Revising Computer Science Learning Objects from Learner Interaction Data

L.D. Miller, Leen-Kiat Soh, Beth Neilsen, Erica Lam, Ashok Samal
Department of Computer Science and Engineering
University of Nebraska
Lincoln, NE 68588
1-402-472-3200
{lmille, lksoh, bneilsen, elam, samal}@cse.unl.edu

Kevin Kupzyk, Gwen Nugent
Center for Research on Youth, Family and Schools
University of Nebraska
Lincoln, NE 68588-0342
1-402-472-2448
{kkupzyk2, gnugent1}@unl.edu

ABSTRACT

Learning objects (LO) have previously been used to help deliver introductory computer science (CS) courses to students. Students in such introductory CS courses have diverse backgrounds and characteristics requiring revision to LO content and assessment to promote learning in all students. However, revising LOs in an ad hoc manner could make student learning harder for subsequent deployments. To address this problem, we present a systematic revision process for LOs (LOSRP) using proven techniques from educational research including Bloom's Taxonomy levels, item-total correlation, and Cronbach's Alpha. LOSRP uses these validation methods to answer seven questions in order to diagnose what needs to be revised in the LO. Then, LOSRP provides guidelines on revising LOs for each of the seven questions. As an example, we discuss how LOSRP was used to revise the content and assessment for 16 LOs deployed to over 400 students in introductory CS courses in 2009. Lastly, although initially designed for LO revision, we briefly discuss how LOSRP could be used for assessment revision in intelligent tutoring systems.

Categories and Subject Descriptors

K.3.2. [**Computer and Education**]: Computer and Information Science Education.

General Terms

Experimentation

Keywords

learning objects, intelligent tutoring system, systematic revision process

1. INTRODUCTION

LOs have previously been used successfully to teach introductory computer science courses to students [3][7][8][9]. For example, students who used the LOs achieved significantly higher scores than students in the control group [9] and rated the LOs highly in terms of usefulness and appropriateness [8]. However, such introductory courses contain students with a wide variety of backgrounds. For example, at our university, students in introductory computer science (CS) courses run the gamut from non-majors taking a single, required CS course to CS and computer engineering majors to honors students in a hybrid CS and business program. These students have also been shown [8][10] to have very diverse characteristics in terms of motivation, self-efficacy, and programming background. This diversity makes it difficult to create LOs that are appropriate for all prospective students. Content and assessment for LOs may be easy for students with some backgrounds but difficult for other students. A question so difficult that students cannot choose the correct answer from the distracters does not promote learning. The same is true for questions that are so easy that the distracters are completely obvious. Thus, it is often necessary to evaluate and revise the LO content and assessment. However, unlike existing Science, Technology, Education, and Math (STEM courses), "computer science courses do not yet have a similar set of validated assessment tools" [12]. This can make it extremely difficult for content developers who try to revise the LOs. Doing this in an ad hoc manner runs the risk of accidently making them worse.

We present a systematic revision process for LOs (LOSRP) using the assessment data collected from actual deployment data. This process was developed after extensive collaboration with educational experts. LOSRP uses proven assessment validation tools from educational research including item-total correlation [13], Cronbach's Alpha statistic [6], and Bloom's Taxonomy [4] levels. These methods are used to provide the answers for seven separate questions that diagnose what part of the LO requires revisions: (1) Does the assessment use multiple mark questions instead of true/false or multiple-choice? (2) Is the item-total correlation on the assessment questions below a threshold? (3) Are specific distracters in the questions not being chosen? (4) Are specific distracters chosen more often than the correct answer? (5) Is the Cronbach's Alpha statistic for the assessment below a threshold? (6) Do the questions fail to cover all the Bloom's Taxonomy levels? (7) Are objectives of the content and the questions inconsistent? All these questions are commonly used for assessment validation in the educational field [2]. To the best of our knowledge, this is the first use of such questions for CS LO revision. Second, LOSRP provides guidelines for revising the LO to address each question listed above. An extensive example of using this systematic revision process on LOs deployed to introductory CS students is also presented in this paper.

