Proposal Addendum

This addendum was written to address specific shortcomings identified by the doctoral committee during the defense of my dissertation proposal. It addresses four underrepresented aspects of the initial proposal, namely the threats to validity that the experiment faces, a more in-depth description of the pilot work that will occur prior to the main study, a more detailed plan of data analysis, and a list of possible publish-worthy units that may arise out of this project.

Threats to Validity

In this section, I outline my threats to validity using the methodology outlined in *Experimental and Quasi-Experimental Designs* (Shadish, Cook, & Campbell, 2002), which break down threats to validity into four categories: statistical conclusion, internal, construct, and external. This section attempts to relate my plan of study to each of these four categories.

Statistical Conclusion Validity

According to Shadish, Cook, and Campbell (2002), statistical conclusion validity refers to the "quality of inferences that can be made about the correlation/covariation between the treatment and outcome." Threats in this category are concerned with things such as low power, unreliable measures, and extraneous variance. Along these lines, this study suffers from the potential of an **unreliable treatment implementation**, meaning that the treatment will likely not be implemented in a standardized manner across study participants. In my study, the treatment is the usage of the SPE by students. Given that usage is left to individual student discretion, it is highly likely that SPE usage will vary widely among individual students. To account for this threat, I plan to capture SPE usage metrics (see the section on data analysis) that can be used to explore how various levels of SPE engagement affect my outcome variables.

Internal Validity

Internal validity refers to the whether or not an observed covariation between variables represents an actual causal relationship between said variables. Threats in this category are concerned with proving temporal precedence and eliminating the possible effects of other covariates. This proposal suffers from the following threats to internal validity:

History. Over the course of a given semester, the outcome variables of self-efficacy, retention, and grades are likely to be affected by sources outside of my intervention. While the nature of my experiment ensures that this threat cannot be eliminated, I have done my best to mitigate this issue by having the study performed in the same class taught by the same instructor using the same materials during the same period of time (the fall) in each academic year.

Maturation. It can be argued that any observable changes in self-efficacy, retention, and grades are a result of having progressed through the CptS 121 curriculum rather than the introduction of an SPE. After all, one might argue that a person's programming self-efficacy is likely to increase after having taken a course in programming. Thankfully, this notion has been shown to be false by prior research (see, e.g. Hundhausen, Agrawal, Fairbrother, & Trevisan, 2010), which found that self-efficacy has a tendency to actually decrease in introductory courses. Furthermore, I attempt to mitigate the effects of maturation by using similar groups (introductory computing students) across treatment conditions. However, like history, the threat of maturation cannot be completely removed in my selected design.

Attrition. Given CptS 121's historically high dropout rates, attrition is likely to be a threat to the overall internal validity of my study. As such, it could be argued that any observable gains in outcome variables are simply a result of lower-performing students dropping out of the course. I am unaware of how this issue might be combated in my study's design.

Construct Validity

Construct validity refers to the appropriateness of match between a high order construct (e.g. self-efficacy) and its measurement particulars (e.g. a survey) within a given study. My proposal suffers from the following threats to construct validity:

Mono-operation Bias. The constructs of *sense of community* and *peer learning* suffer from mono-operation bias in that they are only measured using a single instrument. Ideally, these constructs should be captured using more than one instrument; however, given the fact that data collection of the control group has already been completed, this threat will remain present in any analysis concerning these two constructs.

Mono-methods bias. All of my affective measures (self-efficacy, sense of community, peer learning) suffer from mono-methods bias because they are collected solely through survey. This issue is not as troubling as it might first seem because I am using well-established surveys, and there is general agreement that such instruments do accurately capture my construct.

Confounding construct with levels of construct. As outlined in my proposal, an SPE is a composition of many subcomponents. Should there be no observable effect on my outcome variables, it is possible that the learning theories underlying the SPE were not present in sufficiently high levels to produce an effect. Normally, this can be accounted for by introducing additional conditions in which the treatment is varied at a finer level. However, given the time required to introduce each variation, I believe that designing such an experiment is beyond the scope of a single dissertation.

Novelty. Should noticeable differences be observed within the treatment, it is possible that the novelty of introducing something new into the curriculum, rather than the SPE, was the source of improvement. Given the length of the intervention (a semester), it is likely that the novelty of an SPE will wear off before the end of the study. However, it is impossible to completely explain away this threat to validity.

