8-16-1996

On Developing a Standard, Abilities-based Measurement Tool for Task Analysis: A Preliminary Experiment using Computer Programming Tasks

J. Wayne Spence
Department of Business Computer Information Systems, University of North Texas, spence@cobaf.unt.edu

On Developing a Standard, Abilities-based Measurement Tool for Task Analysis: A Preliminary Experiment using Computer Programming Tasks

J. Wayne Spence
Department of Business Computer Information Systems
University of North Texas
E-mail: spence@cobaf.unt.edu

Introduction

Among the difficulties with research which exists in the IS domain are the variety of approaches and measurement techniques utilized, making replication and "consensus building" more difficult. This paper offers a preliminary test of an instrumentation technique which is both simple to employ and easy to replicate.

The domain to which the preliminary instrument applies is task analysis--a common problem faced by software designers as they develop and install new products for individuals and groups of users. Task analysis attempts to identify and isolate specific components of a problem; those components for which technology holds the promise of improvement are especially important to IS developers.

A number of alternatives have been suggested in terms of how one should approach task analysis (e.g., Companion and Corso, 1982). The approach taken in for this instrument's development could be labeled the abilities, thinking skills or cognitive process approach in that it seeks to identify and isolate the activities (and therefore the implied processes) undertaken by a subject endeavoring to solve a problem. However, it is not sufficient to only identify these activities. To establish a more complete analysis, one must seek the activity, its duration, and its intensity. This instrumentation, because of its developmental nature, seeks only to identify an activity and its intensity. An interesting follow-on to this instrumentation is also the establishment of a measure of cognitive load (a.k.a. cognitive demand, mental effort, and so on). This issue has also been identified in the IS literature (e.g., Vessey & Galletta, 1992) as being of interest in establishing the "difficulty" of a task.

The Development Process

On the assumption that, in some measure, the characteristics of a task can be established on the basis of the skills and abilities utilized to engage the task on a human level, one needs to identify the relevant theoretical foundations of thinking skills and abilities. For this researcher, this process began by examining ***Dimension of Thinking*** (Marzano et al., 1989), which identifies and categorizes a number of thinking skills and the consequences of building those skills in an educational domain. This work has since been expanded to incorporate a broader skill set (Marzano et al., 1993). Using these lists as a foundation, other sources of this concept were sought out (e.g., Stiggins, Stager. . .) and the list of skills was elaborated and expanded. (It should be pointed out that the intent of the instrument development was to establish how an individual engages a problem, thus oral discourse and other forms of human exchange or group dynamics were not considered). Finally, with the assembled list in hand, an attempt was made to consolidate the items into a more manageable set. To this end, a thesaurus from a commonly available word processing software product was consulted. Each term was entered and if the thesaurus identified one (or more) of the other terms in the list, the terms were combined (or eliminated if the expression was overly redundant). This exercise resulted in a list of 30 items (or phrases) which appear on the instrument.

The next phase of development involved the establishment of how to identify and measure the presence of these items in the solution of a task. Of course, one could simply ask if an activity type had been employed; however, it is possible that the content of that employment could vary significantly within a task. Thus, it was decided to establish three sets of measures for each activity type based on Simon's (1962) intelligence-design-choice decision model. Thus, subjects were asked to supply information concerning a task relative to how they learned about the problem (intelligence), the strategy they used when they worked on the problem (design), and how they went about developing their solution (solved by) to the problem (choice).

Finally, the mechanism of measurement was chosen. As indicated above, to be complete, this mechanism should identify the type of activity, its duration, and its intensity. For purposes of this development, subjects were asked only to identify activities used and their intensity. This both simplified instrument design as well as reduced the response time requirement. Thus, subjects were asked to identify and rank the top five activities for each phase (learned about, worked on, and solved by) of their problem development.

Subjects

All subjects (n = 15) were volunteer, junior or senior, information systems majors in the college of business administration at a moderately large university in the southwestern US. All subjects were enrolled in the same courses (one semester apart) and had reasonably equivalent educational backgrounds in computing and business experience (verified through demographics). Subjects received compensation for their participation. The tasks completed by the subjects were an integral part of the courses in which they were enrolled and nominal extra credit was awarded for completing and returning the assessment instrument.

The Tasks (Treatments)

Two tasks were utilized for purposes of instrument evaluation. The problems involved computer programming and could be designated as a data validation routine, whereby data inaccuracies were identified and reported. Each subject completed the programming assignment, followed by the task analysis form. For Task 1, the assessment was completed on a pencil and paper response form. (These forms were collected so that respondents could not review them in preparation for the response to Task 2.) For Task 2, the assessment was completed electronically and the results returned by e-mail. Both tasks were roughly equivalent in terms of the level of coding sophistication required for correct solution and coding activities were separated by approximately two-months. However, Task 1 required a solution using a microcomputer based environment, whereas Task 2 required the solution to be performed in a mainframe environment.

