COMPLETING THE CYCLE: MEANINGFUL COURSE EVALUATIONS

Timothy J. Ellis

Abstract: Good educational practice, to say nothing of standards imposed by every regional accrediting association, mandate that an educational institution perform ongoing evaluations of academic programs. Perhaps the most prevalent method in which schools address this mandate is to use a general-purpose ‘smile sheet’ that attempts to measure the student’s satisfaction with the course.

This type of evaluation is certainly flawed. Although there have been a number of studies supporting the reliability and validity of these general-purpose instruments, they do not provide the granularity of feedback necessary for the instructor to make decisions regarding changes necessary or desirable in future iterations of the course.

This presentation will discuss the results of a study that tested the efficacy of a system for evaluating courses at the design level based upon the test matrix approach commonly used in software engineering. Three questions are addressed: 1) what data are necessary for meaningful design-level assessment; 2) how can that data be gathered; and 3) how can that data be interpreted.

Index Terms – assessment of instruction, program evaluation, course design

INTRODUCTION

Both good educational practice and the mandates of regional accrediting associations and professional licensing organizations require an educational institution to perform ongoing evaluations. Despite this mandate being clearly articulated, there is not a widely accepted, detailed method for addressing it. There is little agreement on the constitute elements and appropriate methodology for meaningful evaluation of educational programs. What data is necessary for a meaningful assessment of educational services, how can that data be gathered, and, once gathered, how can it be analyzed, interpreted, and integrated in the decision making process.

Gathering Data

There are at least three different types of evaluation of educational offerings: at the program or major level, at the specific course-offering implementation level, and at the course design level. Accrediting bodies, for example, focus primarily on assessment of the program or major as a whole. This level of assessment is concerned with capabilities and potential of program graduates and includes factors such as [1] “…student portfolios, including design projects; nationally-normed subject content examinations; alumni surveys that document professional accomplishments and career development activities; employer surveys; and placement data of graduates” (p. 2). Responsibility for this level of assessment is generally shared by administration and faculty and is conducted as a post hoc review of graduate success indicators. Although evaluation at this level is a necessary component of an assessment plan, it typically lacks the granularity necessary for an instructor to make decisions regarding the courses she or he teaches.

Granularity can certainly be added to the assessment process by conducting in-process evaluations of a specific course implementation. Typically categorized as “classroom assessment”, this level of evaluation is [2] “… learner-centered, teacher-directed, mutually beneficial, formative, context-specific, ongoing, and firmly rooted in good practice” (p. 4). Responsibility for this level of assessment falls to the individual instructor and focuses on the effectiveness of a specific course offering. The goal of classroom assessment is to ensure that a specific class offering is “working” for the students enrolled and make necessary adjustments when it is not. Classroom assessment is conducted by the instructor using both informal and formal processes including performance on assignments, questions asked, and class surveys. Although evaluation at this level is also a necessary component of an assessment plan, it is very implementation-specific and serves as only a single data-point. A greater breadth of perspective is necessary to allow the instructor to determine if the course as planned – not just a specific course offering – is effective.

The necessary perspective for the instructor to assess a course plan and design can be provided by meaningful end-of-course evaluations. Unfortunately, this type of assessment is almost always accomplished using a university-wide, standardized form [3] “… used to serve multiple purposes; for feedback to lecturers for improving their teaching, for publication to the student body for course registration decisions, for administrative course evaluation, and for promotion and tenure decisions” (p. 2). Although such student ratings of instructors and courses have been used extensively and well tested in terms of reliability and validity [4, 5], by the very fact that they are general purpose, multifunction instruments, they lack the focus and specificity necessary to provide meaningful insight into the design of the course [6].
Steps such as design, coding, and testing. In terms of course design, these activities can be considered the initial stages of a design process, much like those in software engineering which involve planning and designing a course in that the process involves analyzing, planning, designing, implementing, testing, and evaluating the course to ensure it meets the intended learning outcomes.

The process of planning and designing a course is in many ways similar to developing a computer product in that it is an ordered, systematic process with identifiable tasks and milestones [11]. As in the case of software engineering, there are a number of different theories of the proper instructional design process [12, 13]. For the purposes of this discussion, borrowing the waterfall model from software engineering to describe the processes involved in planning and designing a course would be appropriate.

