A Framework for Measuring Student Learning Gains and Engagement in an Introductory Computing Course: A Preliminary Report of Findings

Billy Lim, Bryan Hosack and Paul Vogt
Illinois State University, Normal, USA
bllim@ilstu.edu
bhosack@ilstu.edu
wpvogt@ilstu.edu

Abstract: This paper describes a framework for measuring student learning gains and engagement in a Computer Science 1 (CS 1) / Information Systems 1 (IS 1) course. The framework is designed for a CS1/IS1 course as it has been traditionally taught over the years as well as when it is taught using a new pedagogical approach with Web services. It enables the new approach to be compared with the traditional way of teaching the courses in terms of student self-assessment of learning gains, student assessment of their engagement with the subject matter, and researcher assessment of student learning gains as measured by performance on a researcher-designed examination. The framework includes a comprehensive pre-test and post-test for students in the control and treatment sections to complete, a common assessment exam module for all students to take, and a faculty survey for the instructors to complete. This enables the researchers to answer many questions regarding the effectiveness of the Web service approach, including "Do students using the Web service approach perform better in the common assessment exam module?" and "Do students and faculty members find the Web service approach more engaging?" Results from the first semester of a 3-year multi-university study are discussed.

Keywords: learning gains, introductory computing course, web services, learning engagement, SALG

1. Introduction

The topic of the measurement and evaluation of a teaching technique is important for all learning paradigms. Measuring and evaluating the success of a pedagogical approach in any project is crucial as it allows one to determine if the given approach is indeed effective, with objective measures to accompany the claim. As it is often quoted, "If it can't be measured, it can't be managed" (Deming, 2012).

In a recently completed 2-year pilot research project (Lim and Hosack, 2009) at a large Midwestern university in the United States through a National Science Foundation grant, there was a clear need for a measurement and evaluation model so that the new pedagogical approach could be properly managed and assessed. In the project, the researchers were trying to assess the effectiveness of a newly devised pedagogical approach for teaching Computer Science 1 (CS1) / Information Systems 1 (IS1), the introductory computing course for computer science and information systems majors respectively. Specifically, the project was to examine CS1/IS1 courses as they had been traditionally taught over the years as well as how they were taught using a new pedagogical approach with Web services technology. Namely, the main objective of the pilot project was to measure the effectiveness of the newly proposed *service-oriented paradigm* to teaching CS1/IS1 in terms of student exam performance. The promising results achieved in the exploratory study encouraged the researchers to expand the study and now branch out to multiple universities so that the approach can be tested in various sites for its effectiveness using the new framework for instruction and its assessment.

The Web service approach to teaching CS1/IS1, an approach that integrates the use of burgeoning Web services technology throughout the courses, has been shown to increase student performance in the final exam score in a recent study. The approach is also more interesting to the students as it allows for more sophisticated apps to be built, where students can build mashups that involve Google Maps, YouTube, Twitter, etc. in their first programming course.

The bourgeoning Web service technology and the approach for using it in teaching CS1/IS1 are recapped in Section 2 of this paper. And, as detailed in Section 3, the model used for assessing the approach is an effective one. The Web service approach has been shown to allow students to

Reference this paper as: Lim, B., Hosack, B. and Vogt, P. "A Framework for Measuring Student Learning Gains and Engagement in an Introductory Computing Course: A Preliminary Report of Findings'" The *Electronic Journal of e-Learning* Volume 10 Issue 4, 2012, (pp428-416), available online at www.ejel.org

perform better in a common final exam (Hosack et al., 2011). A great deal was learned from the pilot project, including some shortcomings/pitfalls on the assessment model as we planned for expanding the approach to include multiple sites. Based on the pilot study, three improvements can be considered by researchers facing similar situations when testing a new pedagogical approach. The assessment model from the pilot study was revised into a new framework that is more standardized and comprehensive can be established are described below. This framework is currently being tested in an expansion study (Lim and Hosack, 2011) that involves multiple institutions given the initial evidence of success in the pilot study. The three improvements are categorized below.

First, the original instrument for assessing student learning gains was developed in-house and not based on a standardized, widely used instrument that has been tested extensively. To address this, a revised instrument that is based on SALG (Student Assessment of Learning Gains) (Seymour et al., 2000, www.salgsite.org) has been developed. SALG is a nationally validated pre- and post-survey of students' self-assessment of their knowledge before and after a course. Because it has been used in numerous courses over many years, it can provide the basis for measured comparisons of student learning. SALG initially targeted the field of chemistry but has since been generalized to work with various disciplines. For example, Anderson (Anderson, 2006) uses SALG in the Nutrition and Food Science area. The SALG instrument can be adapted to address a particular set of skills, in this case computer programming, while retaining its reliability and validity.