Note that the systematic process described here has broader application than just to LOs. Using the same principles, this process can also be used for content and assessment revision in Intelligent Tutoring Systems. We see two possible applications. First, intelligent tutoring systems (ITS) could use our process to improve the assessment by selecting appropriate distracters for questions and appropriate questions for the assessment. Second, the ITS could also use this process to diagnose whether students are struggling with the content learning objectives or the assessments questions.

The remainder of this paper is organized as follows: Section 2 provides related work on assessment revision in current computer science and education research and background on the education methods used to answer the questions. Section 3 discusses the design and deployment of the LOs. Section 4 discusses the guidelines for addressing the questions in more detail and provides examples of LO content and assessment revision. Finally, we provide conclusions and discuss future work on improving the LO revision process.

Note that in this paper we focus primarily on the assessment revision for the LOs. The statistics used in LOSRP consider only the assessment data. For Question (7), the consistency between the content objectives and assessment questions is measured only qualitatively. The LOSRP is by no means complete or optimal. However, this is the first step towards a systematic revision process for CS LOs. Also, LOSRP could make an impact in automating this process for intelligent tutoring systems as noted above. The LO revision process should also include revision of the tutorial and exercises of the LOs in addition to the assessment. For example, factors such as empirical usage behavior that captures the interactive sessions (e.g., questions that are too long or too time consuming and inconsistent correlation between time spent on tutorial concepts and the corresponding questions) can be used to revise content from time and learner attention viewpoints. To illustrate, if the average view time spent on a concept in the tutorial is much longer than the average time spent on viewing the set of questions associated with that particular concept, then perhaps there should be more questions for that concept or the discussions on that concept should be shortened. We will consider this in more detail in the future work.

2. RELATED WORK AND BACKGROUND

This section first discusses assessment revision in current CS and education research. Then, it provides background on the educational methods used in the LOSRP.

2.1 Assessment Revision

Previous work on assessment revision in CS education research has considered many of the questions used in LOSRP. This includes using multiple-choice questions (MCQ), revising the question distracters, using Bloom's Taxonomy to create the questions, and making sure that the questions cover the learning objectives. However, none of this previous work takes an integrated view in which all these approaches are used together. Further, none of this previous work involves revising LOs for CS. Instead, previous work focuses on ITS or developing the curriculum for CS courses. We provide summaries on such previous work below. Table 1 categorizes current CS education research on assessment revision based on our seven questions.

Tew and Guzdial [12] claim that CS lacks strong assessment validation found in STEM courses. Such assessment validation is a necessary step in creating reliable assessments. The authors give a language independent assessment for introductory CS courses. This assessment is based on content commonly presented in textbooks. It contains MCQs developed using experts in CS educations. The authors emphasize using distracters and content objectives coverage for revising the initial questions. However, no pilot study has yet been conducted to evaluate this assessment. LOSRP also considers well-defined statistical measures such as item-total correlation and the Bloom's Taxonomy when revising the questions.

Burge & Leach [5] provide a tool for evaluating how well students understand the learning objectives for introductory CS courses. This tool consists of an Excel spreadsheet that maps learning objectives to assessment questions and evaluates the average score. It was used as part of the ABET accreditation process for the CS program. This tool only considers the raw averages for assessment scores. Our LOSRP uses validation methods such as item-total correlation and Cronbach's Alpha to provide more detailed information on the assessment questions.

Starr et al. [11] use Bloom's Taxonomy (BT) to categorize all the concepts in introductory CS courses. In a case study, the learning objectives and assessment were revised to cover multiple BT levels. In particular, the questions required certain levels of expertise to complete. The authors also determine whether content objectives and assessment questions are consistent through comparing their BT levels. At this time, LOSRP focuses more on assessment than content revision. We use BT levels and several methods to evaluate the assessment questions.

Agarwal et al. [1] use the NetCoach ITS to present introduction to testing concepts to CS2 students. This content is organized in a series of linked hypermedia pages. NetCoach provides recommendations by varying color of the links between the pages (e.g., red link means prerequisites for content are not satisfied). However, students can ignore this advice and visit the page. After students are done navigating through the content NetCoach gives an assessment using MCQs. The authors revised the assessment questions based on an analysis using BT and item-total correlation. LOSRP also considers the distracters in the questions and Cronbach's Alpha when analyzing the scores.

