Self-efficacy Concepts and the Evaluation of Instruction

David R. Haws
Boise State University

Abstract

As instructors, we need to do more than provide our students with an opportunity to absorb content. We also need to help them to develop the continuing motivation to devote large portions of their time to gaining proficiency—in our own course (a micro-goal), in future courses (a macro-goal), and professionally (a mega-goal). As Bandura[1,2] has pointed out, intrinsic motivation is a function of the beliefs we have in our personal efficacy to achieve. A well-prepared final (keyed to a pre-test) can assess the learning of cognitive skills. A similarly referenced performance test can assess the learning of psychomotor skills. But we should also be seeking to achieve and demonstrate learning within the affective domain, including an increase (or at least a maintenance) of student self-efficacy belief. This would require a deeper evaluation, in terms of the four levels (Reaction, Learning, Behavior and Results) proposed by Kirkpatrick[3].

This paper will discuss a method of instruction evaluation that develops a course-specific, quantitative expression of change in self-efficacy belief keyed to actual performance on diagnostic and final exams. This system is being developed by the author, and was used to evaluate courses in Soil Mechanics, Strength of Materials, and Reinforced Concrete Design at Boise State University during the 1998 calendar year. For clarity, only examples from the Strength of Materials course will be discussed.

Introduction

Four levels of assessment are critical to a well-triangulated instructional evaluation. Kirkpatrick[3] identifies these levels as Reaction, Learning, Behavior and Results. **Reaction** evaluations (Level 1) ask students to "react" to instruction by sharing personal perceptions of the learning experience. **Learning** evaluations (Level 2) try to quantify isolated "learning" as the algebraic difference between the comprehension revealed on post- and pre-tests. **Behavior** evaluations (Level 3) reveal the affective response to learning within the context of performance—whether the learner modified personal "behavior" following instruction, by either using or enhancing (as opposed to simply ignoring) the newly acquired skills and knowledge. Finally, **Results** evaluations (Level 4) seek to determine the extent to which those behavior modifications targeted by the learning experience have "resulted" in the fundamental goals they were designed to achieve.

These four levels of evaluation represent a hierarchy of complexity ranging from micro-level interests, to macro and mega-level concerns[4]. All four levels are meaningful and need to be included in any systemic, systematic approach. Results evaluations (Level 4) are something new in engineering education, entering largely through Engineering Criteria 2000. Reaction evaluations (Level 1) have been around in the form of Student Perception of Teaching (SPT) surveys since the "Berkeley Slate" used in the 1960's. While the post-tests associated with Learning evaluation (Level 2) are much more common in engineering education than pre-tests (sometimes called diagnostic exams), these also are fairly familiar devices. The approach to evaluation that I want to discuss in this paper relates more to the third or Behavior level of evaluation—what attitudes relative to the future use of learning have been instilled by instruction, possibly as unintentional outcomes?

You might note here that attitudes are also learned, in the classic sense, and together with cognitive and psychomotor skills form a system of internal controls on behavior. Behavior is also subject to external controls, which need to be acknowledged by the instructional designer. However, while *Instructional Technology* is used to affect *internal* controls, *external* controls are more appropriately the province of *Performance Technology*. External controls are independent of instruction. While they are important for summative instructional decisions, formative decisions benefit from an instructional evaluation focused on internal controls.

Obviously, all four levels of evaluation may reveal features that are more realistically attributable to external controls than to the unit of instruction being evaluated. For example, the student may be "reacting" to the time of instruction (too early or late in the day). "Learning" may be impaired by the student's work schedule, emotional trauma, or lack of learning readiness. External circumstances might control the learner's "behavior," denying an opportunity to apply learning until it has been obscured by time. Or behavior (for example, designing a more convenient form of personal transportation) might have unanticipated side effects (air pollution and the depletion of non-renewable hydrocarbon resources) which obscure the desired "result" (increased mobility for the working class).