Second, while measuring student learning gains was an objective of the pilot project, measuring student engagement was not. Given the nature of Web services, which allows for the wealth of information on the Web to be harvested easily through API (application programming interface) calls from one's computer program, it would be remiss for the new framework to not capture student engagement. Students are expected to be more engaged with the Web service approach as they are interacting with activities that they often personalize to make them more interesting and relevant (e.g., find all 3D movies that are playing in *my* hometown (zipcode xxxxx), display all comments from *my* favorite YouTube video, etc.).

There have been many different efforts in the literature on engaging student learning using a variety of approaches. They include the application of "gamification" to eLearning to engage learners where the theory behind gaming design is applied to build engagement interactive materials such as eLearning (Raymer, 2011), the study of how learning community participation affects student engagement (Pike et al., 2011), the research on curiosity, interest and engagement in technology-pervasive learning environments (Arnone, 2011), just to name a few. The proposed framework allows the researchers to assess whether the Web service approach represents another means to actively engage students in learning the fundamentals of computer programming.

Third, the assessment model was applied to one institution only when the pilot research project was conducted. In the expansion project that involves multiple, collaborating institutions, a framework that supports the assessment on multiple sites is desirable. The proposed framework includes a common assessment exam module, approved by the instructors involved, that allows for comparisons across universities. Also, instructors' own reactions to the new method of teaching programming are measured.

The paper details a framework for measuring student learning gains and engagement in an introductory computing course. The framework allows for a new pedagogical approach to teaching CS1/IS1 to be measured against the traditional approach that has been used for many years. Data such as student self-assessment of learning gains, actual gains in exam performance, and student engagement can be captured for analysis. It is the framework used in a longitudinal study that provides methodological insight into student learning using the new learning paradigm. Together with the framework, the results from the first semester of a 3-year multi-university study are also discussed.

The remainder of the paper is organized as follows. Section 2 recaps the Web service approach to teaching introductory computing course. The assessment model used in the Web service approach is described in Section 3. Section 4 discusses the new framework that improves on the one used in the pilot study. A discussion of the implications of the framework is presented in Section 5. Finally, the summary and conclusions are given in Section 6.

2. Web service approach to teaching introductory computing

Web services technology is a burgeoning technology that has received tremendous amount of attention in the software industry. After mainframes in the 60s, PCs in the 80s, applications in the 90s, and the Internet in the 00s, Web services have been considered a disruptive technology based on the 5th wave of computing. According to Gartner, a reputable research firm, in its “Hype Cycle for Application Architecture, 2010” report (Gartner, 2010), basic Web services are plotted on the “plateau of productivity” portion of the curve, which Gartner defines as a state that “…real-world benefits of the technology are demonstrated and accepted. Tools and methodologies are increasingly stable as they enter their second and third generation” (Table 1).

Given the importance of Web service technology, it is imperative that the technology be introduced in today’s IT curricula. Also, the researchers have shown that Web services can be integrated into the curricula early, i.e., in introductory computing courses (Lim et al., 2005). As a result, this research targets CS1 and IS1, courses that are designed to introduce the basic problem solving and program design skills that are used to create computer programs. The objective is to stimulate student learning of the materials by not introducing concepts in an abstract, boring, and contrived fashion. Instead, the emphasis is on developing modules/scenarios that are creative, novel, and engaging. The key idea is to develop a module that when presented to a student, he/she would think: "let's see what happens when I try …."

To give a sense of how the Web service approach is used, a sample module comparing the Web service and traditional approaches for a typical topic covered in CS1 or IS1 is presented in Table 1 below. This topic, along with various other topics, can be easily enhanced so that students are exposed to the state-of-the-art technology. The topic is presented with a typical delivery mechanism using the traditional approach, then augmented with the Web services approach, and finally followed by an example depicting the Web service approach to the topic.

In the following selected module, the topic presented is “Sequence, Iterative, and Decision Structures.” The learning objectives aim to reinforce the concepts behind the fundamental control structures of sequencing, looping, and decision making. Upon completion of this module, students should be able to ascertain the order in which the various tasks need to be carried out, to apply the appropriate looping structure to iterate over a collection of data, and to impose the necessary conditions to filter the data for display purposes. In the table below, three sections (Typical Delivery, Web Service Delivery, and Example) are presented.

In the “Typical Delivery” section, a typical approach used for discussing the topics of “Sequence, Iterative, and Decision Structures” is discussed. One example is: “Process a collection of numbers (from the user), determine which one is the largest, and finally display it.” Clearly, the sequential aspect of this is that one needs to read the input first before one can decide and then display the largest. The iterative aspect is that one needs to establish a loop to go through the list. The decision aspect is that as each number is processed, an if/else statement is needed to keep track of the largest (so far).

In the “Web Service Delivery” section, a comparable scenario to the above is described. The idea here is to cover the same topic(s), but using Web services as the delivery mechanism. With Web services, the possibilities are endless and one can be creative in incorporating the topic(s) at hand in a way that engages the students more. For example, instead of processing a random list of numbers, the students can be processing a set of numbers representing the populations of all the 50 states via a Web service. Now, the numbers have meanings and the processing seems more interesting as it ties in with their general knowledge about the US geography/society.