The waterfall model entails four distinct steps: analysis, design, coding, and testing. In terms of course design, these steps might be described as:

- **Analysis by developing learning outcomes.** Learning outcomes form the essence of a requirements specification for a course [14, 15]. The learning outcome must clearly specify what the student should be able to do, the conditions under which the student should produce the desired behavior, and how well the student must be able to perform it. In addition to focusing on the specific behavior, learning outcomes must also reflect the level of cognitive activity expected of the student.

- **Design by matching outcomes with activities.** One or more activities must be identified for each learning outcome. The activities might include lectures, reading assignments, individual or group projects, exams, or laboratory assignments. The exact nature of the activities selected is dependent on a combination of factors including course topic, level, student profile, and learning environment [12]. Regardless of the activities selected, it is imperative that each activity be associated with at least one learning outcome and each learning outcome be facilitated by at least one activity.

- **Code by assembling and ordering the activities.** The sequencing and timing of the assignments and other activities planned for a course is, of course, essential. Some activities are prerequisite for others; some require lengthy gestation periods; and some offer little flexibility in terms of scheduling. All these factors must be considered when assembling a course that will indeed be a synthesized whole.

- **Test by evaluating the course.** As in the case of testing software, there are two basic questions to be answered in evaluating the course: were the goals met (validation), and did the course function as planned (verification). Also, as with testing software, it is essential to incorporate feedback from the end users – in the case of a course design, the students – in the testing process.

**Gathering Assessment Data**

Typically, the data for course design evaluations are derived from a standardized, anonymous, multipurpose questionnaire that is distributed by the school administration at the end of a term. This type of questionnaire is inadequate for a number of reasons. Since the questionnaire is not specific to the course and is administered by the school administration, the student may not associate significance to the questions and answer in haste or base ratings on factors other than those asked [7]. There is, further, a strong indication that responses on this type of instrument are more reflective of the student’s happiness with the course or even the school as a whole than of the student’s opinion of the specific item being queried [9].

To assess the design of a course effectively one must directly verify and validate the specifics of that course [6, 10]. In general, there are two essential questions that must be answered: have the learning outcomes identified for the course been attained (validation), and did the course function as planned (verification)? Figure 1 illustrates examples of these verification and validation techniques.
of each type of question, taken from an evaluation for a graduate-level educational database systems course.

Interpreting and Integrating Data

Data are of value only when they can be processed into information and used to support decisions. The raw data from a course assessment can be quite misleading. For example, an average rating of 3.0 on a 4-point scale might seem to be quite satisfactory until one realizes that the average rating on that scale is 3.75. Taken in that context, the 3.0 rating would seem to indicate that design changes are necessary, but are they really? Is the size of the difference adequate to indicate a meaningful difference?

This problem of interpretation can be addressed by treating the results of the course assessment as experimental data. From this perspective, one would view assessment as a test of a hypothesis [6]. In order to test the hypothesis, one must make a comparison. What should be compared to what? Remembering that the assessment is targeted at course design, and not just a single course offering, one set of factors in the comparison should be all the assessment data related to the course design being evaluated. Further, since comparisons are meaningful only when based upon common factors, the data for the course-design being evaluated should be compared to data derived from similarly structured evaluations of other course-designs. In specific, the data related to the attainment of each learning outcome for the course design being assessed should be compared to the data derived from the assessments of learning outcomes from all other course-designs evaluated. Similarly, the data related to each pedagogical tool being assessed should be compared to the data derived from the assessments of pedagogical tools from all other course-design evaluations.

Finally, the most appropriate statistical instrument to use to test the hypotheses must be determined. Historically, data from student evaluations of courses and instructors is not normally distributed [9]. Although processes to normalize the data are available, the rather small number of responses typically found with course assessments, coupled with the fact that the data is really nominal even though numeric values can and have been assigned to the ratings, makes parametric tests of questionable value. The Chi Square test – a non-parametric test of statistical significance by comparing observed grouping of results across categories with expected groupings – offers promise as an instrument to analyze course-design assessment data.

RESULTS

Over a two-year period, course-design assessment instruments based upon measuring the level of student confidence in attaining learning outcomes and perception of efficacy of the pedagogical tools used to facilitate attainment of those outcomes were used to evaluate the design for four different graduate-level computer technology courses. Figure 1 illustrates the questions from the end of course survey used in assessing a graduate-level course in educational database systems.

A total of seven course offerings were included in the study. The first five course offerings were used to establish an expected distribution of ratings across the categories of both learning outcome attainment and pedagogical tool efficacy. Table 1 presents the distribution of the rating scores from the set of five course offerings used to establish the expected distribution.

