomological net for the indicators may differ Indicators are not required to have the same antecedents and consequences	Nomological net for the indicators should not differ Indicators are required to have the same antecedents and consequences

A review of the items on a commonly used spreadsheet CSE scale and on a GCSE scale developed using the Marakas et al. (1998) development framework should serve to illustrate the formative, rather than reflective, nature of the CSE/GCSE construct:

Spreadsheet CSE (Gist, Schwoerer, and Rosen, 1989)

1. I am capable of typing and entering numbers into a cell.
2. I am capable of writing a formula for addition.
3. I am capable of writing a formula for division.
4. …

General Computer Self-Efficacy (Johnson and Marakas, 2000)

1. I believe I have the ability to describe how a computer works.
2. I believe I have the ability to install new software applications on a computer.
3. I believe I have the ability to identify and correct common operational problems with a computer.
4. …

The complete versions of these instruments are included in the appendix to this manuscript.

Recall that indicators of a formative construct need not be correlated, covary, or be interchangeable. A review of the items in each scale indicates conformity to these properties, thus suggesting both CSE and GCSE are formative indicators. It is possible, even likely in many cases, a person responding to the instrument might be capable of installing software, but not capable of accurately describing how a computer works. Or, a person might be capable of entering numbers into a cell, but not be capable of writing a formula. Given this, and as suggested by the literature addressing the validation of measures of formative constructs (c.f. Straub et al., 2004; Diamantopoulos and Winklhofer, 2001; Jarvis et al., 2003), we argue that validation of CSE and GCSE instruments must use techniques appropriate for formative constructs rather than the commonly adopted techniques associated with reflective constructs.[2]

While, to date, no formal review of the IS literature has been conducted to ascertain the extent to which appropriate validation techniques have been used with formative indicators, a review by Podsakoff et al. (2003) in the context of the leadership literature revealed that out of 138 studies, 65 presumably had formative indicators, and none used techniques appropriate for formative indicators, either at exploratory or confirmatory stages. We suspect a similar situation to exist with the development and validation of CSE and GCSE measures in the IS literature. As such, we offer guidance in the application of the appropriate validation and assessment techniques in the course of describing the development of the CSE and GCSE measures used in this study. This guidance will be discussed in greater detail in the next section of the paper.

Construct Isolation and Stability of Measurement Over Time

In addition to the necessity for task or application alignment with the CSE measure to constrain possible reuse of existing measures, the results of recent studies, such as those cited in the examples above, indicate that stability of previously developed measures of CSE may deteriorate over time, thus rendering them less effective. This situation stands in direct contradiction to the accepted and encouraged practice regarding instrument reuse. Intuitively, one explanation for this potential lies with the nature of the domain in which these measures are being used — that of computing and computer-related technologies. It is widely accepted that this domain possesses an evolutionary momentum far greater than most other domains commonly explored in the social and behavioral sciences. Further, an individual can neither fully develop the skills necessary to use a given technology, nor develop a reliable and informed task-specific self-efficacy assessment about

[2] We gratefully acknowledge the input provided by the senior editor and one of the anonymous reviewers with regard to understanding this important issue.

it, until the technology actually exists. As such, the commonly rapid introduction of new computing technologies presents a significant challenge to individual users of new technologies, as well as to the study of CSE.

To further exacerbate these challenges, the differences between the skills necessary to use two distinct applications — such as a spreadsheet and a word processor — are often considerable and have significantly changed since their early conceptualizations. It then seems reasonable the measure for each should distinctly differ and, over time, as the skills required become different, either change to reflect those differences or logically be rendered less effective — if not ineffective. Because of this, it is possible that the commonly held practice and belief of reusing existing measures of a construct may not be as readily applicable to CSE as it is to other constructs that exist in more stable social and behavior domains. In short, we are trying to measure an evolutionary construct using a static method. Evolving constructs need evolving instrumentation, and such instrumentation should be a clear reflection of the evolution of the context.

An additional issue that may further contribute to the potential for error in CSE measurement can be found with the practices associated with the development and adoption of existing CSE measures. Despite widespread use throughout the IS research community, the bulk of the commonly adopted CSE measures have not been rigorously compared to each other or to measures of closely related constructs as recommended by Bandura (2001):

"…Perceived self-efficacy should … be distinguished from other constructs such as self-esteem, locus of control, and outcome expectancies" (pg. 7).

Finally, evidence suggests that when existing CSE measures are used, they are often dramatically altered from their original form in an attempt to make them more relevant and conforming to changes in the computing domain. While this may appear to be a form of update and revalidation, the common alteration is that of removing irrelevant items from an old measure without any consideration to adding new items that reflect relevant, evolutionary changes in the domain or to revalidation using approaches appropriate for formative indicators as discussed above. A recent study by Rainer, Laosethakul, and Astone (2003) serves as an example of this practice:

"Computer self-efficacy was operationalized using the Computer Self-Efficacy Scale developed by Murphy et al. (1989). This 32-item scale measures perceptions of computer self-efficacy on 5-point Likert scales…Previous analysis of the CSE (scale) demonstrated three underlying latent constructs: beginning computer skills, more conceptual computer skills, and mainframe computer skills. The mainframe computer skills construct was not included in these two studies because the pilot tests indicated that students did not use mainframe computers" (pg. 109). …"The exploratory factor analyses indicate that the CSE exhibits acceptable construct validity. With Cronbach alphas greater than .80, the two constructs of the CSE also demonstrate adequate reliability…" (pg. 110).

Similar alterations of the Murphy, Coover, and Owen (1989) scale (another widely respected and adopted measure of CSE) have been found in CSE studies (cf. Torkzadeh and Koufteros, 1994). A quick review of several of the *advanced skills* items contained in the Murphy et al. (1989) scale suggests many of the remaining items are highly cross-domain in nature and no longer reflect a valid measure of what would be considered to be advanced computing skills in today's increasingly networked environment (see Table 2).

Table 2: Advanced Skills Items in Murphy et al., (1989) Measure of CSE
• Understanding terms/words relating to computer hardware
• Understanding terms/words relating to computer software
• Describing the function of computer hardware (keyboard, monitor, disk drives, computer processing unit)
• Understanding the three stages of data processing: input, processing, output
• Learning to use a variety of programs (software)
• Troubleshooting computer problems
• Writing simple programs for the computer
• Getting help for problems in the computer system
• Using the computer to organize information
• Using the user's guide when help is needed

This example is in no way intended to suggest the Murphy et al. (1989) scale is, or was, flawed with regard to its measure of the construct of interest. In fact, quite the contrary. Rather, it serves as an example of how much the computing domain has evolved since its original development less than two decades ago. The high reliability measure of the scale (as reported in the Rainer et al. 2003 example above) simply indicates the extent to which the respondent can answer the same or approximately the same questions the same way each time (Cronbach, 1951). It does not, in any way, indicate the items are validly measuring any degree of CSE estimations. When the Murphy et al. (1989) scale was first developed, the items accurately reflected the current state of the domain. Over time, however, the domain changed dramatically while the measure remained static (except for the commonly identified need to remove mainframe skills). In other words, despite its acceptable reliability coefficients and its long history of adoption within the literature, its validity with regard to accurately measuring the construct of interest has been diminished over time as a result of the substantial changes within the domain of interest. This suggests the possibility that the effectiveness of the instrument may also have become diminished over time, thus rendering it a less desirable choice for measuring the CSE construct. In fact, a closer inspection of its items suggests it may actually be better described as a GCSE, rather than a CSE, measure.

Given this, the CSE construct appears to be an exception to our common practice of initial development, validation, and reuse of widely published measures. By revisiting the foundational theory behind the construct, the merits of this proposition can be more easily seen. Bandura's position regarding this proposition is made clear in a recent paper focusing on efficacy scale development:

> "There is no all-purpose measure of perceived self-efficacy. The *"one-measure-fits-all"* approach usually has limited explanatory and predictive value because most of the items in an all-purpose measure may have little or no relevance to the selected domain of functioning. …scales of perceived self-efficacy must be tailored to the particular domains of functioning that are the object of interest" (Bandura, 2001, p. 1).

The desire to adopt existing measures of CSE is understandable, given that both significant time and effort necessary to develop new measures weigh heavily against the convenience of using an existing one. In addition, there exists a clear lack of incentive to spend precious time to develop one tailored to the research at hand. To wit, one need only review the editorial policies of our premier research journals to realize that the development and validation of a new measure is considered simply to be the product of good practice and is no longer viewed, in and of itself, as a significant contribution to our body of knowledge worthy of scarce journal space. While we wholeheartedly agree with this editorial policy in both spirit and substance, this position nonetheless provides additional disincentive to develop a new measure of CSE tailored to the specific task or application domain of interest, thus further motivating the adoption and reuse of existing measures.

Scale Development

In preparation for this study, we developed measures of CSE, in conformance with the CSE instrument development framework proposed by Marakas et al. (1998) (discussed in detail below), across several common application domains (Windows, word processing, database, and Internet) as well as the general computing domain (GCSE). For each new scale, the instrument development process focused adherence to the CSE framework, paralleled the steps outlined by Straub (1989) and Straub et al. (2004), and focused specifically on the guidelines for the development of formative constructs (cf. Diamantopoulos and Winkelhofer, 2001; Jarvis et al., 2003).

For formative indicators, content validity is considered to be the most important aspect of instrument development (Rossiter, 2002; Jarvis et al., 2003; Diamantopoulos and Winkelhofer, 2001). Given the importance of content validity to the development of formative scales, we make special note of this process. Unlike reflective indicators, where the goal is to randomly select items from the universe of potential items representing the construct (Cronbach, 1951; Straub, 1989), items for formative indicators should be drawn such that the entire scope of the variable as described by the construct is represented (Diamantopoulos and Winkelhofer, 2001; Jarvis et al. 2003). Although a single item may be removed from a set of reflective indicators without materially affecting the quality of the measure, the removal of an item from the measurement of a formative construct may actually serve to alter the meaning of the construct. This fact, alone, suggests support for the contention that reuse of CSE instruments may be problematic.

First, we three investigators independently developed a comprehensive list of tasks unique to each software application (i.e. saving a file, communicating numeric information, etc.) and then aggregated the list items to remove duplicates. Second, we discussed each item to ensure that it was consistent with the CSE framework. Those items determined not consistent with the framework, were either eliminated or rewritten to conform to the framework. Third, consistent with the call of Boudreau et al. (2001), we had several content experts external to the principal research team review the items and provide further suggestions for refinement.

Pilot testing of the scales also consisted of several stages. In stage one, students from an introductory IS skills course (not included as subjects in the main study) were given the scales and encouraged to provide feedback on the items. In the second stage, we analyzed the initial scale items to assess their empirical fit with the theoretical constructs using techniques appropriate for formative constructs, and we removed items that poorly represented the construct. We repeated this process through two additional instrument administrations with students in an introductory IS skills course. This resulted in a final set of scales with 7 – 10 items for each construct.

