*Information Systems Research*
Vol. 17, No. 3, September 2006, pp. 228–246
ISSN 1047-7047 | EISSN 1526-5536 | 06 | 1703 | 0228
DOI 10.1287/isre.1060.0096


# Reconceptualizing System Usage: An Approach and Empirical Test

Andrew Burton-Jones
Management Information Systems Division, Sauder School of Business, University of British Columbia,
2053 Main Mall, Vancouver BC, V6T 1Z2 Canada, andrew.burton-jones@sauder.ubc.ca

Detmar W. Straub, Jr.
Department of Computer Information Systems, J. Mack Robinson College of Business Administration,
Georgia State University, Box 4015, Atlanta, Georgia 30302, dstraub@gsu.edu

Although DeLone, McLean, and others insist that system usage is a key variable in information systems research, the system usage construct has received little theoretical scrutiny, boasts no widely accepted definition, and has been operationalized by a diverse set of unsystematized measures. In this article, we present a systematic approach for reconceptualizing the system usage construct in particular nomological contexts. Comprising two stages, definition and selection, the approach enables researchers to develop clear and valid measures of system usage for a given theoretical and substantive context. The definition stage requires that researchers define system usage and explicate its underlying assumptions. In the selection stage, we suggest that system usage be conceptualized in terms of its structure and function. The structure of system usage is tripartite, comprising a user, system, and task, and researchers need to justify which elements of usage are most relevant for their study. In terms of function, researchers should choose measures for each element (i.e., user, system, and/or task) that tie closely to the other constructs in the researcher's nomological network.

To provide evidence of the viability of the approach, we undertook an empirical investigation of the relationship between system usage and short-run task performance in cognitively engaging tasks. The results support the benefits of the approach and show how an inappropriate choice of usage measures can lead researchers to draw opposite conclusions in an empirical study. Together, the approach and the results of the empirical investigation suggest new directions for research into the nature of system usage, its antecedents, and its consequences.

*Key words*: system usage; theoretical conceptualization; measure; performance; IS success
*History*: Laurie J. Kirsch, Senior Editor; George Marakas, Associate Editor. This paper was received on September 1, 2005, and was with the authors $1\frac{1}{4}$ months for 1 revision.

## 1. Introduction

The system usage construct has played a central role in information systems (IS) research since the 1970s (Barkin and Dickson 1977). Many researchers have studied antecedents to usage and, over time, the field has progressed toward a general model of these antecedents (Venkatesh et al. 2003). Others have studied the impact of usage on individual performance. They report the link to be strongly positive (Doll and Torkzadeh 1998), weakly positive (Goodhue and Thompson 1995, Igbaria and Tan 1997), insignificant (Lucas and Spitler 1999), or negative (Pentland 1989, Szajna 1993). The usage construct itself, however, typically escapes scrutiny in such studies of antecedents and consequences. This has fuelled calls for closer examination of the usage construct (Chin and Marcolin 2001, DeLone and McLean 2003).

Similar in intention to Melone's (1990) conceptual work on the user satisfaction construct, the current study undertakes a theoretical assessment of individual system usage. Despite its centrality in IS research, the system usage construct has received scant theoretical treatment to date. Apart from Trice and Treacy's (1986) brief conceptualization and a short discussion by Seddon (1997), we are unaware of any in-depth, theoretical assessment of the construct. Without theoretical grounding, it is not surprising that past studies have arrived at mixed conclusions about the link between system usage and individual performance.

We begin with a review of how system usage has been conceptualized at a high level in four IS domains: IS acceptance, IS implementation, IS success, and IS for decision making. Next, we shift to the lack of theory and lack of validation in prior usage studies. To remedy these problems, we advance a new, staged approach for reconceptualizing system usage, an approach that enables researchers to build measures that are better contextualized, more complete, and more valid. We then empirically investigate the approach by conceptualizing system usage in terms of its link with individual task performance. While other conceptualizations could be developed, a mapping between system usage and performance is key, given mixed results in this area and disagreements about its conceptualization (DeLone and McLean 2003). In addition, the present work (1) clarifies the richness of system usage measures, (2) demonstrates how different aspects of usage can be integrated to derive a more complete, contextualized construct, and (3) reports on an empirical test that validates the proposed approach.

## 2. Implicit Conceptualizations of System Usage in Past Research

Few constructs in IS have had as long a history as system usage (DeLone and McLean 1992, 2003). Figure 1 depicts the high-level conceptualizations of system usage in four research domains: IS success, IS acceptance, IS implementation, and IS for decision making.

In the IS success domain, researchers have measured usage as an independent variable (IV) or mediating variable leading to downstream impacts in order to determine how IT benefits individuals or organizations (DeLone and McLean 1992). In the IS for decision-making domain, system usage is primarily a dependent variable (DV), as in Barkin and Dickson's (1977) model (see Figure 1). Researchers typically study IS characteristics that improve user decision making. In the IS acceptance domain, researchers study system usage as a behavior determined by social and cognitive variables, with the goal of finding variables that explain most variance in usage. Theories employed to specify the range of antecedents include the theory of reasoned action, the theory of planned behavior, and the theory of social learning. Finally, system usage is a key DV in IS implementation research, that is, determining characteristics of IT implementations that lead to greater use of the final system. Although researchers sometimes choose measures specific to their research domains (e.g., use of information from IS in the decision-making domain, DeLone and McLean 1992; or use of IS to support specific tasks in the IS success domain, Doll and Torkzadeh 1998), researchers across these communities generally deploy similar usage measures. Long-standing measures include: features used, tasks supported, extent of use, use or nonuse, heavy or light use, frequency of use, and duration (see Table 1).

Despite this long-standing investigation of system usage, studies of its relationship with other constructs often report weak effects. With system usage as a DV, researchers have carefully examined a large number

**Figure 1** Past Conceptualizations of the System Usage Construct

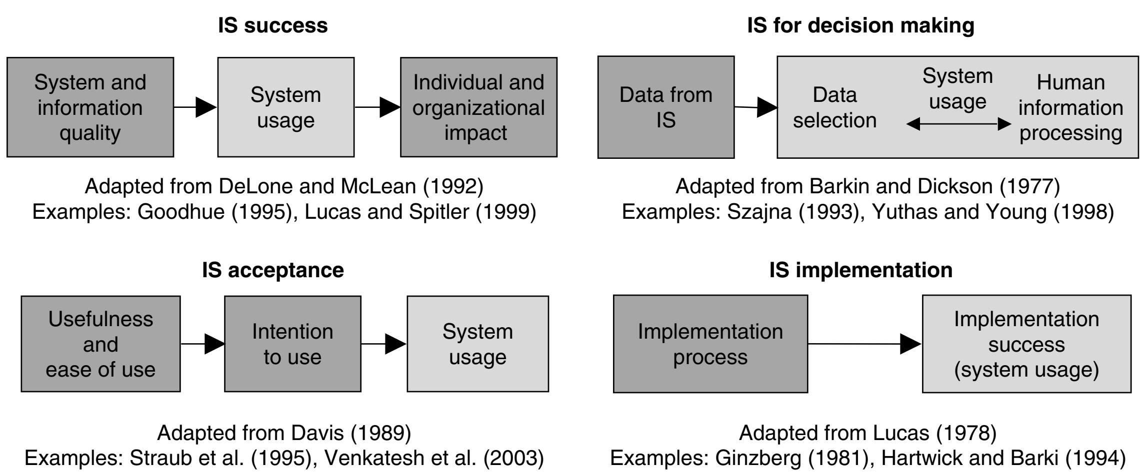


**Table 1 The Diversity of System Usage Measures Employed in Past Research**[†]

| Broad dimension | Individual measures | Used as IV | Used as DV |
|---|---|---|---|
| System usage measured as the use of information from an IS | | | |
| Extent of use | Number of reports or searches requested | ✓ | ✓ |
| Nature of use | Types of reports requested, general versus specific use | ✓ | |
| Frequency of use | Frequency of report requests, number of times discuss information | | ✓ |
| System usage measured as the use of an IS | | | |
| Method of use | Direct versus indirect | | ✓ |
| Extent of use | Number of systems, sessions, displays, functions, or messages; user's report of whether they are a light/medium/heavy user | ✓ | ✓ |
| Proportion of use | Percentage of times use the IS to perform a task | | ✓ |
| Duration of use | Connect time, hours per week | ✓ | ✓ |
| Frequency of use | Number of times use system (periods are: daily, weekly, etc.) | ✓ | ✓ |
| Decision to use | Binary variable (use or not use) | | ✓ |
| Voluntariness of use | Binary variable (voluntary or mandatory) | | ✓ |
| Variety of use | Number of business tasks supported by the IS | ✓ | ✓ |
| Specificity of use | Specific versus general use | | ✓ |
| Appropriateness of use | Appropriate versus inappropriate use | ✓ | ✓ |
| Dependence on use | Degree of dependence on use | ✓ | ✓ |

[†]Developed from a sampling of 48 articles in major IS journals in the period 1977–2005 (Burton-Jones 2005).

