Incorporating the Perspective of Multiple Stakeholders Into the Evaluation of Educational Technology

Kenneth A. Rath
and
Alan R. Peterfreund
Peterfreund Associates
30 Boltwood Walk
Amherst, MA 01002
Voice: 413 256 6169   Fax: 413 256 6648
Email: krath@peterfreund.com   alan@peterfreund.com
Website: http://www.peterfreund.com

David M. Hart
Center for Computer-Based Instructional Technology
Department of Computer Science
University of Massachusetts, Amherst
140 Governor's Drive
Amherst, MA 01003-4610
Email: dhart@cs.umass.edu
Website: http://ccbit.cs.umass.edu/ccbit/

Abstract

This paper discusses the evaluation of OWL, a web-based homework system, in the context of an evaluation framework incorporating the perspective of multiple stakeholders. The framework considers five outcome categories (learning, performance, quality, resource allocation, and inspiration) from the perspective of student users, faculty users, other faculty members affected by the program, the department as a whole, the institutional leadership, the project leadership, and the funding agency. The authors' experiences with the OWL system as implemented in the Art History and Mathematics departments at a large state university provide examples illustrating the framework in action.

Key words: Evaluation, Learning technologies, Project deployment, Web-based homework, College students

The Importance of a Multiple Stakeholder Perspective

Every year, many, many innovations are funded and developed across the country and beyond with the goal of improving student learning. The vast majority of these, at least from our experience—finding the actual numbers is probably impossible—are well thought-out and really do help students learn better, at least to some extent. But the vast majority also never succeed on a wide scale, either sustaining only a local focus and having no effect beyond the original population or petering out as interest and funding fade and the project initiators move on to other things. Why? Why do perfectly good innovations that have such potential to do so much good so rarely make it past the initial stages?

The answer, from what we have seen, is that most project leaders fail to satisfy, or even consider, the concerns of the whole range of people who have a stake in the project. And without the buy-in from the whole range of stakeholders, a project is unlikely to reach its full potential.

Most project leaders who wish to develop some new and wonderful educational innovation begin with the stated goal of improving student learning, typically within a particular domain. They may also consider the impact of the new innovation on the pedagogy of the course or situation into which it will be introduced, but their focus rarely moves beyond that.

Unfortunately for them, there are other individuals with a stake in the success of the project beyond the students and instructors who will be using it. These stakeholders include other instructors who are not using the innovation but may become future customers of it after it has been refined, the academic department or institution that has to fund and support this innovation after the project money has run out, the instructors who are receiving the affected students into their classes and have a vested interest in what those students are learning, the administration who would be responsible for promoting the innovation to the next generation of instructors, the project team themselves who must maintain their enthusiasm for the project if it is to be a continued success, and the funding agency whose goals must be met if further funding, especially for dissemination of the product, is desired. Failure to engage any one of these groups can sound the death knell for a given project.

This paper examines this problem from our experience with a specific innovation, the OWL (Online Web-based Learning) system, as it was developed in the Art History and Mathematics Departments at a large public university. Because of this, the stakeholders we mention relate to the college setting and the field of educational technology and may not translate well to all projects, but we believe that the underlying points are important for all projects that hope to be successful.

The two examples we are using are both instances of projects that could not be said to have been wholly successful. In both cases, key stakeholders were not fully engaged in the project *through no fault of the quality of the product* and, as a result, both products have only had a limited impact when they had the potential to be much greater. Our discussion of the successes and setbacks that both projects met will be made with candor in the hope that our experience will better inform the activities of others. We hope to illustrate the importance of the multiple stakeholder perspective for any project and to demonstrate that only by examining projects from this complete perspective can one fully understand the project's successes and failures.

It should be noted that the lessons learned and codified in this paper came through our years of experience with both successful and unsuccessful projects—we did not know going into the projects that there would be difficulties and limited successes. It is our hope that our travails can better help others develop projects with an eye toward making them successful.

As evaluators, we are also interested in what activities must be performed in order to demonstrate that the concerns of the various stakeholders are being bet. The latter portion of this paper will address these concerns.

Background of the OWL Projects

At the onset of the Mathematics and Art History projects, OWL was a preexisting system, originally developed by the Center for Computer-Based Instructional Technology (CCBIT) at the University of Massachusetts Amherst in 1997 to support the Chemistry Department's need to replace their existing computer-based homework system, which was out of date for a number of reasons. As part of their efforts to build OWL and expand its capabilities, the project leaders applied for and received a grant from the U.S. Department of Education-FIPSE to expand OWL both in breadth, incorporating several disciplines, and reach, introducing it to several campuses beyond the home institute. Through the auspices of this and other, subsequent, grants, OWL was introduced into Mathematics in 2000 and Art History in 2001.

OWL is a web-based homework platform that supports a number of different homework question types, including multiple choice, multiple-multiple choice, short answer, and entering equations. It also allows for more graphically complex homework modules to be added on as external components (such as the molecule building modules created for Chemistry), as well as information pages, surveys, and one-shot exams. Authors can add their own questions, determine the number of trials students can have at a given question, provide feedback at a variety of levels, and parameterize components of questions so that students see different numbers each time they retry a problem. New developments include a tutoring system that allows the depth and quantity of questions seen by a student to match that student's needs and a proposed plan to incorporate OWL with in-class electronic response systems. More information about the specific capabilities of OWL can be found at http://ccbit.cs.umass.edu/owl/.

Currently, OWL is used by well over 10,000 students annually in more than a dozen departments across multiple campuses and in supported subjects ranging from Foreign Languages and Art History to Chemistry and Computer Science. Of these, its use in some disciplines was funded by extensive national grants, while others were small-scale projects developed by individual instructors. The database of questions and special modules developed for chemistry has already been picked up by a national distributor, and plans are underway to pick up the databases for Organic Chemistry, Physics, and Computer Science as well.

OWL continues to evolve in many ways, and new proposals are written every year to provide start-up funds to increase its capabilities—and, presumably, its pedagogical impact—and to expand the range of subjects supported. We expect that new developments will occur by the time this article is published that will make the above list obsolete.

It would appear, then, that OWL is a robust product with a good deal of potential to be disseminated and impact a large number of students. And of the departments whose OWL offering is mature and was supported by a major federal grant, Chemistry, Organic Chemistry, Physics, Computer Science, Mathematics, and Art History, the first four have been very successful at disseminating their products. But the latter two have not—in fact, they have had no lasting adaptation beyond their home campus, or had the OWL system spread to other courses within their department beyond those originally targeted, This paper will explore why.

OWL in Art History

With the funding of a separate U. S. Department of Education-FIPSE grant, the project team in Art History developed a series of computer-based modules for two introductory-level courses that were essentially self-contained lessons on important topics in an introductory Art History course. These modules were intended to be distributed to students on a CD and did not have any homework-style activities associated with them. Initial evaluation demonstrated that their learning impact was not much different from that of a library-based slide viewing activity, and it was determined that the main reasons were that students were not held accountable for doing the activity and that there was little that was interactive about the experience. Because of this, the project leaders decided to incorporate the materials into OWL.

The Art History OWL materials began to be used in 2001, and from 2001 to 2003 seven units were developed that covered a wide variety of topics, including the fundamental principles of the architecture of a Greek temple, linear perspectives in Italian Renaissance paintings, and African ritual artifacts, among other topics. These keep the instructional format of the original CD-based modules, but introduce questions designed to test student understanding of the material and have expanded the interactivities. Examples of such activities include the ability to build a hypothetical Greek temple (where the gods become upset if it is not done in a pleasing fashion) and adjusting the perspective in a painting.

Our evaluation of the use of OWL in Art History began in 2001 and continued through 2003, when the funding for the project elapsed.

OWL in Mathematics

OWL was introduced into Mathematics by one professor under the auspices of the original DOE-FIPSE OWL grant in 2000. In this first semester, OWL was only used in that instructor's sections (two of the four in the class) of Introductory Calculus I, a course aimed at students majoring in business, management, and the life and social sciences. In previous offerings, this class either no homework or intensely laborious, hand-graded assignments (there are over 100 students in every section). These first OWL materials were relatively unsophisticated, containing only simple multiple-choice questions.

In 2001, a new course chair took over, and with him came a revision of the curriculum, turning to a more conceptual rather than computational focus, complete with a change in textbook and a new NSF grant to support further OWL development. From that point a large number of more sophisticated math problems were developed, many making use of the innovation of adding web*Mathematica* to the system (information about this product can be found at http://www.wolfram.com/products/webmathematica/index.html). This addition allowed for dynamic graphs and parameterized equations, as well as giving the students the opportunity to enter equations of their own as answers.