Validity Assessment

Scales were assessed and, later, hypotheses tested using partial least squares (PLSGraph 3.00). Traditionally, scales with reflective indicators are assessed as to their construct validity (including discriminant, convergent, nomological, etc.) and reliability (Straub, 1989; Straub et al., 2004; Bagozzi, 1980; Boudreau et al., 2001) using commonly accepted techniques such as factor analysis and Cronbach's alpha, among others. In contrast, however, because formative indicators do not need to co-vary with each other, a conventional investigation of reliability becomes both inappropriate and moot (Jarvis et al., 2003). In addition, it is difficult to test for construct validity for formative indicators because no pattern of covariance is necessary (Bollen and Lennox, 1991), leading some to argue that construct validity for formative indicators is also inappropriate, and the establishment of content validity is sufficient (Rossiter, 2002). Others, however, have argued that although traditional metrics for determining construct validity are less important for formative indicators, construct validity can still be investigated (Petter et al., 2006; Loch et al., 2003).

Construct Validity

To establish construct validity, we followed a process originally employed by Loch et al. (2003). The first step in this process was to assess the relative contribution of each item to the overall construct via the significance of the weight to the overall construct using a simple t-test. When this was completed, we found that several items for each construct had relatively low weights and others displayed weights that significantly contributed to the construct (i.e. had a significant t-test value). Diamantopoulos and Winkelhoffer (2001) argue that the appropriate way to address this issue is to refine the scales by eliminating items until all items significantly contribute to the construct, while retaining complete representation for the construct. Upon review of the scale items and their weights, we found two issues. First, some items were very similar in phrasing and represented the same aspect of the construct (a plus for reflective indicators, but not for formative indicators). Second, when scale items are highly correlated, it becomes "difficult to separate the distinct influence of the individual (items) on the latent variable" (Diamantopoulos and Winkelhofer, 2001, p. 272). Care must be taken not to remove items that materially represent the construct, though, because if content validity exists and the items are representative of the construct, "we should face dire consequences by removing any one of them," (Bollen and Lennox, 1991, p. 308) since such removal may serve to change the nature of the construct.

This issue becomes especially relevant to computer self-efficacy because, although there is no theoretical reason why one skill co-varies with another across all participants, given that most individuals learn a group of similar skills and learn these together, it is likely that these classes of skills co-vary with each other. For example, consider two statements from database CSE: "I believe I can create a query using a database program," and "I believe I have the ability to create a database table using a database program." Although we freely admit that there are multiple ways to create a table in the realm of database manipulation, formation of a set of queries is often used to do this. As such, it is logical that similar skills can underlie both statements. Given this, in situations where removal of an item was indicated, we consistently retained the broader statement (creating a database table). Overall, the final items retained represented a combination of those that contributed strongly to the construct, as well as those that were believed to be theoretically necessary for construct completeness (see Table 3).

Table 3: Weights for Retained CSE Items

Construct	Weight	Std Error	t
Word Processing			
… move a block of text using a word processor	.381	.088	4.21***
… manipulate the way a paragraph looks using a word processor.	.183	.089	2.16***
…add a footnote to a document using a word processor.	.248	.075	3.28**
… merge information from two documents using a word processor.	.361	.084	4.41***
Internet			
…download the information from another computer to my computer using the Internet.	.264	.094	2.81**
…connect to another computer using the Internet.	.180	.076	2.48*
…transfer files from my computer to another computer using the Internet.	.272	.086	3.16**
…locate information on another computer using the Internet.	.230	.072	3.15**
…subscribe to a newsgroup.	.281	.082	3.47***
Spreadsheet			
…manipulate the way a number appears in a spreadsheet.	.405	.172	2.66**

…use and understand the cell references in a spreadsheet.	.252	.171	0.80
…use a spreadsheet to communicate numeric information to others.	.231	.150	1.72
…write a simple formula in a spreadsheet to perform mathematical calculations.	.006	.158	0.36
…use a spreadsheet to display numbers as graphs.	.230	.120	2.07*
Database			
…specify a primary key using a database program.	.216	.151	1.65
…create a database table using a database program	.258	.184	1.43
…add or delete a specific record from a database using a database program.	.468	.151	2.98*
…understand a query written in a database program.	.149	.206	0.72
General CSE			
…describe how a computer works.	.143	.054	2.79**
…install new software applications on a computer.	.163	.066	2.50*
…identify and correct common operational problems with a computer.	.251	.064	3.94***
…unpack and set up a new computer.	.211	.054	3.88***
…remove information from a computer that I no longer need.	.151	.073	2.17*
…use a computer to display or present information in a desired manner.	.339	.075	4.40***
Windows CSE			
…group programs together using Windows.	.328	.046	7.07***
…create an icon for a program.	.150	.049	3.16**
…arrange icons so that I can conveniently access them.	.103	.049	2.05*
…delete a file that I do not need using Windows.	.371	.056	6.71***
…change monitor settings using Windows.	.272	.051	5.38***

NOTES: $p <= 0.05$ ** $p <= 0.01$ *** $p <= 0.001$

All items begin with the phrase "I believe I have the ability to…"

Convergent and Discriminant Validity

Establishing discriminant validity for formative constructs is somewhat more difficult than for reflective constructs, because formative indicators need not co-vary. To address this, Loch et al. (2003) detailed an approach advanced by Trochim, (2001) for determining the discriminant validity of formative constructs using a modified multitrait–multimethod (MTMM) approach (Campbell and Fiske, 1959). In this method, a weighted score is created for each formative item (termed "indicator" in PLS) by multiplying the raw score by its associated weight obtained in PLS (as shown in Table 3 above). We then created a composite score for each formative construct. From this, we created a correlation matrix, wherein items for the formative constructs are compared to other related constructs.

For convergent validity, items should be highly correlated with other items measuring the same construct, understanding that there may be some violations of this heuristic, especially for those items that do not contribute heavily to the formation of the construct (Petter et al., 2006). For discriminant validity, however, all items should be more highly correlated with items within the construct than with items from other constructs. Finally, the number of violations of the comparison parameters must be low and limited to chance (i.e. < 5.00 %) (Campbell and Fiske, 1959). As such, whenever a matrix is presented in this paper, the legend of that matrix will indicate the relative presence or absence of violations. To make these assessments, the new spreadsheet and general measures were assessed with respect to theoretically related constructs: computer anxiety and outcome expectancy.[3]

As can be seen in Table 4, the individual items for both the CSE and GCSE scales were more highly correlated with their own construct than with either outcome expectancy or anxiety, providing evidence of both discriminant and convergent validity.

[3] Assessing the discriminant validity of the remaining new scales is done as part of the testing for H3.

Table 4: Convergent and Discriminant Validity Analysis for Spreadsheet CSE, GCSE, Outcome Expectancy, and Computer Anxiety

Measurement Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
SPD1	1.00														
SPD2	.63	1.00													
SPD3	.66	.83	1.00												
SPD4	.59	.66	.74	1.00											
SPD5	.56	.78	.80	.75	1.00										
GCSE1	.42	.37	.36	.30	.28	1.00									
GCSE2	.34	.51	.50	.51	.45	.45	1.00								
GCSE3	.04	.31	.31	.42	.36	.16	.58	1.00							
GCSE4	.38	.46	.45	.42	.37	.31	.49	.35	1.00						
GCSE5	.09	.36	.34	.46	.45	.07	.44	.65	.23	1.00					
GCSE6	.44	.58	.58	.59	.55	.29	.57	.55	.63	.46	1.00				
Spreadsheet CSE	.78	.90	.93	.86	.89	.38	.53	.34	.48	.40	.63	1.00			
GCSE	.40	.61	.60	.64	.58	.51	.82	.74	.73	.64	.85	.64	1.00		
Computer Anxiety	-.30	-.37	-.36	-.40	-.30	-.33	-.42	-.40	-.50	-.29	-.49	-.40	-.57	1.00	
Outcome Expectancy	.03	.05	.03	.11	.11	.06	.24	.14	.07	.26	.18	.08	.22	-.04	1.00

NOTES: SPD: new spreadsheet, GCSE: new general computer self efficacy, OE: outcome expectancy, A: computer anxiety, Violations < 5.00%

Investigating the Nomological Net

The next step in the process was to confirm that the scales fit well within the broader CSE – GCSE nomological net. To do this, we tested a structural model focusing on the relationship between the application-specific CSE scales and the GCSE scale by examining path coefficients. Bootstrapping (100 subsamples) was performed to determine the statistical significance of each path coefficient. Results indicated that the theorized relationships between an individual's CSE estimations and his or her GCSE were supported (Table 5), with CSE explaining over 60% of the variance in GCSE (R^2 = .65). Results further indicated that word processing, Internet, and spreadsheet efficacies affected GCSE estimations, but database CSE did not. This is most likely due to the fact that at the time of the administration, subjects had not been exposed to the database component of the course, but they had been exposed to the other course software.

Table 5: Path Analysis Results for CSE – GCSE Relationship

Construct	Beta	t	P-value
Word Processing	.306	5.49	.000
Internet	.421	8.13	.000
Spreadsheet	.199	3.41	.000
Database	.060	1.41	.159

Hypotheses

As a guide to the development of effective measures that accurately reflect these different levels of efficacy, Marakas et al. (1998) offered a simple framework for the development of CSE instruments (see Table 6). The components of the framework closely follow the tenets of the original theory as proposed by Bandura and are rooted within the theoretical and empirical literature related to the construct.

Table 6: Marakas et al. (1998) Framework for the Construction of CSE Measuring Instruments
1. All questions must **focus on the subject's perceived ability to perform a specific task** without regard to outcome expectations or derived benefits. 2. All questions must elicit estimations of ability within a **task-specific rather than a general context.** 3. Specific questions must **avoid** ability assessments that include **cross-domain or general-domain skills**. 4. The **level of analysis (LOA)** of the requested estimation of perceived ability **must agree** with the level of analysis of the task and subsequent performance measure. 5. The ordering of questions must **avoid inappropriate or unnecessary anchoring** with regard to perceived rather than actual increasing levels of task difficulty or complexity.

Framework Component #1: Focus on Subject's Perceived Ability

The definition of CSE as a perception of an individual's ability to perform a specific task suggests that any measurement must be constructed in terms of judgments by the subject regarding his or her ability to perform the required task without regard to any related benefits or outcomes resulting from the performance (Bandura, 1997).

"An efficacy expectation is a judgment of one's ability to execute a certain behavior pattern, whereas an outcome expectation is a judgment of the likely consequences such behavior will produce. The expectation that one can jump six feet is an efficacy judgment; the social recognition, applause, trophies, and self-satisfactions anticipated for such a performance constitute the outcome judgments" (Bandura, 1978, p. 240).

In other words, estimations of ability to perform must be isolated from expectancies of the potential outcomes (successful completion of a task or job) or rewards associated with performing well (fame, recognition, promotion, etc.). In addition, by developing a scale that contains aspects of outcome expectancy, it will be positioned inappropriately as an antecedent of performance. Bandura argued that, "when differences in efficacy beliefs are controlled, the outcomes expected for given performances make little or no independent contribution to prediction of behavior" (Bandura, 1997, p. 24). In addition, Johnson and Marakas (2000), arguing from the perspective of Bandura (1997), found that outcome expectancies are best viewed as a consequence of performance. Thus, the important predictor of performance is self-efficacy, not outcome expectancies. The risk is that when the constructs overlap in their measurement, the combined scales will not as effectively capture the relationships between the constructs. From this, our first hypotheses are derived:

H1a: Questions that focus on the subject's perceived ability to perform a specific task without regard to outcome expectations or derived benefits will be more effective in isolating the CSE construct than those that do not.