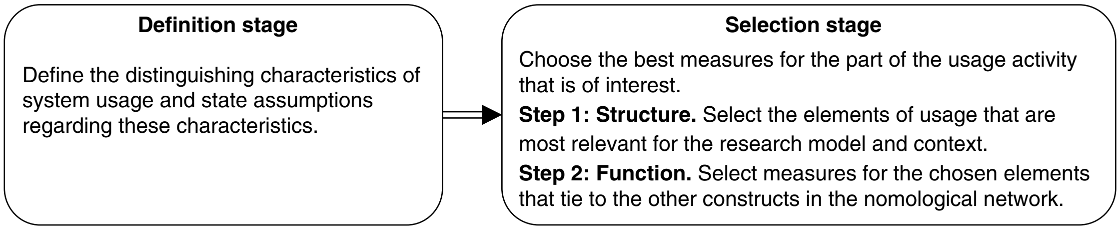
of antecedents (Adams et al. 1992), but explained variance is in a middling range, averaging around 30% (Meister and Compeau 2002). With system usage as an IV, studies of its downstream effects (e.g., on performance) report mixed results. Although progress has been made at an organizational level (Devaraj and Kohli 2003), findings at an individual level report that system usage can increase (Doll and Torkzadeh 1998), decrease (Pentland 1989, Szajna 1993), or have no effect on performance (Lucas and Spitler 1999).

## 3. The Need to Reconceptualize System Usage

Why is there a need to reconceptualize system usage? A review of the literature prompts two concerns: no theory and poor-to-no validation. Our evidence draws upon 48 empirical studies of individual-level system usage (Burton-Jones 2005). The chief limitation in past conceptualizations of system usage has been the atheoretical manner in which usage measures have been chosen. While many researchers carefully use theory to choose antecedents to usage (e.g., the theory of reasoned action; Venkatesh et al. 2003), few discuss how theory informs their choice of usage measures. With the exception of early decision-making studies that drew on information-processing theory (Barkin and Dickson 1977), we found no studies that expressed a strong theoretical basis for system usage, its appropriate empirical indicators, or its relationships with other constructs. The consequences of this dearth of theory can be seen in Table 1, which illustrates the diversity of usage measures in past research. In all, 14 broad measures as well as many minor variants are listed. In the presence of strong theory, diversity of measures is desirable (Campbell and Fiske 1959), but in its absence an abundance of measures is problematical and may have lured researchers into believing that there is no problem with measuring usage (Srinivasan 1985). We believe that the problems stemming from lack of theory are now coming to light, evident in the persistence of mixed results and lack of consensus on how to conceptualize system usage in IS success models (DeLone and McLean 2003).

In addition to this theoretical lacuna, our review found almost no validation of the usage construct. Surprisingly, while user satisfaction (the complement of system usage in DeLone and McLean's model), has undergone extensive instrument development (Doll and Torkzadeh 1988) and validation (Chin and Newsted 1995), work on system usage has not. Most studies select one or two usage measures from the many available (Burton-Jones 2005). A minority of studies use three or more measures and factor-analyze them to arrive at a composite measure of usage (Igbaria et al. 1997). Even in these instances, however, measures of system usage are chosen for their

**Figure 2** Staged Approach for Defining System Usage and Selecting Usage Measures




appearance in past empirical studies rather than for theoretical reasons.

## 4. A Staged Approach for Reconceptualizing System Usage

We suggest that the lack of theory underlying measures of usage in past research and the lack of validation of such measures manifest deeper problems:

- There is no accepted definition of the system usage construct in the IS literature.
- There is no accepted approach for selecting the relevant content of usage for any given study context.

To overcome these problems, this paper presents, for the first time, a systematic approach that enables researchers to reconceptualize system usage (Figure 2). By "reconceptualizing system usage," we mean that system usage is not the type of construct that can have a single conceptualization or measure. Unlike constructs that are strictly unidimensional or multidimensional with specific, known dimensions, we believe that relevant measures and dimensions of system usage will vary across contexts. In this light, having diverse conceptualizations of usage (as in past research, per Figure 1) is desirable. What is needed is a way to make such conceptualizations much more precise and explicit. Thus, while we believe that there cannot be a single, generally accepted conceptualization of system usage, we believe that there is great value in having an accepted approach for systematically developing conceptualizations of usage for specific contexts and selecting usage measures in a theoretically rigorous way. This paper presents such an approach in two stages (see Figure 2).

### 4.1. Defining System Usage

To conceptualize system usage, one must define it. Surprisingly, the IS field has no generally accepted definition of system usage. Granting that other definitions could be proffered, we propose that system usage is an activity that involves three elements: (1) a user, i.e., the subject using the IS, (2) a system, i.e., the object being used, and (3) a task, i.e., the function being performed.[1] Drawing on each element and recognizing that any IS comprises many features (Griffith 1999), we define individual-level system usage as an individual user's employment of one or more features of a system to perform a task. This definition has two implications. First, it distinguishes system usage from related but distinct constructs. For example, it suggests that system usage is distinct from information usage. In contrast to DeLone and McLean's (1992) definition of system usage and others (Table 1), we suggest that information usage is a useful construct, but it is not identical to system usage. System usage is also distinct from a user's decision to use or subsequent dependence on an IS and from user adoption. Even though many IT acceptance researchers have utilized such constructs as proxies for system usage (Table 1), one must not confuse a proxy for a construct. Finally, system usage is not an evaluation. Evaluations such as quality of use (Auer 1998) and appropriate use (Chin et al. 1997)

[1] Any definition of usage must rely on assumptions. Our assumptions about the elements of usage are as follows:

- A *user* is an individual person who employs an IS in a task. This implies that although users are social actors (Lamb and Kling 2003), we assume that it is possible to study user behavior at a purely individual level.
- An *IS* is an artifact that provides representations of one or more task domains. This implies that ISs provide features that are designed to support functions in those task domain(s) (Griffith 1999).
- A *task* is a goal-directed activity performed by a user. This implies that task outputs can be assessed in terms of predefined task requirements (Zigurs and Buckland 1998).

are useful constructs, but they do not measure system usage; instead, they measure the degree to which one's usage *corresponds* with another construct such as expected use or system "spirit" (Chin et al. 1997). Our definition implies that if one is to measure system usage itself, one must quantify it, not evaluate it.

The second implication of the definition is that it clarifies the content of system usage. Because system usage is a complex activity involving a user, IS, and task over time, it has a broad "universe of content" (Cronbach 1971). As Table 2 shows, one could use lean or rich measures to measure this content. Lean measures would attempt to capture the entire content of the activity in an omnibus measure such as use/nonuse, duration of use, or extent of use (see Table 2, Columns 1–2). Although such lean measures can be convenient, they are unfortunately inexact because they do not refer to the aspect of usage that may be most relevant in a specific context and it may not be clear to a respondent what part of the usage activity is actually being measured. Dubin (1978) called such omnibus measures of complex constructs "summative units" and warned against their employ (p. 66). In contrast to lean measures, rich measures incorporate the nature of the usage activity (see Table 2, Columns 3–6). To employ rich measures, one must have a way to select relevant content; this leads to the second stage: selection.

### 4.2. Selecting Content Valid, Contextualized Measures: A Two-Step Approach

As Table 2 shows, system usage always involves a system, but we argue that researchers can and should choose relatively rich measures to capture more or less of its use in a particular context. Some researchers may only be interested in the extent to which the system is used, without capturing much of the user or task context (Table 2, Model 3). Others may wish to include the user context by measuring the degree to which a user employs a system (Table 2, Model 4) or include the task context by measuring the degree to which the system is employed in the task (Table 2, Model 5). None of the approaches in Columns 3–6 is inherently superior. Rather, researchers must choose appropriate measures for their objective, theory, and methods.

Methods can, at times, be very restrictive. For example, a researcher may wish to use a very rich measure to capture all three elements of usage (per Table 2, Model 6). Although it is theoretically feasible to construct a single measure that captures each element of usage (i.e., system, user, and task), it is difficult to do so methodologically because the richness of the activity being measured makes it difficult to construct and cognitively difficult to respond to such a measure in practice. Here, a methodological compromise is to combine measures for the system, user, and task aspects of usage and create an aggregate higher-order construct to capture the entire activity (Law et al. 1998). We proffer an example of this strategy later in the paper.

As Figure 2 shows, this reasoning leads us to suggest a method for selecting measures of usage in future research. In other words, system usage can be attributed with a precise definition, but the definition refers to a broad range of content, only a subset of which will be relevant in a specific study. As different subsets will be relevant in different studies, one cannot create a single measure of usage, but one can define an approach for creating measures in such a way that they capture the most relevant content for a specific context (i.e., are content valid, yet contextualized). To define such an approach, we draw on Cronbach and Meehl's (1955) classic description of construct validity. According to Cronbach and Meehl, a construct's meaning is defined partly by its internal structure or makeup and partly by the other constructs in its nomological network. This implies a two-step method for selecting measures of system usage (per Figure 2):[2]

1. *Structure:* Select the elements of usage (i.e., the user, system, and/or task) that are most relevant for the research model and context.

2. *Function:* For the selected elements of usage, select measures of these elements that tie closely to the other construct(s) in the proposed nomological network.

