Creating Meaningful Assessments for Professional Development Education in Software Architecture

Elspeth Golden
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, PA
egolden@cs.cmu.edu

Len Bass
Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA
ljb@sei.cmu.edu

Abstract
Extensive instructional materials have been developed and used for courses in specific software architecture topics offered at the Software Engineering Institute (SEI) at Carnegie Mellon University, to support the instructional goals laid out by the creators of the SEI’s professional education program and the designers of the individual courses. However, to date, these courses have lacked any assessment component, certification for the course being granted solely on attendance. For an assessment component to be meaningful, it must derive from and support these instructional goals, determine which goals must be assessed and how to assess them, and determine how best to assess whether those goals have been achieved through application of the instructional materials, lectures, and activities which are included in each two-day course. In order to ensure that the course assessments target the intended learning goals, we developed content for low-stakes assessment components grounded in education theory, combining current knowledge of educational psychology and the software engineering domain to create evaluations that will effectively determine whether participants in this course have learned what the curriculum developers and the instructors intend them to learn.

1. Introduction
This paper describes the process of creating meaningful assessment components for two professional development courses in software architecture, offered by the Software Engineering Institute (SEI) at Carnegie Mellon University. The methodologies used in our project, and described in this paper, have the potential to help create meaningful assessments for other professional development courses in the SEI, and elsewhere in the software engineering domain. The methodologies we used included participation in and observation of sample classes, semi-structured interviews with the course instructors, and iterative reviews of the instructional materials, as well as much work to ensure the creation of assessment content that could be utilized within the framework of the SEI’s context for professional development education.

2. Background
Professional development education in software engineering presents different challenges from those inherent in undergraduate and graduate software engineering education. Working software engineers and architects require regular retraining in order to keep pace with the ever-present changes in methods and technologies [5]. Yet it is not feasible for most working professionals to spend the same amount of time on reeducation that higher education students spend in school. Software engineering professionals, therefore, are more likely to try to self-
educate, and if they can, to participate in brief courses on specific topics within their desired domains. Some large development organizations sponsor such courses at their own sites, from time to time, e.g. if a sufficient number of engineers need retraining for a particular project [6]. For the more esoteric, or higher-level, software engineering disciplines, of which software architecture is one, few professional development courses exist to train software engineers tasked with architectural responsibilities. Although most software architects are senior engineers, with many years of professional experience, relatively few of them have had formal education in software architecture design, since software architecture was not a part of most software engineering curricula until relatively recently.

One critical challenge for professional development education is the selection of curricula, which can significantly benefit students in a very limited instruction period, often days. This is a sufficiently difficult problem in selecting software engineering curricula for higher education, in which a topic may be investigated in depth over an entire semester. [3] An undergraduate or graduate degree in software engineering indicates a good deal of time devoted to in-depth study and learning, and many assessments of the student’s proficiency.

Professional development education must be even more selective in choosing curricula and materials, since there is far less time to spend in the classroom, and also far less time to spend in assessing student gains in learning. The Software Engineering Institute (SEI) at Carnegie Mellon University offers three certificate programs in software architecture to professional development students. Each of these certificates is based on completion of between two and six courses in the SEI’s Software Architecture curriculum, over a period of two years. Each course takes place over two consecutive days, with six hours of instruction per day, with class sizes ranging from 12 to 30 students.

In their present embodiment, these software architecture certificate programs have no assessment component, the certificates being granted solely on the basis of attendance in specific sequences of courses. While course attendance itself is valuable and necessary to the learning process, an assessment of student learning is a critical step in verifying the impact of the courses on individual students, as well as providing a gating function that the SEI may use to determine whether a certificate should be granted.

Our project, therefore, sought to develop assessment components for two courses: Software Architectures: Principles and Practices (SAPP), and the Architectural Tradeoff Analysis Method Evaluator Training (ATAM Evaluator) course. The first of these is a required component of each of the three certificate programs in software architecture offered by the SEI. The two courses, taken in sequence, fulfill the requirements for the ATAM Evaluator certificate program. The assessment for these two courses would not be required in order for a student to progress to additional courses, but would be a required component of qualification for the ATAM Evaluator certificate.