In addition to the basic set of OWL problems, the project team developed eight special assignments, called Discovery Modules. These were designed to have students discover difficult mathematical principles through a series of scaffolded steps that focused on both conceptual and computational aspects of the problems. The eight subjects chosen for this treatment were ones that had historically given students the most trouble, and, as an attempt to add relevance to the problems, multiple versions were created for many that focused on particular problems from life sciences, finance, and pure mathematics.

Beyond that, the grant supported the development of Tutors, modules that assigned a varying number and sophistication of problems based upon students' answers to earlier questions. Students who demonstrated difficulties with the problem were given more practice and feedback until they started to catch on, while students who performed well at the beginning could complete the assignment with a minimum of work. For reasons that will be explained later, the Tutors were never fully incorporated into the class and remain unused in Mathematics on any wider scale (although the software basis has been adopted by the Physics department and used to considerable effect).

We began evaluating the use of OWL in Mathematics in 2000 with its first use, and continued through the end of the grant in 2003.

The Stakeholders and What They Want

Both of the projects described above—and all projects, for that matter—had a number of people who were interested in, or potentially affected by, their outcomes. These stakeholders can be grouped into rough categories whose members share similar concerns. In order to truly evaluate the success of a program, it is necessary to understand the perspectives of each of these groups of stakeholders (Patton, 1997), and the success of a project can be measured by the extent to which their concerns are met.

The major groups of stakeholders that we identified from our OWL evaluation efforts were:

- The students
- The course instructors (which includes the graduate teaching assistants)
- Other instructors in the department not teaching the course
- The host department itself
- Faculty members in other departments who are teaching courses that build off of the affected course (this was especially important in Mathematics)
- The university administration
- Faculty members at other institutions that might be interested in adopting the innovation
- The developers of the project (who were often distinct from, or a subset of, the course instructors)
- The commissioners or funding agencies

Of course, this list may be differ somewhat for other projects, but it is likely that most categories of stakeholders would apply to any innovation at the postsecondary level and, to a great extent, in other educational settings as well.

But what do these stakeholders want, and what would a project have to achieve in order to be successful (by which we mean accepted) in their eyes?

Students

In general, the most important goal college students have is to successfully complete the course, generally with a minimum of pain. Unfortunately, for many students, learning the material is less important than getting the grade they want or need, especially in introductory-level, required classes that do not fit into their major. If a course is more interesting to the students, the learning goal becomes more important, but in the authors' experience as evaluators the performance goal has always overwhelmed anything else. The amount of pain they are willing to undergo in order to obtain their desired grade can vary greatly, depending on the importance of the class to their plans and their tolerance for abuse, but is another significant factor.

Therefore, from the students' perspective, a successful innovation is one which either helps them improve their grades or decreases the amount of pain that they have to suffer in order to make it through the class. Fortunately, increasing what students learn generally has the effect of increasing their performance on assessment measures (unless grading is done on a curve) and making material more interesting generally both increases learning and decreases the pain associated with doing the work.

Of course the students are not a homogeneous body and do not all approach a course in the same way. Each individual brings a different mix of previous experiences, interests, familiarity with the subject matter, learning style preferences, learning skills, and motivation, among other important factors. An innovation that might be at just the right level and fit the preferred learning style of one student may be seen as an onerous burden and a waste of time by another. Still, some generalizations can be made about the student body as a whole, and understanding both the mix of students and the commonalities between them is essential for understanding how an innovation plays out in the course.

OWL was used in two different classes in Art History, but both were introductory, general education classes whose takers had a similar makeup. Because they were general education courses that fulfilled a humanities requirement, their takers were from a wide variety of majors, with a very small proportion hailing from Art or Art History and an equally small proportion expecting to take another course in Art History beyond that one. Because the specific course was not required of the students, most had chosen it because of some general interest in art or art history. As such, most students were more interested in learning the interesting stories and such behind the art and less interested in developing a set of critical art historical skills that they could use in more advanced classes. Still, because they had elected to be there, students tended to be generally tolerant of and interested in exposure to new material, although they were still concerned that it should help them achieve an acceptable grade.

The OWL materials did a good job of engaging student interest, and the students regarded them positively in that regard, but in many cases they were presented more as supplemental than integrated pieces of the course, which meant that their effect on the students' overall grades was minimal. Because of this, student reactions to the system were mixed—it was interesting but additional work, and did not meet the students' overall goals very well.

In Mathematics, the situation was very different. Introductory Calculus I is a required course for many majors, but not for majors that are expected to be math intensive such as Physics or Engineering—they are given a more intensive math experience in a different series of classes. Because of this, students who are interested in math for its own sake tend to take the more advanced version of the course, leaving the students in the affected course less than enthusiastic about the material. Additionally, whether true or not, there is not a sense among most students that the calculus they are learning is in any way essential (or even important) to being successful in their majors. Thus most students in the class see it as a hurdle they have to make it over rather than an opportunity to learn interesting, or even important, material.

Because of this situation, an innovation that attempts to make the material more interesting is less valuable in the Mathematics class than one which reduces the amount of pain they have to suffer in order to get the grade they want. In fact, when multiple versions of the Discovery Modules were made available to appeal to students with different interests, most students did not even realize that fact and just did all of the assignments. On the other hand, they liked OWL because they could do the work, get full credit, and have 10% of their grade secured.

Students also perceived that OWL helped them learn the material better than traditional homework with a comparable amount of work, with the corresponding positive effect on class grades. Because of this, reactions toward OWL tended to be very positive as it was meeting the students' goal without much additional pain.

Course Instructors

Students are often not primarily concerned with their learning of the material nor, perhaps surprisingly, are course instructors. From our experience, most instructors do want the students to learn the material, although there are variations on what that means. Some instructors believe that any student *can* learn the material and would like to do what they can to reach those students who are not currently succeeding, but other instructors believe that they are teaching difficult material that most students are incapable of understanding and are interested in winnowing out those students who are not fit to stay in the major. The latter group can be very resistant to many new innovations.

More importantly, course instructors have a limited amount of time and energy that they are willing to pour into a single class. If they had unlimited time they could integrate all manner of interesting innovations into their course, but with the limited resources at their disposal any innovation that takes a substantial time and energy commitment on their part is likely to be met with resistance, regardless of its effect on students. True, greater potential benefits may merit greater sacrifices, but there are limits.

A successful innovation, then, is one which can inspire instructors to adopt it and integrate their classes on the basis of its potential benefits but which does not overburden them so much as to force them to but in a less than adequate amount of effort. If that does not occur, innovations can be adopted without full buy-in, not allowing them to achieve their potential effects or even making them completely ancillary and unconnected to the rest of the course.

With OWL, one of the major attractions for instructors is that it allows them to introduce homework into their classes, a perceived benefit for the students, without having to grade it or provide feedback—OWL does all of that for them. Optimally, however, there is much more that can be done with OWL. It can tell the instructor what topics students are having difficulty with, thus allowing lesson plans to be modified based on the students' needs. And, when the material in OWL is closely paired with what is being discussed in class and tested on exams, it can be a teaching as well as an assessment tool. The degree to which these features are used depends very much on how much effort instructors are willing to put into using the system.

In Art History, most of the faculty users were not part of the project team but had nonetheless agreed to support the project by using the modules in their classes. In many cases, however, they did not use the full collection of modules but, rather, only those that were directly related to the material they planned to teach. As their lectures were already planned out through years of refinement, they did not make the OWL activities part of the class in any meaningful way, assigning them merely as extra assignments with a question on the final exam related to the content of one.

Buy-in from these instructors was not very high. In part, this was because integrating the modules more fully into their classes would require a wide-scale course restructuring, which they were unwilling to do due to the time requirement on their parts and the lack of a clear benefit for doing so.

Another major piece of this reluctance to buy in at a greater level was that they did not feel any sense of *ownership* of the modules. These were interesting activities, but fully developed and requiring no action on their part to use and very little additional explanation on their part for the students to understand the material. Because of this, they could easily let the OWL work stand on its own and have nothing to do with it.

In Mathematics, OWL was introduced into a single course (at least at first—one professor adopted it for his sections of the second-semester course as well). This course has several sections—four or five, depending on the semester, until the budget cuts of 2003 led to larger classes and fewer sections—and each instructor teaches one or two of them. The instructors have very limited freedom, however: the tests are identical for all sections, as was the homework, once the new course chair took over in 2001. Because the course chair was the leader of the project team, OWL was quickly introduced into all of the sections and became the main method for assessing student progress beyond the three exams (although some sections continued to do quizzes in addition to OWL).