A Level 3, Behavior evaluation of instruction usually requires a waiting period of 3 to 6 months, during which the desired behavior has a time to develop. However, if we primarily want a formative evaluation of behavior subject to the components of *instruction* that impact on affective learning, rather than absolute behavior subject to both internal and external controls, then we may appropriately focus on the motivational components of instruction. Cognitive and psychomotor skills, which also exercise internal control on behavior, can be addressed adequately at the Learning level. An appropriate use of a Behavior level evaluation, would be to assess the impact of instruction on attitude changes due to the learner's access to intrinsic motivation (extrinsic motivation is also developed by instruction but fades quickly and should be ignored at the Behavior level). This could be done at the end of instruction, rather than after a more traditional time lag, *factoring out* external controls and placing the formative components of the Behavior level evaluation on a fast track (you would still want to look at behavior subject to external controls for summative purposes).

Motivation, Behavior, and the Perception of Appropriate Complexity

Bandura[1,2] has proposed a model of motivation based on social cognitive learning theory. In this theory a reciprocal relationship exists between the learning environment (constrained opportunities to achieve), the individual's self-efficacy belief (a belief in the ability to achieve personally), and achievement through enactive mastery (as opposed to chance occurrence). Individuals are motivated to exert the effort to achieve when they believe that their efforts will be met with success. When people believe that they can achieve, and extend their efforts, they *increase* their probability of success. When people succeed through a reasonable exertion of effort (enactive mastery), they *increase* their self-efficacy belief, which *increases* their willingness to extend effort toward achievement, which *increases* their probability of future success. When people doubt their self-efficacy to achieve, they fail to exert the effort required, which *decreases* the probability of achievement. Failure to achieve can *decrease* self-efficacy belief, which *decreases* the willingness to exert effort, which *decreases* the probability of future success.

Individuals can also extend their self-efficacy belief through vicarious experience, by observing similar models exerting effort, coping with obstacles and achieving success. Here, model similarity is a critical issue. If the vicarious experience is limited to an obviously superior model (the *mighty* professor) effortlessly achieving the desired outcome (an error-free derivation on the blackboard), those learners who fail to identify themselves with the model, may also fail to generalize modeled achievement with their own potential for success. Because of the increased opportunity for recognizing a model as "similar," multiple models enhance the probability of increased self-efficacy belief through vicarious experience. Therefore, instruction should provide for a variety of models, including expert-models, peer-models, and perhaps even self-models[5]. To enhance learner identification, modeled behavior also needs to reference effort and coping, which are natural features of achievement (regardless of whatever fiction gets played out on the classroom blackboard).

Alternately, individuals may have their self-efficacy increased by verbal persuasion. The effectiveness of verbal persuasion is a function of how the individual perceives the source of influence. Successful persuasion is achieved through a valued opinion: from an expert (valued because of expertise), from someone respected (valued because of prominence), or from someone trusted (valued because of past loyalty or trustworthiness). While there is a hierarchical relationship between the factors affecting self-efficacy belief (enactive mastery is more effective than vicarious experience, which is generally more effective than verbal persuasion), all of these elements are and should be components of effective instruction.

Achievement (and reciprocally self-efficacy) is also greatly affected by the perceived complexity of the content, method and media of instruction (the learning environment). Achievement that was perceived as too easy doesn't do much to improve self-efficacy belief. Similarly, achievement that required unusually large expenditures of effort will also fail to raise self-efficacy belief, because the individual will doubt his or her efficacy to mount a similarly intense campaign in the future. Furthermore, certain instructional media (for example, television) and methods (like lecturing) are considered by some to be too simple, and will not elicit sufficient effort for adequate cognitive processing. Other media (for example, dense textual materials) and methods (like independent reading) may be perceived by others as too difficult, and similarly restrict learner effort.

While perception of what is easy and difficult varies greatly between individuals, it is important that the complexity of a learning task be seen to match the individual's level of expertise. Csikszentmihalyi[6] has developed a theory of optimal experience that he refers to as Flow—or the ability to become engrossed in an activity, enjoying it to the extent that intrinsic motivation is associated with the activity itself. When a well-prepared individual is just beginning to learn within a content area, the tasks are fairly simple and match the available skills. If the tasks were significantly more complex, or the skills significantly less developed, the individual would suffer anxiety and a loss of intrinsic motivation. However learning is by definition, a dynamic process. As the learner's skill level improves, new learning tasks must be sought to match the increasing level of expertise, otherwise boredom will result, with a decrease in interest and motivation. The instructional problem, then, is to monitor the student, provide an adequate series of increasingly complex instructional experiences, allow for enactive mastery to occur, provide multiple, similar models, and deliver acceptable verbal persuasion (*being* either expert, prominent or trustworthy). All this has to occur within appropriate media, using appropriate methods, and with the intent of not only transferring content, but increasing self-efficacy belief both concerning the subject area, and in general.