2.2 Educational Statistics

Item-total correlation [13] is used to compare the scores on individual questions in the assessment. Item-total correlation measures the degree of consistency between the score for a single item (i.e., question) with that for the other questions. It uses the correlation between the scores of an individual item and the sum of the scores of the remaining items. For example, if high-scoring students tend to score worse on a particular question than low-scoring students, then the item-total correlation for that particular question would be low. Item-total correlation of 0.30 or higher is generally acceptable in educational research [13].

Cronbach's Alpha [6] is used to estimate the internal consistency of the assessment questions. It increases with the inter-correlation between scores on the assessment questions—similar questions should elicit similar responses. Thus, it can be used to measure the reliability of the assessment questions as a single latent construct (i.e., how closely related the questions are). A Cronbach's Alpha value of 0.70 or higher is necessary for an assessment to be considered valid [6].

Blooms Taxonomy provides a categorization on question difficulty based on the learning objectives [4]. For cognitive learning objectives this includes six ordered processes from easiest to hardest: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. Knowledge questions are generally easy requiring only the memorization of facts. Comprehension questions usually require students to describe or explain certain concepts or demonstrate their understanding, while Application questions usually require students to apply what they understand in a situation or to a problem where its solution is not apparent. Answering Analysis questions often involve analysis, categorization, comparison, and differentiation of different solutions or options. Synthesis and Evaluation questions require an explanation about responses (e.g., why or why not?) making them less suitable for online systems such as LOs. Thus, when revising assessment questions for LOs it is important to include questions covering the first four levels of the Bloom's Taxonomy: Knowledge to Analysis.

Table 1: Assessment Revision in Recent CSE Research. MCQ stands for multiple-choice questions, IT for item-total correlation, ND for distracter not chosen, DC for distracter chosen often, CA for Cronbach's Alpha, BT for Bloom's Taxonomy, and O for learning objectives.

Paper	MCQ	IT	ND	DC	CA	BT	O
Tew & Guzdial [12]	✓		✓	✓			✓
Burge & Leach [5]							✓
Starr et al. [11]						✓	✓
Agarwal et al. [1]	✓	✓				✓	
LOSRP	✓	✓	✓	✓	✓	✓	✓

3. LO DESCRIPTIONS

Sixteen LOs were created to cover a wide range of concepts in introductory CS courses. This includes a range of concepts (see Table 2) from basic concepts on arrays and numeric data to more advanced concepts on sorting and recursion. Each of these LOs follows the same general format. First, the LO contains a tutorial with a list of learning objectives followed by a set of pages explaining the concept using text and figures. Each page of the LO covers a succinct section on the concept equivalent to several pages in a traditional textbook. Second, it contains a set of 1-4 interactive exercises for the students to practice. These exercises require several steps to complete and students receive feedback on their progress at each step. In particular, students are given hints if they become stuck in an exercise. Students are able to repeat each exercise as often as desired. Finally, the assessment consists of between 4-14 questions depending on the amount of content in the tutorial. These questions are designed to measure whether students have learned the content in the tutorial and exercises that was specified by the learning objectives. Originally, assessments included true/false, multiple-choice and multiple mark questions.

In 2009, the LOs were used in five offerings of introductory CS courses at the University of Nebraska, Lincoln. The courses contained over 400 students allowing us to collect a considerable amount of data resulting in more reliable results for the assessment validation methods used in LOSRP. This data also allows us to support additional research on active learning [9], elaborative feedback [8], and automatic metadata generation [8][10]. Interested readers in the details of the software design and research should consult those papers.

All the LOs are built in compliance with the SCORM standard. This makes the LOs usable on any SCORM-complaint learning management systems (LMS) such as Blackboard and Moodle. This greatly simplifies the deployment of the LOs.

The downside to SCORM is that the LMS generally tracks only the student's final score on the assessment. This makes it impossible to evaluate distracters or to use validation methods such as item-total correlation that require data on the answers to the individual assessment questions. Thus, to provide the required assessment data, we have modified each LO to include additional software called the ***LO Wrapper***. It employs a modification of the Easy SCO Adapter [http://www.ostyn.com] to record student assessment data using the SCORM API. The wrapper also records student interactions with the LO (e.g., page navigation, clicks on a page, etc.). It transmits all recorded data (i.e., assessment and interactions) using JavaScript to our remote server.