Because OWL was so centralized, the other instructors left its maintenance to the course chair and his graduate teaching assistants, spending their time fielding student questions and, as time progressed and they became more aware of its importance as the students' main source of problem practice, discussing the questions from OWL in class. By any reasonable measure, OWL had become well integrated into the program, both because the benefits of homework were well accepted and because the course chair had written both the OWL problems and the exams, thus making it clear that the OWL material was directly related to course assessments and thus student performance.

Unfortunately, the course chair and main driver behind the project retired in 2003 and a new course chair came on, hired from a different institution. For him, learning the OWL system and adopting his tests and teaching methods to it were a massive investment in time and energy, and he was unwilling to make the sacrifice. Instead, he continued to use OWL as it had been under the old system but made his own tests, which were not as closely related to the OWL material, and did not allow OWL to impact his teaching. Although the veteran instructors continue to use OWL, it is no longer as integrated into the course as it used to be.

Our experience has shown that the more capability a system has of being modified to fit the particular situations of a class, the more the instructor and students will benefit from it when those modifications are made, but the fewer instructors will be willing to use it to its full capacity as a product of having to spend the time to make those modifications in order to do so. This is the Catch 22 of educational innovations and one of the major reasons why many excellent innovations do not succeed in being adopted on a wide scale.

Other Departmental Instructors

Other instructors in the department, although not directly affected by having the innovation introduced into their classes, are nonetheless affected by the innovation.

First, they often receive students into their courses who have been serviced by the course into which the innovation has been introduced. Any changes to those students' level of preparation and thus readiness to learn the content of the subsequent courses are of interest to these instructors, most of whom would prefer not to have to spend time reiterating material covered in pervious courses. If an innovation increases students' understanding of the building blocks for their class, they benefit.

Second, if an innovation is capable of being modified to suit a number of different course situations, as is the case with OWL, these instructors can become its next adopters. As such, they are interested in a system that can be adapted to their needs with relatively low effort on their parts.

And third, although most instructors would not realize it as an issue, an innovation in one class can affect the students' expectations of what a course in the department should offer. If a subsequent course fails to provide the student benefits present in the course with the innovation, student opinions of the subsequent course may be lowered, with corresponding problems for the instructor.

Because OWL use in Art History ended up being so ancillary to the class, and because most of the students taking the courses using OWL never took another Art History course, there was no expectation that the students would be coming out of the courses with different skills than they had previously, either on the part of the project team leaders or the receiving faculty. Because of this, the Art History instructors were able to ignore OWL's presence in the class an teach their classes as they always had.

Additionally, the culture in this particular Art History department was such that there was no compulsion, either internal or external, for faculty members to work together on pedagogy. Because of this, no faculty members other than those actually using OWL expressed any interest in adopting it. This lack of enthusiasm, or even interest, on the part of fellow faculty members in the department had important consequences on the project leaders' own motivation to continue with the project.

In Mathematics, the students from Introductory Calculus I feed directly into the second semester of that sequence, but after that most do not continue taking math courses. The students going into that second semester course may know more math—discerning whether or not that is the case has been a challenge since there is no general pre-test—but they definitely come in with the expectation that there will be support comparable to that from OWL the previous semester. As a function of this pressure, one instructor, who teaches in both courses in the sequence, has introduced OWL into the second semester. Because of the time commitment and lack of grant support, however, none of the other instructors have adopted it.

The problems that led Introductory Calculus I to adopt OWL are still present in many of the other introductory courses: large classes, and not enough instructor time to make homework grading tenable. But OWL is a complex system and takes both time and money to implement. Unfortunately for OWL from a business perspective, there are a number of commercially available products, often offered free to departments when they choose specific textbooks, that, although less robust than OWL, do not require the time it would take to develop a new set of questions and instructional materials for a new course. The demonstrations of OWL to these faculty members have always been met with interest and approval, but no one has been willing to overcome the time hurdle necessary to adopt it. As of now, there are no plans to expand OWL's use in the department beyond current limits.

The Host Department

Separate from the teaching concerns of its faculty, the department as a whole is primarily concerned with money, and once the initial grant period has come to an end they are the ones who will have to foot the bill for any costs associated with maintaining the innovation. Because of this, their interest lies in having an innovation that is monetarily sustainable, either because it does not cost significantly extra to keep going or because it can be funded by charging the students in some way (which could antagonize the students).

Another aspect of money, however, are the number of students who make it through the class. Departments, at least at the campuses we have worked with, receive money based on the numbers of students who complete their classes. If an innovation increases the number of students making it through a class without dropping out, the department could get more money, which may make it worth while paying a bit to keep the innovation going.

A third issue is reputation. With more students going through the courses, and especially with more students choosing that subject as a major, the department gains in reputation, which is directly related to more funds coming in. Also, if the department's reputation grows outside of the school they will attract better and more students, leading to more money. So if an innovation increases the department's reputation in the eyes of the students or of other institutions, it may also be worth the investment.

If, on the other hand, the departmental leadership has not heard anything about an innovation until the project leaders come by asking for money to continue its use, that innovation's future might well be in jeopardy.

In many departments, the introduction of OWL has saved the department money by replacing the human resources that had been used in grading and managing homework in the past, freeing those personnel up for other duties or allowing courses to continue as they had in the face of faculty retirements. This was not the case for either Art History or Mathematics. In both cases, the OWL homework was supplemental to the existing course model rather than in place of existing homework.

In Art History, there has been practically no recognition of the OWL project at any time, and that did not change at the end of the grant. Still, OWL continues to be used in those classes because maintenance costs are minimal and have been borne by the Center for Computer-Based Instructional Technology (CCBIT), the team who created and maintain the OWL software. But because the continued support for OWL has been minimal, there are no funds to allow further modification of materials or development of new modules in the absence of future grants.

The situation in Mathematics is practically identical, but the consequences are even more problematic. At the end of the project's funding, there were a number of materials in need of further reform and development, especially the yet-to-be-tested Tutors. Because of the minimal funding level and the lack of a strong supporter from within the department now that the project leader has retired, it is unlikely that the further developments will ever be made. In both cases, the departmental stakeholders were not sufficiently engaged to generate enthusiasm for the projects continuation.

Faculty Members in Other Departments

Some courses, including the Art History courses into which OWL was introduced, do not impart skills that are explicitly built upon by courses offered in other departments. Others, including Introductory Calculus I, do. In that case, students in upper level courses in their home departments are expected to understand the basic concepts of calculus and be able to apply them to their own field.

In situations where other courses build off the knowledge of the course affected by the innovation, the instructors of those courses are primarily interested in the learning of the students. They want students coming into their classes who are prepared to use what they learned previously with a minimal effort on their part—essentially, they are the same in this regard as faculty members in the same department teaching upper-level courses. A successful innovation, in their eyes, is one which enhances the readiness of students to take their classes.

Faculty in several departments reported to us that they offered upper level classes that made use of some calculus—the reason they required their students to take the course—but that students, historically, were not prepared to do calculus in those classes without a significant review. There was a sense that the amount of calculus being required of the students was minimal due to their poor preparation, but that if matters changed the instructors would happily enhance the calculus components of their classes.

In response to these needs, the project team began modifying some of their materials to reflect specific problems in the various fields, especially within the context of the Discovery Modules and the unfinished Tutors, in the hopes of increasing the topics' relevance to the students and increasing their interest in learning the material. Unfortunately, due to the impending retirement of the project leader and fiscal problems within CCBIT, the momentum on those changes faltered and they ended up having minimal impact. More on this situation will be discussed when examining the Project Team.

The willingness to entertain and act upon feedback from faculty receiving the calculus students did have the effect of enhancing the awareness of OWL and increasing interest in it, with the result that several projects have been undertaken to develop OWL materials for other departments, including Psychology.

The University Administration

As with the departmental leadership, the university administrators are primarily concerned with increasing the institution's reputation and revenue intake. What this means for projects is that they are interested in finding ones which they can leverage for either increasing the attractiveness of their institution to students and donors or building off them into future grants. On the other side, large-scale projects that span several departments may need institutional support to continue after grant-based funding ends, so it is very important that the administration be impressed with the project in order that it continue.

OWL in general has impressed the administration at the university greatly. It has advocates (and users) in the higher echelons of the leadership structure, and several university-wide proposals and grants have included OWL components. Its creation has also significantly enhanced the reputation of the Computer Science department as an entity for educational research. At one point, the funding for CCBIT was removed because of massive budget cuts and the incorrect perception that OWL was paying for itself, but when the facts were revealed a number of individuals strongly advocated for its continued funding, efforts which, do date, have been successful.

Because there are so many successful OWL projects on the campus, it is difficult to say that any particular project has had an independent impact on institutional thinking. It is true that university-wide proposals have been written with the intent of leveraging the Mathematics materials for wide-scale tutoring help, but that has generally been through the advocacy of the people at CCBIT rather than by the specific Math project team. The same is also true to a lesser extent for Art History.