Methods

The evaluation method uses three basic instruments. The first instrument is the self-efficacy evaluation form. It consists of a series of instructions and ten engineering "problems" relevant to the course being evaluated, four of which represent pre-requisite skills that will appear on the diagnostic exam discussed below. The remaining six problems derive from the course content, four of which will appear on the final exam. The last question probes the student's general self-efficacy belief relative to anticipated performance in Strength of Materials and similar courses. Each problem, along with the last question, is accompanied by a rating scale from 0 to 10. The individual is instructed to respond, not by solving the problem, but by indicating personal beliefs about his or her ability to solve the problem in an open-book test situation. The instructions state that a rating of zero would indicate that the student has no idea of how to solve the problem, while a rating of ten would mean absolute certainty in the ability to solve the problem correctly. Other ratings would indicate the degree of certainty/uncertainty with which the student would attempt to solve the problem. Students were told to think of the self-efficacy scale as a percentage of credit that they would expect to get for the problem if they were to attempt to solved it on an exam, for example 9 = 90% (this created some problems which will be addressed in the discussion section). For future self-efficacy evaluations, students will be acclimatized to the self-efficacy scale by relating it to the physical ability to jump a given distance[7], or how many times out of ten do they think they would succeed. This may help them to develop a more accurate rating of their abilities, particularly with entry-level skills.

Students are informed that the problems on the self-efficacy evaluation are randomly selected, and appear in no particular order. The instructions explain that some of the problems deal with understanding that will be developed in the course (that they are not expected to be familiar with) while other problems deal with pre-requisite skills that they are expected to have mastered (they are given no other clues concerning which problems belong in which categories). They are also told that the evaluation will not have any effect on their grade, and that the instructor will not review the evaluations until after the course grades have been turned in (I held to this commitment not to review the self-efficacy evaluations until the end of the semester for 1998 classes, but this basically doubles the time delay for the evolution of methods and will not be used in subsequent courses). They are then encouraged to be as honest as possible in their self-rating.

The second instrument is a diagnostic exam. The diagnostic exam consists of 8 problems testing ability with prerequisite skills only. Students are instructed that it will have no bearing on their grade, but that the items on the diagnostic exam will need to be cleared before the student will be allowed to proceed. These are short problems taking 5 to 20 minutes each to solve. They are graded on a scale from 0 to 10, and discussed individually with each student. The student then has the opportunity to review additional material and make corrections, before the exam is placed in the student's file. The third instrument is a fairly standard final, containing four problems and requiring two to three hours to complete.

During the first day of class, the syllabus was discussed, after which the self-efficacy evaluation was introduced and administered as a pre-test. The diagnostic exam was then made available in the testing center and completed within 2 days. On the last day of class, the same self-efficacy evaluation was reviewed and administered again, this time as a post-test. This was followed by the administration of the final, within a period of one week. Each problem on the final was graded on a scale of 0 to 75, converted to an equivalent score from 0 to 10, and the course grades were assigned. A simple matrix was then constructed to aid in the computation and comparison of the four scores on the three instruments. In particularly, the following questions were examined:

- What was the change in self-efficacy belief between the pre-test and the post-test (both with specific problems, and in general with the final question)?
- How accurately were students able to predict (with the pre-test) their performance on the four problems that were also included on the diagnostic exam?
- How accurately were students able to predict (with the post-test) their performance on the four final problems?

The first of the two measures of self-efficacy change is the general expression of overall self-efficacy belief indicated by the student's answer to the last question on the self-efficacy evaluation form: "How would you rate your ability to learn the material contained in this course and similar courses?" The second expression of self-efficacy could be determined from the average change in self-efficacy belief relative to the learning objectives that were contained in the course content (excluding the problems that were addressing pre-requisite skills). Because it was anticipated that self-report might under or over estimate the actual achievement levels, and that the degree of under or over estimation might also be expected to change over the course of the semester in significant ways, the diagnostic and final exams could be used to establish a general trend in the student's perception of his or her own capabilities. In other words, the pre and post-test self-efficacy evaluations indicate a level of self-efficacy belief, and the correlation with actual problems on the diagnostic and final exams, gives us a sense of how accurate the student's reported self-efficacy beliefs were, and how the accuracy of this self-perception changed during the course of the semester.

