

Evaluating the Effectiveness of Learning Interventions: An Information Systems Case Study

Daniel L. Moody

Department of Software Engineering
Charles University
Prague, Czech Republic
moody@ksint.ms.mff.cuni.cz

and

School of Business Systems,
Monash University
Melbourne, Australia
dmoody@infotech.monash.edu.au

Guttorm Sindre

Department of Computer and Information Science
Norwegian University of Science and Technology
Trondheim, Norway
guttors@idi.ntnu.no

and

Department of Management Science
and Information Systems
University of Auckland, New Zealand
g.sindre@auckland.ac.nz

Abstract

Currently, there is no standard instrument for evaluating learning effectiveness. While final examinations and end-of-semester course evaluation surveys can be used to do this, they are not designed for this purpose, and there are inherent problems using them in this way. This paper describes a survey instrument, called the Learning Effectiveness Survey, which can be used to evaluate and improve the effectiveness of learning interventions. Learning effectiveness is evaluated in the context of the learning goals of the course (short term learning), and in the context of the overall educational programme and future working life (long term learning). The instrument also provides feedback on the intervention and how it could be improved. A case study is described in which the instrument is used to evaluate the use of peer reviews as a learning activity in a requirements analysis course. The instrument was found to have relatively high validity, but reliability was below acceptable levels. Some interesting results were also found on the determinants of learning. In particular, attitude was found to have no effect on short term learning, but was found to be the primary determinant of long term learning.

Keywords

Quality assurance, requirements analysis, IS education, measurement, evaluation,

1. Introduction

1.1 Evaluating the Effectiveness of Learning Interventions

The question of how to evaluate the effectiveness of learning interventions is a problematic one. That is, how do we determine whether a change to a course has been successful in improving student learning? In most cases, it is left to judgement as to whether the change was effective or not. However such judgements are susceptible to cognitive biases, such as selective observation—the teacher may seek out evidence that confirms what they want to believe, while ignoring or downgrading contrary evidence (Neuman, 2000). Clearly, there is a need for more systematic way of evaluating learning effectiveness.

1.2 Performance Based Assessments of Learning

Theoretically, the best way to evaluate learning effectiveness is to measure improvement on achievement tests (Cashin, 1995). In a university context, this means measuring changes in performance on final examinations. There are two alternative ways of doing this:

- Between-years (longitudinal) comparison: comparison in achievement between one year and the next.
- Within-year (two group) comparison: comparison between two randomly selected groups within the one year.

The first approach represents a quasi-experimental design. However it is difficult to make any sensible comparison between exam results from year to year, because of the number of potential confounding variables:

- There will be differences in student characteristics which may provide an alternative explanation for any differences (*selection bias*).
- It is usually not practicable to use the same exam from year to year, so differences in the exam itself may represent a possible confounding variable (*instrumentation*).
- The graders are likely to be (at least partly) different from year to year—differences in results may be therefore attributable to the leniency/harshness of the graders.
- In most undergraduate courses, there will be a tendency to normalize grades, which will tend to obscure any real differences in performance.
- There may be learning effects as a result of giving the same course another time (or boredom effects of giving it too many times!).

The second approach is a true experimental design, and is the only way to show a causal link between the intervention and student learning. However there are a number of practical problems applying such a research design in a university context:

- There is likely to be problems of diffusion of treatments, as it is difficult to isolate groups from each other.
- There is likely to be problems of selection bias, as it is difficult to randomly assign students to groups.
- If groups are run at different times, in different locations or using different instructors, these all represent potential confounding variables.
- It raises issues of equity and fairness, as a result of the fact that students in one of the groups may receive an unfair advantage.

Perception Based Assessments of Learning

The other alternative for measuring learning effectiveness is via student perceptions. In a university context, this is typically done using end of semester course evaluation surveys. Such surveys have become pervasive in higher education, and are increasingly promoted as the principal avenue for collecting information for assessing teaching effectiveness of university staff (Snare, 2000). Research shows that these are the most widely used source of information for evaluating teaching effectiveness (Seldin, 1993).

However the collection of such information tends to be primarily for the purpose of producing management information, performance assessment and making tenure/promotion decisions (Cashin, 1990). The extent to which it is used by (or useful to) staff to improve the learning process is questionable (Comrie & Lim, 2001). Traditional course evaluation instruments have been criticised as being poor measures of teaching effectiveness as there is often little, if any, connection between changes in teaching and the ensuing ratings (e.g. Hinton, 1993; Langbein, 1994; Wilson, 1998; Wiese, Seymour & Hunter, 1999; Seymour, Weise, Hunter & Daffinrud, 2000; Snare, 2000). For the purpose of process improvement, traditional student evaluations provide little useful feedback as to how to enhance student learning (Seldin, 1993; Snare, 2000).