Finally, in the “Example” section, a specific scenario that details how the “Web Service Delivery” section can be implemented is given. In Table 1, the example is about finding the warmest temperature in the entire U.S. by zip code, plotting the area on a map, and getting a route to go to the warmest area from a given location. The Web services that help in achieving the result are pointed out in this segment.

Table 1: Sample module for web service approach

Module Name:	Sequence, Iterative, and Decision Structures
Typical Delivery:	These topics are typically covered by traditional discussion of scenarios that (1) necessitate a certain ordering be imposed in order to solve a problem (e.g., read the input values before processing them), (2) require a loop be used (e.g., processing a collection of numbers to find the average), and (3) need an if-else structure be employed (e.g., find the largest and smallest numbers from a collection of numbers).
Web Services Based Delivery:	Instead of merely processing a collection of numbers or strings that may be meaningless (and boring to the students), one could present a scenario where the goal is to solve a problem by using the three fundamental structures and some existing Web services that can be composed to form a solution for the problem.
Example:	A plausible scenario here is to discuss a problem where one wishes to find out the warmest temperature in all the U.S. by zip codes at a particular moment in time. Further, the warmest area of the country needs to be plotted on a map. Lastly, get a route to go to the warmest area from a given location. This scenario, which is much more interesting for today's freshmen, may seem intractable in the traditional computing environment. But there exist various publicly available Web services that can be composed together to solve this problem rather effortlessly. There exists one that retrieves all the US zip codes (Remotemethods, 2008), another one that finds the temperature given a zip code (xMethods, 2008), yet another one that plots a particular area on a map given a zip code, and one that plots the route given two endpoints (Google Maps, 2008). Thus, one can cover the sequence, iterative, and decision structures all in one shot using above example. 

3. Evaluation of the pilot study of the web service approach

The pilot study employed a quasi-experimental design. The advantage of this design is that the control and experimental groups were comprised of real students in real classes in which real professors field-tested the teaching method. This design offered advantages in external validity (or generalizability) and ecological validity (similarity of the study setting to the settings in which it will be applied). The disadvantage was that student experimental participants could not be randomly assigned to treatment groups as they might have been, for example, in a learning laboratory experiment. Random assignment provides greater internal validity (certainty about causal attributions). The advantages in external and ecological validity more than compensate for the disadvantage in terms of internal validity. A laboratory approach would be less appropriate because it could not deliver the teaching method in a realistic fashion. The method studied in the project is designed for use in regularly scheduled classes and requires implementation over a period of several weeks. That is why field experiment used pre-existing groups of students and their instructors. The experimental group was the students of instructors who adopted the WS approach; the comparison group was students whose instructors did not use the WS approach.

The above discussed advantages of this design led to its adoption in the expansion study as well as in the pilot. However, quasi-experiments and field experiments pose special measurement problems.

Methods of measurement of outcomes appropriate for laboratory experiments (mean difference between control and experimental groups) must be supplemented in this type of design. Chiefly, this means the addition of control variables (covariates) to ensure that outcomes are not attributable to variables other than differences in treatments. To accomplish this multiple regression techniques were used to control for students' majors, class ranks, genders, and cumulative grade point averages before treatment. These covariates were used for two reasons: (1) to establish that the control and experimental sections/groups were initially comparable at the group level (they were) and (2) to control at the individual level for variables that could confound results. The same strategy was used for both the pilot study and for the current expansion study. The main differences between the pilot and the expansion study are in the measurement of the outcomes.

In the pilot study the main outcome measure was scores on the common final examinations given in the IS and the CS courses over 4 semesters both in experimental classes and in comparison group classes. The results (described in detail elsewhere (Hosack et al., 2011)) were encouraging. Even after controlling for other variables that could explain differences among students in learning outcomes, the 222 students in the treatment group classes scored about 5 points higher (out of 100) on average than the 364 students in the comparison group classes ($p = .03$). Also, considerable qualitative evidence indicated that the students in the treatment group classes found programming to be more interesting and engaging. This was enough for us to want to see whether these positive results would hold beyond a single university. We applied for and received funding from the NSF to expand our research over a period of three years to a group 24 universities. The next section describes the procedures used for evaluating results from this much larger group.

4. Framework for measuring student learning gains and engagement in the expansion study

The same quasi-experimental design is being used for the expansion study, and for the same reasons: greater external and ecological validity. And the same regression-based models are employed to assess outcomes. But on the basis of what was learned in the pilot study and because there will be a much greater number of cases, more extensive and rigorous measures of student gains in learning of and engagement with the subject matter are being employed. To better measure any learning gains that could be attributed to the new methods of instruction, a pretest and a posttest (see the documents in Appendix A) are used. These take the form of the Student Assessment of Learning Gains (SALG), which is a nationally validated pre- and post-survey of students' self-assessment of their knowledge before and after a course. Because it has been used in numerous courses over many years it can provide the basis for measured comparisons of student learning. Instructors using the SALG can, while retaining the format, adapt the questions to their particular learning goals and add questions as needed. Students are asked about skills and their understanding of concepts at the beginning of the class and at its conclusion. For example, the questions in section 2 of the pre-survey are strictly parallel to those in section 3 of the post-survey: each set of questions asks about the same skills.