<table>
<thead>
<tr>
<th>Learning Outcomes</th>
<th>% of Total Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>0%</td>
</tr>
<tr>
<td>I don’t think so</td>
<td>1%</td>
</tr>
<tr>
<td>I think so</td>
<td>37%</td>
</tr>
<tr>
<td>Yes</td>
<td>62%</td>
</tr>
</tbody>
</table>

The final two course offerings included in this study were used as a pilot test of the analysis of the assessment data. Both of the course offerings were instances of the same course-design of a graduate level course in educational database systems and were taught by the same instructor. Table 2 presents the results of the Chi Square analysis of the design for this course.

A couple of points must be addressed in reviewing the distributions and Chi Square results. First, since there are several cells with fewer than five instances, one of the assumptions of the Chi Square test is violated. Violation of this assumption increases the potential for Type I errors. Since the purpose of this analysis is to identify potential faults with the design of a course, the elevated alpha level is not of great concern since it would only entail re-evaluating a design that perhaps did not necessarily warrant re-evaluation.

Secondly, as can be seen in Table 2, one of the learning outcome Chi Square calculations did in fact produce a statistically significant result. The confidence level for outcome five did not match the expected distribution. It is not clear, however if that result should be interpreted as a flaw in the design of the course. An examination of the learning outcome – Produce meaningful original work in the area of applying computerized database management systems in educational settings – indicates that it addressed a very high level of scholarly work. Perhaps a lower confidence level for this specific learning outcome is in fact appropriate. The data from any assessment process can be used only to support, not replace, decision making.
Educational Database Systems
Course Evaluation Survey

I want to promote the very best learning experience for the students in my classes as possible. The only way I can do so is to receive meaningful feedback from the people who have completed the course. Please take a few minutes to complete the following survey. It should take no more than 15 minutes to answer the questions, and your responses will help keep the class current, challenging, and enjoyable.

There were six learning outcomes for this course. Please indicate how well you though you attained each of the outcomes by indicating how confident you feel about doing the following tasks:

1) Plan, develop, and document an educational database management system.
   - Nope, can't do it
   - I don’t think so, but it sounds familiar
   - I think so, after some research
   - Yes, no problem

2) Identify and analyze the technological impediments to implementing computerized database management system solutions in an education setting.
   - Nope, can't do it
   - I don’t think so, but it sounds familiar
   - I think so, after some research
   - Yes, no problem

3) Analyze educational database implementations, identifying strengths and weaknesses based upon a review of appropriate literature.
   - Nope, can't do it
   - I don’t think so, but it sounds familiar
   - I think so, after some research
   - Yes, no problem

4) Evaluate problem areas in the education field and analyze the appropriateness of a computerized database management system as a solution.
   - Nope, can't do it
   - I don’t think so, but it sounds familiar
   - I think so, after some research
   - Yes, no problem

5) Produce meaningful original work in the area of applying computerized database management systems in educational settings.
   - Nope, can't do it
   - I don’t think so, but it sounds familiar
   - I think so, after some research
   - Yes, no problem

6) Work effectively as a member of an educational database development team by collaboratively developing a requirements document, project schedule, entity-relationship diagram, and data dictionary and collectively produce an educational database application.
   - Nope, can't do it
   - I don’t think so, but it sounds familiar
   - I think so, after some research
   - Yes, no problem

There were five instructional tools used during this course. Please indicate how important each of the tools was in helping you attain the course outcomes.

1) Online discussion forums
   - No value
   - Of some value
   - Of significant value
   - Vital

2) Technical briefs
   - No value
   - Of some value
   - Of significant value
   - Vital

3) Computer Managed Instruction paper
   - No value
   - Of some value
   - Of significant value
   - Vital

4) Group practical application
   - No value
   - Of some value
   - Of significant value
   - Vital

5) E-Mail to professor
   - No value
   - Of some value
   - Of significant value
   - Vital
The evaluation of educational services that is mandated by both good educational practice and accreditation standards is multi-tiered. When done at the highest level, evaluation entails assessing the major or program as a whole to ensure that the graduates have been exposed to the necessary experiences to make them viable in the greater marketplace. When done at the lowest level, academic evaluation entails examining a specific course offering to ensure that the students enrolled in the class are in fact benefiting from the course through distribution of grades, feedback from instructors who teach the same students in subsequent courses, and, sometimes, performance on standardized tests, to list just a few. Just as in the case of testing a new software product, however, it is essential that the targeted end users – the students – be a vital part of the evaluation team. Any assessment of a course design must include feedback directly from the students regarding their level of confidence in attaining the learning outcomes specified for the course and their perceptions of the efficacy of the pedagogical tools in facilitating attaining those outcomes.