[2] Another way of referring to these steps would be to refer to the "structure" step as "conceptualization" and the "function" step as "measurement." This would underscore that conceptualization precedes measurement and, by corollary, measures are meaningless without clear conceptualizations. We thank Reviewer 1 for this insight.

**Table 2 Rich and Lean Measures of System Usage**

| Richness of measures | 1. Very lean | 2. Lean | 3. Somewhat rich (IS) | 4. Rich (IS, User) | 5. Rich (IS, Task) | 6. Very rich (IS, User, Task) |
|---|---|---|---|---|---|---|
| Type | Presence of use | Extent of use (omnibus) | Extent to which the system is used | Extent to which the user employs the system | Extent to which the system is used to carry out the task | Extent to which the user employs the system to carry out the task |
| Domain of content measured* | Usage | Usage | Usage: System; User, Task | Usage: System; User, Task | Usage: System; User, Task | Usage: System; User, Task |
| Example | Use/nonuse | Duration; extent of use | Breadth of use (number of features) | Cognitive absorption | Variety of use (number of subtasks) | None to date (difficult to capture via a reflective construct) |
| Reference | Alavi and Henderson (1981) | Venkatesh and Davis (2000) | Saga and Zmud (1994) | Agarwal and Karahanna (2000) | Igbaria et al. (1997) | |

*Lean measures reflect usage alone; rich measures reflect its nature, involving the system, user, and/or task.

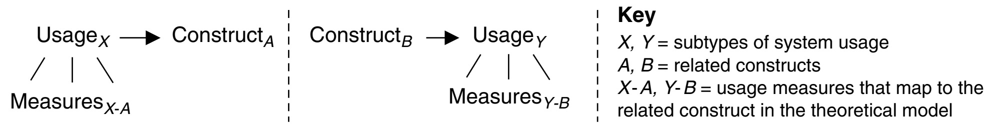
Four points might be raised in relation to this two-step method.[3] First, one might argue that the method is too relaxed because it could allow a researcher to use the term "system usage" when his or her measures only capture part of the construct. Although we recognize this criticism, we believe that system usage is so complex that researchers should be able to focus on just one or two elements of it (per Table 2, Columns 3–5), as long as they can justify what parts they select based on the context of their study. Certainly, researchers should be careful in such instances to keep in mind that they are only measuring part of the usage activity. The two-step method can help researchers (and readers) maintain such awareness.

Second, one might argue that because the two-step method enables different researchers to select different elements and measures of usage, this could lead to a proliferation of measures that all refer to system usage, but refer to different content, thus hindering cumulative progress. We believe that this criticism would be mistaken because diversity can help progress (Robey 1996), especially via diversity of measures (Boudreau et al. 2001), and this is perfectly consistent with how construct linkages are studied through such techniques as meta-analysis (Hunter and Schmidt 1990). Paraphrasing Landry and Banville (1992), what is needed is disciplined diversity. The distinction between the two-step method and existing practice is that the method enables researchers to select and validate their usage measures from a theoretical basis. We show this in Figure 3 by using subscripts to denote subtypes of usage appropriate for different contexts. By clarifying the subset of usage being measured, and theoretically justifying one's measures, cumulative progress will improve.

Third, one may question whether the two-step method should allow researchers to measure system usage via existing measures or whether all new measures should be created from scratch. A very strict view might state that because past studies offered no detailed definition and conceptualization of usage from which to build valid measures, perhaps all existing usage measures are invalid. We believe such a view would be extreme. Thus, the two-step method allows researchers to use existing measures (or create new measures) as long as they can justify the measures based on the study context. Certainly, some measures of system usage in past research are not valid measures of usage (per §4.1). Even so, we believe that some usage measures in past research, and even some measures that were not explicitly created to measure usage, can serve as valid usage measures. Consider cognitive absorption (CA). In Table 2, we list CA as a way to measure a user's engagement with an IS during use. However, Agarwal and Karahanna (2000) introduce CA to the IS literature not as a measure of system usage per se, but rather as an antecedent to perceived ease-of-use (PEOU) and

[3] We thank Reviewer 1 for helping us to articulate and address these important issues.

**Figure 3 Contextualizing the System Usage Construct**




perceived usefulness (PU). In their study, PEOU and PU were antecedents to intention-to-use, which they in turn conceptualized in an omnibus fashion (per Table 2, Column 2). In other words, the two-step method suggests that researchers can recast CA from (a) an indirect antecedent to an omnibus conceptualization of intention-to-use an IS to (b) a rich measure of system usage itself. Given our belief that omnibus conceptualizations of usage are not very useful (per §4.1), we believe that it is not only viable to recast CA in this way, but also useful to do so.

A final issue that could be raised in relation to the two-step method is the need to balance completeness with parsimony. That is, there will be times when a researcher believes that all elements of usage should be selected and multiple measures of each one should be taken but the practical realities of data collection require that she settle for less. As such, we recognize that it may sometimes be difficult to employ measures that the two-step method suggests are optimal. We do not believe that this is a weakness of the method. On the contrary, we believe that the method can serve to highlight the gap between what should be measured and what has been measured in a given study. By so doing, the method can help guide a research program in which researchers gain a deeper, more integrated understanding of system usage over time.

## 5. Empirical Investigation of the Staged Approach for Reconceptualizing Usage

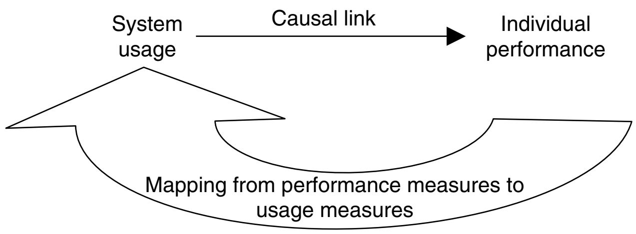
If the staged approach for reconceptualizing system usage is beneficial, a researcher should obtain more persuasive and meaningful results in a study of system usage if he follows the approach than if he does not. We empirically investigate this proposition to provide indicative (albeit, not definitive) support or refutation for the utility of the approach. To limit the scope of the investigation, we adopt the definition of system usage that we proposed in the definition stage above, and we focus on the benefit of following the steps in the selection stage.

To investigate the approach empirically, we must choose a theoretical and substantive context. The theoretical context that we chose was the relationship between system usage and short-run, individual task performance.[4] This is an important context because DeLone and McLean's IS success model (1992) suggests a link between system usage and individual task performance, but past studies of this link report mixed results (Pentland 1989, Lucas and Spitler 1999) and several scholars have called for more research to determine which usage measures are appropriate in this context (Chin and Marcolin 2001). The substantive context that we chose for the empirical investigation was analysts' use of spreadsheets for financial analysis. This is a crucial practical context because spreadsheets are among the most common end-user applications and decision-support tools in practice (Carlsson 1988, Panko 1998).

Following Figure 2, the first step when selecting usage measures is to define its structure. Because usage involves an IS, user, and task, the relevance of each element should be judged in light of the theoretical context. As financial analysis is a complex, cognitive activity, not a simple, mechanical task (Goodhue 1995), we expect that each element is relevant and, thus, a very rich usage measure is required (Table 2, Model 6). The second step is to choose measures for its elements that relate theoretically to the other constructs in its nomological network. There are two constructs in our case: system usage and performance. Therefore, we select usage measures by chaining backwards from performance measures to usage measures (per Figure 4).

[4] Past research suggests a distinction between the causes of short-run and long-run performance (March 1991). Although both are clearly important, we limit our empirical investigation to short-run performance.

**Figure 4** **Selecting Usage Measures via Nomological Network Analysis**




Following the logic in Figure 4, the next sections (§§5.1–5.4) describe the nature of individual task performance and propose measures of system usage that relate to it theoretically. Specifically, we identify a type of system usage (exploitive system usage) that relates theoretically to performance and we demonstrate how a very rich measure of this type of usage can be formed by combining two rich measures of system usage: *cognitive absorption* (that captures a user's employment of an IS, per Table 2, Model 4) and *deep structure usage* (that captures the use of the system for the task, per Table 2, Model 5). As proposed above, if the staged approach for measuring usage is beneficial, measures selected according to the approach should be superior to other measures. Thus, after defining performance and its related usage measures below, we report on an experiment designed to test whether our ability to explain the relationship between individual system usage and short-run task performance improves when richer measures are used. In other words, we test whether explanations are strongest (in terms of the amount of variance explained and the interpretability of relationships) when a very rich measure is employed (i.e., exploitive system usage), less strong when a rich measure is employed (i.e., cognitive absorption or deep structure usage alone), and even poorer when a lean measure is employed (i.e., duration).