3. Designing the assessments

3.1. Consolidation vs. separation of exams

Since there are two required courses in the ATAM Evaluator certificate track, the assessment could take one of two forms:

a) a single, consolidated assessment, covering the content of both courses, or

b) two separate assessments, one covering the SAPP course and one covering the ATAM Evaluator course.
Although there is some overlap in the knowledge being tested between these two courses, we determined that it was advisable to create two separate assessments, one for each course in the certificate track. Such a separation yields multiple benefits. Since the SAPP course is also a required course in the other two certificate tracks at the SEI, separating the assessments would enable later reuse of the individual assessments to accommodate future assessment-based certificates in the SEI’s software architecture course program. Additionally, students would benefit, as they would be able to take the assessments immediately, or soon, after completing each course, which is preferable to a situation in which as much as two years might elapse between the completion of one course and the next. Multiple assessments for a single certification, each assessment covering materials learned in a separate course, are extant elsewhere in professional development educations, e.g. Cisco’s CCNA certification program.

3.2. Goal-aligned assessments: What to test

For an assessment of student learning in a course to be meaningful, the assessment should be aligned with the instructional goals of the course, and the instructional materials and activities used during the course itself. The instructional materials and activities should also be aligned with the instructional goals. If these factors are all present, a course may be created in which goals for learning are established at the outset, in which the instructional materials and activities are selected and/or designed to support these goals, and assessments measure student achievement in learning those goals, as expressed in the materials and activities used for instruction [4]. Goals, instruction, and assessment are each interconnected, and may be traced from one to the next. [Fig. 1]

![Figure 1. The Goal-Instruction-Assessment Triangle](image)

The process we developed to create a goal-aligned assessment for two courses with existing curricula, goals, and instructional materials is described in Section 4 below.
3.3. Delivering the assessments - Timing

Initially, we had conceived of these assessments as paper-based examinations, administered by the course instructors during the final hour on the second day of each course. In a traditional classroom environment, this might be readily achieved. However, observation of the courses in practice revealed that this was unlikely to be a practical or desirable method of administering the assessments, for several reasons.

Firstly, in each offering of one of these two-day courses, professionals enrolled in the course must travel to the SEI for the duration of the course, generally interrupting their regular work schedule to do so. As a result, in each course, there are at least one or two students who must leave a little early on the second day to meet a departure time at the airport, rendering it impossible for those participants to take an exam during the final hour of the course. This is a typical circumstance in professional development courses, and requires a different assessment than would be feasible for full-time students.

Secondly, the two courses are tightly structured to utilize the full instruction time over the two days each course takes to administer. It is fair neither to the instructors nor to the students to shorten the already cramped teaching time.

Thirdly, not all students in each course desire the certificate. However, some students who do desire the certificate may not reach this decision until after completing the course and leaving the course site. It is desirable for the SEI to grant all students in the courses the opportunity to take the certificate assessments, provided the students have attended the required courses.

Lastly, and perhaps most importantly, administering the exam during the last hour of the course admits no opportunity for reflective learning or additional study. Either or both of these may be necessary for deep learning to take place [1], and they are certainly available to those students who sit for some professional certification examinations in related domains.

We therefore recommended that the assessment for each course should be available at any time after the student has completed that course, and before the two-year limit for the student to complete the certificate track has expired. The course curriculum is currently fixed, so that the same assessment should be appropriate from one offering of the course to the next.

3.4. Delivering the assessments - Format

Assessments for the SAPP and ATAM Evaluator courses could have been delivered in one of several common forms:
- Paper-based assessment,
- Oral assessment, or
- Computer-based assessment.

Oral assessment, while offering opportunities for a rich understanding of the student’s level of learning, was ruled out immediately, as being too costly in terms of instructor time, as well as prone to subjective differentiation of results. Paper-based assessments, while very easy to distribute, still have a high time-and-effort cost to the instructors, or whoever will be performing the grading.