The SALG questions were also used to measure student engagement, which is an important learning outcome in its own right and which, as postulated, will be an important predictor of student learning. The quality of the analysis was also improved by the addition of a new background variable: self-efficacy (questions 3.1 – 3.12 on the pre-survey). Self-efficacy has repeatedly been shown in numerous studies to account for a good deal of the variance in performance and learning in a wide range of contexts (Bandura, 2006). A final predictor variable, to be coded and analyzed by the principle investigators, is fidelity of, or intensity of, implementation of the new curriculum. It is to be expected that with dozens of faculty participants, there will be differences in the rigor with which the new curriculum is implemented and that these differences will have an important impact on the student learning and engagement variables that are the main outcomes of the study.

In addition to using the SALG assessments the researchers have designed an assessment test module of objective questions to be taken by students in both the control and experimental classes at the end of each semester. The questions measure student knowledge of programming concepts and skills. This common module of objective questions will allow comparisons across universities. The questions have been reviewed at a workshop with the first cohort of faculty participants; in the judgment of that group as well as of the principle investigators the questions have extensive face validity. The use of objective questions with a large N of student participants will enable the

 ©Academic Publishing International Ltd

researchers to use more advanced analytic techniques to measure student outcomes in the study, specifically: (1) propensity score matching to simulate experimental attribution of cause and (2) item response theory (specifically differential item functioning or DIF) to conduct subgroup analyses of responses to particular questions in the module. A final outcome measure is an instructor assessment of students' learning as well as of the instructors' own reactions to the new method of teaching programming. The combination of these factors yields the causal model shown in Figure 1.

The narrative in parts 3 and 4, on evaluating the pilot and the expansion study, strongly implies a causal model. To make the implied model, the causal diagram graphic is presented in Figure 1. The main independent or predictor variable is WS vs. Traditional, which are the treatment and comparison groups respectively. The thick arrow between the predictor and student engagement means that it is anticipated that WS will positively affect engagement, and student engagement, in turn will positively influence learning. This is shown by the second thick arrow, the placement of which indicates that student engagement is a mediating variable. While it is possible that the treatment could influence either student engagement or learning without having a comparable effect on the other, and this will be tested for, it is unlikely. Next, it is assumed that background variables (cumulative GPA, academic major, gender, etc.) will influence learning independent of the main line of causal variables, which is why that cluster of variables enters the model from the "outside." The same kind of external influence on learning will probably be exerted by self-efficacy. Finally, the two measures of student learning, the SALG survey and the objective test module, are on the far right of the model; the arrows point from those measures to the variable, learning, that they measure. A line between SALG and TEST has no arrow heads because, while they will probably be correlated, there is no causal relation between them. Rather they are two indicators of the same latent variable (learning).

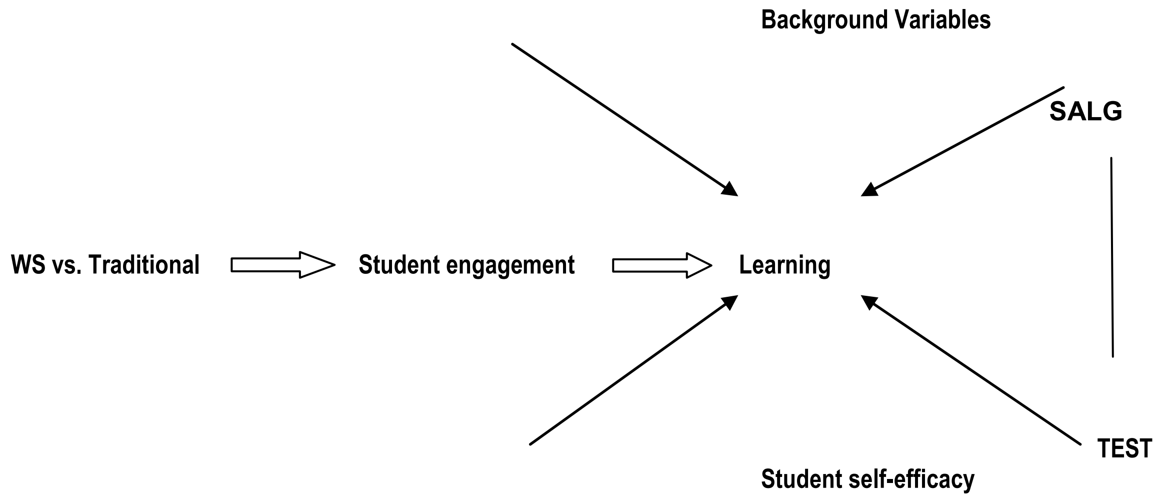

Figure 1: Causal diagram

5. Discussions

Based on lessons learned from an earlier pilot study, we concluded that there are three key improvements that can be made to a multi-site pedagogical study. First, using the SALG instrument to collect the student's perceptions of gains they have or have not made under the new approach gives a more reliable assessment. Second, included in the SALG instrument, were questions on student self-efficacy and engagement. As outlined in Section 2, the approach allowed students to personally connect with the material that also had real world applications in their future academic and professional careers. Finally, we incorporate in the framework a means to manage the data collection and assessment across multiple sites as part of a longitudinal study.