### 5.1. Defining Individual Task Performance

In the performance measurement literature, job performance comprises two dimensions: task performance and contextual performance (Sonnentag and Frese 2002). *Task performance* consists of behaviors carried out to complete a job (Meister 1986); *contextual performance* consists of behaviors that contribute to the social and psychological climate in which a job is performed (Sonnentag and Frese 2002). Both are measured via assessments, but assessments of task performance are job specific, while assessments of contextual performance are not (Sonnentag and Frese 2002). Thus, when studying short-run task performance, one's measures must reflect the task under consideration. Assessments of task performance can be made in two ways: assessments of behavior, or outcomes (Campbell 1990, Sonnentag and Frese 2002). These can differ in complex scenarios such as group work where an individual's output is not under her complete control (Beal et al. 2003). However, for the purpose of this empirical investigation, we assess performance as an outcome because the individual user has complete control of her own work on her spreadsheet, i.e., her output does not depend on other people.

The outcome of one's task performance can be assessed in terms of effectiveness (Campbell 1990). Other assessments such as efficiency can also be made (Beal et al. 2003), but for reasons of scope we focus on effectiveness alone in this study. Thus, consistent with the performance measurement literature, we measure individual task performance as an assessment of individual task output in terms of its effectiveness, i.e., the degree to which it meets the task goals.

### 5.2. Mapping from Individual Task Performance to Existing Usage Measures

When we examine the relationship between individual task performance and the extant usage measures in Table 1, only 2 of the 14 measures are, according to the literature, related theoretically to task performance. Specifically, Szajna (1993) explains the benefit of examining the nature of information used (e.g., the benefit of requesting particular types of reports and by being more specific), while Nance (1992) explains the benefits of appropriate use. However, neither of these measures complies well with the definition of system usage because Szajna's (1993) measures relate to information usage rather than system usage, and Nance's (1992) measure was an evaluation of use rather than a measure of usage itself. Thus, new usage measures are needed that can map to performance. Following the logic in Figure 4, we next outline the type of use—exploitive usage—that is most conducive to a clear mapping to short-run task performance.

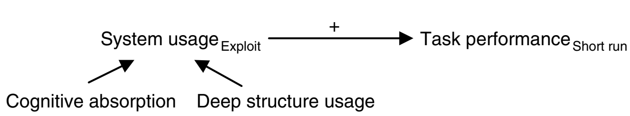
### 5.3. Types of System Usage that Relate to Individual Task Performance

Two types of system usage can drive individual task performance: exploitation and exploration (March 1991). *Exploitation* refers to routine execution of knowledge, whereas *exploration* refers to the search for novel or innovative ways of doing things. A balance between these is necessary for long-run performance, but exploitation is preferred in the short run because it has more predictable, immediate benefits (March 1991). In this illustrative case, the task is short, so our measures of usage must refer to exploitive use. Exploitive usage refers to usage that implements and executes one's knowledge of one's system and task. In a recent article, Subramani (2004) examined the benefit of measures of exploitive use at an organizational level, but no such measures have ever been employed at an individual level. Moreover, Subramani's (2004) measures of exploitive use refer only to the extent to which a system was used in specific tasks (per Table 2, Model 5). Because a very rich measure of usage is needed in our theoretical context (per Table 2, Model 6), the next section demonstrates how such a measure was constructed.

### 5.4. A Model of System Usage and Individual Task Performance

Figure 5 presents the proposed theoretical model for the empirical investigation. Individual task performance is envisioned as a reflective construct measured in terms of effectiveness. System usage is modeled as an aggregate higher-order construct with two subconstructs that together capture a user's employment of the system (cognitive absorption) and use of the system in the task (deep structure usage). Cognitive absorption represents the extent to which a user is absorbed when using the system (Agarwal and Karahanna 2000); deep structure usage represents the extent to which features in the system that relate to the core aspects of the task are used (DeSanctis and Poole 1994). We use these subconstructs to measure exploitive system usage, i.e., the extent to which the user exploits features of the system to perform the task. The subconstructs may be, but need not, be highly correlated. As such, they form the very rich, higher-order construct of system usage per Table 2, Model 6 (Edwards 2001, Law et al. 1998).[5] To maintain focus on the proposed approach, we do not present a detailed outline of each subconstruct here. Instead, these details are provided in the online supplement.[6]

**Figure 5** A Contextualized Model of System Usage and Individual Task Performance




*Notes.* The subscripts Exploit and Short run indicate that this theoretical model specifies the relationship between exploitive usage and short-run task performance (per Figure 3).

## 6. Empirical Test of the Staged Approach for Reconceptualizing System Usage

If the staged approach is beneficial, measures created with the help of the approach should perform more effectively than measures not selected with the help of the approach. Thus, in the present case, our ability to explain the relationship between system usage and short-run task performance would be best if we employ a very rich measure of system usage (i.e., exploitive system usage), less strong if we employ a rich measure of system usage alone (i.e., either cognitive absorption or deep structure usage), and poorest if we employ a lean measure of system usage such as "duration of use." We used an experiment to test this proposition. A field study would have increased external validity, but lab experimentation provided a tighter test of the proposed measures of usage (Calder et al. 1981).

### 6.1. Task and Design

The task required user subjects to build a spreadsheet model in MS Excel to determine the best approach for financing an asset purchase. The task enabled a strong test of the theoretical model as the analysis was cognitively engaging, which allows variation in

[5] The logic for constructing aggregate higher-order constructs is similar to the logic for constructing formative lower-order constructs, but in aggregate constructs each subconstruct can itself be reflective (Edwards 2001).

[6] An online supplement to this paper is available on the *Information Systems Research* website (http://isr.pubs.informs.org/ ecompanion.html).

cognitive absorption, and the system (Excel) contains features that directly support the task, which allows variation in deep structure usage. Because our interest is in the importance of rich measures but not in specific values on these measures, we adopted a free simulation design rather than a factorial design (Fromkin and Streufert 1976). Free simulations allow values of the IVs (e.g., cognitive absorption and deep structure usage) to vary freely over their natural range. This gives an insight into nature of the IV → DV relationship as well as the range over which it occurs.

### 6.2. Subjects

Subjects were 229 students in an intermediate accounting course in a university in the southern United States. The accounting course integrated a series of spreadsheet-based business analysis assignments into the intermediate and prerequisite introductory accounting course. Students were graded on four assignments in the introductory class and four in the intermediate class. Data were collected during an end-of-semester case exam worth 10% of the student's grade. The case used the same general format as previous assignments and involved accounting concepts learned during the course (present value, asset financing, and risk versus return). To the greatest extent possible, therefore, the system and task were believed to enable exploitive use by our subjects. Completion of the posttest instrument was voluntary; 177 were returned (response rate of 77%). Six cases with nonsensical responses were removed, leaving a full data set of $n = 171$.

### 6.3. Instrumentation—IV

Table 3 shows scales used to capture the IV for self-reported usage. To measure cognitive absorption, we adopted Agarwal and Karahanna's (2000) prevalidated scale of focused immersion. Given that Agarwal and Karahanna model cognitive absorption as a reflective, higher-order construct, the dimensions can be considered to be interchangeable (Edwards 2001, Jarvis et al. 2003). Thus, we selected just one of the dimensions (focused immersion) to balance parsimony of measurement with completeness of conceptualization.[7]

**Table 3 Measurement Scales**

| Construct | Items |
|---|---|
| Cognitive absorption (adapted from Agarwal and Karahanna 2000) | 8. When I was using MS Excel, I was able to block out all other distractions.<br>11. When I was using MS Excel, I felt totally immersed in what I was doing.<br>14. When I was using MS Excel, I got distracted very easily.*<br>21. When I was using MS Excel, I felt completely absorbed in what I was doing.<br>24. When I was using MS Excel, my attention did not get diverted very easily. |
| Deep structure usage (new scale) | 15. When I was using MS Excel, I did not use features that would help me analyze my data.*<br>17. When I was using MS Excel, I used features that helped me compare and contrast aspects of the data.<br>20. When I was using MS Excel, I used features that helped me test different assumptions in the data.<br>27. When I was using MS Excel, I used features that helped me derive insightful conclusions from the data.<br>30. When I was using MS Excel, I used features that helped me perform calculations on my data. |
| Objective measure of performance (reflective) | This relatively objective scale allocated a single percentage score based on marks for the following components: 1. identifying the problem; 2. building a flexible model; 3. correctly analyzing the data; 4. identifying solutions; 5. highlighting impacts; 6. creating a focused report; and 7. giving clear recommendation. The scale was created independently from the research by task experts and was assessed by independent coders. |

*Notes.* All self-report items used a nine-point strongly agree–strongly disagree Likert scale. Items reflect their sequence in the questionnaire.

*Negatively worded items were used to check for response bias. However, recent studies suggest that negatively worded items can change a construct's meaning (Motl and DiStefano 2002). Therefore, we excluded the two negatively worded items (Items 14 and 15) from our tests. This had no substantive effect on the results.