We have also developed a remote application for our project called the ***MetaGen***. MetaGen runs offline on our server and handles all the data logging, data extraction and data analysis for our project using separate modules. The *data logging module* uses PHP to store the student interactions received from the wrapper in a MySQL database. The *data extraction module* queries the database and processes the data into our dataset. The *data analysis module* computes all the assessment validation methods used to answer the LOSRP questions and also runs data mining techniques (such as feature selection and association rule mining) to support automatic metadata generation [8][10].

4. LO REVISION PROCESS

As previously discussed, LOSRP uses seven questions to diagnose whether LO content or assessment needs revision. In this section, we explain the seven diagnosis questions in more detail. We also provide the LOSRP guidelines on revising LOs for each question. Throughout, as a comprehensive example, we discuss how LOSRP was used to revise our LOs after the 2009 deployment. This includes examples of actual LO content and assessment questions where appropriate. The revisions done to the LOs are summarized in Table 2. Finally, we provide general guidelines on revising both the LO content and assessments.

4.1 Multiple Mark Questions

Multiple mark questions have multiple correct answers and all of them must be selected in order to get the question correct. The problem with these questions is that students can choose a subset of the correct answers and still get the question wrong. Consequently, multiple mark questions in the LOs were either removed or replaced with MCQs. The easiest way to change the multiple mark questions into MCQs is to retain all the multiple mark answer choices, labeling them using numerals, and then offering answer choices that are differing groups of numeral answers. For example, the following multiple mark question in the Arrays LO was converted into the MCQ shown below:

Original: Multiple Mark Question
Which of the following are true of arrays? (Select all that apply)
A. Arrays and loops can be used together to simplify code. (correct)
B. Arrays help avoid the problems associated with memory garbage.
C. Arrays can make it easier to keep track of variable names. (correct)
D. Arrays can help group related variables together. (correct)

Revised: Multiple Choice Question
Which of the following are true of arrays?
1. Arrays and loops can be used together to simplify code.
2. Arrays help avoid the problems associated with memory garbage.

3. Arrays can make it easier to keep track of variable names.
4. Arrays can help group related variables together.
   A. 1 and 2
   B. 2 and 3
   C. 1, 2, and 3
   D. 1, 3, and 4 (correct)
   E. 2, 3, and 4

Table 2. Summary of Revisions to the 2009 Deployment LOs. Entry gives number of questions revised for each LO and number or revisions based on LOSRP guidelines.

LO Name	**Multiple Mark**	**Low item-total correlation**	**Not chosen Distracters**	**Distracters chosen too often**	**New Questions Added**	**Total**
Advanced Logic	3	0	0	0	4	**7**
Advanced Recursion	0	1	0	1	5	**7**
Algorithms	0	6	1	2	4	**10**
Arrays	0	0	0	0	4	**7**
Conditionals	2	2	0	2	4	**8**
Debugging	0	2	1	1	5	**7**
Functions	3	6	2	1	0	**8**
Logic	2	0	0	0	2	**4**
Looping	3	0	0	0	5	**8**
Non-OO Prob Analysis	0	5	1	2	3	**8**
Numeric Data	1	7	1	2	3	**10**
OO Problem Analysis	0	0	1	0	5	**6**
Recursion	1	2	1	0	4	**7**
Searching	2	1	2	0	0	**3**
Sorting	1	3	0	0	0	**3**
Variables & Constants	0	2	2	2	4	**8**
Total	18 16%	37 33%	12 10%	13 11%	52 46%	**111**

4.2 Low Item-Total Correlation

Item-total correlation is a measurement of the correlation between the scores of an individual item and the sum of the scores of the remaining items [13]. The item-total correlation on a question is low when high-scoring students get the question wrong more frequently than low-scoring students. *An item-total correlation below 0.3 is considered low*. Generally, this occurs when either the question or the related content are ambiguous to the students. It could also be caused by ambiguity in the answers as discussed later (see Section 4.4). To address this issue, we revised the questions (see example below) and related content (see Section 4.7) to make sure that the wording was clear and unambiguous. For example, the wording in this Numeric Data question was clarified:

Original: Low Item-Total Correlation
The four integer data types are essentially the same.
A. TRUE (correct)
B. FALSE

Revised: Clarified Question
The four integer data types are essentially the same, ***and there are not reasons to choose one type over another type.***
C. TRUE (correct)
D. FALSE

4.3 Distracters Not Being Chosen

Distracters not being chosen by any student are indicators that they are not pertinent to the question so students do not consider them as possible answers [13]. If a distracter was not chosen by any student, it needs to be changed or replaced. If every student got the question correct and none of the distracters were chosen, the question should be removed or made more challenging because it is not providing any valuable information regarding students' abilities. For example, the distracter in this Debugging question was never chosen so we replaced it with a distracter t more relevant to the specific question being asked.

Original: Distracter Not Chosen
What is the purpose of a debugger's breakpoint?
A. It generates a list of variables containing their value
B. It stops execution of a program at the breakpoint (not chosen)
C. It pauses the program at the breakpoint (correct)
D. It skips execution of the statement at the breakpoint

Revised: New Distracter Added
What is the purpose of a debugger's breakpoint?
A. It generates a list of variables containing their values
B. It causes the program to exit at the breakpoint (new)
C. It pauses the program at the breakpoint (correct)
D. It skips execution of the statement at the breakpoint

4.4 Distracters Being Chosen Too Often

Distracters being chosen too often are indicators that the question is either unclear or the answer choices are ambiguous. Ideally, all distracters should be chosen in equal proportions. A distracter that is chosen too often should be compared with the correct answer. If the two answers are similar, one or both should be elaborated upon to distinguish them. For example, in this Variables & Constants question, point-biserial correlation [13] (.059) indicated that one distracter was often chosen by high-scoring students so we revised the answers to make the correct answer unambiguous.

Original: Distracters Chosen Too Often
Which of the following constant declarations is <u>incorrect</u>?
A. final double KS_TAXRATE
B. int MAX_MILES = 100000
C. final double PI = 3.14159 (chosen too often)
D. A and B (correct)
E. A, B, and C

Revised: Question Answers Reworded
Which of the following constant declarations is <u>incorrect</u>?
1. final double KS_TAXRATE
2. final int MAX_MILES = 100000
3. final double PI = 3.14159
4. int METERS_IN_MILE = 1600
   A. 1 & 2
   B. 1 & 4 (correct)
   C. 1 & 3
   D. 2, 3, and 4
   E. All are correct

4.5 Cronbach's Alpha

Cronbach's Alpha [6] is a statistical measure of internal consistency or reliability of the set of items contained in the LO. *Any value above 0.7 is considered an acceptable consistency.* Unfortunately, several LOs had Cronbach's Alpha values below 0.7. This was most likely due to the limited number of questions in each LO. In an attempt to increase the Cronbach's Alpha value, it was decided that each LO should have at least ten questions to increase the reliability of the assessment questions as a single latent construct [6]. For every LO that had fewer than ten questions, new questions were added to the LOs assessment. The number of questions added to each LO is given in Table 2.

4.6 Bloom's Taxonomy Coverage

LO assessment questions were categorized into the first four levels of BT: Knowledge, Comprehension, Application, and Analysis. This categorization was performed independently by two researchers and was then compared to ensure accuracy. Table 3 gives the BT Levels for questions in the 2009 deployment. We analyzed the BT coverage for each LO to determine which categories the new questions should fit in to. As stated earlier, it was decided that each LO should have at least ten questions, so we strived to write the new questions such that the ten questions would be equally spread across the first four levels of BT. However, because the LOs are intended for introductory CS courses we focused on adding knowledge and comprehension questions because they better assess the student learning necessary for an introductory course. Additionally, given the number of LOs that needed revision, we prioritized the LOs with the most "need" based on the insufficient number of questions, Cronbach's Alpha, and item-total correlation values.

Table 3. BT Coverage for Assessment Questions in 2009 Deployment Before and After Revision. KN: knowledge, CO: comprehension, AP: application, AN: analysis. A single entry indicates that no additional questions were added.