H1b: Questions that focus on the subject's perceived ability to perform a specific task without regard to outcome expectations or derived benefits will be more effective in predicting performance than those that do not.

Framework Components #2 and #3: Task Specific Context, Cross-Domain Skills, and Level of Analysis

Recent research into the CSE construct reflects a clear move toward the identification and measurement of both a specific and a general level of computer efficacy. Nonetheless, the degree of specificity of the ability estimation must be driven by the specificity of the task.

"…self-efficacy scales must be tailored to activity domains and assess the multifaceted ways in which efficacy beliefs operate within the selected activity domain. The efficacy scales must be linked to factors that, in fact, regulate functioning in the selected domain" (Bandura, 2001, p. 3).

If a subject is asked to estimate his or her ability to perform a skill that can be applied in a variety of task situations within the knowledge domain of computer use, such estimation of CSE will necessarily be formulated more at the general domain level than at the task-specific level. Further, if a subject is asked to estimate his or her ability regarding a computer-related task that requires significant skills from outside the computing domain, or that suggests multiple contexts such as outside help or a variety of context conditions, the isolation of the CSE construct becomes impaired. The outcome of this lack of parallelism will be a weakening in the observed relationship between CSE and performance as well as a reduction in the predictability of future task-specific performance based on prior measures of CSE. Instead, when multiple and cross domain skills are necessarily a part of the skill set of interest, multiple measures of efficacy will be required, each of which must be designed to isolate the perception of skills for each specific domain and context. To further investigate these issues, two additional hypotheses are offered:

H2: Questions that elicit estimations of ability within a task-specific rather than a general context will be better predictors of performance within that task-specific domain than those that do not.

H3: Questions that avoid ability assessments that include cross-domain or general-domain skills will more effectively isolate the CSE construct than those that do not.

Framework Component #5: Avoiding Inappropriate or Unnecessary Anchoring[4]

Bandura (1997) points out while every set of items relating to the measurement of SE must begin somewhere, the preferred format is one that minimizes any anchoring influence. The items should, therefore, be ordered randomly such that no inappropriate inference regarding increasing (or decreasing) task complexity is present. Ideally, several sequences of the items should be tested during development and validation for the presence of order or anchoring bias. Berry, West, and Dennehey (1989) found that ordering questions in descending order of implied complexity produced higher SE estimations than either ascending or random ordering. In addition, the work of Cervone and Peake (1986) demonstrated the ease with which SE estimations can be manipulated by simply manipulating the initial anchor values for estimations conducted by subjects. As such, we will test the following hypothesis to investigate the effect of anchoring:

H4: Instruments where the ordering of questions avoids inappropriate or unnecessary anchoring with regard to perceived or actual increasing levels of task difficulty or complexity will better isolate the CSE construct than those that do not.

Degradation of Measures over Time

Though not included in the original Marakas et al. (1998) framework, the potential for degradation of measures of CSE over time stands as an important component of CSE measure development, adoption, and reuse. Given the relative volatility of the domain in which CSE perceptions and estimations are formed, the following hypothesis will be tested to better inform our understanding of this issue:

H5: As domain evolution results in new technologies and new applications, as well as new functionalities to existing technologies and applications, newer CSE measures developed to reflect these evolutionary changes will be better predictors of performance than older CSE measures developed prior to such domain evolution.

current CSE/GCSE Instruments

Table 7 contains a summary listing of the existing and commonly adopted measures of the CSE construct. An exhaustive review and comparative analysis of this entire set of existing measures is clearly beyond the scope of this study. As such, our evaluation of the effectiveness of the current measures of CSE in this study was limited to those that are both widely regarded and commonly adopted by researchers investigating CSE. Based on a thorough review of the literature, we identified a set of 13 measures of CSE and GCSE for comparison.

Table 7: Summary of Existing Measures of CSE and GCSE

CSE	GCSE
Developed By	Developed By
Hill, Smith, and Mann (1985)	Hill, Smith, and Mann (1987)
Gist et al. (1989)	Miura (1987)
Martocchio and Webster (1992)	Jorde-Bloom and Ford (1988)
Delcourt and Kinzie (1993)	Gist et al. (1989)
Mitchell et al. (1994)	Murphy, Coover, and Owen (1989)
Smith (1994)	Burkhardt and Brass (1990)
Busch (1995)	Martocchio (1992)
Compeau and Higgins (1995a; 1995b)	Russon, Josefowitz, and Edmonds (1994)
Christoph et al. (1998)	Compeau and Higgins (1995b)
Johnson and Marakas (2000)	Henry and Stone (1997)
Kuo and Hsu (2001)	Young (2001)

[4] Framework component #4 refers to a focus on level of analysis. Since this study used the individual level of analysis for all scales, the component was not tested herein.

Methodology

Framework-Conforming Instrument Development Process

Of the measures chosen for this study and outlined in the previous section, only the measures from Johnson and Marakas (2000), and those developed for this study, were found to be constructed in strict conformance to the proposed CSE framework described above.[5] To explore the framework across a variety of task-specific domains, we developed measures of CSE, in conformance with the proposed framework, across several common application domains (Windows, word processing, database, and Internet) as well as the general computing domain (GCSE).

As discussed previously, and to every extent possible, our instrument development process paralleled the steps outlined by Straub (1989) and Straub et al. (2004), with initial stages focusing on the development of content valid items and the later stages focusing on ensuring both discriminant and convergent validity using techniques appropriate for validating instruments measuring formative constructs.

Subjects

A total of 476 subjects from multiple sections of an introductory software skills course conducted at three U.S. universities participated in the research study (along with 57 graduate students randomly selected from one of the test sites).[6] The purpose of the course was to provide introductory software training on using the Windows operating system, the Internet (email and web skills), and applications contained within the Microsoft® Office suite (Excel, Word, and Access). The subjects in this research were judged to be moderately computer literate, with over 60% of the subjects indicating their experience with computers ranging from “fair” to “a lot”. The average age of the sample was 21.8 years (sd = 3.44), with a range of 18 to 44 years. The sample was slightly gender imbalanced, with males representing 58% of the sample. All aspects of the study were conducted identically at each study site using multiple trained facilitators who were unaware of the hypotheses under test. Post-test analysis revealed no significant differences with regard to demographics or descriptives across either study sites or facilitators. To ensure against variations associated with the individual sites, hypotheses were tested for each of the sites. Given that no significant differences were found in either this regard or with regard to the demographics and descriptives, it was deemed appropriate to pool the responses across the three sites for the final analyses.

Measures

We used multiple scales in this study. The scales included both new CSE scales developed for this study, existing CSE scales, and scales shown to be related, but distinct, to CSE. We selected existing CSE scales based upon their wide adoption in the literature. Table 8 contains a complete list of the scales used in this study, and the Appendix contains the final items for the newly developed scalesfor the study.

Table 8: Scales Utilized in Study

Construct	Source	Items
Computer Anxiety	Heinssen, Glass, and Knight (1987)	4
Outcome Expectancy	Compeau and Higgins (1995a)	8
Spreadsheet CSE	Martocchio and Webster (1992) Compeau and Higgins (1995) Johnson and Marakas (2000)	6 10 9
GCSE	Martocchio (1992) Gist et al. (1989) Compeau and Higgins (1995a)	6 30 10
GCSE Word-processing Database Windows Internet	Developed for this study	7 7 10 10 9
Technology skills	Russon, Josefowitz, and Edmonds (1994)	13

[5] The items for each of the new instruments developed using the proposed frameworks (as well as those which were not retained after the analysis and validation) are listed in the appendix to this document.
[6] The subject pool of 476 subjects was used to test the hypotheses 1, 2, 3, and 5 while the 57 graduate subjects were used to test hypothesis 4 only.

Performance Task	Johnson and Marakas (2000) Yi and Davis (2003)	---

Administration of the Instruments

Given the rather lengthy nature of the various instruments, administration occurred in three consecutive phases over three class periods to minimize subject fatigue. The first administration contained the measures of CSE and GCSE constructed as described above (Windows, word-processor, database, spreadsheet, Internet CSE, and GCSE). The second administration contained the new spreadsheet and GCSE instruments, computer anxiety, outcome expectancy, and existing GCSE instruments (Compeau and Higgins, 1995a; Gist et al., 1989; Martocchio, 1992; Murphy et al., 1989). The third administration captured the new spreadsheet and GCSE instruments and existing spreadsheet measures (Gist et al., 1989; Compeau and Higgins, 1995a; Martocchio and Webster, 1992). All items were randomized in the administration of all instruments, and all subjects had the same exposure to the applications under study at the point at which the questionnaires were completed.

Analysis and Results

Hypothesis Testing

H1a: Questions that focus on the subject's perceived ability to perform a specific task without regard to outcome expectations or derived benefits will be more effective in isolating the CSE construct than those that do not.

Given the results suggesting the new scales are strong with respect to content, convergent, discriminant, and nomological validity, we proceeded with the investigation of the hypotheses. To investigate H1a, we utilized the modified MTMM described previously to assess whether or not the constructs were isolating CSE equally. In this analysis, we compared two widely adopted CSE scales (Compeau and Higgins, 1995a; Martocchio and Webster, 1992), the new spreadsheet instrument, (Johnson and Marakas, 2000), and outcome expectancy (Compeau and Higgins, 1995a). The instrument developed by Martocchio and Webster (1992) is targeted toward specific use of the software of interest (i.e. "I expect to become very proficient in the *use of Excel*"), whereas the instrument developed by Compeau and Higgins is more targeted toward an unspecified task requiring the use of a software package (i.e., "I could complete the job using the software package if I could call someone for help if I got stuck"). For the latter instrument, the focus is more heavily on using the software to complete an unspecified task in a variety of support environments. This framing may serve to invoke an expectation of outcome (…complete the job…) and to require cross domain skills (unless the "job" is specifically focused on use of the software with no other domain skills required) and, thus, to require cross-domain efficacy estimations. As a result, the Compeau and Higgins instrument may be more likely to capture levels of CSE at a more general rather than task-specific level (this will be investigated further in the next section). Also, recall that the Johnson and Marakas (2000) spreadsheet measure was designed to isolate only those skills unique to spreadsheets with no cross-application skills present. Table 9 contains the results of this analysis.

The results of this analysis show that all of the items loaded strongly on a CSE composite score (e.g. M-W CSE) and also exhibited strong discriminant validity from the outcome expectancy measure (Table 9). In addition, in the presence of the outcome expectancy scale, the items from the CSE scales did not discriminate among the new spreadsheet measure, Compeau and Higgins' measure, and the Martocchio and Webster measure, suggesting that each was capturing CSE and not outcome expectancy. We believe this provides support for H1a.

H1b: Questions that focus on the subject's perceived ability to perform a specific task without regard to outcome expectations or derived benefits will be more effective in predicting performance than those that do not.

Given the results of H1a, it was necessary to create a new scale that purposively "overlapped" CSE and outcome expectancy into each item. We developed this scale specifically to illustrate the importance of isolating the CSE construct from expectations of outcomes and modeled it after Martocchio and Webster's (1992) spreadsheet measure[7] and Compeau and Higgins' (1995a) measure of outcome expectancy. We developed the following questions for this intentionally "mixed" CSE measure:

1. I believe that my Excel skills will increase my job effectiveness.
2. I am confident that I can use Excel to make my job easier.