SALG enable a better assessment of student learning gains while new questions on self-efficacy and engagement better capture the student's response to the new pedagogical approach. Combining the measurement of these three areas adds to the robustness of the SALG instrument and provides some insight into the student's perceptions of the approach. The impact from being able to see multiple facets of the student's viewpoint is exciting. The combination of factors also allows us to triangulate a student's perception of the approach and get a better idea of whether or not it was a success beyond measuring an improvement in grades the common question exam items across CS and IS curricula.

 ISSN 1479-4403

Pairing a strong data collection model with the appropriate pedagogical approach is equally important. The approach outlined in Section 2 in the pilot study promoted self-efficacy and engagement while improving learning both from the student's perception and also in the quantitative data. Since, the instruments used in the pilot study did not consider all aspects of learning gains, self-efficacy and engagement, the lesson learned in finding a good match between approach and instrument can't be stressed enough.

The causal model in Figure 1 will be applied across multiple institutions and this will provide a better means to evaluate the approach's effectiveness across universities and different student demographics. Based on findings in the pilot study a modular approach to collecting data will be applied. The structure of the SALG instrument provides the ability to collect data on students and courses regardless of content using an online survey as a centralized collection point. Based on experience, a set of questions targeting core programming concepts has been developed. This question set can be tailored based on the concepts covered in a particular course allowing the experiment to adapt to each institution's curriculum while applying the Web service approach. Online completion of the common programming questions again allows for centralized data collection. When evaluating multiple institutions, the framework allows for the adjustment for quarter or semester scheduling and small (possibly a single instructor and section) or large (possibly multiple instructors and multiple sections) programs. The use of the SALG instrument and modular common questions makes this possible. Thus, data collection may spread over multiple semesters with the same instructor teaching comparative sections or multiple instructors running parallel comparative sections. Finally, using control background data across institutions allow for us to ensure that we can control for variances in participant populations across institutions.

Data from the first semester of the expansion study

As of this writing we have collected data from the first semester of the expansion study. The analysis of these data is approached in the spirit of "training data," or "statistical learning," in which early analyses are used to improve the collection and analysis plans for subsequent data (Berk 2006; Hastie et al, 2009). We have focussed initially on the measurement properties of the variables and on planning for analysis of data from upcoming semesters. Analyses of outcomes comparing control and experimental groups are of limited value at this stage. Only 5 professors at 4 universities teaching approximately 120 students participated in the project in the first semester; and because of the fairly high dropout rate in those courses, complete data are available for only 90 of those students. The number of valid cases from the first semester is not sufficient for the analysis of outcome variables (because of inadequate statistical power), but it is fully sufficient for discussing the measurement properties of the main variables.

Variables

The study uses five types of variables:

(1) Outcome variables:
The primary outcome of interest is student learning. This is measured in two ways, which allows us to cross-validate findings—with the student assessment of learning gains (SALG) and with the content knowledge survey (CKS), which is an objective test of their knowledge. For example, students are asked on the SALG to self-report their understanding of particular concepts (see Douglas et al., 2012). Their understanding of the same concepts is tested on the CKS. Each measure helps validate the other and enhances the investigators' ability to interpret data related to the main outcome variable—student learning.

(2) Predictor or independent variable:
The independent variable is whether student participants are in a control or experimental section.

(3) Control variables:
The background variables gender, academic major, class rank, cumulative grade point average, and student experience with programming are used as controls as is the student self-efficacy measure described above (see Zajacova, et al., 2005).

(4) Mediating variable:

The mediating variable postulated to influence the outcome in the study is student engagement (Ahlfeldt et al. [2005] and Carini et al. [2006]). This means that it is believed that WS instruction influences student learning directly and also indirectly by increasing student engagement.

(5) Additional pedagogical variables:
Also included are three categories of pedagogical variables in which students describe what was helpful in their learning of the subject matter: How much did each of the following help your learning: (1) *activities*, such as participating in discussions and attending lectures, (2) *assignments*, such as lab work and quizzes, and (3) *resources*, such as textbooks and sample programs? For each of these categories students are given both a 5-option scale (ranging from no help to great help) and an open-ended question to which they can type a response. (These are questions 6.1 – 6.5, 7.1 – 7.5, and 8.1 – 8.4 on the post-SALG).