LO Name	KN	CO	AP	AN
Advanced Logic	0/2	0/1	3	3/4
Advanced Recursion	0/2	1/3	3	1/2
Algorithms	2/4	3	0/2	1
Arrays	2/3	2	1/3	1/2
Conditionals	0/3	2/3	3	1
Debugging	2/6	0/1	0	3
Functions	10	0	1	3
Logic	2/4	0	2	1
Looping	0/2	3/4	0/1	2/3
Non OO Problem Analysis	1/2	2/4	3	0
Numeric Data	3	1/2	3	0/2
OO Problem Analysis	1/4	0/1	2/3	1
Recursion	1/2	2/4	2	1/2
Searching	2	5	0	4
Sorting	4	2	2	2
Variables & Constants	2/3	2/3	2	0/2

4.7 Learning Objectives Coverage

Several learning objectives are listed at the beginning of each tutorial to describe what students should learn by the end of the LO. Similar to the BT categorization, the assessment questions were also categorized by which learning objective they fulfilled. For the LOs that did not have at least ten questions, we strived to write the new questions so that the questions are equally spread across the learning objectives. In some cases, there were already questions that did not fit the learning objectives so new content was added to align the content and the assessment. For example, in the Non OO Problem Analysis LO, additional content on the divide-and-conquer approach was added.

Original: No Tutorial Coverage for Question

The principle of this divide-and-conquer approach to problem analysis is to divide a large problem into smaller problems and handle each smaller problem with its own

A. TRUE (correct)
B. FALSE

Revised: Added Tutorial Content

As you can see, there are 6 specific tasks that must occur when the submit button is clicked, and each one can be handled by its own separate module (divide-and-conquer), instead of trying to lump everything together within the body of the "button clicked" function. This is useful for several reasons:

1. You can focus on implementing each small sub-problem individually instead of trying to work on the entire problem at once.
2. When other programmers look at your code, they will clearly be able to tell what part of the code handles which sub- problem.
3. If you come across errors when testing your program, it will be easier for you to isolate where the problem occurs if each step of the big problem clearly has its own module.

4.8 How to Examine and Write Questions

Here we provide general guidelines on revising LO questions. It is important to make sure that the assessment questions and the tutorial content use the same vocabulary and are consistent throughout. If this is not the case, the question or the tutorial content should be adjusted to preserve consistency. Next, the question itself should be examined. It is important to have multiple people look at each question and provide feedback about the clarity and fairness of a question. Finally, examine each of the answer choices for a question. If any of the incorrect answer choices are similar to the correct answer or if any of them do not have a specific reason why they are incorrect, either the correct answer choice should be clarified to make it the only correct answer or the incorrect answer should be changed to emphasize how it is incorrect.

There are several ways to handle the questions after they are examined. These include removing the question, changing the answer choices, clarifying the question, or determining that the question is acceptable as written. If a question is examined and more than one reviewer deems a question incorrect or inappropriate it might be reasonable to remove the question entirely from the LO assessment. Before removing a question, however, it is important to consider the purpose of a question and what objective it is based on. If a question is removed, another question based on the same objective should be added.

When determining what new questions to write, the writer should analyze both BT and learning objective coverage. Ideally the assessment questions for each LO will cover the first four level of BT and every learning objective, which should be representative of the material covered in the LO. A good principle to follow is evenly spread the assessment questions across the first four levels of BT and cover all the learning objectives for each LO. In the revision process, we strived to write questions that specifically fulfilled a BT level and correspond to a specific learning objective that needed more questions. However, revising BT coverage is less important on LOs with (1) sufficient number of questions, (2) high value of Cronbach's Alpha, and (3) high item-total correlation values.

5. CONCLUSIONS & FUTURE WORK

In this paper we present a systematic revision process called LOSRP for LOs that uses assessment validation methods from educational research including Bloom's Taxonomy [4], item-total correlation [13] and Cronbach's Alpha [6]. These validation methods are used to answer seven questions in order to diagnose what needs to be revised in the LO. LOSRP also provides guidelines on revising LO content and assessment for each diagnosis question. For all diagnosis questions that elicit a "yes" response, the appropriate guidelines (➡) should be used to revise the LO content or assessment.

Does the assessment use multiple mark questions instead of true/false or multiple choice?