The scale for deep structure usage was created afresh. We defined *deep structure usage* as use of fea-

[7] As Reviewer 1 noted, we could be criticized here for not practicing what we preach because we did not include all five dimensions of cognitive absorption in Agarwal and Karahanna's (2000) study. However, as we note in §4.2, we believe that researchers must always balance parsimony with completeness. Because cognitive absorption is a reflective higher-order construct and focused immersion was the only subconstruct in Agarwal and Karahanna's (2000) study that was measured with items that referred to being absorbed, we believe that this decision is justified.

tures in the IS that support the underlying structure of the task. Items were first constructed by one of the researchers. Because the deep structure scale needed to be task centered, an independent domain expert (in this case, a course instructor) selected activities that described the task's underlying structure: analyzing data, testing assumptions, and deriving conclusions, and consistent with the approach in recent research (Subramani 2004), the items were adapted for this task domain (see Table 3). The items were set at a mid-range of task specificity to ask about the class of features used (i.e., deep structure features), but not specific features (e.g., present value functions). This balanced the need for broad applicability with the need for focused questions (Jasperson et al. 2005, Griffith and Northcraft 1994).[8]

To test the need to capture each component of usage (per Table 2), the instrument also included a lean, omnibus measure of duration of use. To measure duration of use, subjects were asked, "About how many minutes did you spend doing the case?" As the case was entirely computer based, this captured both usage duration and task duration. As with the self-report measure for deep structure usage, we also obtained an objective measure of usage duration.[9]

Once the self-reported items for system usage were drafted, a Q-sort exercise was used to improve construct validity (Moore and Benbasat 1991). The 10 usage measures and measures from other published scales were randomized and given to eight doctoral students. They were asked to separate items into bundles for each construct and to name each construct. Their feedback supported the validity of the scales; minor changes were made based on their feedback.

### 6.4. Instrumentation—DV

To reduce common method bias between IV and DV, an objective scale for overall task performance was developed independently from the research (Table 3). This scale assessed the degree to which an individual's output met the task requirements. Two independent coders rated participant performance using the scale, and the interrater reliability was high ($\text{ICC}(2, 2) = 0.87$).

### 6.5. Procedure, Pretest, and Pilot Test

In the experiment, subjects read the instructions (five minutes), performed the analysis task in MS Excel (90 minutes), and completed the questionnaire (15 minutes). To validate the procedure and instruments, we conducted a pretest with four students and a pilot test with 38 students.

## 7. Results of the Empirical Investigation

Data analysis proceeded in two steps. We first examined the descriptive statistics and the proposed measurement model, then the posited structural model. Both steps were performed using partial least squares (PLS). PLS was used in preference to LISREL software because LISREL is not suited to testing higher-order molar constructs in the presence of only one DV (Edwards 2001).

### 7.1. Descriptive Statistics

Table 4 details study descriptive statistics. The data for minutes included some values greater than the allowable time (90 minutes). As the results did not vary when these cases were deleted, all were kept in the final analysis. The variance inflation factors (VIF) from a regression of minutes, CA, and deep structure (DS) usage against performance ranged from 1.01–1.27, indicating no significant multicollinearity (Mathieson et al. 2001). Skewness and kurtosis and the normal probability plot from a regression of the three variables on performance-supported normality. We also checked for outliers, which were not an issue.

[8] A potential risk with the deep structure usage scale is that it assumes that subjects have knowledge of the system's deep structure. As each subject had previously completed eight similar cases, we believe this assumption is reasonable. However, to control for this risk, we obtained protocol data from a subsample of 46 users during the case by having Screen-Cam software video record their sessions. Two independent coders then coded the protocols, rating the degree to which each user employed the system's deep structure. The results for this data are stronger, but lead to the same conclusions as the self-report data regarding the value of rich measures versus lean measures of system usage (see online supplement).

[9] As an anonymous reviewer noted, self-reported measures of duration can be problematic. To control for limitations in the measure, we objectively measured usage duration by viewing the protocols of system usage referred to in Footnote 8. The objective data led to the same conclusions as the self-reported data (see online supplement).

**Table 4 Descriptive Statistics**

| Construct/Item | *N* | Mean | Std. deviation |
|---|---|---|---|
| Performance | 166 | 81.01 | 15.87 |
| Minutes | 166 | 81.07 | 19.99 |
| CA1 | 171 | 5.96 | 1.89 |
| CA2 | 171 | 5.78 | 1.80 |
| CA4 | 171 | 5.94 | 1.65 |
| CA5 | 171 | 5.73 | 1.68 |
| DS2 | 171 | 6.11 | 1.73 |
| DS3 | 171 | 6.08 | 1.68 |
| DS4 | 171 | 6.09 | 1.59 |
| DS5 | 171 | 6.98 | 1.56 |

*Notes*. Performance measured on a 0–100 scale. CA (cognitive absorption) and DS (deep structure) used a 1–9 scale.

## 7.2. Measurement Model

Tables 5–6 report tests of instrument validity and reliability. Table 5 supports scale validity because each item loaded on its construct significantly ($p < 0.01$) and more highly than 0.70 (Hair et al. 1998) and loaded less highly on the other constructs. Table 6 provides further support for construct validity because the square root of the average variance shared between each construct and its indicators is higher than 0.50 and in all cases is higher than the variance it shares with the other constructs (Fornell and Larcker 1981). In terms of reliability, Table 5 indicates that both scale reliabilities are higher than Nunnally's (1967) minimum guideline of 0.60. Therefore, the results suggest that the validity and reliability of the data are adequate for testing the structural model.

**Table 5 Loadings, Cross-Loadings, and Reliability**

| Item | CA | DS | Minutes | Reliability |
|---|---|---|---|---|
| CA4 | **0.84** | 0.45 | 0.11 | CA: Cron. $\alpha = 0.81$ |
| CA2 | **0.81** | 0.36 | 0.03 | CR = 0.69 |
| CA5 | **0.81** | 0.36 | −0.01 | |
| CA1 | **0.73** | 0.45 | 0.06 | |
| DS4 | 0.53 | **0.86** | 0.03 | DS: Cron. $\alpha = 0.82$ |
| DS3 | 0.31 | **0.82** | −0.04 | CR = 0.70 |
| DS2 | 0.39 | **0.81** | −0.02 | |
| DS5 | 0.38 | **0.73** | −0.01 | |
| Minutes | 0.06 | −0.01 | **1.00** | NA |

*Notes*. All item-to-construct loadings are significant ($p < 0.05$). Loadings, cross-loadings, and composite reliability (CR) obtained from PLS; Cronbach's alpha obtained from SPSS software.

**Table 6 Interconstruct Correlations and Average Variance Extracted**

| | Cognitive absor. | Deep structure | Minutes | Performance |
|---|---|---|---|---|
| Cognitive absor. | **0.80** | | | |
| Deep structure | 0.51 | **0.81** | | |
| Minutes | −0.08 | −0.03 | **1.00** | |
| Performance | 0.37 | 0.46 | −0.29 | **1.00** |

*Note*. The bolded values on the diagonal are the square root of each construct's average variance extracted (AVE) and should be higher than 0.50.

## 7.3. Structural Model and Nomological Validity

Table 7 reports results for nomological validity. As PLS does not provide overall goodness-of-fit tests, one examines $R^2$ values (Mathieson et al. 2001). To test the effect of modeling usage as a higher-order construct, we tested two models (Edwards 2001): one included both subconstructs as independent components, and a second formed a higher-order construct using the factor scores of cognitive absorption and deep structure usage as formative indicators.

In a separate analysis (not shown to conserve space), we tested another higher-order model using the method of repeated indicators (Chin et al. 2003, Lohmoller 1989). The results were consistent with those using the formative model. Finally, we ran each model in PLS and stepwise regression controlling for important predictors of task performance and the results did not change (see online supplement). Overall, three findings in Table 7 are noteworthy:

1. The lean usage measure (duration) has a significant negative relationship with performance.

2. The rich usage measures (cognitive absorption and deep structure usage) both positively affect performance and each yields more than twice the variance explained by duration.

3. A very rich measure of usage (exploitive usage) that captures the user, system, and task aspects of use yields almost three times the variance explained by the lean measure, and the results are similar whether usage is modeled as a higher-order construct or as a combination of components.

As several models in Table 7 are nested, one can statistically compare the degree to which each usage measure explains performance. Table 8 shows the results of this test. Consistent with our predictions, the results suggest that excluding either the user or task aspects

**Table 7 PLS Structural Models**

| Measurement approach | Model | Results |
|---|---|---|
| Extent of use (omnibus): Table 2, Model 2 | Minutes → Performance | $B_M = -0.29$, $t = -4.00^{**}$, $R^2 = 0.087$ |
| Extent to which the user employs the system: Table 2, Model 4 | Cognitive absorption → Performance | $B_{CA} = 0.42$, $t = 7.11^{**}$, $R^2 = 0.178$ |
| Extent to which the system is used to carry out the task: Table 2, Model 5 | Deep structure usage → Performance | $B_{DS} = 0.47$, $t = 8.97^{**}$, $R^2 = 0.218$ |
| Extent to which the user employs the system to carry out the task: Table 2, Model 6 | Component model<br>Cognitive absorption → Performance<br>Deep structure usage → Performance | $B_{CA} = 0.25$, $t = 3.39^{**}$<br>$B_{DS} = 0.34$, $t = 4.77^{**}$<br>$R^2 = 0.264$ |
| | Higher-order model*<br>Usage → Performance<br>Cognitive absorption → Usage<br>Deep structure usage → Usage | $B_U = 0.51$, $t = 9.57^{**}$<br>$\text{Weight}_{CA} = 0.83$, $\text{Weight}_{DS} = 0.90$<br>$R^2 = 0.262$ |

B: the coefficient between an antecedent and performance.
*Weight*: the weight of the subconstruct on the higher-order usage construct in PLS.
*The higher-order model was constructed using factor scores for each subconstruct. We also ran this test using averages rather than factor scores; the results were not substantively different.
**All $t$-values significant at $p < 0.01$.

of use leads to a significant (small-to-medium) reduction in $R^2$. Excluding the rich usage measures altogether and relying solely on a lean measure (duration) leads to a large reduction in $R^2$ and a change in the direction of the relationship between usage and performance. Although speculations can be drawn post hoc regarding possible reasons for a negative relationship between duration and performance, we do not believe that this relationship is highly interpretable, i.e., speculations could also have been made if we had found the relationship to be positive.