One problem with traditional course evaluation instruments is that they are based on a “student-as-consumer” model. They focus on what students liked or didn’t like about the course, rather than how well learning goals were achieved. Another problem is that they tend to take a “one size fits all” approach. A standard form is generally used to enable comparison between different courses and different teachers. As a result, the questions are not adaptable to the particular course being evaluated or teaching methods used (Cashin, 1990). This makes the instrument bureaucratically convenient but less useful for evaluation and process improvement. In general, the effectiveness of any educational programme can only be sensibly evaluated in terms of its learning goals (Bloom, 1984; Gagne, Briggs & Wager, 1992).

1.3 Objectives of this Paper

The objective of this paper is to develop an instrument to evaluate the effectiveness of learning interventions. The paper is structured as follows:

- Section 2 reviews instruments for evaluating learning effectiveness previously published in the literature.
- Section 3 defines the theoretical basis for the instrument.
- Section 4 defines the survey instrument. This includes definition of an underlying theoretical model and development of survey items to operationalise the model.
- Section 5 describes a case study where the survey instrument was used to evaluate the introduction of quality reviews in a requirements analysis course.
- Section 6 summarises the findings, contributions and further research.

2. Previous Research

2.1 Overview

Surprisingly, we were not able to find a standard instrument in the literature for evaluating learning effectiveness. Literature searches and web searches revealed very few standard instruments of any kind for course evaluation. It would appear that most institutions develop their own course evaluation instruments to suit their own purposes. We were only able to find two previous instruments specifically developed to measure learning effectiveness.

2.2 Student Assessment of Learning Gains

The Student Assessment of Learning Gains (SALG) instrument was developed to address the limitations of traditional course evaluation surveys (Seymour et al, 2000). The items in the instrument were empirically derived based on analysis of qualitative responses by students. The instrument consists of a set of suggested questions which can be customised to the needs of a particular course. Some limited empirical testing of the instrument is reported, but this is largely anecdotal, in the form of comments by “satisfied users” (Seymour et al, 2000). However there is no formal analysis of reliability or validity, and only positive feedback is reported.

2.3 Student Opinion Survey of the Learning Process

Another instrument, called the Student Opinion Survey of the Learning Process, has also been proposed to address the limitations of existing course evaluation instruments (Snare, 2000). Unlike SALG, this consists of a set of standard questions that can be applied across different courses. No empirical validation of the instrument is reported, although it has been used in the host institution for a number of years.

2.4 Conclusion

Both of the instruments reviewed have serious weaknesses in terms of survey design. In particular, both define a set of *ad hoc* survey items which do not measure any underlying theoretical constructs. This makes evaluation of validity and reliability and interpretation of results problematic (Cashin, 1995).

3. Theoretical Foundations

3.1 Learning Goals

In developing any educational programme, it is important to first define *learning goals*. These can be used to help guide the selection of teaching methods and learning activities that are most appropriate to achieve these goals (Bloom, 1984; Gagne et al, 1992). The effectiveness of any educational programme can only be sensibly assessed in the context of its learning goals. We define learning goals as “particular *knowledge, skills* or *attitudes* that participants should have at the end of the learning episode”. We distinguish between three different types of learning goals:

- Knowledge: “what facts and concepts participants should understand”.
- Skills: “what tasks participants should be able to perform”.
- Attitudes: “what attitudes, beliefs and motivation participants should possess”.

3.2 Short Term vs Long Term Learning

Learning is an ongoing process. University courses are not undertaken in isolation, but in the context of some larger educational programme (e.g. a university degree or diploma) and in preparation for working life. While learning can be evaluated within the context of the objectives of a particular course, it should also be evaluated in the wider context. We therefore distinguish between the following concepts:

- Short term learning (internal validity): was the course successful in achieving its stated learning goals? This relates to the effectiveness of the course as a standalone unit of education.
- Long term learning (external validity): did the course contribute to the student’s overall learning experience? This addresses the issue of *relevance*—a course may be effective in achieving its learning goals, but the learning goals themselves may be of little long term value.

3.3 Evaluation vs Improvement

One of the desired outcomes of any evaluation process is improvement. In this context, it is desirable that the results of the evaluation can be used to make changes to the learning process in order to improve learning effectiveness. In general, quantitative items (numerical scale based questions) are most useful for evaluation purposes, while qualitative items (open ended questions) are most useful for improvement (Seldin, 1993; Cashin, 1995).

4. Development of the Instrument

This section describes an instrument for evaluating learning effectiveness called the Learning Effectiveness Survey. Given the problems in using examinations to evaluate learning effectiveness (Section 1), we decided to use a *perception based* approach—to ask students to evaluate how much they learned. While performance based assessment is preferable on theoretical grounds (i.e. to demonstrate a “real” improvement in learning as opposed to a perceived improvement), we concluded that a perception based approach was the only workable approach in a university context. Student perceptions of their learning have been found to be highly correlated with scores on achievement tests (Cohen, 1981; 1986; Marsh, 1987; Feldman, 1989; Seymour et al, 2000).