Scale Reliabilities

When using an existing scale, even one that has repeatedly been tested for reliability, such as the SALG, one must still conduct one's own tests of the reliability of the scales in the study using the responses of participants in the study. Reliability is not an invariant property of scales. It always refers to and varies from one sample to the next and/or from one investigator to the next. The scale reliabilities and much other descriptive information about the scales used in this study are provided in Table 2.

Table 2. Item scales: descriptive statistics and reliabilities

Scale	Mean of Items[1]	Alphas[2]	Scale Means(N)[3]	Std dev of Scale	N
Pre: understanding concepts	2.21	.919/.920	13.28 (6)	5.61	127
Pre: programming skills	2.13	.945/.944	12.75 (6)	6.34	122
Pre: learning attitudes, self-efficacy	4.01	.914/.919	44.07 (11)	7.78	128
Post: understanding concepts	3.60	.878/.882	21.58 (6)	5.75	85
Post: programming skills	3.84	.835/.860	17.20 (5)	4.15	86
Post: learning attitudes, self-efficacy	3.44	.959/.960	27.54 (9)	8.58	91
Post: Student engagement in learning	3.09	.725/.688	30.85 (10)	6.09	81
Post: Content knowledge survey (test)	.589	.693/.676	11.79 (20)	3.43	89

Notes:
1. The items are on a scale of 1-5 except on the content knowledge tests where the range is 0 (incorrect) to 1 (correct).
2. Alphas are reported in the natural metric and for standardized items: natural/standardized.
3. The number in parentheses is the number of items in the scale.

The scales and the variables they measure are discussed above. Here we summarize their measurement properties and present some summary statistics. The column entitled "Alpha" presents two versions of Cronbach's alpha, each of which indicates a very similar level of reliability or consistency of measurement. The reliabilities for all the measures except the last two are very high. The last two—post student engagement and the CKS—are only acceptable. That is because each of those measures is almost certainly more than one scale; each tests students on several domains of subject matter or course participation. The presence of multiple domains will be tested with confirmatory factor analysis after data are gathered from a sufficiently large number of participants.

The column "mean of items," gives the average score (on a 1 to 5 scale) given by a student in response to a self-assessment question. For example, the 2.21 in the first row, indicates that the average rating of his/her understanding of various concepts was a bit over 2—or "just a little." Comparing that to the post understanding of concepts, 3.60, we can see that students said that on average they understand concepts on a range from "somewhat" to "a lot." (See the Pre and Post SALG questionnaires in the appendix for the full text of the questions.)

The scale means are in essence the item means multiplied by the number of items. So, for pre understanding of concepts multiply the 2.21 times 6 items to get the scale mean of 13.28. This statistic is interesting mainly in how it relates to the standard deviations. In all cases, the standard deviations are sizeable as compared to the scale means. This indicates that there is variance sufficient for analysis in the answers the students gave to the scale questions. In sum, all the measures that are important to the analysis of data in this project are "well behaved."

6. Summary and conclusions

We present an approach to better evaluate pedagogical approaches to teaching in the IT classroom. While our approach focuses on lessons learned from an earlier pilot study teaching introductory programming to CS and IS students using Web services, we feel this approach is easily tailored to other methods of teaching as well. The instruments proposed for use in this study can be used to target a variety of skill sets. The key lesson learned was finding a good match between the instrument, background data and the approach. Based on a relatively successful pilot study, we learned the importance of this and were able to improve and share our knowledge. Additionally, based on this lesson we developed a strategy the uses modular and flexible data collection from a variety of sources to work with multiple institutions over time. Implementation of this approach will be conducted over multiple sites and years. Based on the results we'll receive feedback on the success of our research plan outlined here. In attempting to answer the questions "Do students using the Web service approach perform better in the common assessment exam module?" and "Do students and faculty members find the Web service approach more engaging?" across many instructors, institutions and students we have found a set of tools that are applicable not only to these questions but other pedagogical questions not only in the IT field but other disciplines as well.

We also learned several methodological lessons in the early stages of the transition from a single-campus study to an investigation involving a score of universities and professors. Based on what has been learned, the PIs have made the following adjustments.

1. Concerning collecting data from many sites. In the single-university study with the cooperation of the Registrar, it was easy to collect complete and reliable data on students' background variables including: gender, major, class rank, and cumulative GPA. The reporting of the same data from the other universities in the expansion study has been much more difficult and sometimes impossible (largely due to interpretation of IRB regulations). The remedy, not a perfect one, will be to ask students to self-report these data on the pre-survey questionnaire. When data from official documents are available they are used and compared to student self-reports.

2. Concerning attrition. The attrition rate was fairly high from the pre-data collection at the beginning of the semester to the post data collection at the end. For most measures in the first round there were 120+ responses; in the second round about 90 completed the survey and exam. (See the specifics in Table 2.) This level of attrition (around 25%) is not uncommon in introductory programming courses. Understanding the causes of attrition is now an additional goal of the study. After several semesters of data collection, information about which students drop out and which persist should be fairly extensive.