This is a potentially important issue. It is critical that students be willing to extend their effort to achieve mastery of learning objectives, but it is also important that the objectives they tackle be within their zone of proximal development[8]. If a student's self-efficacy belief is unrealistically high, he or she will be encouraged to attempt learning tasks for which they are not prepared to succeed. The subsequent, inevitable failure can do significant damage to self-efficacy belief.

Therefore, we want self-efficacy belief to be high enough so that students will extend appropriate effort, but realistic enough to keep students focused on appropriate learning experiences and being met with anticipated success at anticipated levels of exertion. A reasonable threshold for self-efficacy belief above actual ability, is something that may be developed with this or similar methods, over an extended period of time. However, the breadth of a student's zone of proximal development, may be course specific (as well as individually specific) and generalization may be difficult to support.

Growth in self-efficacy was determined as the post-test average less the pre-test average, considering only those items that were contained within the course objectives (problems 1,2,4,5,8, and 9). In addition to the self-efficacy evaluations, each student completed a learning log, which was available for triangulation purposes. Since the Strength of Materials course was student-paced, the time spent on individual learning objectives was also tracked. Neither of these features will be discussed here.

Results

The class originally consisted of 12 individuals. One student dropped half way through the course. A second student was called to active military duty, and completed the course by correspondence. A third student had to return to Alaska toward the end of the semester, and so was unable to complete the post-test in a timely manner. Consequently, the data for these three individuals are incomplete. However, the results sited contain all available data, including partial data. While I realize this would be an unacceptable research design, I am not testing any hypotheses here, and want only to determine whether the method is feasible, what kinds of results I might expect, and how the evaluation method might be improved. In a way, then, this should be considered an initial try-out[9], and therefore the small group of learners is appropriate.

With two exceptions, all of the students reported self-efficacy beliefs that grossly overestimated their ability to solve the problems on the diagnostic exam. Several of these students had performed marginally in the past, and their stated self-efficacy on individual problems had little reference to reality. Overestimates of ability on the diagnostic ranged from 117 to 300 percent. The average over estimation was 168.5%. As stated earlier, this can have a dangerous effect on self-efficacy belief over the near term, if the student is allowed to fail. Individuals with particularly high over estimates of self-efficacy should be flagged as requiring special help to "gently" gain a more realistic view of their own competencies.

There were two exceptions to this general trend of over estimating self-efficacy to use prerequisite skills. They were both non-traditional students with a fair amount of engineering related work experience, and underestimated their performance on the diagnostic by 75% and 81%. These scores were low enough to indicate that special help might be needed to bolster lagging self-efficacy belief. By the end of the course, the students indicated self-efficacy beliefs that were much closer to their actual performance. The worst overestimate was 148%. A third of the class stated a self-efficacy rate that corresponded with their actual performance within 4%, and the class average was an overestimate of 101% (it could be argued that this is too low, as will be discussed in the following section).

In terms of general self-efficacy belief, no one indicated a rate less than 8 on the pre-test (which would correspond roughly to a B letter grade, or higher). The average on the pre-test was 8.7, and the average on the post-test was 8.3, which would indicate a small decrease in self-efficacy belief. One individual, who performed very marginally throughout the semester initially rated his self-efficacy belief at 10, and revised this on the post-test to 7. One individual increased his efficacy rating from 8 to 9 between the pre and post-tests, one individual decreased from 8 to 7 and two individuals decreased from 9 to 8. Everyone else remained constant at 9. Self-efficacy growth ranged from a high of 8.5 to a low of 2.7 (average 4.5). Of course, these numbers don't mean much, without a larger database for comparison, but they appear to be rather low. Some of the reasons that might partially account for this will be discussed in the following section.