[7] We chose to modify the Martocchio and Webster measure for convenience purposes only and that we are in no way suggesting that their measure suffers from these characteristics.

3. I am confident that I can learn Excel well enough to perform my job better.
4. I can learn to use Excel to increase my job effectiveness.

As can be seen, these items are clearly focused on the outcomes associated with using Excel rather than the successful use of Excel, but could be construed by a researcher to capture CSE, since they are similar in construction to other extant measures of the construct.

Table 9: Comparison of Spreadsheet CSE Instruments

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1. CSE1	1.00																								
2. CSE2	-.61	1.00																							
3. CSE3	.68	-.63	1.00																						
4. CSE4	-.64	.59	-.77	1.00																					
5. CSE5	-.51	.47	-.59	.72	1.00																				
6. CSE6	.62	-.55	.75	-.84	-.78	1.00																			
7. CSE7	-.48	.43	-.51	.58	.58	-.61	1.00																		
8. CSE8	.39	-.54	.59	-.57	-.59	.52	-.55	1.00																	
9. CSE9	.30	-.22	.45	-.65	-.74	.72	-.54	.44	1.00																
10. CSE10	-.60	.49	-.69	.64	.50	-.70	.59	-.48	-.50	1.00															
11. SPD1	-.60	.35	-.69	.64	.46	-.64	.37	-.33	-.46	.74	1.00														
12. SPD2	.48	-.28	.69	-.59	-.42	.60	-.29	.35	.40	-.67	-.80	1.00													
13. SPD3	.58	-.40	.69	-.63	-.59	.65	-.35	.37	.46	-.56	-.72	.65	1.00												
14. SPD4	.57	-.31	.58	-.56	-.56	.50	-.33	.30	.49	-.55	-.72	.66	.74	1.00											
15. SPD5	-.41	.44	-.52	.48	.31	-.48	.34	-.43	-.28	.45	.46	-.47	-.52	-.30	1.00										
16. MW1	.33	-.21	.39	-.25	-.15	.22	-.09	.25	.20	-.33	-.50	.39	.38	.36	-.33	1.00									
17. MW2	-.10	.20	-.14	.10	.08	-.11	.17	-.14	-.05	.00	-.05	.07	-.01	.04	.16	-.12	1.00								
18. MW3	.52	-.35	.66	-.55	-.37	.53	-.29	.29	.40	-.64	-.79	.78	.69	.68	-.54	.42	.01	1.00							
19. MW4	.56	-.28	.37	-.31	-.35	.28	-.30	.21	.28	-.27	-.33	.29	.44	.43	-.22	.31	-.13	.33	1.00						
20. MW5	.43	-.32	.51	-.50	-.56	.55	-.36	.35	.54	-.41	-.55	.57	.62	.58	-.42	.26	-.04	.63	.46	1.00					
21. MW6	.46	-.36	.58	-.61	-.58	.62	-.42	.36	.47	-.53	-.50	.66	.66	.58	-.48	.22	.02	.52	.39	.59	1.00				
22. C – CSE	.14	.33	.20	.29	.10	-.03	.08	-.06	.05	.08	-.08	.14	.11	.10	.01	.20	-.02	.13	.22	.13	-.03	1.00			
23. S – CSE	.00	.19	.06	-.01	-.13	.02	.10	-.11	.08	-.02	-.08	.35	.19	.34	.53	-.06	.27	.09	.10	.12	.24	.15	1.00		
24. M-W CSE	.63	-.39	.65	-.59	-.56	.58	-.41	.38	.50	-.56	-.67	.70	.75	.70	-.50	.53	.02	.71	.75	.78	.79	.17	.20	1.00	
25. Out. Exp.	.02	-.03	.10	-.10	-.08	.15	-.03	.10	.17	-.02	-.13	.11	.18	.00	-.08	.26	-.02	.02	.24	.10	.18	.10	.14	.26	1.00

Violations: < 5.00% for both tables

Table 10. Comparison of New Spreadsheet CSE and Compeau and Higgins CSE

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. CSE1	1.00																
2. CSE2	-.86	1.00															
3. CSE3	-.63	.69	1.00														
4. CSE4	-.63	.62	.63	1.00													
5. CSE5	-.47	.55	.54	.60	1.00												
6. CSE6	-.52	.51	.54	.70	.67	1.00											
7. CSE7	.71	-.71	-.60	-.47	-.56	-.66	1.00										
8. CSE8	-.48	.46	.53	.62	.54	.77	-.65	1.00									
9. CSE9	-.33	.34	.48	.55	.72	.76	-.54	.58	1.00								
10. CSE10	-.50	.49	.44	.39	.54	.64	-.68	.50	.67	1.00							
11. SPD1	-.47	.51	.49	.50	.46	.44	-.56	.42	.49	.54	1.00						
12. SPD2	-.39	.42	.52	.54	.49	.41	-.42	.38	.43	.35	.74	1.00					
13. SPD3	-.43	.47	.39	.45	.40	.30	-.37	.35	.37	.45	.76	.53	1.00				
14. SPD4	-.40	.46	.38	.52	.52	.42	-.45	.38	.52	.56	.87	.67	.85	1.00			
15. SPD5	-.52	.54	.54	.50	.45	.38	-.43	.37	.39	.48	.76	.65	.67	.75	1.00		
16. Compeau	-.69	.72	.68	.81	.77	.84	-.73	.71	.78	.83	.63	.55	.54	.65	.60	1.00	
17. Spreadsheet	-.55	.58	.49	.55	.47	.48	-.51	.43	.40	.52	.81	.62	.70	.81	.83	.65	1.00

As a test for H1b, we randomly administered three scales to subjects asked to perform a specific spreadsheet task originally developed by Yi (1999) and later used in several CSE studies (Johnson and Marakas, 2000; Yi and Davis, 2003).[8] The first scale was Martocchio and Webster's (1992) spreadsheet CSE, the second scale was the measure of outcome expectancy developed by Compeau and Higgins (1995a) specifically for their 1995 study, and the third scale was the contrived OE-CSE scale above.

Path analysis using PLS was performed with all scales being represented as independent variables and performance serving as the dependent variable. Although the model explained over 60% of the variance in performance, only the "pure" Martocchio and Webster (1992) CSE scale (the original scale without the added outcome expectancy items) had a significant relationship with performance ($t = 2.16$, $p < .05$), with neither the contrived OE-CSE scale ($t = 1.28$, $p = .20$) nor the Compeau and Higgins (1995a) outcome expectancy measure ($t = 0.51$, $p = .61$) explaining any significant variance in performance variable. Thus, we found support for H1b.

H2: Questions that elicit estimations of ability within a task-specific rather than a general context will be better predictors of performance within that task-specific domain than those that do not.

To address H2, we undertook two types of analyses. First, we completed a modified MTMM analysis (as discussed above and detailed by Trochim, 2001) of the new spreadsheet measure (Johnson and Marakas, 2000) and the Compeau and Higgins (1995a) scale. The subjects were framed to consider "the software package" referenced in the Compeau and Higgins (1995a) questions as a spreadsheet package. A review of the items contained in the Compeau and Higgins scale suggests it may be more appropriate at the general estimation level (i.e. "I could complete the job using the software package if…). Results of the modified MTMM analysis indicate that, while strong correlations exist between the scales, the new CSE scale exhibits both convergent and discriminant validity, suggesting that can the two scales are not equivalent (Table 10)[9]

Next, we compared these two scales to a spreadsheet performance measure (as used in H1b). The results indicate that both the Johnson and Marakas (2000) spreadsheet CSE instrument and the Compeau and Higgins (1995a) CSE instrument were significant predictors of performance on the task ($\beta = .54$, $p < .001$ and $\beta = .28$, $p < .05$, respectively). There was, however, a large difference between the two instruments with regard to the amount of variance explained in the performance variable. While the spreadsheet CSE measure (Johnson and Marakas, 2000) explained 29% of the variance in performance, the GCSE measure only explained 8% of the variance (Compeau and Higgins, 1995a). Overall, the task-specific instrument was able to explain almost four times as much variance as was the general instrument, thus providing support for H2.

H3: Questions that avoid ability assessments that include cross-domain or general-domain skills will more effectively isolate the CSE construct than those that do not.

We also tested Hypothesis 3 using two separate analyses. First, we compared the GCSE scale developed by Compeau and Higgins (1995a) to the scale developed by Gist et al. (1989). As previously discussed, one of the key aspects of the Compeau and Higgins' scale is that, while focusing on computer self-efficacy, it also frames the respondent to focus on an unknown task (i.e. respondents were to assume they "were given a new software package for some aspect of your work") (Compeau, 1995a, p. 210). This suggests successful completion of the task could require two types of ability estimations: those specifically related to the software under study and those required to complete the job or task. To adequately answer the questions the respondent must, therefore, consider a collection of cross-domain skills in forming his or her SE estimations.

In contrast, the Gist et al. (1989) scale focuses on a small set of software-specific spreadsheet skills (i.e. "…typing and entering numbers into cells," "…writing a formula for addition," etc.). Given this, we would expect the instruments to differ significantly in their respective ability to isolate the construct. Results of the modified MTMM indicated that, while there was correlation between all items, the majority of the items were most highly correlated with those other items assessing the same construct (Table 11). Interestingly, the items from Gist's scale were those that focused on what might be characterized as broader computing skills like copying ("calling up the command to copy," "telling a computer what to copy," and "telling

[8] A sample of the items contained in the performance measure can be found in the appendix of this manuscript. For further details, the reader is referred to Johnson and Marakas (2000) and Yi and Davis (2003).

[9] Note that this same relationship was also demonstrated above using the new measure of GCSE as part of initial divergent validity testing (Table 8).

a computer where to copy"), rather than being specifically related to spreadsheet tasks. Also, per previous discussion regarding convergent and discriminant validity, note the number of violations reported for Table 11 is significant — 40%. This condition is the result of each measure having at least one very low item/construct correlation. Given the previous evidence with respect to the Compeau and Higgins instrument, suggesting its more general nature, the results illustrate how instruments that include skills that may be cross-task or cross-domain (such as copying or typing or saving files) can cause the instrument to less than accurately isolate the construct.

To further illustrate this, we compared the new instruments, developed specifically to avoid any cross-task or cross-domain items, to each other (Table 12). As can be seen for all scales, the items correlated more strongly with their associated composite score than with the other composite scores. Even more interesting, however, was that although the items for GCSE and Windows CSE were more strongly correlated with their composite factors, they still had very high correlations on the other composite score (.55 - .72 for each). This suggests that despite the fact the scales were developed to assess different skills for GCSE (ex. "identify common operational problems," "unpack and set up a computer," and "remove information from a computer that I no longer need") and Windows (ex. "delete an icon that I no longer need," "change system settings," and "group programs using Windows"), the evidence from the analysis indicates that users who are heavily Windows-based may see the totality of these skills as more general computing skills and, as such, not clearly distinguish between the Windows environment and the general computing environment. In contrast, the correlations with the other composite scores were lower, suggesting they were more clearly distinguishing between the more general skills and the application specific skills (e.g. spreadsheet, database, etc.) The combined results from these analyses provide support for H3.