Test-Retest Results

The first evaluation performed was to compare the responses provided on Task 1 with Task 2 to establish test-retest reliability. For each phase of the problem development, responses for individual respondents were compared. If the subject identified the same item on the two assessments, it was adjudged a match and was added to the item-match count. Next, the maximum number of responses (up to 5) were counted and the match count divided into the maximum number of matches for a "match average." Thus, the match average is a representation of the percent of identical responses provided by the participants on the two assessments. The match average for "Learned About" was 49.6%, "Worked On" 55.7%, and "Solved By 37.2%. Overall (combining all phase), identical responses rate was 47.3%. Thus, an early assessment of test-retest reliability of the instrument suggests that the instrument is reasonably reliable. It should be noted that only identical items were considered matched. If highly similar items were selected, they were not considered an exact match. Further, the lower average for "Solved By" can be explained due to the difference in environment (microcomputer versus mainframe) for the two tasks. Further, the highest average was attained in the "Worked On" category, indicating a higher degree of internal consistency for the respondents that either "input" or "output" activity, which could be influenced by environmental factors.

Further Analysis

Further analysis of respondent results are useful with respect to at least two different aspects of task analysis: the mix of activities employed on the task and the cognitive load produced by the task. First, table 1 provides insight into the most frequently ranked activities for each task (by frequency of ranking). Note that the items listed for both tasks are highly similar. Thus, by implication, this suggests that the two tasks were indeed highly similar. However, upon further inspection one should note that the total frequency count presented in table 1 only accounts for a reasonably small percentage of the total possible frequencies. This further implies that, although common elements do exist, the development of solutions to these tasks have significant differences at the individual level.

Finally, if the two tasks are indeed similar, one would expect their cognitive loads to be similar. Based on Bloom's Taxonomy (Bloom et al., 1956), a Delphi technique (employing seven raters) was used to map the 30 activities into Bloom six layers of cognitive domain. It is posited that the higher level of Bloom's taxonomy in which one operates, the higher the cognitive load. (Bloom indicates that one can reach a higher level only by relying on elements existing at lower levels.) Based on this technique, the activities were recoded into the cognitive load mapping. The results of this process are shown in table 2. The cognitive load values are not significantly different.

Conclusions

This paper has demonstrated the usefulness of a new means of task analysis and the establishment of a standard means of measuring cognitive load. While these results are preliminary in nature, the consequences of establishing instrumentation for this purpose are extremely significant. First, it promotes a standard mechanism for task analysis. Having such a mechanism should promote understanding and make future research more directly comparable. Second, it identifies specific activities accomplished by users, pointing out specific domains through which software development could directly benefit users. Lastly, it has the capacity to establish a measure of cognitive load whereby IS researchers can determine whether a product

Table 1: Selected Activity Type by Task Phase

Element	Task 1 (n = 15)		Task 2 (n = 9)	
	Activity *	Frequency	Activity *	Frequency
Learned About	Analyze Break apart Design Outline Perceive Demonstrate Explain	7 6 6 6 6 4 4 ** 39 of 75	Explain Recall Analyze Outline Break apart Summarize Translate	6 5 5 5 3 3 3 ** 30 of 43
Worked On	Break apart Design Analyze Develop	10 10 8 6	Break apart Design Analyze Explain	7 7 4 3

	Categorize Compare <u>Integrate</u>	5 4 4 ** 47 of 73	<u>Integrate</u> Perceive	3 3 ** 27 of 41
Solved By	Conclude <u>Develop</u> Synthesize Integrate Analyze Perceive Summarize	6 6 6 5 4 4 4 ** 35 of 63	Break apart <u>Develop</u> Design	4 3 3 ** 10 of 43
Overall	<u>Analyze</u> <u>Break apart</u> <u>Design</u> <u>Perceive</u> Develop Conclude <u>Integrate</u> Synthesize	19 18 18 13 12 11 11 10 **112 of 211	<u>Break apart</u> <u>Design</u> <u>Analyze</u> Explain Recall <u>Integrate</u> <u>Perceive</u> Demonstrate	14 12 10 10 8 7 7 6 **74 of 126

* Underlined processes appear in the list for both tasks.

** Total represented frequency of total response frequency.

Table 2: Cognitive Loads by Task Phase

Cognitive Load Element	Task 1 (n = 15)	Task 2 (n = 9)

	Mean	Std. Dev.	Mean	Std. Dev.
Learned About	47.93	10.02	42.33	12.93
Worked On	53.47	8.52	46.11	9.35
Solved By	57.79	14.61	51.22	17.25
Overall	53.06	11.05	46.55	17.42

renders problem solution easier or more difficult. Finally, as put by Companion and Corso (1982), the theory with a field can be expanded by development and use of tools and techniques capable of assessing the attributes of that domain. This paper speaks to all of these issues.

References

References available upon request from the author.