External Validity

External validity refers to whether or not the results obtained in a given experiment can be generalized across a variety of people, settings, treatment variables, and measurement variables. My proposal suffers from two threats to external validity:

Interaction of causal relationship over treatment variations. It is possible that any effect (or lack thereof) may be a result not of the SPE but of other factors that accompanied the intervention. For example, we know that engagement is a key component to increasing retention (Tinto, 1993) and that expert moderation is an important aspect to the success of online learning systems (see, e.g. Hundhausen, Agarwal, & Trevisan, 2011). Perhaps, then, it is not the intervention of the SPE but the additional contact time that caused the change in outcome variables.

Interaction of causal relationship with setting. Because I am only testing the SPE in a single setting, it could be argued that an SPE is only effective at Washington State University because of its particulars (programming language, class size, instructor, homeworks, etc.). This claim can be refuted by running additional studies in different settings, but again, I believe this course of action to be beyond the scope of a single dissertation.

SPE Pilot Work

Based on the current development progress of the SPE, I expect to have a working prototype by June 2013. This timetable allows me to test the tool and obtain feedback from students during the subsequent summer semester. Feedback will be obtained through observational lab studies, questionnaires, and informal class discussions.

Observational Lab Studies

Near the deadline of each course deliverable, I will offer a subsection of the class an opportunity to work on their homework assignment for an hour under observation in the HELP Lab. The offer will be made to a different subset of students for each assignment, which should result in approximately 2-3 observations per assignment. I plan to use the lab study to explore whether or not students are able to successfully integrate the SPE into their learning process. For example, I am curious to learn how quickly students access SPE features after encountering a programming issue, and once accessed, if students are able to use the SPE to solve their issues. A semi-structured exit interview will be conducted after each observation and will be used to further elicit any thoughts that students might have regarding the SPE.

Student Questionnaire

Near the beginning and end of the course, I will ask students to complete a general questionnaire regarding OSBIDE. I hope to use these surveys as an indicator of how well students accept OSBIDE and integrate it into their learning process. Additionally, I am interested in what kinds of features students themselves deem important and how these might change over the course of the semester. I list possible survey questions in Appendices A and B.

Informal Class Discussions

My experience with other classroom software has taught me that students do not hesitate to complain about malfunctioning or poorly designed software. Likewise, students tend to request more of software that they like. I would like to document these naturally occurring conversations so that they can serve as motivation for future software development.

Data Analysis

Before discussing my data analysis strategy, I would first like to revisit the theoretical foundations on which I have based my SPE. My thesis utilizes the theories of Situated Learning Theory (Lave & Wenger, 1991), Self-Efficacy Theory (Bandura, 1997), and Vincent Tinto's model of student attrition (Tinto, 1993). In Situated Learning Theory, learning occurs as learners participate in a social learning community. According to Bandura, a learner's self-efficacy is affected by events both witnessed and experienced by the learner. Tinto's model of attrition suggests that engaging students in social communities is one of the most effective methods of reducing student attrition. All three theories place a heavy emphasis on student engagement and interaction, and see such interaction as a catalyst for success. The importance of engagement has led to the development of four general SPE design goals, listed in Table 1. These goals are further broken down into specific activities to take place within the SPE.

Goal	Corresponding SPE Activities
1. Observe others' activities and progress within the class [2,3]	1.1. View activity feed 1.2. Relate posts in activity feed to current user 1.3. Relate posts in activity feed to other users
2. Observe others' problem solving processes [2,3]	2.1. View activity feed details view, which houses conversations about a given issue 2.2. View code compile "diffs"
3. Observe others' social interactions [2,3]	3.1. View others' conversations within activity feed post details 3.2. View others' conversations within OSBIDE chat
4. Participate within social learning community [1,3]	4.1. Create posts in activity feed 4.2. Create posts in activity feed details view 4.3. View / participate within OSBIDE chat 4.4. Respond to others' posts within OSBIDE
Supporting Theories	
[1] Situated Learning Theory [2] Self-Efficacy Theory [3] Tinto's model of student attrition	