### Interpreting the Data

It is important not to confuse evaluating the course at the design level with evaluating an implementation of that design in a specific course offering. Even well designed courses might not work well for a specific grouping of students. Although instances when a course with a proven track record of success simply “does not work” should not be ignored, they should likewise not be given undue weighting. Treating the data derived from assessment of a course design as if it were experimental data being accumulated over time, and analyzing it with accepted statistical instruments such as Chi Square offers promise as an effective decision support approach.

### Implications for Future Research

This study might best be viewed as a proof of the conceptual validity of conducting a separate, controlled, design-level assessment of college courses. The methodology employed serves as a starting point for this type of assessment, but several very important areas for continued research are clearly indicated:

1. Is there a true “baseline” against which course designs can be compared? This study used the data derived from evaluations of three course designs spread over five course offerings to set a baseline for assessing a fourth course design through two course offerings. Further research into the basis upon which a course design can be evaluated would be of value.

2. The instrument used to gather data for design-level assessment focused exclusively on the elements of the pedagogical tools in facilitating attaining those outcomes.

### Necessary Data

The learning outcomes the students are to attain and the set of assignments and activities – pedagogical tools – selected to help the students accomplish those outcomes are two essential elements in the design of a course. Any analysis of a course design, therefore, must include a direct assessment of how well the course as designed worked in helping students attain the learning outcomes.

### Gathering Data

There are a number of sources for data regarding how well the course as designed does in fact help students attain the desired learning outcomes – the instructor’s evaluation of the course, the students’ feedback, feedback from instructors who teach the same students in subsequent courses, and, sometimes, performance on standardized tests, to list just a few. Just as in the case of testing a new software product, however, it is essential that the targeted end users – the students – be a vital part of the evaluation team. Any assessment of a course design must include feedback directly from the students regarding their level of confidence in attaining the learning outcomes specified for the course and their perceptions of the efficacy of the pedagogical tools in facilitating attaining those outcomes.

<table>
<thead>
<tr>
<th>Outcome #</th>
<th>No</th>
<th>I don't think so</th>
<th>I think so</th>
<th>Yes</th>
<th>Chi Square</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>15</td>
<td>3.435</td>
<td>0.179</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>19</td>
<td>1.433</td>
<td>0.488</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>27</td>
<td>3.282</td>
<td>0.194</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>23</td>
<td>0.355</td>
<td>0.837</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>13</td>
<td>6.395</td>
<td>0.041</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>26</td>
<td>4.656</td>
<td>0.097</td>
</tr>
</tbody>
</table>

### Pedagogical Tools

<table>
<thead>
<tr>
<th>Tool #</th>
<th>No value</th>
<th>Some value</th>
<th>Significant value</th>
<th>Vital</th>
<th>Chi Square</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>8</td>
<td>11</td>
<td>15</td>
<td>3.831</td>
<td>0.28</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>4</td>
<td>14</td>
<td>16</td>
<td>0.362</td>
<td>0.948</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>4</td>
<td>8</td>
<td>8</td>
<td>0.878</td>
<td>0.831</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>20</td>
<td>3.959</td>
<td>0.266</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>21</td>
<td>5.382</td>
<td>0.146</td>
</tr>
</tbody>
</table>
course, not on the course take as a whole. In a well designed course, the whole is greater than the sum of its parts. Research into how to structure the assessment instrument to gauge the course as a whole as well as its component parts would be quite interesting.

3. Other means of analyzing the course design assessment data likewise offer potential for future research efforts. For example, rather than doing a comparative analysis of the assessment of a course design against a baseline, a probability estimate using Bayesian belief analysis offers interesting potential. Research investigating other methods of analyzing assessment data is indicated.

4. Student feedback is only one aspect of a comprehensive evaluation of a course design. Other sources of meaningful data include the impressions of the course instructor and of instructors who later teach the same students. Feedback from the ultimate employers of the students would also be of great value. Research in how to gather, analyze, and use data from these other sources would be valuable.

REFERENCES