Overall, the administration on the campus , or at least significant parts of it, have been very supportive of the OWL projects, which have done well at meeting their needs.

Faculty at Other Institutions

Although they are not actually a stakeholder until efforts begin to disseminate the innovation, the perspective of potential faculty users beyond the home campus is very important. What, then, do these individuals want from a product? Like potential users on the home campus, they are interested in a product that can increase student learning while requiring only a modest investment of time, energy, and resources on their part to significantly do so. The resource question is especially important because, unlike instructors at the host institution, the adopting institutions will not have the benefit of the project members being in close physical proximity to work through any problems that might arise.

As stated previously, some OWL projects have been very successful at wooing faculty instructors at other institutions into becoming product users, but the same cannot be said of the Mathematics and Art History products. The reasons are different for each.

At first glance, it would seem that incorporating the Art History materials into another class would not take a substantial amount of time or effort on the instructors part. These materials are stand-alone lessons with all the components bound up into one package—all the instructor need to is assign them and pull out the results to see how well the students did on them.

But, as always, matters are more complicated than they appear. Because they are stand-alone lessons, they have a particular pedagogy associated with them, which is the pedagogy of the project team and may not necessarily reflect the goals of the potential adoptee. Also, they cover specific materials and in a specific way. Even if a course is already covering Greek architecture, the Greek temple module may not cover the material to the desired amount of depth or within the desired overall context. In order to make the modules fit into their course, many instructors would have to put a substantial amount of effort into framing that effort.

And, finally, the modules are sophisticated multimedia presentations. What that means is that there is a substantial amount of computer infrastructure necessary in order for them to be usable: students need to have access to computers that have high speed internet connections and can display sophisticated graphics and audio files, and these computers must be convenient enough to use that the students are not subjected to any particular hardship when trying to do their homework. Even with today's sophisticated computer facilities, these conditions are not met at all campuses.

Thus, despite continued advocacy on the part of the project team, there have been no major adopters of the Art History OWL materials on other campuses.

In mathematics, the pedagogy-fit issues are less of a barrier to adoption since the materials do not contain much in the way of direct instruction but there is a substantial amount of effort necessary in order to familiarize oneself with the materials and understand how they relate to the class one is teaching in order that the homework assignments can match the sequence of coursework. This is not an insurmountable barrier—the model has been very successful in Chemistry, where the OWL materials are made available alongside a textbook and closely keyed to the content of that book, but no successful overtures were made by the Mathematics project team in order to create a similar matching.

The technology factor is still an issue, but the OWL materials in Mathematics are not as sophisticated as those in Art History—they do not contain audio files and high-resolution graphics, for example—which makes that much less of a barrier. Certainly, many schools have adopted the Chemistry system.

From the Chemistry example, it seems that there is no reason why the Mathematics OWL could not have been widely disseminated using a similar model. But it was not, and that is mainly because the project team never took the necessary steps to try to make it so that it could be. In that project, the project team's momentum had already fizzled by the time the dissemination question was being bantered about.

The Project Team

Ultimately, what the project team wants is to create a successful product that they can be proud of. But creating that product takes time, and in the meantime the team needs to maintain their initial enthusiasm and remain inspired to continue.

What it takes to keep up the momentum varies tremendously based upon the personalities of the individuals on the project team. In some cases, minor setbacks such as a less-than-enthusiastic reception by students or a lack of predicted positive effects regarding an initial version of a product can serve as fuel for the project team to work harder to make the necessary improvements. But in other cases, the lack of success can be demoralizing and lead to a greater amount of inertia that must be overcome in order to move forward.

Once the innovation has been developed and tested, there are still potential setbacks that could impact the team's morale. The innovation must be advertised and potential users convinced. New proposals for additional grants may be submitted. Setbacks in any of these areas could equally spur a team forward or stop momentum, depending on the team. And then there are the team members' other responsibilities, which also need attending to.

From the project team's perspective, a successful project is one for which they are able to maintain their enthusiasm. This means that the desired goals of the project are either being met or seem to be attainable, that the work required by the project continues to be fun rather than burdensome, and that the amount of time demanded by the project does not significantly take away from other preferred or necessary activities.

In both of the projects discussed here, the project team consisted of two sets of people: the department experts who were creating specific instructional materials and questions and the software experts at CCBIT who were responsible for modifying and enhancing the capabilities of OWL to meet the needs of the department experts' visions. The perspectives of each side were somewhat different, and the dynamics of how they interacted differed by project.

From the project team's perspective, the Art History project could be said to have been successful throughout the development and implementation phases. Most of the materials had already been developed prior to the partnership with CCBIT and the decision to incorporate them into OWL, so the bulk of the work in the latter portion of the project fell upon the shoulders of the CCBIT programmers, with the leader from the department filling a mainly advisory and testing role. For them, incorporating the multimedia aspects of the lessons was an enjoyable challenge, and work proceeded on pace.

The difficulty came in when the project leader applied for additional funding to create new materials and disseminate the old and was unable to acquire it. Funding in education in the arts is always scarce, and budget cuts made it more so. For a time, the project faltered and work on the system ceased. As of the present, the project is still running at the same status that it was in 2003.

The scarcity of outside funding has not kept the project leader from applying for more, and a new project has recently opened up that is bridging work done by another group with the OWL software and the project's Art History team's pedagogy. The effects this renewed activity will have on the old project remain to be seen.

Overall, though, it seems safe to say that the Art History project team's needs were mostly met. They were able to create a product of which they remain proud and are still interested in expanding, and they were able to successfully continue work on it at a high level throughout the life of the project. The same cannot be said for the Mathematics team.

The difficulties experienced by the Mathematics project team stemmed from one major issue: the massive budget cuts at the university caused by the state's budget crisis in 2002-2003. Because of these cuts two important things happened: the project leader in the Mathematics department chose to accept his pension and retire and the post-doctoral fellow at CCBIT who was doing most of the programming work on the project could no longer be funded under the revised CCBIT budget.

Unfortunately for the Mathematics project, the project leader, who was also the course chair, was the major driving force behind the project. Other faculty members in Mathematics were interested and attended project meetings, but he was the only one who had any real interest in doing the work. All of that changed when he accepted retirement. In the last months of his involvement, his focus on the project was lessened by the knowledge that he would be leaving before its final fruition, which meant that he moved most of the work onto his graduate assistant rather than tending to it himself. He was also unable to find someone to really take over the reins after he left: the new project leader was such in name only and had neither the time nor the inclination to give it the attention it deserved. So, after Spring 2003, with a number of goals, including the Tutors, still unrealized, there was no one to helm the project. This, more than anything else, was responsible for the project stalling as it did.

But the story on the CCBIT side was almost equally disastrous. The budget cuts at the university level led to the decision to dispense with CCBIT's funding entirely. Although they were able to find advocates at the administrative level who successfully argued to retain some of the budget and thus sustain OWL and many of the CCBIT staff, they were forced to downsize the programmer who had most strongly been associated with the Mathematics project. From the time he received his notice, his morale was severely weakened and very little was accomplished, including the expected software framework for the Tutors—what was done was eventually accomplished by another programmer not originally part of the Mathematics team, but it was too little too late for that material to be fully realized.

Virtually nothing novel was introduced into the Mathematics OWL material for the entire last year of the funding, and although the existing materials continued to be used in the course (and still do), no one remained with the vision to push for its continued development and dissemination. That situation remains to the present time, and although there have been a number of OWL proposals that would have strongly benefited by a Mathematics component it has not been included because there is no one willing to put in the work.

The Funding Agency

Finally, there is the funding agency to consider. They are the ones who provide the funds for the project, and one would expect that they would be very interested in being assured that they get their money's worth. Sometimes this is true. At other times, surprisingly, it is not. It all depends on the funding agency.

Most of the OWL-related projects have been funded by grants from large, federally-sponsored funding agencies, primarily NSF and FIPSE. In both of these agencies, until recently, the model for how funding was doled out was as follows. First, a project team had to submit a proposal that successfully struck the interest of a series of review boards. If the project looked good and exciting on paper, and if there were enough available funds, it would be funded. The project team would receive the money, go through the process, and be asked to submit annual reports on a regular basis. Projects might get advise and visits from funding agency personnel, but for the most part it took a spectacular failure or complete lack of action to convince them to remove promised future funding. The worst that could happen was that a major failure, as shown on the final report, (submitted by the project team, not representatives of the funding agency) was that the teams chances of receiving future funding *from that agency* could be at risk.