Discussion

Students aren't always the best judge of what they know[10]. Coming to an understanding of what you know, what you don't know, and how much practice you need to gain proficiency, is part of the process of developing self-regulatory competence. Most of the students came to the course with a fairly poor understanding of what they knew, and left with a much more accurate picture. This may have been due to the extensive amount of testing required in the course (this was a PSI course, and some students spent as much as 60 hours combined on the diagnostic exam, all of the learning objective exams, the module exams, and the final exam). Winfrey and Weeks[11], noted in their study of self-modeling, that students who received self-modeling (here, they viewed videotapes of themselves performing on the balance beam) developed a much more accurate understanding of the level of their own performance. The extensive amount of testing may have acted similar to self-modeling in this respect.

The problem with the change in general efficacy belief, as indicated on the final question of the self-efficacy evaluation may have been a function of student familiarity with skewed distributions. The 0 to 10 scale was designed for a normal distribution, but the grading scale (that students are perhaps most familiar with), is skewed with a C cut-off typically at 70%. Perhaps even worse, a C is significantly lower than the average grade. Although the self-efficacy scale ranged from 0 to10, by relating it to test scores, I effectively shortened the general scale to a range of from 7 to10, because no one would be taking the course if they expected a grade less than C. In fact, it might even be worse than this. Since none of the students rated their final self-efficacy at 10 (including the students who received A letter grades), the actual scale might range from 7 to 9. This is probably too small, and the range of reported values should be changed (perhaps ranging from 0 to 100).

While I would not like to think that my students are actually suffering a loss of general self-efficacy belief in my course, this might well be the case. Because the course was taught using PSI methods, the workload was very high. Students in my PSI courses almost routinely comment on the excessive amount of work required to master all of the objectives listed in the syllabus. For reasons mentioned earlier, this might well affect a decrease in self-efficacy belief. Paulo Freire[12] has long commented on the "banking" nature of formal education (instructors making and students mildly accepting deposits). This banking analogy might also extend to student motivation, except that instead of making deposits, we are making motivational withdrawals (which the students mildly permit). Students come to formal education excited by the prospect of learning new things, and buoyed by the possibilities. By the time we're through with them in college, many will vow never to return (or worse yet, never to return to learning). This may well be due to the slow attrition of motivational reserves indicated by the rating of general self-efficacy belief on the final question of the post-test.

Another possible reasons for the lower-than-expected growth in self-efficacy on individual problems, is the initial self-efficacy belief expressed on the pre-test relative to those skills that were to be learned in the course. The first problem on the self-efficacy evaluation concerned the stress/strain relationship for a uniaxially applied load, which they were not expected to solve with their entry-level skills. The problem asked them to calculate the final length of a survey chain under a given tension. However, most of the students had taken the surveying course the previous semester, and rated this problem higher than 0. Only four students rated their self-efficacy to solve this problem at 2 or less, while the average on this question for the remaining students was 8.25. Two additional problems dealt with the plotting of shear force and bending moment diagrams (which was not considered an entry-level skill) and the calculation of shear and bending stress on a simply supported beam. The majority of students, however, rated their self-efficacy to solve these problems at a level of 7 or above, indicating that they may have seen the problems in a previous course (possibly not at our school, since several of these individuals were transfer students).

It is, of course, also possible that some students do already know many of the learning objectives in Strength of Materials. If this were the case, I would not be the first one to "teach" students something they already know[13]. However, if this condition is demonstrated by diagnostic exams and self-efficacy evaluations over a period of years, then we need to change the curriculum. Alternately, the way that these problems were presented, as simply supported beams, may have looked very familiar to the majority of students (from Statics) and so they felt that they would probably be able to solve them. Five students rated their ability to calculate uniaxial stress at 5 or above, perhaps because the force was fairly easy to solve for, or perhaps because they had already been introduced to the topic of axial stress. The majority of the students gave a non-zero rating to the majority of the problems associated with non-prerequisite skills. This may be due to their long familiarity with the concept of partial credit, and may possibly be corrected by doing a better job acclimatizing them to the self-efficacy scale. Finally, the initial, general self-efficacy (as indicated by the answer to the last question on the self-efficacy evaluation) may have been high enough to create a ceiling effect.