Table 11: Illustration of Cross-Domain Skills and Factor Structure

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1. GCSE1	1.00																					
2. GCSE2	-.61	1.00																				
3. GCSE3	.68	-.63	1.00																			
4. GCSE4	-.64	.59	-.77	1.00																		
5. GCSE5	.51	-.47	.59	-.72	1.00																	
6. GCSE6	-.62	.55	-.75	.84	-.78	1.00																
7. GCSE7	-.48	.43	-.51	.58	-.58	.61	1.00															
8. GCSE8	.39	-.54	.59	-.57	.59	-.52	-.55	1.00														
9. GCSE9	.30	-.22	.45	-.65	.74	-.72	-.54	.44	1.00													
10. GCSE10	.60	-.49	.69	-.64	.50	-.70	-.59	.48	.50	1.00												
11. GIST1	.34	-.18	.40	-.57	.47	-.49	-.30	.30	.47	.44	1.00											
12. GIST2	-.44	.52	-.65	.52	-.36	.51	.30	-.55	-.33	-.54	-.24	1.00										
13. GIST3	.46	-.23	.45	-.49	.58	-.45	-.38	.26	.53	.44	.64	-.38	1.00									
14. GIST4	-.42	.18	-.38	.47	-.47	.41	.24	-.21	-.41	-.37	-.59	.36	-.78	1.00								
15. GIST5	.38	-.19	.35	-.44	.42	-.33	-.18	.39	.38	.31	.56	-.24	.61	-.55	1.00							
16. GIST6	.56	-.37	.73	-.51	.47	-.59	-.32	.41	.47	.63	.28	-.75	.52	-.43	.30	1.00						
17. GIST7	.51	-.27	.56	-.47	.39	-.52	-.28	.25	.41	.60	.43	-.54	.54	-.48	.45	.66	1.00					
18. GIST8	.40	-.31	.55	-.51	.41	-.49	-.30	.40	.46	.53	.37	-.46	.38	-.37	.61	.51	.61	1.00				
19. GIST9	-.30	.14	-.25	.36	-.29	.31	.13	-.12	-.29	-.24	-.53	.20	-.54	.51	-.55	-.27	-.43	-.36	1.00			
20. GIST10	.56	-.41	.69	-.56	.39	-.56	-.24	.48	.36	.64	.47	-.74	.50	-.48	.39	.78	.64	.50	-.39	1.00		
21. Compeau GCSE	.73	-.40	.88	-.69	.59	-.70	-.42	.58	.55	.78	.47	-.62	.52	-.47	.47	.79	.66	.60	-.31	.76	1.00	
22. Gist CSE	.59	-.40	.77	-.62	.52	-.65	-.36	.49	.53	.72	.50	-.73	.58	-.49	.50	.90	.78	.74	-.32	.87	.85	1.00

GCSE: Compeau and Higgins, GIST: Gist
Violations: 40%

Table 12: Composite Score Correlations of New Task Specific CSE Instruments

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1. GCSE1	1.00																			
2. GCSE2	.56	1.00																		
3. GCSE3	.53	.57	1.00																	
4. GCSE4	.45	.67	.45	1.00																
5. GCSE5	.52	.55	.48	.55	1.00															
6. GCSE6	.55	.57	.55	.49	.63	1.00														
7. WIN1	.47	.64	.56	.55	.57	.59	1.00													
8. WIN2	.47	.61	.54	.51	.46	.54	.65	1.00												
9. WIN3	.39	.49	.41	.36	.48	.51	.50	.60	1.00											
10. WIN4	.47	.51	.50	.53	.72	.61	.56	.48	.54	1.00										
11. WIN5	.43	.63	.53	.60	.56	.56	.62	.60	.47	.56	1.00									
12. WP1	.37	.40	.38	.44	.51	.54	.45	.40	.51	.53	.44	1.00								
13. WP2	.39	.37	.40	.40	.52	.51	.39	.39	.42	.55	.39	.75	1.00							
14. WP3	.41	.35	.39	.43	.47	.48	.45	.41	.44	.54	.43	.57	.63	1.00						
15. WP4	.40	.42	.41	.42	.47	.51	.53	.45	.45	.55	.43	.57	.60	.66	1.00					

16. I1	.48	.58	.55	.50	.46	.47	.54	.52	.33	.43	.50	.34	.33	.32	.40	1.00				
17. I2	.45	.54	.48	.49	.45	.45	.51	.46	.35	.45	.53	.34	.35	.34	.36	.69	1.00			
18. I3	.41	.47	.47	.38	.36	.49	.54	.51	.31	.39	.49	.33	.36	.38	.43	.55	.61	1.00		
19. I4	.42	.51	.55	.41	.37	.43	.54	.53	.35	.37	.51	.35	.31	.33	.39	.72	.64	.61	1.00	
20. I5	.45	.50	.49	.47	.44	.47	.51	.50	.40	.44	.52	.34	.35	.33	.43	.67	.71	.59	.60	1.00
21. S1	.44	.43	.43	.40	.43	.49	.45	.41	.37	.46	.47	.43	.44	.46	.43	.42	.40	.33	.40	.41
22. S2	.28	.27	.24	.27	.27	.27	.31	.28	.24	.27	.26	.28	.27	.26	.24	.19	.24	.22	.22	.20
23. S3	.42	.38	.39	.38	.44	.48	.47	.38	.39	.46	.44	.44	.44	.48	.46	.36	.31	.32	.32	.37
24. S4	.37	.37	.32	.40	.42	.45	.48	.34	.37	.45	.43	.47	.40	.49	.44	.31	.27	.28	.29	.32
25. S5	.39	.42	.38	.37	.38	.44	.46	.39	.39	.41	.43	.41	.39	.44	.40	.34	.32	.29	.36	.36
26. DB1	.31	.34	.35	.28	.26	.31	.40	.38	.31	.29	.39	.22	.23	.28	.31	.36	.33	.30	.45	.40
27. DB2	.35	.45	.36	.34	.30	.32	.40	.35	.29	.34	.34	.27	.28	.31	.34	.41	.35	.25	.40	.35
28. DB3	.32	.43	.36	.31	.29	.33	.41	.37	.32	.33	.37	.30	.28	.33	.36	.39	.38	.30	.41	.40
29. DB4	.35	.41	.35	.30	.33	.38	.38	.31	.31	.35	.36	.29	.29	.35	.38	.32	.33	.26	.37	.31
30. GCSE	.75	.84	.76	.79	.78	.79	.72	.66	.56	.70	.71	.56	.54	.53	.55	.65	.61	.55	.57	.60
31. Windows	.55	.72	.63	.63	.68	.70	.82	.83	.78	.76	.81	.58	.53	.56	.59	.57	.57	.56	.57	.59
32. Word Proc.	.46	.46	.46	.50	.58	.60	.54	.48	.53	.64	.50	.84	.86	.85	.84	.41	.41	.44	.41	.43
33. Internet	.52	.62	.60	.54	.49	.55	.62	.59	.42	.50	.60	.40	.40	.40	.48	.86	.87	.80	.84	.85
34. Spreadsheet	.45	.44	.41	.43	.45	.50	.51	.43	.41	.48	.47	.48	.46	.49	.46	.37	.37	.34	.37	.38
35. Database	.37	.45	.39	.34	.33	.38	.44	.39	.34	.36	.40	.30	.30	.36	.39	.41	.39	.30	.45	.40

	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
21. S1	1.00														
22. S2	.45	1.00													
23. S3	.77	.43	1.00												
24. S4	.74	.38	.80	1.00											
25. S5	.72	.36	.73	.72	1.00										
26. DB1	.47	.20	.44	.43	.50	1.00									
27. DB2	.49	.21	.47	.47	.51	.67	1.00								
28. DB3	.50	.20	.46	.47	.53	.71	.85	1.00							
29. DB4	.54	.22	.51	.49	.54	.64	.82	.78	1.00						
30. GCSE	.55	.34	.52	.49	.50	.39	.45	.43	.45	1.00					
31. Windows	.54	.34	.53	.51	.52	.44	.42	.45	.42	.83	1.00				
32. Word Proc.	.52	.31	.54	.53	.49	.31	.35	.37	.39	.64	.67	1.00			
33. Internet	.46	.25	.40	.35	.40	.43	.42	.44	.38	.71	.68	.50	1.00		
34. Spreadsheet	.85	.75	.88	.83	.80	.46	.48	.49	.52	.56	.57	.55	.44	1.00	
35. Database	.56	.23	.52	.52	.58	.83	.93	.92	.91	.48	.48	.39	.46	.54	1.00

GCSE: General CSE, **WIN**: Windows CSE, **WP**: Word Processing CSE, **INT**: Internet CSE, **SPD**: Spreadsheet CSE, **DB**: Database CSE, Violations: none

H4: Instruments where the ordering of questions avoids inappropriate or unnecessary anchoring with regard to perceived or actual increasing levels of task difficulty or complexity will better isolate the CSE construct than those that do not.

To investigate the effects of anchoring and task complexity ordering, we utilized a 13-item subset of the instrument developed by Russon et al. (1994), since it measures a broad set of technology skills in a manner easily allowing for ordering based on perceived difficulty or complexity. The Russon CSE instrument was originally published using a task difficulty hierarchy, and studies that have used it have often faced equivocal and, sometimes, counterintuitive results.

Three versions of the instrument were administered (increasing complexity, decreasing complexity, and randomized) to 57 graduate students with a variety of technology backgrounds (these subjects were randomly recruited from only one of the three original study sites). The mean score for CSE was 88.51 (sd = 10.57). As suggested by the Marakas et al. (1998) framework, we observed an ordering effect, with those subjects receiving items in decreasing complexity scoring the highest, those receiving randomized items in the middle, and those receiving items in increasing complexity scoring the lowest (Table 13). While the group means were not significantly different, the mean values were consistent with previous research (Berry et al., 1989) and with the hypothesis under test.

Table 13: Comparison of Instrument

Order of Items	n	M	SD
Decreasing Difficulty	21	89.3	9.9
Randomized	20	88.0	8.5
Increasing Difficulty	16	86.4	14.1

Given the overall strong efficacy levels and the potential for a ceiling effect impacting results, we investigated a subset of the data focusing specifically on the differences between the randomized group and the increasing difficulty group. To do this, we removed those with a CSE greater than 90 out of 100 (very high CSE), leaving a usable sample of 19 (11:8). We removed high CSE subjects to ensure that no ceiling effect created by the high CSE subjects would be observed. Investigation of these two groups revealed a significant difference (p =.06) in the groups (at a more conservative .10 significance level). Given the small sample size with which this measurable effect was found, we felt the results provided encouraging evidence of the impact ordering of difficulty can have on responses to CSE. To investigate this further, we conducted an additional study using a completely different instrument.

An instrument developed and validated for this study, using the same procedures described earlier in this paper, asked the subject to estimate his or her ability to remember a sequence of two-digit numbers. Each item increased the number of two-digit numbers the subject was asked to remember by one additional number. For example, the first item would be "I believe I have the ability to memorize and recall one (1) 2-digit number;" the second item would be "I believe I have the ability to memorize and recall two (2) 2-digit numbers," etc. Three different versions of the instrument (ascending, random, and descending) were randomly administered to a sample of 38 graduate and undergraduate business majors at a large Midwestern business school. Results are found in Table 14.