Times are changing, and now there is a demand for greater accountability of grantees, as in every field of education, but both of the projects discussed her were funded under the old model.

So what is a funding agency looking for? Essentially, they want to be the supporting agency for as many successful projects as possible, based on the amount of money they spend. But the larger funding agencies use the strategy of funding what seems be promising projects with the expectation that a large proportion of them will not get very far; from their perspective (which is hard to fault) even the failures can add to the general fund of knowledge on the subject. So their immediate goals, after the original proposal is accepted, are to have periodic reports submitted on time that show some sort of progress, as defined by the goals of the proposal. And at the end of the project they would like to see a final report that shows the team has done the bulk of what they promised they would do, with the hope that they realized some of the expected effects.

If a project wants to acquire follow-up funds from the same agency, however, the results of the project become much more important. In that case, they have to demonstrate that the project, if funding continues, will further the overall goals of the agency. Receiving additional funds after the initial grant is much more difficult.

This second fact became abundantly clear with the Art History project. When the project was originally funded, the stated goals of DOE-FIPSE were to enhance and, more importantly, *change* the postsecondary educational experience in the direction of higher interactivity and inquiry. In theory, in order for the project to have been successfully refunded, it would have had to meet those criteria, which it did not; the Art History modules ended up being to ancillary to the classes where they were used to create any change in the students' classroom experiences. The project's chances of continued funding were hurt even further by a shift in FIPSE's focus from clearly postsecondary endeavors to ones that branched postsecondary and K-12 or community college environments, a response to the change in the political climate of education under the Bush Administration. Additionally, FIPSE, as with many other agencies, prefers to fund new projects and has relatively little money set aside for the furtherance of existing projects.

So, although the DOE-FIPSE administration was satisfied with the product as such, it was not enough in line with their goals, especially as they became reconfigured with the change in the political climate, to receive further funding. In essence, the concerns of the funding agency had not been fully met.

The Mathematics project never applied for further funding—its leadership broke down before that was a viable option. As it stood, it was able to meet the minimal demands placed upon it by NSF—annual reports showing some progress every year (albeit minimal in the last year) and no spectacular failure that would cast a blight over all those connected to it. To date, there has been no indication that the results of that project have had any impact on the ability of other projects put forth by members of the team to get funded.

Summary of the Two Projects

In both cases, there were successes and failures, and each can be attributed to either meeting or not meeting the needs of the various stakeholders. The table below summarizes the relationship of each project's history to stakeholder concerns.

Stakeholder	Art History	Mathematics
Students	Work is interesting and not aversive but does not tie closely into class and thus is not closely related to grades – MIXED	Work has better feedback and is less frustrating than other homework; is directly related to success on exams – MET
Course Instructors	Allowed for structured homework in class but is difficult to integrate with other lessons – MIXED	Allowed for structured homework to support lessons when none was previously available; not easy to continue integration under new course leadership – MIXED
Other Departmental Instructors	Did not spark interest due to specificity of content and difficulty of creating new materials – NOT MET	Sparked interest due to need for less costly homework models but too costly to create new materials for new courses – NOT MET
Host Department	Did not spark interest due to independence of individual instructors and specificity to a general education course – NOT MET	Did not spark interest due to expense of dissemination and specificity to an introductory course – NOT MET
Faculty in Other Departments	Not relevant	Addressed their concerns and incorporated impact into project design – MET
University Administration	Entire OWL system seems to improve student learning and becomes a focus for several university-wide proposed initiatives which include Mathematics as a component – MET	
Faculty at Other Institutions	Material is too pedagogy-specific and technology-intensive for easy adoption – NOT MET	Insufficient dissemination efforts by project team lead to no viable model for dissemination – NOT MET
Project Team	Team maintained enthusiasm for the product and continues to pursue additional opportunities – MET	Team leadership and support imploded with retirements and layoffs; nearly all momentum lost in last year – NOT MET

Funding Agency	Products fit the requirements for initial funding but were not enough in line with agencies goals for additional funds – MIXED	Able to produce reports of progress on time; did not attempt re-funding – MET

One of the major reasons that so many concerns were either not met or met on only a mixed basis was that few of these concerns were considered from the origins of the two projects. That is not to say that considering the concerns of all stakeholders would necessarily have led to untrammeled success in both projects—some things, like university budget cuts, are nearly impossible to take into consideration—but it is likely that there would have been fewer disappointments.

Setting Goals from a Multiple Stakeholder Perspective

When deciding the course of a project, it is useful to start with a set of goals. For many projects, these goals tend to be simple ones such as "the innovation will increase student learning of the material" or "the innovation will successfully teach better than a traditional textbook."

Both of the projects discussed in this paper had similar goals to these, although not so concretely laid out. Both had some vague idea of increasing student learning of the subject matter by increasing their time on task outside of the classroom and by increasing the interest value of the homework assignments they would be completing. And, incidentally, these goals were met in both cases (though some might debate that the Math OWL homework really increased student interest).

Yet neither project could be determined an overwhelming success. Why? Because the goals of the project were not directed toward the perspective of all the stakeholders.

In order to optimize a project's success, it may be more helpful to design a series of goals around the concerns of all of the stakeholders in the project. On the table below, we have listed some potential goals that meet the concerns of each group of stakeholders, as outlined above.

Stakeholder	**Potential Goals to Meet Their Concerns**
Students	Will increase student learning and thus performance in the class Will be more valuable and time-efficient than other modes of instruction
Course Instructors	Will not take substantial effort to be integrated into a course Will increase student learning and thus limit the need for spending class time on review
Other Departmental Instructors	Will increase student learning and retention of material, making them more prepared for subsequent classes Will not take substantial effort to expand the innovation to additional departmental courses
Host Department	Will be less costly to maintain than the previous method of instruction Will make taking courses in the department more attractive to students
Faculty in Other Departments	Will increase student learning and retention of material, making them better prepared to use it in classes outside the department Will incorporate examples of how material can be used in other settings Will not take substantial effort to transplant key aspects of the innovation into courses outside the department

University Administration	Will be expandable to further grants, allowing more money to come into the institution Will increase students' interest in coming to the institution Will recruit donors to give more money
Faculty at Other Institutions	Will increase student learning over traditional models Will be able to be implemented with a minimal amount of effort and outside support
Project Team	Will allow the team to tackle interesting problems Less enjoyable work will be distributed such that it does not become overly burdensome Will increase student learning of and interest in the material
Funding Agency	Will increase student learning Will meet project milestones in a timely fashion Will be in line with the overall goals of the funding agency, making additional funding desirable

These particular goals, obviously, are drawn from the specific experiences of the two projects discussed in this paper, and are likely not all appropriate to any given project. Still, goals of these types should be conceivable for most projects.

Our experience has shown that goals can be grouped into several broad categories. These are:

- Learning goals: These hope to increase the rate of learning and retention of material. Increasing the interest value of material or the time spent on the task are goals of this type.
- Performance goals: These hope to increase measures of performance, which is distinct from but related to learning. Some examples might be increasing student test performance or progression through the courses of a major, or increasing the amount of material successfully covered by an instructor.
- Quality goals: These relate to the product's features and their usability. Examples of quality goals would be creating an innovation that is easily customizable by an instructor or that can be readily understood by a student.
- Resource goals: These relate to the amount of time, money, space, etc., that is necessary for the innovation to succeed. Some goals of this type might include the product being cost-effective enough that no significant maintenance will be necessary once funding has ended or that the time required to learn it by a new faculty user will not be so great as to be significantly higher than what would go into creating a similar lesson or other part of the course.
- Progress goals: These relate to the meeting of the milestones of a project within a specified time period, including the dissemination.
- Inspiration goals: These relate to improving or continuing the interest that the stakeholders have in producing, using, or disseminating the innovation.

Once goals have been set, the next thing to do, of course, it to take steps to meet those goals. But how would one know whether or not one is actually being successful at meeting the goals? That's where evaluation comes in.

Determining If the Stakeholders' Concerns Are Being Met

A good evaluation plan is one that provides for activities to assess each of the goals of a given project. As the evaluators of many different projects, including the two OWL projects described here, we were confronted with the task of evaluating these various goals and relating them to the various stakeholders.

Learning Goals

How do you know that your students are learning better and retaining more information with the introduction of a new innovation? Most commonly, projects will examine proxies for learning such as grades on quizzes, exams or overall course performance. But there are a number of issues with using such proxies.

First, grades may not directly measure the skills that your innovation is trying to get across, or they may test these skills in conjunction with other skills from the course that are not covered by the innovation. This becomes more problematic as the amount of material going into the grade increases: quizzes can be pointedly directed at specific information, but exams must cover large pieces of the curriculum.