Overall, the results strongly support the proposed two-step method for selecting usage measures. The implications of these results are considered next.

## 8. Discussion

This paper presents a systematic attempt to define, conceptualize, and measure the system usage con-

**Table 8 Impact of Excluding Usage Measures**

| Test | Models compared: Full model | Models compared: Partial (nested) model | Change in $R^2$ | Effect size†† |
|---|---|---|---|---|
| Impact of not measuring the IS, user, and task (only measuring duration) | Perf = CA, DS, Mins (Table 2, Models 6 and 2) | Perf = Mins (Table 2, Model 2) | $0.25^{**}$ | $f^2 = 0.37$ Large |
| | Perf = Usage†, Mins (Table 2, Models 6 and 2) | Perf = Mins (Table 2, Model 2) | $0.24^{**}$ | $f^2 = 0.36$ Large |
| Impact of measuring the user/IS but not the IS/task | Perf = CA, DS (Table 2, Model 6) | Perf = CA (Table 2, Model 4) | $0.09^{**}$ | $f^2 = 0.12$ Small-medium |
| Impact of measuring the IS/task but not the user/IS | Perf = CA, DS (Table 2, Model 6) | Perf = DS (Table 2, Model 5) | $0.05^{**}$ | $f^2 = 0.06$ Small-medium |

**Sig. at $p < 0.01$, †Usage is formed with the factor scores of CA and DS as formative indicators, per Table 7.
††Each construct's effect size ($f^2$) can be calculated by the formula $(R^2_{full} - R^2_{partial})/(1 - R^2_{full})$ (Mathieson et al. 2001, Chin et al. 2003). According to Mathieson et al. (2001), multiplying $f^2$ by $(n - k - 1)$, where $n$ is the sample size (171) and $k$ is the number of independent variables, provides a pseudo $F$ test for the change in $R^2$ with 1 and $n - k$ degrees of freedom. An effect size of 0.02 is small, 0.15 is medium, and 0.35 is large (Cohen 1988).

**Table 9 Research Contributions and Implications**

| Element of research | Contribution |
|---|---|
| Staged approach | Provides a way to explicitly conceptualize and reconceptualize system usage in IS research |
| Definition stage | Enables researchers to distinguish between system usage and other constructs, and to specify the content of the system usage construct |
| Selection stage | Enables researchers to select measures of system usage that minimize errors of inclusion and omission |
| Empirical investigation of the staged approach | Provides initial, indicative evidence that the staged approach is beneficial and provides a validated measurement model of system usage for an important, practical domain |

struct. It contributes in four ways, summarized in Table 9, which we discuss in turn.

The core contribution of the research is that it provides a way for IS researchers to explicitly reconceptualize system usage. Although system usage has long been a central construct in IS research, past conceptualizations of it have remained implicit, there is no accepted definition of the construct, and there is no standard approach for selecting or validating its measures. The staged approach that we advance acknowledges the complexity of system usage and the diverse contexts in which it can be studied (per Figure 1), and presents the first systematic approach for conceptualizing and contextualizing system usage. This assists IS research in two ways. First, it provides a way for researchers to select precise measures of usage and thereby obtain more meaningful findings about the relationship between system usage and its antecedents and consequences in specific contexts. Second, by encouraging researchers to explicate the theory and assumptions behind their choice of usage measures, the approach should support cumulative research in IS by enabling researchers to achieve a more integrated understanding of system usage across different contexts (e.g., by supporting meta-analyses of usage research).

Each stage of the proposed approach offers additional concrete contributions. For example, the definition stage enables researchers to distinguish system usage from other constructs. In past research, many studies have employed other constructs (e.g., dependence on use, or information usage, per Table 1) as proxies for system usage (Goodhue and Thompson 1995, Szajna 1993), or conversely have employed system usage as a proxy for other constructs, e.g., IT acceptance (Trice and Treacy 1986). There is nothing inherently wrong with this practice except that in past research it has been implicit rather than explicit, and there remains a lack of evidence for which proxies are accurate and which are not. Consider the technology acceptance model (TAM). TAM explains IT acceptance, but the DVs in TAM are usage intentions and usage behavior (Davis 1989). It is not clear that either of these constructs completely captures the notion of acceptance (Trice and Treacy 1986). Our research highlights the need for researchers to provide systematic evidence for which usage measures, if any, are valid proxies for related constructs and to determine which other constructs, if any, are good proxies for system usage.

The selection stage contributes by providing a way to reduce errors of inclusion and omission when measuring usage. Errors of inclusion occur when a researcher includes irrelevant aspects of usage or another construct in his usage measures, while errors of omission occur when a researcher omits key elements of usage from his usage measures (e.g., omitting cognitive absorption from measures of exploitive use). The two-step selection method reduces the possibility of both errors. It also strongly cautions against the employment of lean or omnibus usage measures. Lean measures such as use/nonuse, duration of use, and extent can increase errors of inclusion and omission because they obscure (1) what constitutes usage, and (2) what part of usage the researcher intends to measure. Thus, subjects who respond to a lean usage measure may have a broader or narrower view of usage than intended by the researcher, leading to systematic errors in their responses. Lean measures can also risk errors of inclusion by simply reflecting different constructs. For example, task duration can often equate to usage duration, as it did in our empirical test. The error of inclusion in our investigation was so strong that it changed the direction of the estimated relationship between usage and performance. As omnibus measures have such significant limitations, we suggest that they be used very cautiously if at all.

The proposed two-step method provides a way to move beyond lean measures by advancing a systematic way to create contextualized usage measures.

For too long, we submit, IS researchers have studied system usage without specifying and theoretically justifying the type of usage being studied. Nevertheless, it is clear that different types of usage could be more relevant in different contexts. The two-step method provides a way for researchers to develop contextualized usage measures by specifying which elements (i.e., system, user, task) and which measures of usage are most relevant for a theoretical context. For example, user cognitive absorption may be a highly relevant metric for line managers who depend on performance outcomes from employees' usage, but it may be of little relevance to system administrators who must make decisions based on how systems are actually being utilized (i.e., in terms of system load) irrespective of employee cognitions. Moreover, even if the elements of usage are the same in two contexts, different measures may be needed. For example, when studying long-run rather than short-run performance, one would need additional measures to capture exploratory and exploitive use (March 1991, Subramani 2004). Thus, much research is needed to systematically identify managerially relevant subtypes of usage, define appropriate measurement models for these contexts, and theorize the antecedents and consequences of these subtypes of use, rather than, or at least in addition to, studying the antecedents and consequences of system usage in general.

Despite these contributions, we should note two criticisms of the approach.[10] First, one might argue that the approach is obvious. That is, of course research will improve if researchers define, conceptualize, and measure constructs carefully. It is hard to argue against this. Yet, if the approach is so obvious, why are there no accepted definitions, detailed conceptualizations, or rigorous approaches for selecting measures of usage in the literature? In our view, Zigurs (1993, p. 117) correctly identifies the reason when she noted that system usage is a "deceptively simple" construct. Thus, while our approach may be obvious to some, most researchers (including ourselves) have not given the measurement of system usage enough attention in past studies. We believe that the proposed approach can help improve this situation.

Second, one might argue that the approach provides insufficient guidance to help researchers use it easily in practice. In other words, although we demonstrated its use in one context, it would have been useful if we had shown how it could apply to a wide range of other contexts (e.g., different theories, users, tasks, and systems). For example, what usage measures should a researcher select if she is employing learning theories to study analysts' use of a decision support system (DSS) in a hospital? Because the approach offers no simple recipe, it would be a nontrivial exercise to identify the appropriate measures in this case. Nevertheless, we believe that the approach can give useful guidance. Consider Devaraj and Kohli's (2003) impressive study of usage and performance in hospitals. They measured analysts' use of a DSS in hospitals via three measures: CPU time, number of reports retrieved, and number of disk inputs/outputs (I/Os). Our proposed approach would have required researchers to consider other parts of usage, such as the user's engagement with the system and the use of features that supported certain tasks. It would be interesting to see if these measures would have led to different conclusions to those in Devaraj and Kohli's work. Another interesting point is that Devaraj and Kohli aggregated individual usage data to measure usage at the organizational (hospital) level. Our approach is limited to the individual level. Nevertheless, it is possible that researchers could extend the approach to support other levels of analysis. For example, drawing on multilevel theory, one could argue that although system usage at an individual level involves a user, system, and task (per Table 2), system usage at a collective (e.g., group or organizational) level may be more than the sum of its parts, e.g., the interdependencies that occur among users may be another important element of usage (Burton-Jones 2005). In short, in the case of Devaraj and Kohli's (2003) hospital setting, the proposed approach would suggest alternative measures of usage at an individual level and potentially other additional measures of usage at an organizational level. Therefore, while our approach does not provide a simple recipe, we believe that it

[10] We thank our anonymous review team for helping us to identify these criticisms.

can be used and extended in a systematic way to reexamine existing studies and plan new studies.