Table 14: Descriptive Statistics for Each Scale Order

Order	N	Min	Max	μ	sd
Random	13	2.7	9.6	6.5	2.5
Ascending	14	3.1	10.00	6.5	2.8
Descending	11	4.2	9.1	7.1	1.5

As shown in Table 14, the results were similar to those obtained in the test of the Russon et al. (1994) measure. Those receiving the decreasing difficulty instrument (descending) scored highest, and those receiving the increasing difficulty version (ascending) scored lowest. As with the previous test, however, the differences between the mean scores were not significant.

Given that a measure of CSE is equally important for predicting variance as well as relative changes in estimation levels, we conducted an analysis of the explained variance across the three measures (ascending, descending, and random ordering) as a final test. The appropriate test in this instance is Levine's test for homogeneity of variance (Millikin and Johnson, 1994). This test uses the absolute values of the deviations from the group means as data and computes a t-statistic between two groups. Here, a significant p-value implies non-equality of variances and, thus, dissimilarity in performance. The results of

Levine's test indicate that there is a difference in variance between the groups (Levine Statistic = 7.95, $p < .001$). Results further indicate that there were anchoring effects, with a significant difference in variance between both the random and descending groups ($F = 7.05$, $p < .05$) and the ascending and descending groups ($F = 18.27$, $p < .001$). No difference was found between the random and ascending groups ($F = 1.30$, p = n.s.). Collectively, these results indicate that while measures of CSE may be moderately resistant to all but the most severe of ordering effects, care must be taken to not anchor individuals via the administration of the instrumentation because both the variance in scores and overall scores themselves may be problematically impacted. The collective results obtained in the test of this hypothesis indicate support for H4.

H5: As domain evolution results in new technologies and new applications, as well as new functionalities to existing technologies and applications, newer CSE measures developed to reflect these evolutionary changes will be better predictors of performance than older CSE measures developed prior to such domain evolution.

Hypothesis 5 suggests that newer measures designed to accurately reflect the current nature of the task domain will be better predictors of performance than older measures developed during a prior period of domain evolution. As originally shown in the analysis of H3, the Gist et al. (1989) spreadsheet CSE measure focuses on skills directly related to the performance of common spreadsheet tasks. In the analysis of H3, it was also shown to be clearly separated from the GCSE measure proposed by Compeau and Higgins (1995a; 1995b). This same separation of constructs was observed in the testing of H2 between the Compeau and Higgins GCSE measure (focusing on using Excel to perform a task) and the Johnson and Marakas (2000) spreadsheet measure developed using the framework proposed by Marakas et al. (1998).

Upon inspection, several similarities exist between the Gist spreadsheet measure and the spreadsheet measure developed by Johnson and Marakas with regard to measures of validation, differentiation from a generalized measure, and focus on an identical task-specific domain. The primary difference between the two instruments lies with their comparative age. The Gist instrument has been widely used since its introduction and has been shown to be a strong predictor of performance variance in spreadsheet tasks (Brock and Sulsky, 1994; Henry and Stone, 1994; Igbaria and Iivari, 1995; Lee and Bobko, 1994).

Given that the Gist instrument was developed and validated over a decade and a half before the Johnson and Marakas spreadsheet CSE measure, this creates an ideal scenario for an initial test of H5. If the propositions from which H5 was derived have merit, we should expect to see a difference between the two measures with regard to their individual abilities to predict performance using a current spreadsheet application.

To test H5, we compared the two measures discussed above using a separate regression procedure for each scale. Both the Gist measure ($\beta = .32$, $p < .05$) and the new spreadsheet measure ($\beta = .671$, $p < .001$) were significant predictors of performance on the spreadsheet task. As predicted, however, the Johnson and Marakas measure ($r^2 = .74$) was able to predict just under three times as much variance in the dependent variable as was the older Gist et al. measure ($r^2 = .26$). [10]

To further examine H5, we employed a procedure for measuring the effect size and significance in the change in r^2 between the two scales. First, we calculated the effect size for the change in r^2 by $(r^2_{aggregated} - r^2_{non\text{-}aggregated}) / (1 - r^2_{aggregated})$ (Cohen, 1988). The extant literature categorizes a small effect size as 0.02; medium as 0.15; and large as 0.35 (Cohen, 1988). In this case, our effect size for the change in r^2 between the two scales was 0.64, suggesting a significantly large effect size.

We then conducted a pseudo-F test for the change in r^2 with 1 and (n-k) degrees of freedom. Research has typically employed such tests to ascertain the significance of the change in explained variance realized within nested models by addition of one or more constructs or paths within a model (Chin et al., 2003; Mathieson et al., 2001). For our purposes, the pseudo-F can be calculated by multiplying the effect size by (n-k-1). In this case, the pseudo-F value is 306.8 ($p < .001$). Combined with the above, we believe these results provide support for H5.

Discussion

Computer self-efficacy has become an important variable in the study of information technology. From adoption and use to the learning process, computer self-efficacy has been shown to be an effective predictor of end-user performance. The challenge to research, and to our further understanding of the construct, is the plethora of measures that continue to be utilized, many of which are approaching two decades old. With the domain of IT changing from its historical roots of backroom data processing with few users and text-based processing to a graphically intensive, strategic asset impacting

[10] We are testing standardized betas in this regression. As such, the larger beta is meaningful.

organizations and society, it is critical that our measures evolve to reflect this dynamic environment. As such, the skills and expertise required to successfully leverage information technology have also changed, and the measures used to elicit estimations of ability in this domain must reflect these changes.

Thus, this research was driven by three overall goals. The first was to rigorously examine the effective construction of CSE instruments via the Marakas et al. (1998) framework. The second goal was to provide the research community with a functional understanding of the need to correctly view the CSE construct as formative rather than reflective and to use the appropriate validation techniques in the design of measures of the construct. The third goal was to provide a detailed comparison of the currently utilized instruments of CSE to determine their relative abilities to isolate and measure the CSE construct, as well as their relative effectiveness at predicting performance. Overall, the results suggest the framework for CSE measure development proposed by Marakas et al. (1998) is an effective guide for researchers to follow when developing new CSE measures. Further, our results provide support for the validation of measures of CSE as formative constructs rather than the more commonly employed validation techniques appropriate for reflective constructs. Finally, results indicate the most effective measures of CSE have two key characteristics: they adhere to the base theory as operationalized by the proposed development framework, and they are in keeping with the current state of evolution within the computing domain.

Limitations

Before we can consider the value of any conclusions derived from this study, several limitations bear acknowledgement. First, the subjects across the three test sites were undergraduate and graduate students. Use of student subjects has often been argued to reduce the potential to generalize to a broader and more useful population. One accepted method to mitigate the issues associated with the use of student subjects is to recruit them using characteristics that represent the population of interest and present them with questions and tasks for which they have the requisite skills and knowledge (Gordon et al., 1986). In this case, the student subjects clearly represented a subset of the broader population and possessed the skills and knowledge necessary to perform the software tasks assigned. Further, estimations of self-efficacy are individual in nature and require no generalizable skill or experience to make them. Given this, we believe students serve well as subjects for this study.

A second limitation arises from the limited scope of this study, precluding a comparison across all available measures of CSE and GCSE. We believe, however, our findings with regard to the most widely adopted measures would not materially change with the introduction of lesser known or adopted measures. This issue, however, remains for future research.

Finally, we used only one performance task in this study — spreadsheet skills. It is possible that comparison across multiple performance environments would serve to more closely identify the relative strengths and weaknesses across the various extant measures. We chose this task because of its prevalence in the literature, both with regard to instrument development and behavioral modeling research. While we believe the task to be representative of the items necessary to establish a valid performance measure for spreadsheet skills, we must acknowledge that other common computing skill domains remain unrepresented and should be examined by future research.

Implications to Research and Practice

We believe this research provides several contributions to both research and practice. This study developed multiple new measures of CSE for common office productivity applications (i.e. word processing, spreadsheets, databases, Internet, Windows, and general) that exhibited strong estimates of reliability as well as convergent and discriminant validity. These new measures provide researchers with up-to-date measures to further extend investigations into the CSE construct at both the task specific and general computing domain levels. The measures further provide practitioners with instruments they can use to assess the current skill sets of their employees with an eye toward predicting future performance and assessing the need for various levels of training. Along with assessment tools, the research can also provide both researchers and practitioners with tools to better understand the nature of efficacy estimations in one software sub-domain and its relationship to other similar or diverse domains.

As Bandura (1997) suggests, one of the key aspects of self-efficacy is its generality. Efficacies developed in one area can provide information cues that can transfer to other related domains. Thus, these new measures can be utilized in research contexts to better understand how skills gained in one task domain, such as word processing, impact skills in another related task domain, such as spreadsheet, database, or Internet skills.

A related aspect for future research is investigation of more complex models of performance outcomes. Both self-efficacy research in general, and computer self-efficacy research more specifically, have focused heavily on models in very distinct domains (i.e. overcoming snake phobias, learning spreadsheets, performing memory tasks, etc.). For businesses and information systems, real world tasks are neither as simple nor single domain focused. Rather, they often draw on multiple skill sets and require an individual to be able to perform tasks that span several skill domains. As an example, the use of a computer-based human resource information system (HRIS) for recruiting and selecting new employees for an open position will require at least two types of skills to be successfully completed. First, the individual would need knowledge of the human resources processes and laws, and their various conditions of application, as well as the technical skills necessary to effectively utilize the computer-based system to identify the most qualified applicants. As such, efficacy estimations with regard to such a complex task would require multiple instruments to accurately measure. Any program geared toward improving performance would require interventions focused not only on enhancing computer skills, but on the domain-specific skills as well. Future research should determine how CSE estimations play a role in these types of assessments with regard to possible interaction effects and the extent to which GCSE measures can effectively predict the computer-related performance of the candidate. It is possible under these conditions that increasing individual levels of CSE may serve to improve confidence in content knowledge as well. While there is a temptation to suggest that either one may play a more key role, before we can adequately address such models, we need to better understand the interplay among efficacies for tasks that require skills from multiple domains.

This research makes a second contribution to our understanding of the CSE construct by providing evidence of potential weaknesses associated with adoption of previously validated instruments without thought. While reuse of available instruments is considered to be a hallmark of strong research, the values of previously validated instruments wane when domains and skill sets for these domains evolve. Thus, researchers should seek to determine whether or not the computer self efficacy instruments of interest are indeed still appropriate for the domain of study. If not, then the instrumentation should be updated to reflect the new nature or status of the domain. One of the problems facing researchers is that the development of strong empirically valid instruments is a difficult and time consuming process above and beyond considerations of the study itself. Despite this, we argue that the development and updating of CSE instruments to reflect both the dynamic nature of the skills of interest and their adherence to the tenets of the underlying theory are necessary conditions for effective and generalizable research into the CSE construct. Future research should subject the current measures developed in this study to both rigorous replication and extension and should continually evaluate their appropriateness in the future as skills and software evolve.