3. Concerning scale quality. The scale quality is good for all scales and more than sufficient for measurement. More sophisticated methods, envisioned for use in this study—including structural equation modeling, item response theory, and differential item functioning—are very "case hungry." They require a much larger N of participants. More elementary methods of analysis can be used while additional data are gathered.

4. Concerning research design. The project uses two main designs. Following the first design, a professor attends a workshop and begins teaching classes using WS/SOA methods immediately. Implementing this design requires that the professor obtain the cooperation of one or more institutional colleagues who will agree to serve as control group instructors, and that entails persuading them to have their students complete the SALG and the CKS. Such simultaneous implementation of control and experimental group classes has been rare and is likely to remain so. However, this is probably an advantage in terms of the quality of causal inferences that can be drawn from the study. The second model involves the participating instructor teaching one semester of introductory programming using his/her standard methods. In the second and third semesters (after having attended the workshop) the instructor teaches using WS/SOA methods. The first design controls for time and institutional variables, since the control and experimental sections are taught simultaneously at the same institution. But it does not control for variation among instructors. The second method also controls for institutional variables, and, in addition, it controls for variation among the instructors since each serves as his or her own control. Time is not controlled in this second, longitudinal, approach. But controlling for variation among instructors is arguably more important. In the early days of the study, based on the experiences gained from the pilot study, the PIs encouraged recruits to the study to follow the first model. While we think the first model is acceptable, we now believe that the second, within-subjects model, in which instructors serve as their own controls, is not only more practicable but also a stronger design.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. NSF DUE-CCLI-0837056 and TUES-1123094

References

Ahlfeldt, S., Mehta, S., & Sellnow, T. (2005). Measurement and analysis of student engagement in university classes where varying levels of PBL instruction are in use. *Higher Education Research and Development, 24,* 5-20.

Anderson, A. (2006). "An Assessment of the Perception of Learning Gains of Freshmen Students in an Introductory Course in Nutrition and Food Science," Journal of Food Science Education, Vol. 5.

Arnone, M., Small, R., Chauncey, S., and McKenna, P. (2011). "Curiosity, interest and engagement in technology-pervasive learning environments: a new research agenda," Educational Technology Research and Development, Volume 59, Number 2.

Gartner (2010) "Hype Cycle for Application Architecture, 2010" ID Number: G00200962, August 2010.

Bandura, A. (2006). Guide for constructing self-efficacy scales in *Self-Efficacy Beliefs of Adolescents.* New York: Information Age Publishing.

Berk, R.A. (2006). An introduction to ensemble methods for data analysis. *Sociological Methods and Research, 34,* 263-295.

Carini, R.M., Kuh, G.D., & Klein, S.P. (2006). Student engagement and student learning. *Research in Higher Education, 47*, 1-25.

Deming, E. (2012) Retrieved from http://en.wikipedia.org/wiki/W._Edwards_Deming, January, 2012.

Douglas, J.A., Thomson, G., & Zhao, C. (2012). The learning outcomes race: the value of self-reported gains in large research universities. *Higher Education, 64,* 317-335.

Google Maps API (2008), http://code.google.com/apis/maps/, Retrieved May 16, 2008.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning,* 2nd edition. New York: Springer.

Hosack, B., Lim, B., and Vogt, P. (2011) "Incorporating The Use Of Web Services In Introductory Programming Classrooms," Journal of Informatics Education Research, Fall 2011.

Lim, B., Jong, C., Mahatanankoon, P. (2005) "On Integrating Web Services From the Ground Up Into CS1/CS2," ACM SIGCSE Bulletin, Vol. 37, No. 1, St. Louis, MO, Feb., p. 241-245, 2005.

Lim, B., Hosack, B. (2009). Integrating Service-Oriented Paradigm into Introductory Information Technology Curricula, National Science Foundation, 2009-2011.
Lim, B., Hosack, B. (2011). Service-Oriented Paradigm Across Introductory Information Technology Curricula: An Expansion Study, National Science Foundation, 2011-2014.
Pike, G., Kuh, G., McCormick, A. (2011). "An Investigation of the Contingent Relationships Between Learning Community Participation and Student Engagement," Research In Higher Education, Volume 52, Number 3.
Raymer, R. (2011) "Gamification: Using Game Mechanics to Enhance eLearning," eLearn Magazine, September 2011.
Remotemethods (2008), http://www.remotemethods.com/home/valueman/validati/zipcodes, Retrieved May 16, 2008.
Seymour, E., Wiese, D., Hunter, A., Daffinrud, S. (2000) "Creating a Better Mousetrap: On-line Student Assessment of their Learning Gains," American Chemical Society Symposium, San Francisco, March 27, 2000. Also available at http://www.salgsite.org/docs/SALGPaperPresentationAtACS.pdf
Xmethods (2008), http://www.xmethods.net/sd/2001/TemperatureService.wsdl, Retrieved May 16, 2008.
Zajacova, A., Lynch, S.M., & Espenshade, T.J. (2005). Self-efficacy, stress, and academic success in college. *Research in Higher Education, 46,* 677-698.