In addition to these theoretical contributions, the empirical investigation also offers several contributions. Most importantly, it provides validated usage measure for an important, practical context: the relationship between system usage and short-run performance in cognitively engaging tasks. The investigation also revealed ways to improve empirical studies of usage. For example, it suggests that the two-step approach could be used to help researchers select methods for collecting usage data because objective measures may be more able to measure the system and task aspects of usage, but self-report questionnaires may be more able to measure user states such as cognitions or emotions during usage (see the online supplement) (Hilbert and Redmiles 2000). Our test also highlights the need to determine when higher-order models or component models of usage are appropriate. A higher-order model can never increase statistical explanations of a DV over an optimally weighted combination of components (Edwards 2001). Even so, we prefer the higher-order model because it maps more closely to the theoretical specification of usage in our study. The higher-order model also explained performance to the same degree as the component model in our tests (per Table 7). However a higher-order model may not always be best. If we examined long-run rather than short-run performance, we would have needed to include subconstructs for both exploratory and exploitive use and create a third-order model to capture both subconstructs. Whether such a third-order model of usage could be developed is not clear. Thus, more research is needed to determine when an overarching construct of usage can be constructed and when it is best modeled as a combination of components (per Edwards 2001).

When assessing the contributions of our empirical investigation, it is important to note its limitations. In terms of construct validity, our work could be extended to see if additional aspects of the user, IS, or task contexts could be modeled. For example, we measured user cognitions, but some noncognitive elements such as a user's affective state may have been relevant. In terms of internal validity, we did not consider antecedents to usage or potential mediators between usage and performance such as learning. Such research would be valuable because tying each usage measure to relevant causes and consequences could allow researchers to develop more complete models of IS success (DeLone and McLean 2003). Finally, in terms of external validity, we utilized student subjects and examined just one task; examining a range of tasks in the field would be useful. One could also argue that our deep-structure usage measure lacks external validity because operationalizing it requires researchers to create items that reflect the deep structure of the IS and task under investigation. This problem only occurs at the level of measures, however, not of constructs. For example, Goodhue (1995) measured task-technology-fit (TTF) in the context of data management, and although his measures are not easily generalizable to other domains, TTF is a generalizable construct (Lee and Baskerville 2003). Some even argue that researchers should consider creating even more specific feature-level measures (Jasperson et al. 2005).

Finally, the paper has important practical contributions. There has long been a lack of good metrics in IS practice (Strassman et al. 1988). A recent comprehensive review of IT metrics found 31 widely used metrics in IS practice, of which only two related to system usage (Chidambaram et al. 2005). These two metrics of system usage were lean measures of duration: CPU hours utilized and hours logged per employee. The approach proposed in this paper and its empirical results suggest that such lean metrics will not provide very meaningful insights into how use of an organizational IS leads to important outcomes such as employee performance. We therefore suggest that the proposed approach in this paper can contribute to practice by helping organizations select metrics of system usage that can explain relevant organizational outcomes from using systems (e.g., performance, satisfaction, quality of work life, and so on).

## 9. Conclusion

To overcome the lack of explicit conceptualizations of system usage in past research, the present study advances a staged approach for reconceptualizing it. The first stage, definition, recommends that researchers explicitly define system usage and its assumptions. The second stage, selection, recommends that researchers select usage measures by a two-step method that involves identifying the relevant elements of usage for a research context (i.e., IS, user, and/or task) and identifying measures for these elements based on the other constructs in the nomological network.

In the present study, we demonstrate how such an approach would work by means of an empirical investigation in which we examined the degree to which lean usage measures and rich usage measures explain the relationship between system usage and task performance in cognitively engaging tasks. The results strongly support the staged approach and indicate that inappropriate choices of usage measures can significantly reduce explanations of performance, even causing the estimated relationship between usage and performance to change direction.

Despite acknowledged limitations, we believe the staged approach advanced in this paper helps to clarify the meaning of system usage and the range and dimensionality of past usage measures. Given contradictory results in past studies of system usage and performance, and the centrality of the usage construct in past research, our focused reconceptualization of the construct should enable more informed research into the pathways by which IT impacts individuals at work.

## Acknowledgments

This paper stems from the first author's doctoral dissertation. The authors thank his committee: Jeff Hubona, Mike Gallivan, Arun Rai, and Dan Robey. The paper has benefited from presentations at Bentley College, Georgia State University, Hong Kong University of Science and Technology, University of British Columbia, University of Delaware, University of Georgia, University of Houston, University of Kansas, University of Texas at Dallas, Washington State University, and the International Conference on Information Systems (ICIS) Doctoral Consortium (2004). The authors thank Henri Banki and Rick Watson for helpful comments and are particularly indebted to Dale Goodhue and Elena Karahanna for detailed comments on several drafts. The authors thank Laurie Kirsch (SE), George Marakas (AE), and the three reviewers for *Information Systems Research* for exemplary reviews that greatly enhanced the quality of the paper.

## References

Adams, D. A., R. R. Nelson, P. A. Todd. 1992. Perceived usefulness, ease of use, and usage of information technology: A replication. *MIS Quart.* **16**(2) 227–247.

Agarwal, R., E. Karahanna. 2000. Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quart.* **24**(4) 665–694.

Alavi, M., J. C. Henderson. 1981. An evolutionary strategy for implementing a decision support system. *Management Sci.* **27**(11) 1309–1322.

Auer, T. 1998. Quality of IS use. *Eur. J. Inform. Systems* **7**(3) 192–201.

Barkin, S. R., G. W. Dickson. 1977. An investigation of information system utilization. *Inform. Management* **1**(1) 35–45.

Beal, D. J., R. R. Cohen, M. J. Burke, C. L. McLendon. 2003. Cohesion and performance in groups: A meta-analytic clarification of construct relations. *J. Appl. Psych.* **88**(6) 989–1004.

Boudreau, M.-C., D. Gefen, D. W. Straub. 2001. Validation in information systems research: A state-of-the-art assessment. *MIS Quart.* **25**(1) 1–16.

Burton-Jones, A. 2005. New perspectives on the system usage construct. Doctoral dissertation, Department of Computer Information Systems, Georgia State University, Atlanta, GA.

Calder, B. J., L. W. Phillips, A. M. Tybout. 1981. Designing research for application. *J. Consumer Res.* 8 197–207.

Campbell, D. T., D. W. Fiske. 1959. Convergent and discriminant validity by the multitrait-multimethod matrix. *Psych. Bull.* 56 81–105.

Campbell, J. P. 1990. Modeling the performance prediction problem in industrial and organizational psychology. M. D. Dunnette, L. M. Hough, eds. *Handbook of Industrial and Organizational Psychology*, 2nd ed. Consulting Psychologists' Press, Palo Alto, CA, 687–732.

Carlsson, S. A. 1988. A longitudinal study of spreadsheet program use. *J. Management Inform. Systems* **5**(1) 82–100.

Chidambaram, L., R. W. Zmud, M. Karahannas. 2005. Measuring the business value of information technology (IT): A review and analysis of IT metrics. S. S. Kambhammettu, ed. *Business Performance Measurement: Toward Organizational Excellence*. Le Magnus University Press, Hyderabed, India, 1–16.

Chin, W. W., B. L. Marcolin. 2001. The future of diffusion research. *DATA BASE Adv. Inform. Systems* **32**(3) 8–12.

Chin, W. W., P. R. Newsted. 1995. The importance of specification in causal modeling: The case of end-user computing satisfaction. *Inform. Systems Res.* **6**(1) 73–81.

Chin, W. W., A. Gopal, W. D. Salisbury. 1997. Advancing the theory of adaptive structuration: The development of a scale to measure faithfulness of appropriation. *Inform. Systems Res.* **8**(4) 342–367.

Chin, W. W., B. L. Marcolin, P. 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. *Inform. Systems Res.* **14**(2) 189–217.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Lawrence Erlbaum, Hillsdale, NJ.

Cronbach, L. J. 1971. Test validation. R. L. Thorndike, ed. *Educational Measurement*, 2nd ed. American Council on Education, Washington, D.C., 443–507.

Cronbach, L. J. P. E. Meehl. 1955. Construct validation in psychological tests. *Psych. Bull.* **52**(4) 281–302.

Davis, F. 1989. Perceived usefulness, perceived ease of use, and end user acceptance of information technology. *MIS Quart.* **13**(3) 318–339.