But even with quizzes, it is often difficult to ensure that grade being given is indicative of the degree to which the student has obtained the skill in question. Grades may be assigned on a relative scale (a "curve", if you will) rather than based on absolute scores, which can make comparing across usages almost impossible. Worse, the questions may be assessing different skills and knowledge from what you expect they are; test validity is always something that needs to be assessed, especially when the skills in question are higher-order thinking processes such as inquiry learning.

We encountered the issue of relative grading when trying to assess the impact of learning in the Art History course. There, students were asked to write essays describing what they had learned from the various modules, both in the pre-OWL conditions and with OWL being used. According to the grades students received, there were not differences between the conditions. But closer examination showed that there were no differences between grades on essays given earlier in the year and those given later, even though the professor was certain that most students had become better at employing the Art History skills being asked about. It turned out that the teaching assistants were applying a subjective grading system that normalized the average response to a grade of a B. Clearly, these grades were of no use in determining learning of the skills.

To overcome this problem, we had an outside grader score a selection of essay tests from the final exam using a carefully constructed rubric, and used this score—which students never saw and which had no effect on their grades—to determine whether learning had occurred or not. As it turned out, students in the pre-OWL module conditions scored an average of around 48% of the possible points from the rubric, compared to 67% in the OWL-using condition.

The problem with the grades being invalid measures of the art history skills was also evident with the original essays. There, graders were more inclined to give points for coherency and proper use of the English language rather than the important skills the module was trying to get across—this led to the need to develop a separate rubric in the re-scoring.

But even if you have a good measure that you are confident with, in order to demonstrate learning you need to have something to compare that learning to.

A number of clients we have talked to have naively assumed that giving students a pre-test and then showing learning on an identical post-test is sufficient to demonstrate that learning has occurred. And this is true to some extent: it *does* show that learning occurred. But what it doesn't do is demonstrate that the innovation was a better conduit for that learning than something else might be. Those students might have learned just as well or better if the instructor just handed out worksheets explaining the skill or knowledge to be learned. You don't know.

So you need a control group. This is a huge problem with a lot of innovations. In many cases that we have encountered, the innovation accompanies a new course that has never been offered before or other major changes in the existent course, making it impossible to find a comparable pre-innovation comparison group.

This was certainly the case in the Mathematics use of OWL. OWL was introduced into the course curriculum at the same time as a number of other changes, including a new textbook and a new course chair with a corresponding change in the focus of course exams. There was no way to determine if any changes in learning were a function of OWL or of these other changes, and the new exams with a more conceptual focus were not even comparable to old performance metrics.

In this course, however, there were multiple sections, and in the first semester of OWL use some of those sections were using OWL and some were using older methods of homework (essentially ungraded paper-and-pencil methods). So one would expect that there would be a chance to have a reasonable between-section comparison with the added bonus that all of the students were using the same exams. The problem there was that each section was taught by a different instructor save one, and the one who taught two sections used OWL in both. So differences between sections were as likely to be a function of instructor differences as OWL use. It turned out that the variation in performance between instructors using OWL was just as great as that between OWL- and non-OWL-using instructors: the effect of instructor and the individual class dynamics had a much higher influence on grades, etc., than OWL did.

So it is important to have a control group that is essentially the same on all variables as the innovation group (or at least on all variables that you have some control over), and it is important that both groups use the same measurement tools.

This happy situation was the case in the Art History example given above, and we were able to set up a similar situation in Math when examining the Discovery Modules. There, we had a pair of semesters where two instructors each taught two sections and we were able to convince each to use the Modules in one section but not the other. We then measured learning by examining scores on a subset of multiple choice questions from the final that examined the skills taught, but found no differences.

Thinking that the scores on the exam could have been influenced by studying behavior beyond the use of the Discovery Modules, we set up a second test where students were given one-shot quizzes in OWL that directly related to the contents of the Discovery Modules, asking both conceptual and quantitative questions. The full results of this study are given in Rath, Peterfreund, and Hart (in preparation), but, in brief, we found better scores on the conceptual questions among the Module-using group than among the regular OWL users, but no difference in the quantitative questions—until we looked at gender. At that point, we found that women performed better in the Module condition than men, but in the regular OWL condition the roles were reversed. The gender difference was not, however, present on the questions from the exams. The full implications of this are discussed in the other paper.

There are other measures of learning that we have used in other situations that were not appropriate for these two situations. One example is determining the students' success in subsequent courses that build off the skills taught in the class where the innovation was introduced. Although there was no question of trying to pursue this avenue for Art History—the vast majority of students did not go on to take further courses building off their art historical skills—we did consider doing so for Math. But even though there was a second-semester math course taken by most students, the massive changes that accompanied the switchover to OWL accompanied by nonstandard learning metrics in the second semester class made any potential results of this study meaningless. So we chose not to exert the resources to look into it.

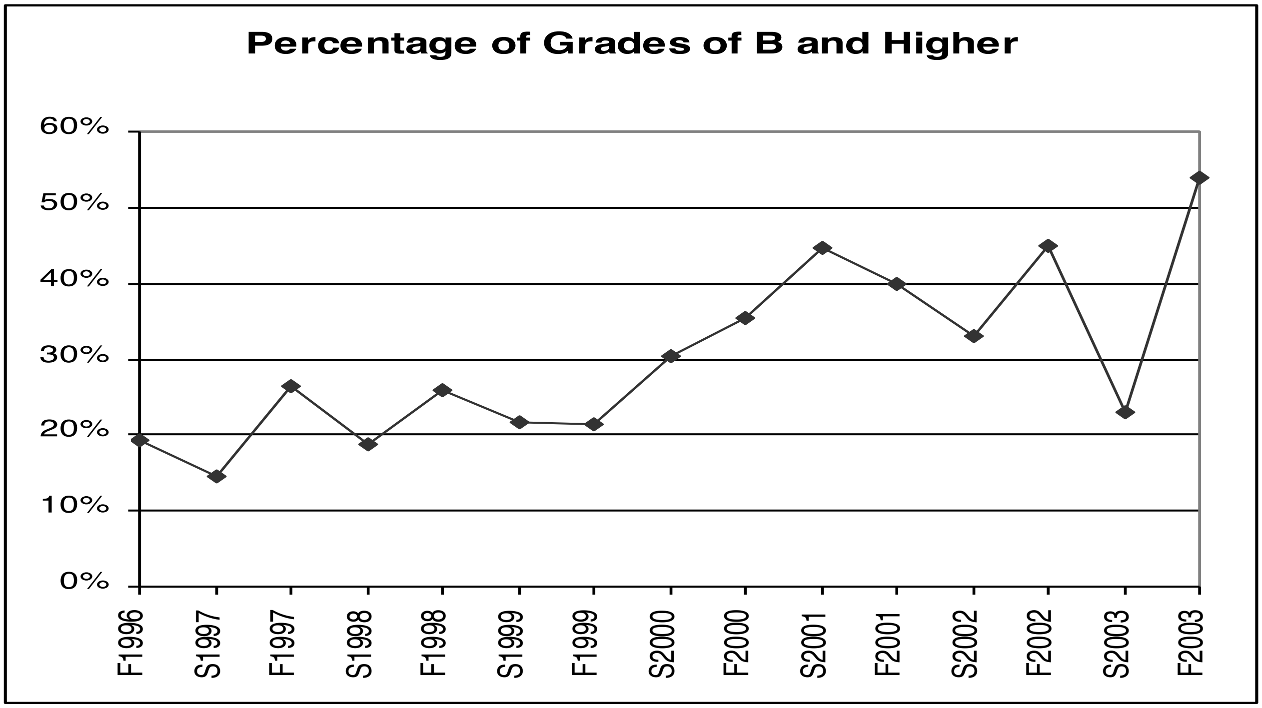
We have examined learning in subsequent courses in other situations with some success. Again, it is important to keep the controls in mind: courses taught be different instructors and using different tests will not give comparable measures. The best situations are ones in which a class contains both students who took the preceding course with the innovation and those without, allowing a relatively clean study. But if one group has had more time between the taking of the supporting course's material than the other, the more recent takers may have an advantage in the class irrespective of their exposure to the innovation, so care is still needed.

Performance Goals

In many ways, achievement of grade-based performance goals can be measured using the same methods described for learning goals—the proxy learning measures of quizzes, exams, and course grades are direct measures of performance. In that sense, many of the same considerations apply when using these measures to examine performance as when using them to examine learning, although the controls can be looser—if you are mostly interested in increasing grades, you may not need to determine that the expected learned skills are directly responsible for that increase.