Table 1: SPE Goals

Having derived a list of design goals from my selected theories, I must now consider how the implementation might manifest itself through a change in students' overall behavior. According to Bandura (1997), providing additional experiential opportunities should help to solidify students' self-efficacy. Given that self-efficacy often starts low for newly learned tasks, it is reasonable to see a general trend of increased self-efficacy among students. Therefore, using a standardized instrument, such as the C++ Self Efficacy Scale (Ramalingam & Weidenbeck, 1998), to capture changes in students' self-efficacy appears to be an acceptable course of action. In a similar fashion, Tinto theorizes that increasing student engagement will reduce attrition. Therefore, it makes sense to examine retention rates obtained from the registrar. However, because many students who take CptS 121 are undeclared freshman, it would be best to reinforce the retention data with additional measurements. Unlike self-efficacy and retention, Situated Learning Theory is not directly linked to any specific attitudinal outcomes. Yet, given that Situated Learning Theory emphasizes the development of communities of practice, it stands to reason that individuals involved in a successful implementation of Situated Learning Theory may feel more like members of a community than might otherwise be expected. Therefore, I have chosen to operationalize Situated Learning Theory through the Classroom Community Scale (Rovai, 2002) and the peer learning component of the MSLQ (Pintrich, Smith, Garcia, & McKeachie, 1991). However, as noted in the section on threats to validity, my study suffers from the threat of unreliable treatment implementation. Therefore, it also becomes important to monitor how specific SPE usage might affect these outcomes. For this reason, it would also be helpful for the SPE to collect usage metrics from students as they work on programming projects throughout the semester. I list possible quantitative measurements in Table 2.

Quantitative Dependent Variable	Source
Self-Efficacy	Survey
Retention	Survey, Registrar
Learning	Course Grades
Peer Learning	Survey
Error Quotient	SPE Logs
Number of times SPE opened	SPE Logs
Number of hours spent within IDE	SPE Logs
Number of activity details page viewed	SPE Logs
Number of comments written	SPE Logs
Average length of comment written	SPE Logs
Time spent in SPE chat room	SPE Logs
Number of chat messages written	SPE Logs
Average length of chat message	SPE Logs
Number of profile picture changes	SPE Logs
Number of compile attempts that had errors	SPE Logs
Number of compile attempts that did not have errors	SPE Logs
Number of steps taken in debugger	SPE Logs
Number of runtime exceptions	SPE Logs
Number of cut/copy/paste events	SPE Logs
Amount of content pasted into the IDE from outside sources	SPE Logs

Table 2: Quantitative Dependent Variables

While the above list of quantitative variables do a good job of capturing SPE usage frequency, they do little to describe the kinds of interactions that students might be having. Therefore, it might be important to perform qualitative analysis on the conversations had within the SPE. To this end, I present a list of proposed questions that can be used to drive any qualitative analysis:

1. How do students appropriate the SPE for learning?
2. How relevant are activity post discussions to the original post's issue?
3. Are students reliably receiving answers to posted questions?
4. Do the answers that a student receives aid that student's learning process?
5. How "deep" are conversations?
   5.1. Do conversations mainly deal with class procedure (due dates, exam times, etc.) or do they discuss programming content in depth?
   5.2. How many conversations include or reference code?
6. What roles do students take within the community?
7. How do community roles change over the course of the semester?

Having presented a list of possible data points, I now turn to specific research questions that I hope to answer.

General Data Analysis Strategy

My data analysis strategy is guided by the following research questions:

RQ1: What effect does an SPE have on students' self-efficacy, sense of community, peer learning, and performance?

In essence, I am asking whether or not students benefited from the SPE. To answer this question, I expect to compare the self-efficacy, sense of community, peer learning, grades, and retention levels of students

across my control and treatment conditions. Because I am measuring multiple dependent variables, I must employ some sort of multivariate analysis. Furthermore, there exists the possibility of the two treatment groups being unequal. I try to account for using a pretest at the beginning of each treatment. Likewise, any data analysis should also account for this possible source of covariation. Given this line of reasoning, it appears that examining for differences would be best accomplished through the use of a multivariate analysis of covariance (MANCOVA). The MANCOVA will consist of two groups (control and treatment) and the dependent variables will be listed for the final measurements of self-efficacy, sense of community, peer learning, grades (homework, exam, and final), and self-reported retention for each group. To control for possible variation of each group, initial measurements of the aforementioned dependent variables will be inputted as covariates into the MANCOVA.

RQ2: To what degree does SPE usage promote self-efficacy, sense of community, peer learning, and performance?

This is a follow-up question to RQ1. Assuming that that the MANCOVA reveals a statistical difference between treatment groups, I must then use post-hoc analysis techniques (e.g. Tukey's HSD or Scheffe's Test; Vogt, 2005) to determine where the statistical differences occurred. For example, I might find that SPE usage promotes higher self-efficacy and sense of community but not higher peer learning and performance.