Possibly one of the most interesting findings in this study is the apparent relationship between the general computing environment and the skills necessary to operate within a Windows environment. While this finding was unanticipated, post-hoc analyses of the results provide a logical basis for it. As discussed previously, we believe the strength of this relationship to be driven by the common computing environment of the subjects. Nonetheless, this raises some interesting questions for future research. Can a more effective measure of GCSE be constructed by including both general computer skills and common Windows-related skills? Within a Windows environment, can any well-developed measure of GCSE substitute for multiple task-specific measures? Would a well-developed measure of GCSE cross load in a similar fashion with a LINUX or UNIX CSE measure when administered to users common to these environments? In studies where the CSE construct is positioned as a mediating or moderating variable to other constructs of interest, is it necessary to use a measure that parallels the subjects' common computing environment? We have only begun to understand the effectiveness of measuring CSE estimations within broader computing environments and future research should begin to focus its attention outside the boundaries of personal computing.

The introduction of the concepts of formative versus reflective constructs as they apply to validation of measures of CSE and GCSE stands as a contribution to future researchers who wish to effectively isolate the CSE construct within their domain of interest. We believe this demonstration of appropriate validation methods for CSE as a formative construct to be the first to appear within the IS literature and believe it will serve to improve the development and validation of future measures of both the CSE and GCSE constructs. That said, we also must acknowledge that this perspective of the CSE construct is relatively new and thus requires additional empirical study to determine whether or not the formative approach is valid in all instances and applications of the construct.

We also believe that the conceptualization of the GCSE domain by Marakas et al. (1998), while plausible and clearly embraced in this research, is but one of several conceptualizations and, therefore, must be investigated by future research. For example, it is reasonable to argue that a general measure of CSE may actually be more of a generic measure of the construct than a weighted product of previous application-specific estimates as suggested by Marakas et al. (1998). Using this approach, a general measure of the construct may be stable over time and would reflect estimates of ability closer to the individual than the tool. Given the presence of multiple conceptualizations, future research should focus on reconciling these perspectives such that we have a clear characterization of the construct upon which to build our understanding.[11]

Finally, this study provided initial validation for the Marakas et al. (1998) framework through an analysis of existing measures of CSE as well as the development and validation of a selection of new instruments. Results of the analysis suggest the framework to be a useful tool in the development of new CSE instruments. Given the dynamic nature of information technology, researchers will continue to develop new measures as the domain evolves. As such, a strong organizing framework will provide researchers with the ability to build higher quality instruments and with an additional tool to assist them in evaluating the quality of current instruments and designing new, and more effective, CSE instruments.

REFERENCES

Agarwal, R., V. Sambamurthy, and R. M. Stair (2000) "The evolving relationship between general and specific computer self-efficacy: An empirical assessment," *Information Systems Research* (11) 4, pp. 418-430.

Bagozzi, R.P.. 1980. "A Conceptual System for Discovering and Testing Causal Relationships in Marketing." *Macromarketing: Evolution of Thought* Boulder, CO Business Research Division, Graduate School of Business Administration, University of Colorado 295-313

Bandura, A. (1977a) "Self-Efficacy: Toward a Unifying Theory of Behavioral Change," *Psychological Review* (84) 2, pp. 191-215.

Bandura, A. (1977b) *Social Learning Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1978) "Reflections on self-efficacy," *Advances in Behavioral Research and Therapy* (1pp. 237-269.

Bandura, A. (1986) *Self-efficacy, social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997) *Self-efficacy: The exercise of control*. New York: W. H. Freeman.

Bandura, A. (2001) Guide for constructing self-efficacy scales, pp. 7-37.

Baroudi, J. and W. Orlikowski (1988) "A short-form measure of user information satisfaction: A psychometric evaluation and notes on use," *Journal of Management Information Systems* (4) 4, pp. 44-59.

Berry, J. M., R. L. West, and D. M. Dennehey (1989) "Reliability and validity of the memory self-efficacy questionnaire," *Developmental Psychology* (25pp. 701-713.

Bollen, K and R. Lennox (1991) "Conventional wisdom on measurement: A structural equation perspective," *Psychological Bulletin*, (110) 2, pp. 305-314.

Bolt, M. A., L. N. Killough, and H. C. Koh (2001) "Testing the interaction effects of task complexity in computer training using the social cognitive model," *Decision Sciences* (32) 1, pp. 1-20.

Boudreau, M., D. Gefen, and D. W. Straub (2001) "Validation in IS research: A state-of-the-art assessment," *MIS Quarterly* (24) 1, pp. 1-16.

Brock, D. B. and L. M. Sulsky (1994) "Attitudes toward computers: Construct validation and relations to computer use," *Journal of Organizational Behavior* (15pp. 17-23.

Brown, S. D., R. W. Lent, and K. C. Larkin (1989) "Self-efficacy as a moderator of scholastic aptitude-academic performance relationships," *Journal of Vocational Behavior* (35pp. 64-75.

Burkhardt, M. E. and D. J. Brass (1990) "Changing patterns or patterns of change: The effects of a change in technology on social network structure and power," *Administrative Science Quarterly* (35pp. 104-127.

Busch, T. (1995) "Gender differences in self-efficacy and attitudes toward computers," *Journal of Educational Computing Research* (12) 2, pp. 147-158.

Campbell, D. T. (1960) "Blind variation and selective retention in creative thought as in other knowledge processes," *The Psychological Review* (67pp. 380-300.

Campbell, D. T. and D. W. Fiske (1959) "Convergent and discriminant validation of the multitrait-multimethod matrix," *Psychological Bulletin* (56) 2, pp. 81-105.

Carlson, R. D. and B. L. Grabowski (1992) "The effects of computer self-efficacy on direction-following behavior in computer-assisted instruction," *Journal of Computer-Based Instruction* (19) 1, pp. 6-11.

Cervone, D. and P. K. Peake (1986) "Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior," *Journal of Personality and Social Psychology* (50pp. 492-501.

Chin, Wynne W., Barbara L. Marcolin and Peter R. Newsted, (2003)."A partial least squares latent variable modeling approach for measuring interaction effects: Results from a monte carlo simulation study and an electronic-mail emotion/adoption study," *Information Systems Research*, 14, 2, June, 189-217.

[11] We wish to acknowledge and thank one of the anonymous referees for providing us with this perspective.

Christoph, R. T., G. A. Schoenfeld Jr., and J. W. Tansky (1998) "Overcoming barriers to training utilizing technology: The influence of self-efficacy factors on multimedia-based training receptiveness," *Human Resource Development Quarterly* (9) 1, pp. 25-38.
Cohen, J., *Statistical Power Analysis for the Behavioral Sciences, (2nd Edition)*, Lawrence Erlbaum, Hillsdale, NJ, 1988.
Compeau, D. R. and C. A. Higgins (1995a) "Application of social cognitive theory to training for computer skills," *Information Systems Research* (6) 2, pp. 118-143.
Compeau, D. R. and C. A. Higgins (1995b) "Computer self-efficacy: Development of a measure and initial test," *MIS Quarterly* (19) 2, pp. 189-211.
Cronbach, L. J. (1951) "Coefficient Alpha and the internal structure of tests," *Psychometrika* (16pp. 297-334.
Davis, F. D. (1989) "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly* (13) 3, pp. 319-339.
Delcourt, M. A. B. and M. B. Kinzie (1993) "Computer technologies in teacher education: The measurement of attitudes and self-efficacy," *Journal of Research and Development in Education* (27) 1, pp. 35-41.
Diamantoppoulos, A. and H.M. Winklhofer (2001) "Index construction with formative indicators: An alternative to scale development," *Journal of Marketing Research*, 38 (May), pp. 269-277.
Gist, M. E. (1987) "Self-efficacy: Implications for organizational behavior and human resource management," *Academy of Management Review* (12) 3, pp. 472-485.
Gist, M. E. and T. R. Mitchell (1992) "Self-efficacy: A theoretical analysis of its determinants and malleability," *Academy of Management Review* (17pp. 183-211.
Gist, M. E., C. Schwoerer, and B. Rosen (1989) "Effects of alternative training methods on self-efficacy and performance in computer software training," *Journal of Applied Psychology* (74) 6, pp. 884-891.
Gordon, M. E., L. A. Slade, and N. W. Schmitt (1986) "The "science of the sophomore" revisited: From conjecture to empiricism," *Academy of Management Review* (11) 1, pp. 191-207.
Harrison, A. W. and J. R. K. Rainer (1992) "An examination of the factor structures and concurrent validities for the computer attitude scale, the computer anxiety rating scale, and the computer self-efficacy scale," *Educational and Psychological Measurement* (52pp. 735-745.
Heinssen, R. K., C. R. Glass, and L. A. Knight (1987) "Assessing computer anxiety: Development and validation of the computer anxiety rating scale," *Computer and Human Behavior* (3pp. 49-59.
Henderson, R. D., F. P. Deane, and M. J. Ward (1995) "Occupational differences in computer-related anxiety: Implications for the implementation of a computerized patient management information system," *Behaviour and Information Technology* (14) 1, pp. 23-31.
Henry, J. W. and R. W. Stone (1994) "A structural equation model of end-user satisfaction with a computer-based medical information system," *Information Resources Management Journal* (7) 3, pp. 21-33.
Henry, J. W. and R. W. Stone (1997) "The development and validation of computer self-efficacy and outcome expectancy scales in a nonvolitional context," *Behavioral Research Methods, Instruments, and Computers* (29) 4, pp. 519-527.
Hill, T., N. D. Smith, and M. F. Mann (eds.) (1985) *Communicating Innovations: Convincing computer phobics to adopt innovative technologies*. Vol. 13. *Advances in Consumer Research*, Provo, UT: Association for Consumer Research.
Hill, T., N. D. Smith, and M. F. Mann (1987) "Role of efficacy expectations in predicting the decision to use advanced technologies: the case for computers," *Journal of Applied Psychology* (72pp. 307-313.
Igbaria, M. and J. Iivari (1995) "The effects of self-efficacy on computer usage," *Omega* (23) 6, pp. 587-605.
Jarvis, C.B., S.B. MacKenzie, and P.M. Podsakoff (2003) "A critical review of construct indicators and measurement model misspecification in marketing and consumer research," *Journal of Consumer Research*, 30 (Sep), pp. 199-218.
Johnson, R. D. and G. M. Marakas (2000) "The role of behavioral modeling in computer skills acquisition: Toward refinement of the model," *Information Systems Research* (11) 4, pp. 403-417.
Jorde-Bloom, P. and M. Ford (1988) "Factors influencing early childhood administrators' decisions regarding the adoption of computer technology," *Journal of Educational Computing Research* (4) 1, pp. 31-47.
Keen, P. G. W. (1980) MIS research: Reference disciplines and a cumulative tradition. *First International Conference on Information Systems, 1980*, pp. 9-18.
Kuo, F.-Y. and M.-H. Hsu (2001) "Development and validation of ethical computer self-efficacy measure: The case of softlifting," *Journal of Business Ethics* (32pp. 299-315.
Lee, C. and P. Bobko (1994) "Self-efficacy beliefs: Comparison of five measures," *Journal of Applied Psychology* (79) 3, pp. 364-369.
Lloyd, B. H. and C. P. Gressard (1984) "Reliability and factorial validity of computer attitude scales," *Educational and Psychological Measurement* (44), pp. 501-505.
Loch, K.D., D.W. Straub, and S. Kamel (2003) "Diffusing the Internet in the Arab world: The role of social norms and technological culturation," *IEEE Transactions on Engineering Management*, (50) 1, pp. 45-63.
Mathieson, Kieran, Eileen Peacock and Wynne W. Chin,(2001). "Extending the technology acceptance model: The influence of perceived user resources," *The DATA BASE for Advances in Information Systems*, 32, 3, 86-112.