Obviously, there are some difficulties with some of the problems on the self-efficacy evaluation, and these problems need to be validated. The four problems noted above need to be re-worded or dropped. Also, graphics should be picked that are not quite so reminiscent of the problems from Statics

How to Use the Method

Once the problems on the self-efficacy evaluation have been validated for a particular course, then a database could be developed to compare self-efficacy growth patterns. For one thing, it would have been very helpful at the beginning of the course, to know which students were rating their ability as high, to solve problems that they shouldn't have been able to do. If they were actually determined to have an ability to solve these problem, then instruction here could be skipped. If they weren't, then they could be flagged as individuals who need a little more external direction. This, of course, was impossible since I announced that I wouldn't review the pre or post-tests until after assigning grades, and I will avoid this in the future.

Conclusions

Since we aren't going to totally eliminate many core courses, the main focus of instructional evaluation should obviously be formative. Sadly, much of the Level 1, Reaction evaluation being done is really an evaluation of *instructors*, rather than instruction, with obvious summative overtones. At Boise State's College of Engineering, we are attempting to separate the two functions and improve the reliability of both. The evaluation of instruction at the university level has generally settled on SPT (Level 1) assessments, with some Peer evaluation, which is really just Level 1 evaluations by peers, rather than by students (it might similarly be argued that most of our industrial advisory boards are really used for Level 1 evaluation, as well). The reasons are fairly obvious. Level 1 evaluations produce a cheap, quantitative picture of how an instructor or a unit of instruction is perceived. If asked to chose between a two or three digit number and a page of qualitative analysis, most administrators will hang their hats on the number every time. The method proposed in this paper might present a similar temptation for summative misuse, being used by administrators or instructors to play a summative game of one-ups-manship, with a method really designed for formative decision. However, as a formative method, I found it revealing and helpful. Finally, while self-efficacy is not an A—K outcome, it might well become trackable by this method, and might become an objective for engineering programs that consider it significant.

Bibliographic Information

1Bandura, A., (1997). Self-efficacy: The exercise of control, New York: W. H. Freeman and Company.

2Bandura, A., (1986). Social foundations of thought and action: A social cognitive theory, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

3Kirkpatrick, D. L., (1996). Evaluating training programs: The four levels, San Francisco: Berrett-Koehler Publishers

4Kaufman, R., (1997). A strategic-planing framework: Mega planning, in Kaufman, R., Thiagarajan, S., & MacGillis, P (Eds.). The guidebook for performance improvement: Working with individuals and organizations, San Francisco: Jossey-Bass.

5Schunk, D. H., & Zimmerman, B. J., (1997). Social origins of self-regulatory competence, Educational Psychologist, 32, 195-208.

6Csikszentmihalyi, M., (1991). Flow: The psychology of optimal experience, New York: Harper Perennial.

7Schunk, D. H., & Hanson A. R. (1987). Self-modeling and cognitive skill learning. New York: American Psychological Association. (ERIC Document Reproduction Service No. ED 292 847)

8Vygotsky, L. (1986/1934). Thought and language, Cambridge, Massachusetts: The MIT Press.

9Dick, W., & Carey, L., (1996). The systematic design of instruction, 4th edition, New York: HarperCollins College Publishers.

10Kinzie, M. B., (1990). Requirements and benefits of effective interactive instruction: Learner control, self-regulation and continuing motivation, Educational Technology Research and Development, 38, 5-21.

11Winfrey, M. L. & Weeks, D. L. (1993). Effects of self-modeling on self-efficacy and balance beam performance, Perceptual and motor skills, 77, 907-913.

12Freire, P. (1996/1972). Pedagogy of the oppressed, New York: Continuum Publishing Company.

13Mager, R. R., & Pipe, P., (1997). Analyzing performance problems: Or you really oughta wanna, Atlanta, Georgia: The Center for Effective Performance, Inc.

Biographical Information

DAVID HAWS received his undergraduate Civil Engineering degree from the University of Utah, with Master's and Doctoral Degrees in Civil Engineering from Brigham Young University. He has also received an undergraduate degree in English from the University of California at Berkeley, and is completing a second Master's degree in Instructional and Performance Technology at Boise State. He has taught full-time for the past 5 years, and has primary interests in individualized instruction, motivation and developing self-regulatory competence in his students.