As with learning, one of the more interesting performance measures is examining patterns over time. We performed this type of analysis in Mathematics, reasoning that, although there was a different mix of instructors every semester, the differences in the mix of instructors should come through as noise but that any change in the general trend that occurred concurrently with OWL and the other course changes should be attributable to one or more of those changes, assuming that the criteria for receiving particular grades (in terms of test scores) did not change over that time period, as the instructors assured us they had not. The results of this are shown on the accompanying figure, which illustrates the proportion of students in the various semesters who achieved a score of "B", "AB" or "A".

Examining this figure, one can see some interesting trends. First, at least in the beginning, there seems to be a tendency for students in the Fall to outperform those in the Spring. This is likely because the best prepared students are ready to take the class their first semester of freshman year, and these students are not represented in the Spring classes. Second, there seems to be a large jump in the percentage of students receiving high grades starting in Fall 2000 and climbing steadily until Spring 2001, where it starts varying considerably. Spring 2000 was the first semester where some OWL was used by one instructor, albeit in a relatively primitive state. In Spring 2001, the new course chair took over and OWL was greatly improved, though still not used in all classes, but by Fall 2001 it had been fully institutionalized. This state of OWL use continued until Fall 2002. In Fall 2003, the new course chair took over, and there is a corresponding fall in grades. The following semester grades were back up again, but this is likely due to the tremendous amount of scaling added to the second exam (more then 10 extra points were given out of scores of 100).



The trends, then, could be taken to say that performance increased through the introduction of OWL. But this is confounded by the effect of the course chair—grades increased with the introduction of the new course chair, and when he left, grades fell again. That fall may be attributable to the fact that the most recent chair was not familiar with OWL and did not utilize it as effectively, or it may be that the first chair who ushered in OWL just gave easier exams. Without sufficient controls, it is impossible to know for certain.

Another interesting measure of performance relates to the instructor and, more specifically, to the ability to effectively make use of time. This is something we always tried to get a handle on with OWL: did OWL allow instructors to offload some of the teaching burden to the homework system, allowing them to cover more and different things in class?

Answering this question does not require the formal evaluation setup of some other evaluation methods. Generally, instructors have a good sense of how much they can cover in a semester, and can tell you if students are coming in better prepared, allowing them to spend less time covering a topic in class (although such things can be tracked objectively by outside observers, a tactic we never used). Instructors are also easily able to tell you if they end up ahead of schedule.

In Art History, the modules did not cover material that was already being taught; rather, they added new material to the curriculum. Because of this, there was no expectation that they would change class time in any appreciable way.

The case in Math was different. There, the use of the Discovery Modules was expected to allow students to learn the material on their own, taking away some of the teaching burden for in-class instruction. Unfortunately, there was never any indication from the instructors that the amount of time they spent on topics was appreciably changed, regardless of the OWL-use situation.

In general, the goal of all performance evaluation is to be able to demonstrate that higher standards of performance are being achieved as a function of the innovation, whether it be grades, use of time, or something else. In our two cases, we were able to demonstrate performance changes in Math grade trends (although these may not be solely attributable to OWL, as explained above), but grades in Art History remained constant, despite demonstrated *learning* gains, due to the subjectivity of the grading. Unfortunately, students in a general education class such as the Art History offering are much less likely to be interested in increasing their chance of learning skills that they don't expect to have to use again than they would be in increasing the likelihood of getting an "A" in the course.

Quality Goals

Our evaluation efforts have typically assessed the quality of from two different perspectives: that of the students users and that of the instructors of the courses.

For assessing students' opinions of the quality of OWL, we have developed a survey that we administer in two parts, one at the onset of the semester and the other at the end[1]. Among the data we collect are student demographics (age, major, class, gender, etc.), interest in the class, and perceptions of what barriers they encounter toward achieving success in the class. We also ask them to rate methods of learning the material, tools to help them learn, and specific aspects of OWL. Responses to these surveys were backed up by interviews with small groups of student volunteers, which generally confirmed and elaborated upon what we found from our surveys.

In Math, we found that students' opinions of OWL tended to be neutral to positive, with ratings hovering in the 3.5 range on a scale from 1 to 5, one being the most negative (the average OWL ratings from all courses surveyed tend to be around 3.7). Students didn't feel overly put-upon by having to use OWL, but they weren't fired up about it either.

More interestingly, we found that students consistently reported that OWL was *the most important* tool for achieving success in the course, with higher ratings than more traditional modes such as lecture, notes, and the textbook. Clearly, regardless of their thoughts about the quality of the experience of using OWL, students saw it as the most important tool available. That it did so well relative to the lecture and textbook speaks well for OWL's usability (or suggests that other class resources are remarkably poor).

We also asked students using the Discovery Modules to rate the quality of these individual assignments through short surveys tailored to each module. Here we found two things. First, students did not rate them all that highly, although more students rated them positively than negatively. This was similar to the ratings of OWL in this class. More tellingly, from the comments we collected we were able to determine that students were not distinguishing the Discovery Modules from their regular OWL assignments. Analyzing the data, we had found that most students completed all of the different versions of the Modules (when there were multiple versions) even though they were only required to do one. It turned out that most students were unaware that there were multiple versions when they were doing them—either they were just doing the homework on autopilot or the different content differentiated the

[1] The authors will happily provide copies of these surveys to interested parties if personally contacted.

assignments enough that they were unable to see the underlying similarities. This may also explain why the module survey responses tended to be so similar to one another and to the OWL results in general.

At any rate, we were able to determine that the Discover Modules did not provide the students with a noteworthy experience. This, coupled with the lack of learning differences found in the classes as a whole, led the instructors to stop using them, starting in Spring 2004.

Compared to Mathematics, opinions of OWL in Art History were very positive, with average scores around 3.8-3.9 on the same five-point scale. The students also gave the individual modules high ratings: over 65% of the students rated them as being valuable experiences in nearly all circumstances (the newest ones introduced in the last semester of evaluation and not yet refined received slightly lower ratings). The students clearly saw the Art History OWL assignments as being valuable and easy to use.

The overall reliance on OWL for success in the course was not nearly as high in the Art History courses as it was in Math, however, because the OWL modules were supplemental rather than providing support for the entire curriculum. OWL use was deemed enjoyable, but not essential.

Although we did conduct some formal interviews, most of the information we gathered assessing the quality of OWL from the instructors' perspective was acquired through informal conversations with the individuals involved. Over the course of the projects, we spoke with most of the Mathematics and all of the Art History instructors and most of their teaching assistants on numerous occasions, asking questions about how easy it was to use the OWL assignments, how well they fit into the established curricula, and how easy it was to create new assignments.

In Art History, the instructors (not counting the project leader) found OWL very easy to incorporate into their classes—all they had to do was assign the modules and collect the grades at the end; the teaching assistants took care of grading, etc., and they were taught by the project leader. For their part, the teaching assistants found the mechanics of collecting grades and answering student questions to be very easy as well. So the quality of OWL's usability was not in question.

Incorporating OWL into the existing class proved to be more challenging. Because the OWL modules were complete lectures, in some cases their content was not different enough from what was already being presented in class to warrant inclusion, and in other cases the content just didn't fit into the class' curriculum. So, except for the project leader, none of the instructors used the whole suite of available modules.

None of the instructors save the project leader even attempted to create their own modules. These modules were quite complicated, involving several interactive multimedia components, and creating new ones required considerable computing resources that could not be obtained in the absence of grant money. Modifying existing modules involved a similar level of effort.

From the instructors' point of view, then, the Art History modules were quality products but with limited utility. Additionally, there was no way that they could be expanded in scope or modified in content to fit differing situations.

In Mathematics, it was also relatively easy to shuffle the task of assigning the OWL assignments off to a teaching assistant who would match their due dates up with the course syllabus and collect grades. In fact, it was so easy that the course chair that succeeded the project leader was able to continue to assign OWL in his classes without so much as looking at a single assignment, thus removing any impetus to incorporate the OWL material into the lectures. In a sense, OWL may have been too easy to incorporate into a class.

Because the OWL materials were designed to support the entire Calculus I curriculum, no instructors found any difficulty in incorporating the available materials into the flow of their course; the materials had been designed in close consort with the syllabus.

Even creating new OWL materials was relatively easy in the Math system, as some instructors ancillary to the project and teaching different courses readily demonstrated. But little was done to change the core set of OWL materials for the studied Calculus course after its initial development because the time required to edit and enhance the entire course would be substantial.

In general, however, the Mathematics instructors rated the quality of their OWL materials very highly on all major fields that we were examining. The issue was the individuals' willingness to take the time to develop OWL, not OWL's ability to developed.

Resource Goals

Our examination of resource goals from the departmental and instructors' perspectives focused on three main areas: the cost of the product, including any savings its use might entail and the cost of sustaining it over time, the time it takes to use, including any time savings over old methods, and the space needs on the campus.