➡Change multi mark questions to multiple-choice questions.

Is the item-total correlation on the assessment questions below a threshold of 0.3?

➡ Add additional content to clarify what the questions are asking.

Are specific distracters in the questions not being chosen? Are specific distracters chosen more often than the correct answer?

➡ Replace the distracters that are not chosen. Change the wording for distracters chosen to differentiate them from the correct answer.

Is the Cronbach's Alpha statistic for the assessment below a threshold of 0.7?

➡ Add additional questions to better evaluate student understanding of the content.

Do the questions fail to cover the first four Bloom's Taxonomy levels?

➡ Add additional questions with different BT levels to promote student learning.

Are objectives of the content and the assessment questions inconsistent?

➡ Add additional questions based on the learning objectives, add additional content covering the learning objectives, or change the learning objectives.

We used LOSRP on sixteen LOs deployed in 2009 to introductory CS courses. Overall, we found that many of our LOs needed to be revised because they did not have enough assessment questions. We rigorously applied all LOSRP guidelines to the revision of the LOs for the 2010 deployment (see Table 2 and 3). The LOs are currently being deployed to 403 students in five introductory CS courses. In the future, we will compare the results on the same LOs from the 2009 and 2010 deployments. We expect to see improvements on all assessment validation methods used not just in the assessment scores for the LOs.

LOSRP currently focuses on assessment revision rather than content revision. Even the best assessment questions ever written would not be effective when the content cannot convey the concepts and learning objectives to the students. Currently, LOSRP only gives qualitative guidelines on revising LO content. This is not complete, but it provides the first step towards a systematic revision process for LOs. In the future, we will add to LOSRP the capability to quantitatively diagnose between problems with the content and the assessment questions. This will involve combining assessment validation methods and natural language processing on student interactions and assessment responses to identify "disconnects" between the LO content/questions.

We would also like to apply LOSRP to ITS. An ITS would gain even more benefit from LOSRP than LOs because it could dynamically revise the assessments rather than waiting an entire year between LO deployments. Additionally, an ITS could use the LOSRP to diagnosis in real-time whether students are struggling with the learning objectives or the assessment questions.

6. ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grants No. 0632642 and 0829647.

7. REFERENCES

[1] Agarwal, R., Edwards, S.H. et al. 2006. Designing an adaptive learning module to teach software testing. *SIGCSE* (2006), 259-263.

[2] Association, A.P. 1986. *Standards for Educational and Psychological Testing*. Amer Psychological Assn.

[3] Balci, B. and Inceoglu, M. 2006. Reusable Learning Objects (RLOs) for Computer Science Students. *Computer Science and Its Applications*. 373-382.

[4] Bloom, B.S. 1956. *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain*. Addison Wesley Publishing Company.

[5] Burge, L.L. and Leach, R.J. 2010. An Advanced Assessment Tool and Process. *SIGCSE* (2010), 451-454.

[6] Cronbach, L. 1951. Coefficient Apha and the Internal Structure of Tests. *Psychometrika*. 16, 3 (1951), 297-334.

[7] Mundy, D.P. 2006. Using Learning Objects to Support Introductory Computer Architecture Education. *IASTED* (2006), 305-310.

[8] Nugent, G., Kupzyk, K. et al. 2009. Empirical Usage Metadata in Learning Objects. *FIE* (2009), 1285-1290.

[9] Nugent, G., Soh, L. et al. 2005. Design, development, and validation of a learning object for CS1. *SIGCSE* (2005), 370-370.

[10] Riley, S., Miller, L.D. et al. 2009. Intelligent Learning Object Guide (iLOG): A Framework for Automatic Empirically-Based Metadata Generation. *AIED* (2009), 515-522.

[11] Starr, C.W., Manaris, B. et al. 2008. Bloom's Taxonomy Revisited: Specifying Assessable Learning Objectives in Computer Science. *SIGCSE* (2008), 261-265.

[12] Tew, A.E. and Guzdial, M. 2010. Developing a Validated Assessment of Fundamental CS1Concepts. *SIGCSE* (2010), 97-101.

[13] Thorndike, R. 1982. *Applied Psychometrics*. Boston: Houghton Mifflin.