We determined that a computer-based, multiple-choice assessment could be self-administered, automatically graded, and would allow for remote testing for the overwhelming proportion of out-of-town students. Additionally, such an assessment would enable ongoing flexibility in choosing from a pool of questions, rather than using an identical set of questions each time. This option would require more initial set-up time than a paper-based assessment,
but would be maintainable, and could be set up to create its own audit trail. We therefore elected to design a multiple-choice, online, computer-based assessment.

3.4. Length and duration of assessments

Any assessment administered remotely must be assumed to be “open book.” If the goal of the assessment is to measure student learning, rather than the ability to look up answers in reference materials, it is desirable to set a time limit for the assessment, with an appropriate number of questions to be answered during that time. We examined duration and content of other multiple-choice assessments, including Cisco’s CCNA course, and the GRE and SAT examinations. We thereby arrived at a recommendation that the assessment for each of the SEI courses should have a maximum duration of 60 minutes, during which time a student would answer between 40 and 50 multiple-choice questions. An assessment duration of one hour per course would result in two hours of testing being required to qualify for the SEI ATAM Evaluator certificate.

4. Creating the assessments

This section describes the process we developed to create the assessment content for the two courses. The content for both assessments was developed using the same process.

4.1. Participation in the courses

The goal of the project was to create assessment components for two existing courses. We began with direct observation of the educational setting, and of how the existing instructional materials were used in that setting. To this end, the first author participated as a student in two offerings of the SAPP course, and one offering of the ATAM Evaluator course. This enabled us to observe, from a student’s perspective, the application of the written instructional materials, and the admixture and dynamics of interactive class activities.

4.2. Elicitation and refinement of instructional goals

The existing instructional materials for each course consisted primarily of printouts of PowerPoint presentations, through which instructors guided the class over the course of the two days. Each day was broken into four 90-minute sessions. The majority of the SAPP course consisted of lectures, with a few group exercises. The ATAM Evaluator course was more evenly divided between lecture sessions and group exercises, with the concepts used in the exercises clearly delineated in the printed instructional materials. Although a copy of a software architecture textbook was also provided to each student in both classes, the book was hardly used during the courses, the key concepts having already been included in the printed materials.

We therefore began our elicitation of a list of instructional goals for each course, separately, using the printed instructional materials. The direct participation experience enabled us to determine which of the concepts and terms in the printed materials had been most emphasized by the instructors during the course sessions. Our preliminary lists contained over 150 instructional goals for each course, ranging from the very general (e.g. the student should be able to identify how architecture design generally fits into any software life cycle) to the highly specific (e.g. the student should know the meaning of the term “sensitivity points” in the context of the ATAM).

Next, we performed semi-structured interviews with each of the course instructors, over
several hours, to review our extracted lists of instructional goals. We also sought to develop insight and receive feedback from the instructors, regarding any intended instructional goals that might not have been apparent to us from the printed materials. In conjunction with the instructors, we categorized each instructional goal in our list as to its importance (primary, secondary, tertiary), and the level of knowledge the student should have about that goal (declarative, procedural, metacognitive).

4.3. Revising instructional goals

Based on the results of the instructor interviews, we were able to eliminate some of our original instructional goals, as not all the goals that might be reasonably inferred from the printed instructional materials, were considered necessary by the instructors. Additionally, some goals were considered by the instructors to be interesting but not important to test; these were classified as being of tertiary importance, and were not included in the final set of instructional goals.

One point of interest revealed by the instructor interviews, from an education research standpoint, was the elimination of metacognitive knowledge as a desired level of student understanding. In a developmental setting, in particular, there is a strong and growing interest in the metacognitive processes in which students engage as they learn a new subject. Here, the instructors were not particularly interested in testing whether the students understood how they had arrived at a particular piece of knowledge; the metacognitive processes were regarded as an adult responsibility of the students, and the instructors were interested solely in testing declarative knowledge and conceptual understanding of the course content, not in the formative learning that is of such pressing importance in K-12 education. Although we do not have enough evidence to generalize broadly from the number of instructors we were able to interview, it would be interesting to explore whether the de-emphasis on metacognition is a common feature of professional development education.