As part of the DOE-FIPSE grant funding the earliest installment of OWL, we performed a complete cost analysis of all the classes using OWL at the time, the results of which are detailed in an upcoming paper (Rath & Hart, in preparation). This was done by interviewing a number of OWL users regarding the amount of time they and other people in their department spent on various tasks both before and after the introduction of OWL, questions that the interviewees were able to answer with a fair degree of specificity. The analysis included a section on OWL in Mathematics for Fall 2000, and the cost model did not change appreciably since then.

In some of the classes examined, notably those in the Physics and Chemistry departments, there were cost *savings* associated with OWL because having the graded homework within OWL allowed departments to cut back on discussion sections rather than having to hire new faculty members to replace those who had been teaching those discussion sections and were now retiring. This was not, however, the case with Mathematics. There, the department had never had the funds to support multiple discussion sections, and prior to OWL's introduction all homework had remained ungraded or been hand-graded by the rare, highly-dedicated instructor.

But OWL had only three areas of significant cost. The first was salaries for members of the project team to author the materials, a cost that was covered by the grant and did not need to be maintained after the database of questions had been created. (Had the course instructors wished to add or modify materials, this would have been an additional cost, but they evidenced no desire to do so.) Second were the costs associated with maintenance of the system—setting and modifying due dates, keeping track of grades, and answering student questions. The burden of these activities turned out to be fairly low and well within the capabilities of the single teaching assistant for the course. The third area was the maintenance of the software and server, a cost borne by CCBIT, the creators of OWL. At the current time, CCBIT is providing the OWL database to Mathematics for free through university funding, but that could change, and there would certainly be a cost for any other institution that wished to take on OWL. The various possibilities for covering this cost are a matter of current discussion at CCBIT.

The cost model for Art History is essentially identical to that in Mathematics, although the effort required of the teaching assistant is considerably less due to the much smaller number of assignments. Again, server and maintenance costs are being provided by CCBIT.

In both cases, once the initial materials were developed, using OWL required very little time on the part of the instructors and teaching assistants. Assignments needed to be given start and due dates, which sometimes had to be changed due to unforeseen circumstances such as snow days and system failure, grades needed to be collected, and student concerns had to be addressed. Although we never acquired exact estimates from the instructors in these two departments, one instructor in Physics estimated that it amounted to around 25 hours a semester for over 100 students. The number and complexity of assignments in Mathematics is similar to that in Physics, so the time there is probably comparable, while that in Art History is probably less. This amount of time is easily addressable in a single teaching assistant's salary, especially when compared to standard pencil-and-paper grading.

Finally, although the Mathematics department did construct a computer center as part of the grant, there does not seem to be a need for more space to support OWL, at least not anymore. When OWL was first implemented in Chemistry, many students still did not have computers or internet access in their rooms, making lab space necessary. Now, however, our studies, conducted by analyzing where student are physically when they log in to OWL, have shown that nearly every student on campus has access to the internet in their dorm room or off campus, making the maintenance of extra departmental computer space unnecessary.

The resources required to implement OWL, therefore, were relatively minimal from a departmental and instructor perspective. We also asked students on our surveys to indicate how much time they spent on OWL on a weekly basis. In Art History, the average reported time spent on OWL was about one hour a week, out of the three and a half spent for the class. In Mathematics, the reported average times were three and four and a half, respectively. So in neither course could the OWL work be construed as placing an undue time burden on the students.

Progress Goals

Measuring progress goals is a relatively simple task, compared to many of the other evaluation methods examined here. Essentially, all the evaluator needs to do is take note of when milestones are met and compare that progress to the progress proposed in the initial plan. This includes the entire scope of the plan: materials development, use in classes, publicity, and dissemination to other users.

In Art History, progress followed the stipulations of the grant proposal in nearly all respects save one: although a subset of the modules were used at a different college, there were no users who permanently agreed to incorporate the modules into the courses they were teaching though the proposal did call for a dissemination of the materials. The reasons behind this lack of dissemination have already been discussed.

The Mathematics project had an even more difficult time meeting their progress goals. Although the basic OWL question set was developed in a timely fashion, the Discovery Modules took longer to develop than planned for, and the Tutors were never fully developed and implemented within the lifespan of the project. After some halting initial efforts to bring in outside users, no real attempts were made to spread the product outside the host institution or even to incorporate it into Introductory Calculus II on a course-wide basis. Although conferences were attended, only the evaluators have made any attempt to publish the findings in peer-reviewed journals. In short, for a variety of reasons discussed earlier, progress toward the stated goals of the project slowed by the middle of the time period and almost stopped entirely by the final semester.

Inspiration Goals

Measuring the extent to which a project inspires people is very difficult to measure, and most of what we have done to capture these outcomes has been the recording of anecdotes. But keeping track of this is important because may be one of the most important indicators of a project's success.

Inspiration can happen at several levels. An innovation can inspire students to become more interested in the subject matter, which may inspire them to take additional courses they wouldn't otherwise have attempted, pursue research in the topic, or even change majors. Faculty instructors can be inspired by developing a greater interest in and excitement toward teaching the course, or in introducing a modified version of the innovation into other courses they are associated with. And, of course, the project team needs to maintain their initial level of inspiration in order to continue taking pleasure in the project and, while increased inspiration as a function of success can lead to the desire to seek funding for the next step of the project or the development of new ideas beyond the project's initial scope.

Gaining an understanding of how inspiration is affecting the project team is often not difficult since the evaluators should work regularly with these people and develop an understanding of how their thoughts about the project are progressing. But understanding the affects on faculty instructors requires interviews, either formal or informal, and student inspiration is often only discovered through chance comments on open-ended survey items and responses to interviews. As not all students, or even more than a very small minority of students, can be expected to be inspired by an innovation to change their educational plans, you are not likely to be lucky enough to get interviews with the affected students unless you conduct a large number (which is very resource-draining and which we have never attempted to do).

In neither case have we encountered any cases of students who have indicated that they have been inspired to pursue math or art history further specifically because of their experience with OWL, although there are always a few people who indicate that they have been inspired by the course experience in general. Given that OWL is a homework system, any other findings would be very unexpected.

In terms of faculty members, one instructor in Mathematics did pick up OWL to be used in his section of Introductory Calculus II, but he was the only one who evidenced any inspiration to carry it farther, and he was a partial member of the project team, having designed the very first OWL assignments in 2000. None of the Art History instructors did more with OWL than assign it in their classes and add a few supporting questions on the exams.

The Mathematics project team essentially fell apart with the impending retirement of the project leader and the downsizing of the lead programmer, and by the last year the only thing that kept development of the tutors going was the desire on CCBIT's part to meet their obligations so they wouldn't look bad in other proposals. After the end of the grant, numerous large-scale proposals have been written that have included a Mathematics OWL component, but no one from the Mathematics Department has stepped forward to offer leadership of that component. OWL-based inspiration has entirely dissipated.

The project leader in Art History, however, has continued to write articles (not yet published) about her accomplishments with OWL and continues to pursue opportunities to expand the OWL materials and visibility—despite all of the setbacks, she is still inspired to continue what she believes is a worthwhile and important project. Unfortunately, educational funding in Art History is scarce, but some promising contacts have been made which may breathe new life into the OWL project.

Predicting and Driving Success

In order for a project to be successful, more needs to be accomplished than the creation of an exciting new innovation. At the project's onset, the groups of stakeholders who will be impacted (or who you want to be impacted) by the innovation need to be identified, and the project team needs to set a set of goals to meet the needs of these stakeholders. The project team must then take steps to meet these goals, while the evaluators undertake activities to determine the extent to which they are being met.

Too many excellent innovations with the potential to truly drive change in education fail by not meeting stakeholder concerns, while some less-deserving products go forward because they have taken the necessary steps to do so. We hope that the lessons we have learned through the two projects detailed here will help you to create project and evaluation plans that are focused to wider-scale successes than just the creation of an innovation as an end onto itself. Not all problems can be predicted and overcome from the onset—at the beginning of the Mathematics project we could not have predicted the budget cutbacks that occurred in the wake of the national recession, for example—but with careful planning some of the more obvious setbacks can be averted before they occur.

Maybe you can change the world.

References

Patton, M. Q. (1997). *Utilization-focused evaluation: the new century text*, Thousand Oaks: Sage.

Rath, K. A., & Hart, D. M. (In preparation). The use of a web-based homework system to decrease costs in large, discussion section-based college classes.

Rath, K. A., Peterfreund, A. R., & Hart, D. M. (In preparation). Gender differences in the reaction to web-based calculus discovery instruction.