Appendix A: :

Student Assessment of Learning Gains: Pre-Survey

Instructions to students

Teachers value students' information feedback and take it into account when improving their courses. Please be as precise as you can in your answers. Please choose "not applicable" for any activity you did not do. You may find one or more questions at the end of each section that invite an answer in your own words. Please comment candidly, bearing in mind that future students will benefit from your thoughtfulness. Remember that this is an anonymous survey: your teacher will never know what any individual student has written.

Concepts	**Not at all (1)**	**Just a little (2)**	**Somewhat (3)**	**A lot (4)**	**A great deal (5)**	**Not applicable (99)**
1. Presently I understand the computer programming concepts of …						
1.1 Objects and Classes						
1.2 Arrays						
1.3 Class Inheritance						
1.4 File Processing						
1.5 Code Reuse						
1.6 Web Services						
1.7 What do you expect to *understand* at the end of the course that you don't understand now? (*open-ended question; students type responses*)						
Skills						
2. Presently I can …						
2.1 write a computer program in a programming language to solve a computer problem						
2.2 write a computer program that involves the use of *repetition* (e.g., loop) statements						
2.3 write a computer program that involves the use of *decision* (e.g., if-else) statements						
2.4 write a computer program that involves the use of *step-by-step* statements						

 ©Academic Publishing International Ltd

Concepts	**Not at all (1)**	**Just a little (2)**	**Somewhat (3)**	**A lot (4)**	**A great deal (5)**	**Not applicable (99)**
2.5 write a computer program that *reuses existing web services*						
2.6 write a computer program that involves the use of *objects/classes*						
2.7 What do you expect to *be able to do* at the end of the course that you cannot do now? (*open-ended question; students type responses*)						
Attitudes						
3. Presently I . . .						
3.1 am enthusiastic about the subject of the course						
3.2 am interested in discussing the subject area with friends or family						
3.3 am interested in taking or planning to take additional classes in this subject						
3.4 have expectations for learning about programming that are positive						
3.5 am confident that I can regularly attend this course's classes						
3.6 am confident that I can regularly attend this course's labs						
3.7 am confident that I can get myself to study programming						
3.8 am confident that I can ask for help if I have programming problems						
3.9 am confident that I can learn to understand programming concepts						
3.10 am confident that I can learn to write computer programs						
3.11 am confident that I can learn to solve computer programming problems						
3.12 Please comment on your present level of interest in the subject of this course and your confidence that you can succeed in it. (*open-ended question; students type responses*)						
Experience						
4.1 How long have you used a computer to surf the Web, word process, make presentations, etc.?						
4.2 If you have programmed before, how long have you done so?						
4.3 What is your class rank? (check the appropriate box)	Freshman	Sophomore	Junior	Senior	Other	
4.4 How old were you on your last birthday? (please write in the number)						
4.5 What is your academic major? (please type in response)						
4.6 Are you Female or Male? (please check the appropriate box)	Female	Male				
4.7 What was your college grade point average before taking this course? (please type in the number; if you do not know it exactly, please estimate)						

Student Assessment of Learning Gains: Post-Survey

Instructions to students

Teachers value students' information feedback and take it into account when improving their courses. Please be as precise as you can in your answers. Please choose "not applicable" for any activity you did not do. You may find one or more questions at the end of each section that invite an answer in your own words. Please comment candidly, bearing in mind that future students will benefit from your thoughtfulness. Remember that this is an anonymous survey: your teacher will never know what any individual student has written.

	Never (1)	Once or twice (2)	Some times (3)	Often (4)	Very Often (5)	Not applicable (99)
1. Your Activities for this Class: During this class and in your preparations for it, how often did you do each of the following?						
1.1 Worked with other students outside of class to prepare assignments						
1.2 Asked questions in class or contributed to class discussions						
1.3 Missed class						
1.4 Attended class without having prepared						
1.5 Attended the weekly labs and did the assignments						
1.6 Discussed course content with the instructor						
1.7 Discussed course content with other students in the class						
1.8 Discussed course content with people not taking the class (friends, co-workers, etc.)						
1.9 Became engaged in class assignments because they could be applied to real problems						
1.10 Applied what you learned to topics of interest to you outside of class						

	No gains (1)	A little gain (2)	Moderate gain (3)	Good gain (4)	Great gain (5)	Not applicable (99)
Your understanding of class content						
2. As a result of your work in this class, what GAINS DID YOU MAKE in your UNDERSTANDING of each of the following?						
2.1 Objects and Classes						
2.2 Arrays						
2.3 Class Inheritance						
2.4 File Processing						
2.5 Code Reuse						
2.6 Web Services						

 ©Academic Publishing International Ltd