RQ3: What kinds of SPE usage promote self-efficacy, sense of community, peer learning, and performance?

It is possible that the way in which students use an SPE will affect the net change in the outcome variables. To investigate this question, I can use structured equation modeling to determine how specific SPE usage (e.g. comments written, messages viewed, depth of thread posts, etc.) affect self-efficacy, sense of community, peer learning, and performance.

RQ4: How does a student's self-efficacy, sense of community, and peer learning contribute to his or her overall performance?

Separate from the contribution of the SPE to computing education research, this question allows me to investigate how self-efficacy, sense of community, and peer learning might influence students' academic performance and retention within computing majors. Again, this question can be investigated through the use of MANCOVAs and post-hoc analysis.

Publishable Components

Given the amount of data being collected as a part of my dissertation, I expect the dissertation to serve as the foundation for several academic papers. I expect to be able to publish separate papers on my high level measures and my set of quantitative metrics. Furthermore, several of my proposed qualitative measures might be paper worthy. One that particularly interests me is that of performing a detailed analysis of community roles and how such roles change over the course of the semester. These papers, which focus on the academic benefit of an SPE, appear to be most appropriate for the SIGCSE group of journals and conferences. Separate from analyzing the education benefit of an SPE, I might also be able to get a publication from SIGCSE regarding OSBIDE's overall UI and design process.

References

Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: Worth Publishers.

Hundhausen, C. D., Agarwal, P., & Trevisan, M. (2011). Online vs. face-to-face pedagogical code reviews: an empirical comparison. In *Proceedings of the 42nd ACM technical symposium on Computer science education* (pp. 117–122). Presented at the SIGCSE 2011, Dallas, TX, USA: ACM.

Hundhausen, C. D., Agrawal, A., Fairbrother, D., & Trevisan, M. (2010). Does studio-based instruction work in CS 1?: an empirical comparison with a traditional approach. In *Proceedings of the 41st ACM technical symposium on Computer science education* (pp. 500–504). New York: ACM.

Lave, J., & Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. New York: Cambridge University Press.

Pintrich, D., Smith, D., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for learning questionnaire* (Technical report No. NCRIPTAL-91-B-004). Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning. Retrieved from http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/23/3c/44.pdf

Ramalingam, V., & Weidenbeck, S. (1998). Development and validation of scores on a computer programming self-efficacy scale and group analyses of novice programmer self-efficacy. *J. Educational Computing Research*, *19*(4), 367–381.

Rovai, A. P. (2002). Development of an instrument to measure classroom community. *Internet and Higher Education*, *5*, 197–211.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin Company.

Tinto, V. (1993). *Leaving college: rethinking the causes and cures of student attrition* (2nd ed.). University of Chicago Press.

Vogt, W. P. (2005). *Dictionary of Statistics & Methodology* (3rd ed.). Sage Publications.

Appendix A: Summer 2013 Beginning of Course Survey

Your instructor has recently introduced OSBIDE, a tool aimed at fostering collaboration and class awareness throughout the course. The ideas behind OSBIDE are continually evolving and we would appreciate any thoughts you might have on the tool and its potential uses. Please answer the questions below:

1. How might a tool like OSBIDE be useful as you work on course assignments?
2. What features of OSBIDE do you envision being the most useful or beneficial?
3. What programming activities of your fellow classmates would you find the most helpful or interesting?
4. Are there any features of OSBIDE that you think are missing or might be helpful to have?
5. In an effort to build community awareness, OSBIDE will expose certain aspects of your programming processes to the class.
    a. Does this aspect trouble you from a privacy perspective? Why?
    b. If so, what steps could be taken to make you feel more comfortable about using OSBIDE or similar tools?

Appendix B: Summer 2013 End of Course Survey

Having used OSBIDE over the course of this class, we would appreciate any thoughts you might have on the tool and its uses. Please answer the questions below:

1. What features or aspects of OSBIDE did you find to be the **most** beneficial?
2. What features or aspects of OSBIDE did you find to be the **least** beneficial?
3. Were there any features or aspects of OSBIDE that were difficult to use? If so, please explain.
4. Did OSBIDE positively contribute to your ability to learn the material presented in class? Explain.
5. Did OSBIDE positively contribute to your ability to feel connected with your classmates? Explain.
6. Would you recommend using OSBIDE in other computer science courses? Why?
7. Are there any suggestions that you would like to make with regards to how OSBIDE might be improved?