The interviews also revealed that the instructors were interested not in evaluating procedural learning, which would imply mastery of particular skills, but in conceptual understanding, i.e. comprehension of concepts introduced during the course. We therefore revised our categories of desired level of knowledge to include only declarative and conceptual knowledge. We reviewed the prioritized lists of instructional goals with the instructors, and determined, in consultation with them, that the assessments would be based on goals that were deemed to be of primary or secondary importance, and would measure a student’s acquisition of declarative or conceptual knowledge of the instructional materials covered during the courses.

The composition of the final set of instructional goals is shown in Table 1 below.

<table>
<thead>
<tr>
<th>Priority</th>
<th>Level of Desired Student Knowledge for Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Declarative</td>
</tr>
<tr>
<td>Primary</td>
<td>Included</td>
</tr>
<tr>
<td>Secondary</td>
<td>Included</td>
</tr>
<tr>
<td>Tertiary</td>
<td>Excluded</td>
</tr>
</tbody>
</table>

Table 1. Composition of final set of instructional goals
4.4. Creating assessment items

Through our process thus far, we had ascertained that every instructional goal we sought to test with the assessment, could be traced directly to the instructional materials. The next step in creating the assessments, therefore, was to create a set of questions for the assessment, such that each instructional goal was addressed by at least one question in the set. The process of prioritizing the instructional goals for the two courses had resulted in lists of 139 instructional goals for the SAPP course (77 of primary importance, and 31 secondary), and 125 instructional goals for the ATAM evaluator course (83 primary, 25 secondary). It was possible, in many cases, to address multiple instructional goals with a single question. This was especially true in cases where one goal was to convey the understanding of a concept, and another goal existed to convey the understanding that the reverse of that concept was untrue, or where a particular goal was a generality, whose specifics were already fully covered in other goals. The final sets of assessment questions included 80 questions for the SAPP course, which included more goals that could be effectively consolidated, and 91 questions for the ATAM Evaluator course.

For each question in each set, we then created four possible answers: one correct answer, one answer close to the correct answer, one further from correct, and one distant or opposite to the correct answer. Since a risk of using multiple-choice questions is that they measure recognition rather than recall [2], and since the goal of the assessment is to measure student learning, we took care not to invent possible answers out of thin air, but to draw all possible answers from the instructional materials used in each course.

5. Conclusions and future work

5.1. Conclusions

During the course of this project, we created content for assessments for two existing professional development courses in software architecture. In so doing, we developed a process for creating these types of assessments, such that the resulting content was of similar weight and type across the two different courses, despite the differences in specific curricula between the two courses. This is especially valuable, as the two courses comprise a single certificate track within the professional development education organization in which they are being taught.

Additionally, the process we developed in the course of this project could potentially be applied to the problem of creating assessment components for other professional development courses with existing curricula and instructional materials, either in the software architecture domain, or in other software engineering or information technology domains. To a teaching organization such as the SEI, the development of such a process could enable them to create assessment components for many, if not all, of the other existing courses in their software architecture program, as well as other programs, readily and in a timely fashion.

5.2. Future work

Initially, we had planned to create an implementation framework for the online assessments, and deliver this framework to the SEI along with the question sets. However, the SEI determined that an assessment delivery framework already existed in another domain area in which they grant certifications, and requested that we deliver the question sets to them, but allow them to deliver the assessment through their existing framework. To date, the
question sets have been delivered to the SEI, but no assessment of students in the courses has yet taken place.

Before the assessments are instituted as a required gating function for receipt of the ATAM Evaluator certificate, the questions should be calibrated. We anticipate that calibration of questions will be performed first with a sample group of SEI personnel, and then with a group of actual students in the two courses. We plan to assist the SEI in calibrating the assessments, based on feedback from the SEI sample group, and the results of early applications to assessment of actual students in the courses.

6. Acknowledgements

The authors would like to thank Bonnie E. John, Sharon Carver, Brian Junker, Rob Wojcik, Felix Bachmann, Linda Northrop, and Jack Mostow for their advice and assistance. This work was supported in part by a Graduate Training Grant awarded to Carnegie Mellon University by the Department of Education (R305B040063). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Department of Education, or the U. S. Government.

7. References