DeLone, W. H., E. R. McLean. 1992. Information systems success: The quest for the dependent variable. *Inform. Systems Res.* **3**(1) 60–95.

DeLone, W. H., E. R. McLean. 2003. The DeLone and McLean model of information systems success: A ten-year review. *J. Management Inform. Systems* **19**(4) 9–30.

DeSanctis, G., M. S. Poole. 1994. Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organ. Sci.* **5**(2) 121–147.

Devaraj, S., R. Kohli. 2003. Performance impacts of information technology: Is actual usage the missing link? *Management Sci.* **49**(3) 273–289.

Doll, W. J., G. Torkzadeh. 1988. The measurement of end-user computing satisfaction. *MIS Quart.* **12**(2) 259–274.

Doll, W. J., G. Torkzadeh. 1998. Developing a multidimensional measure of system-use in an organizational context. *Inform. Management* **33**(4) 171–185.

Dubin, R. 1978. *Theory Building*, rev. ed. Free Press, New York.

Edwards, J. R. 2001. Multidimensional constructs in organizational behavior research: An integrative analytic framework. *Organ. Res. Methods* **4**(2) 144–192.

Fornell, C., D. Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement error. *J. Marketing Res.* **18**(1) 39–50.

Fromkin, H. L., S. Streufert. 1976. Laboratory experimentation. M. D. Dunnette, ed. *Handbook of Industrial and Organizational Psychology*. Rand McNally, Chicago, IL, 415–465.

Ginzberg, M. J. 1981. Early diagnosis of MIS implementation failure. *Management Sci.* **27**(4) 459–478.

Goodhue, D. L. 1995. Understanding user evaluations of information systems. *Management Sci.* **41**(12) 1827–1844.

Goodhue, D. L., R. L. Thompson. 1995. Task-technology fit and individual performance. *MIS Quart.* **19**(2) 213–236.

Griffith, T. L. 1999. Technology features as triggers for sensemaking. *Acad. Management Rev.* **24**(3) 472–488.

Griffith, T. L., G. B. Northcraft. 1994. Distinguishing between the forest and the trees: Media, features, and methodology in electronic communications research. *Organ. Sci.* **5**(2) 272–285.

Hair, J. F., R. E. Anderson, R. L. Tatham, W. C. Black. 1998. *Multivariate Data Analysis*, 5th ed. Prentice Hall, Upper Saddle River, NJ.

Hartwick, J. H., H. Barki. 1994. Explaining the role of user participation in information system use. *Management Sci.* **40**(4) 440–465.

Hilbert, D. M., D. F. Redmiles. 2000. Extracting usability information from user interface events. *ACM Comput. Surveys* **32**(4) 384–421.

Hunter, J. E., F. L. Schmidt. 1990. *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. Sage, Newbury Park, CA.

Igbaria, M., M. Tan. 1997. The consequences of information technology acceptance on subsequent individual performance. *Inform. Management* **32**(3) 113–121.

Igbaria, M., N. Zinatelli, P. Cragg, A. L. M. Cavaye. 1997. Personal computing acceptance factors in small firms: A structural equation model. *MIS Quart.* **21**(3) 279–305.

Jarvis, C. B., S. B. MacKenzie, P. M. Podsakoff. 2003. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *J. Consumer Res.* **30**(2) 199–218.

Jasperson, J., P. E. Carter, R. W. Zmud. 2005. A comprehensive conceptualization of the post-adoptive behaviors associated with IT-enabled work systems. *MIS Quart.* **29**(3) 525–557.

Lamb, R., R. Kling. 2003. Reconceptualizing users as social actors in information systems research. *MIS Quart.* **27**(2) 197–235.

Landry, M., C. Banville. 1992. A disciplined methodological pluralism for MIS research. *Accounting, Management Inform. Tech.* **2**(2) 77–97.

Law, K. S., C.-S. Wong, W. H. Mobley. 1998. Toward a taxonomy of multidimensional constructs. *Acad. Management Rev.* **23**(4) 741–755.

Lee, A. S., R. Baskerville. 2003. Generalizing generalizability in information systems research. *Inform. Systems Res.* **14**(3) 221–243.

Lohmoller, J.-B. 1989. *Latent Variable Path Modeling with Partial Least Squares*. Physica-Verlag, Heidelberg, Germany.

Lucas, H. C. 1978. Empirical evidence for a descriptive model of implementation. *MIS Quart.* **2**(2) 27–41.

Lucas, H. C., V. K. Spitler. 1999. Technology use and performance: A field study of broker workstations. *Decision Sci.* **30**(2) 1–21.

March, J. G. 1991. Exploration and exploitation in organizational learning. *Organ. Sci.* **2**(1) 71–87.

Marcolin, B. L., D. Compeau, M. C. Munro, S. L. Huff. 2001. Assessing user competence: Conceptualization and measurement. *Inform. Systems Res.* **11**(1) 37–60.

Mathieson, K., E. Peacock, W. W. Chin. 2001. Extending the technology acceptance model: The influence of perceived user resources. *DATA BASE Adv. Inform. Systems* **32**(3) 86–112.

Meister, D., ed. 1986. *Human Factors Testing and Evaluation*. Elsevier, West Lafayette, IN.

Meister, D. B., D. R. Compeau. 2002. Infusion of innovation adoption: An individual perspective. *Annual Conf. Admin. Sci. Assoc. Canada* (ASAC), Winnipeg, Canada, (May 25–28) 23–33.

Melone, N. 1990. A theoretical assessment of the user-satisfaction construct in information systems research. *Management Sci.* **36**(1) 76–91.

Moore, G. C., I. Benbasat. 1991. Development of an instrument to measure the perceptions of adopting and information technology innovation. *Inform. Systems Res.* **2**(3) 192–222.

Motl, R. W., C. DiStefano. 2002. Longitudinal invariance of self-esteem and method effects associated with negatively worded items. *Structural Equation Model.* **9**(4) 562–578.

Nance, W. D. 1992. Task/technology fit and knowledge worker use of information technology: A study of auditors. Doctoral dissertation, University of Minnesota, MN.

Nunnally, J. C. 1967. *Psychometric Theory.* McGraw-Hill, New York.

Panko, R. R. 1998. What we know about spreadsheet errors. *J. End User Comput.* **10**(2) 15–21.

Pentland, B. T. 1989. Use and productivity in personal computing: An empirical test. *10th Internat. Conf. Inform. Systems.* Boston, MA (December) 211–222.

Robey, D. 1996. Diversity in information systems research: Threat, promise, and responsibility. *Inform. Systems Res.* **7**(4) 400–408.

Saga, V. L., R. W. Zmud. 1994. The nature & determinants of information technology infusion at an organizational level of analysis. *Acad. Management, Distinguished Poster Session, Technology & Innovation Management*, Dallas, TX (August).

Seddon, P.B. 1997. A respecification and extension of the DeLone and McLean model of IS success. *Inform. Systems Res.* **8**(3) 240–253.

Sonnentag, S., M. Frese. 2002. Performance concepts and performance theory. S. Sonnentag, ed. *Psychological Management of Individual Performance*. Wiley, UK, 3–25.

Srinivasan, A. 1985. Alternative measures of system effectiveness: Associations and implications. *MIS Quart.* **9**(3) 243–253.

Strassman, P. A., P. Berger, E. B. Swanson, C. H. Kriebel, R. J. Kauffman. 1988. *Measuring Business Value of Information Technologies*. ICIT Press, Washington, D.C.

Straub, D., M. Limayem, E. Karahanna-Evaristo. 1995. Measuring system usage: Implications for IS theory testing. *Management Sci.* **41**(8) 1328–1342.

Subramani, M. 2004. How do suppliers benefit from information technology use in supply chain relationships. *MIS Quart.* **28**(1) 45–74.

Szajna, B. 1993. Determining information systems usage: Some issues and examples. *Inform. Management* **25**(3) 147–154.

Trice, A. W., M. E. Treacy. 1986. Utilization as a dependent variable in MIS research. *Proc. 7th Internat. Conf. Inform. Systems*, San Diego, CA, 227–239.

Venkatesh, V., F. D. Davis. 2000. A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Sci.* **46**(2) 186–204.

Venkatesh, V., M. G. Morris, G. B. Davis, F. D. Davis. 2003. User acceptance of information technology: Toward a unified view. *MIS Quart.* **27**(3) 425–478.

Weber, R. 1997. *Ontological Foundations of Information Systems*. Coopers & Lybrand and Accounting Association of Australia and New Zealand, Melbourne, Australia.

Webster, J., J. J. Martocchio. 1992. Microcomputer playfulness: Development of a measure with workplace implications. *MIS Quart.* **16**(2).

Yuthas, K., S. T. Young. 1998. Material matters: Assessing the effectiveness of materials management IS. *Inform. Management* **33**(3) 115–124.

Zigurs, I. 1993. Methodological and measurement issues in group support system research. L. M. Jessup, J. S. Valacich, eds. *Group Support Systems: New Perspectives*. Macmillan, New York, 112–122.

Zigurs, I., B. K. Buckland. 1998. A theory of task/technology fit and group support systems effectiveness. *MIS Quart.* **22**(3) 313–334.