Marakas, G. M., M. Y. Yi, and R. D. Johnson (1998) "The multilevel and multifaceted character of computer self-efficacy: Toward clarification of the construct and an integrative framework for research," *Information Systems Research* (9) 2, pp. 126-163.
Martocchio, J. J. (1992) "Microcomputer usage as an opportunity: The influence of context in employee training," *Personnel Psychology* (45) 3, pp. 529-552.
Martocchio, J. J. (1994) "Effects of conceptions of ability on anxiety, self-efficacy, and learning in training," *Journal of Applied Psychology* (79) 6, pp. 819-825.
Martocchio, J. J. and J. Webster (1992) "Effects of feedback and cognitive playfulness on performance in microcomputer software training," *Personnel Psychology* (45) 3, pp. 553-578.
Millikin, G. A. and D. E. Johnson (1994) *Analysis of Messy Data, Vol. 1, Designed Experiments*. New York, NY: Van Nostrand Reinhold.
Mitchell, T. R., H. Hopper, D. Daniels, J. George-Falvy et al. (1994) "Predicting self-efficacy and performance during skill acquisition," *Journal of Applied Psychology* (79) 4, pp. 506-517.
Miura, I. T. (1987) "The relationship of computer self efficacy expectations to computer interest and course enrolment in college," *Sex Roles* (16) 5/6, pp. 303-311.
Moore, G. C. and I. Benbasat (1991) "Development of an instrument to measure the perceptions of adopting an information technology," *Information Systems Research* (2) 3, pp. 192-222.
Multon, K. D., S. D. Brown, and R. W. Lent (1991) "Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation," *Journal of Counseling Psychology* (38) 30-38.
Murphy, C. A., D. Coover, and S. V. Owen (1989) "Development and validation of the computer self-efficacy scale," *Educational and Psychological Measurement* (49pp. 893-899.
Petter, S., Straub, D.W., and Rai, A. (2006) "Specification and validation of formative constructs in IS research," Georgia State Working Paper Series.
Podsakoff, P.M., S.B. MacKenzie, N.P. Podsakoff, and J.Y. Lee (2003) "The mismeasure of man(agement) and its implications for leadership research," *The Leadership Quarterly*, (14), pp. 615-656.
Rainer, R. K., K. Laosethakul, and M. K. Astone (2003) "Are gender perceptions of computing changing over time?" *Journal of Computer Information Systems* (43) 4, pp. 108-114.
Rossiter, J. R., (2002) "The C-OAR-SE procedure for scale development in marketing," *International Journal of Research in Marketing*, (19), pp. 305-335.
Russon, A. E., N. Josefowitz, and C. V. Edmonds (1994) "Making computer instruction accessible: Familiar analogies for female novices," *Computer in Human Behavior* (10) 2, pp. 175-187.
Shapka, J. D. and M. Farrari (2003) "Computer-related attitudes and actions of teacher candidates," *Computer in Human Behavior* (19) 3, pp. 319-334.
Smith, J. M. (1994) "The effects of education on computer self-efficacy," *Journal of Industrial Teacher Education* (31) 3, pp. 51-65.
Straub, D. W. (1989) "Validating instruments in MIS research," *MIS Quarterly* (13) 2, pp. 147-169.
Straub, D.W., M.C. Boudreau, and D. Gefen (2004) "Validation guidelines for IS positivist research," *Communications of the AIS*, (13), pp. 380-427.
Taylor, S. and P. A. Todd (1995) "Understanding information technology usage: A test of competing models," *Information Systems Research* (6) 2, pp. 144-176.
Torkzadeh, G. and X. Koufteros (1994) "Factorial validity of a computer self-efficacy scale and the impact of computer training," *Educational and Psychological Measurement* (54) 3, pp. 813-821.
Trochim, W.M.K, *The Research Methods Knowledge Base*, Atomic Dog Publishing, Cincinnati, OH, 2001.
Venkatesh, V. and F. D. Davis (1996) "A model of the antecedents of perceived ease of use: Development and test," *Decision Sciences* (27) 3, pp. 451-482.
Vispoel, W. P. and P. Chen. (1990) Measuring self-efficacy: The state of the art. *Annual Meeting of the American Educational Research Association, Boston, MA, 1990*.
Webster, J. and J. J. Martocchio (1995) "The differential effects of software training previews on training outcomes," *Journal of Management* (21) 4, pp. 757-787.
Wood, R. and A. Bandura (1989) "Social cognitive theory of organizational management," *Academy of Management Review* (14) 3, pp. 361-384.
Yi, M. Y. (1999). Developing and Validating an observational learning model of computer software training. Unpublished Doctoral Dissertation, University of Maryland.
Yi, M. Y. and F. D. Davis (2003) “Developing and validating an observational learning model of computer software training and skill acquisition,”, *Information Systems Research*, 14(2), pp. 146-169.
Young, M. R. (2001) "Windowed, wired, and webbed - Now what?" *Journal of Marketing Education* (23) 1, pp. 45-54.

Appendix

All scale items were constructed as in the following example:

		Not at All Confident				Moderately Confident				Totally Confident	
I believe I have the ability to save a file.	YES. . . NO	10	20	30	40	50	60	70	80	90	100

(Retained items are italicized)

General Computer Self-Efficacy

1. *I believe I have the ability to describe how a computer works.*
2. *I believe I have the ability to install new software applications on a computer.*
3. *I believe I have the ability to identify and correct common operational problems with a computer.*
4. *I believe I have the ability to unpack and set up a new computer.*
5. *I believe I have the ability to remove information from a computer that I no longer need.*
6. I believe I have the ability to understand common operational problems with a computer.
7. *I believe I have the ability to use a computer to display or present information in a desired manner.*

Windows Computer Self-Efficacy

1. *I believe I have the ability to group programs together using Windows.*
2. I believe I have the ability to change system settings using Windows.
3. *I believe I have the ability to create an icon for a program.*
4. I believe I have the ability to delete an icon that I do not need.
5. *I believe I have the ability to arrange icons so that I can conveniently access them.*
6. I believe I have the ability to copy/move a file using Windows.
7. I believe I have the ability to change the appearance of Windows.
8. *I believe I have the ability to delete a file that I do not need using Windows.*
9. I believe I have the ability to change time and date of computer systems.
10. *I believe I have the ability to change monitor settings using Windows.*

Spreadsheet Computer Self-Efficacy

1. *I believe I have the ability to manipulate the way a number appears in a spreadsheet.*
2. *I believe I have the ability to use and understand the cell references in a spreadsheet.*
3. I believe I have the ability to enter numbers into a spreadsheet.
4. *I believe I have the ability to use a spreadsheet to communicate numeric information to others.*
5. *I believe I have the ability to write a simple formula in a spreadsheet to perform mathematical calculations.*
6. I believe I have the ability to summarize numeric information using a spreadsheet.
7. I believe I have the ability to use a spreadsheet to share numeric information with others.
8. *I believe I have the ability to use a spreadsheet to display numbers as graphs.*
9. I believe I have the ability to use a spreadsheet to assist me in making decisions.

Word-processing Computer Self-Efficacy

1. *I believe I have the ability to move a block of text using a word processor*
2. *I believe that I have the ability to manipulate the way a paragraph looks using a word processor.*
3. *I believe that I have the ability to add a footnote to a document using a word processor.*
4. *I believe I have the ability to merge information from two documents using a word processor.*
5. I believe I have the ability insert and delete words in a paragraph using a word processor.
6. I believe I have the ability to change the appearance of words or phrases within a paragraph using a word processor.
7. I believe I have the ability to check or improve my grammar in a document using a word processor.

Internet Computer Self-Efficacy

1. I believe I have the ability to create a shortcut to access programs.
2. *I believe I have the ability to download the information from another computer to my computer using the Internet.*
3. *I believe I have the ability to connect to another computer using the Internet.*

4. *I believe I have the ability to subscribe to a newsgroup.*
5. *I believe I have the ability to transfer files from my computer to another computer using the Internet.*
6. *I believe I have the ability to locate information on another computer using the Internet.*
7. I believe I have the ability to send messages to others using the Internet.
8. I believe I have the ability to publish information on the Internet.
9. I believe I have the ability to move from one computer to another using the Internet.
10. I believe I have the ability to navigate through Internet sites.

Database Computer Self-Efficacy

1. *I believe I have the ability to specify a primary key using a database program.*
2. I believe I have the ability to communicate information using a database program.
3. I believe I have the ability to update the database using a database program.
4. I believe I can create a query using a database program
5. *I believe I have the ability to create a database table using a database program*
6. *I believe I have the ability to understand a query written in a database program.*
7. I believe I have the ability to create a field using a database program.
8. I believe I have the ability to summarize information from a database table using a database program.
9. *I believe I have the ability to add or delete a specific record from a database using a database program.*
10. I believe I have the ability to manipulate the information in a field using a database program.

Task Performance Test (Sample of Items)

1. Enter a formula to compute profits (=sales - expenses) for each season in cells B8:E8.
2. Using an appropriate function, compute the total amounts of sales, expenses, and profits of year 2000. The computed amounts should be located in cells F6:F8.
3. Using an appropriate function, compute the average amounts of sales, expenses, and profits of year 2000. The computed amounts should be located in cells G6:G8.
4. Compute YTD (year-to-date) profits. The computed amounts should be located in cells B9:E9.
5. Calculate % change of sales from the previous season. The computed amounts should be located in cells C11:E11.

ABOUT THE AUTHORS

George M. Marakas is an Associate Professor of Information Systems at the University of Kansas. Professor Marakas' research has appeared in many prestigious academic journals including *Information Systems Research, Management Science, International Journal of Human Computer Studies,* and *European Journal of Information Systems*. In addition, Dr. Marakas is the author of five leading textbooks in the field of information systems.

Richard D. Johnson is an Assistant Professor of Management at the University at Albany, State University of New York. He received his Ph.D. from the University of Maryland, College Park. His research interests focus on human resource information systems, psychological and sociological impacts of computing technology, computer self-efficacy, e-learning, and issues surrounding the digital divide. Dr. Johnson's research has appeared in journals such as *Information Systems Research, Journal of the Association for Information Systems, International Journal of Human Computer Studies*, and the *Journal of Applied Social Psychology,* as well as numerous international research conferences.

Paul F. Clay is an Assistant Professor of Information Systems at Washington State University. He received his Ph.D. in Management Information Systems from Indiana University. His research interests include the interaction among technology, knowledge-based tasks and knowledge-based workers, on technology-mediated mobilization of knowledge resources and on improving technology-mediated decision making. His research has appeared in *Information Systems Research* and the proceedings of the *Hawaii International Conference on Systems Sciences*.

Acceptance Information

Detmar Straub was the accepting senior editor for this paper. The manuscript was received on December 28th 2004 and went through one round of revision and was with the author for 11 months.

Journal of the Association for Information Systems

ISSN: 1536-9323