4.1 Theoretical Model

Unlike instruments previously proposed in the literature, the Learning Effectiveness Survey is based on an explicit theoretical model of the learning process. This is summarised in Figure 1:

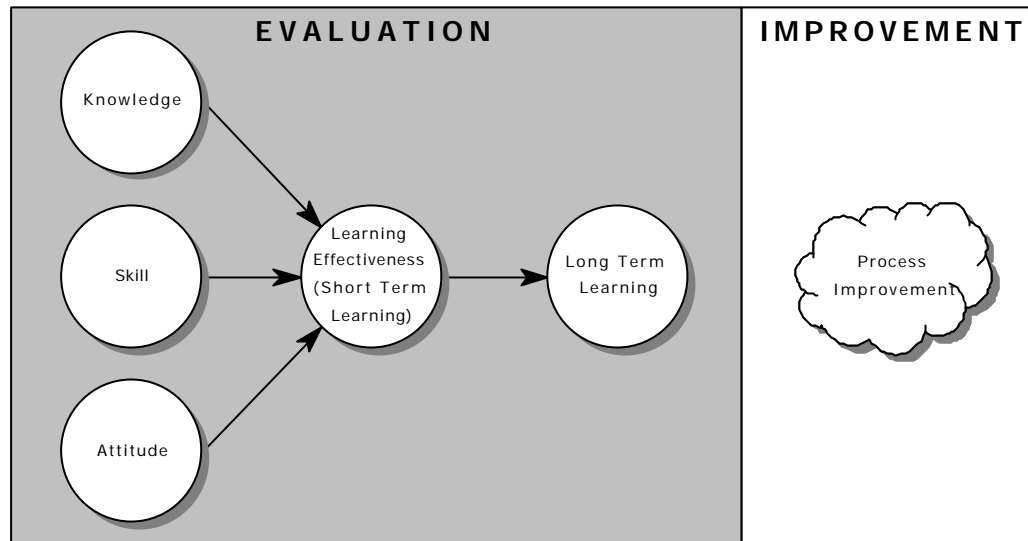


Figure 1. Underlying Theoretical Model

In the diagram:

- Circles represent theoretical constructs (*latent variables*).
- Arrows represent causal relationships between them (*laws of interaction*).
- Process improvement is shown as a cloud to indicate that it is a qualitative construct (and therefore “soft” and fluffy).

The definitions of the constructs are:

- Learning Effectiveness (Short Term Learning): what was the overall effect of the intervention on student learning in the course?
- Knowledge: what was the effect of the intervention on increasing knowledge?
- Skills: what was the effect of the intervention on improving skills?
- Attitude: what was the effect of the intervention on changing attitudes?
- Long Term Learning: what was the effect of the intervention beyond the scope of the course itself (on future courses and future working life)?
- Process Improvement: how could the intervention be modified to more effectively achieve learning outcomes?

The following causal relationships are hypothesised between the constructs:

- Learning Effectiveness will be determined by gains in Knowledge, Skill and Attitude. This reflects the fact that learning effectiveness is defined by how well the learning goals were achieved.
- Long Term Learning will be determined by Learning Effectiveness (Short Term Learning). This reflects the assumption that how effectively people learn during the course will determine their perceptions of the usefulness of the learning beyond the scope of the course.

4.2 Operationalisation of the Model

To operationalise the theoretical model, survey items need to be developed to measure each of the constructs. Multiple items are used to measure each construct, following Churchill’s

(1979) recommendation to use a minimum of two indicators for latent variables. However the theoretical model can only be fully operationalised in the context of a particular intervention and specific learning goals.

Learning Goals

Items used to measure Knowledge, Skill and Attitude must be developed based on the specific learning goals defined for the course. A single item is used to measure each learning goal. In addition, two standard items are defined for Attitude, which evaluate the effect of the intervention on participants' motivation and their enjoyment of the course.

Learning Effectiveness

Learning Effectiveness is measured by a combination of standard items and intervention-specific items. There are two standard items:

- Contribution to Learning: How much did the «*learning activity*» contribute to your learning in the course?
- Relative Effectiveness: How effective was your learning from the «*learning activity*» compared to other learning activities in the course?

Additional intervention-specific items may be developed to evaluate the value of specific aspects of the intervention to learning.

Long Term Learning

Long Term Learning is measured by three standard items which evaluate the contribution to learning beyond the scope of the course itself: to future courses, to practical work and future working life:

- Future Courses: Do you think your experience in the «*learning activity*» will improve your performance in future courses?
- Practical Work: Do you think your experience in the «*learning activity*» will improve your performance in future practical/project work?
- Working Life: Do you think your experience in the «*learning activity*» will improve your performance in future working life?

Process Improvement

Process improvement items must be developed based on the specific intervention being evaluated, and should collect information about how to improve the effectiveness of the intervention. This will generally consist of a combination of closed and open questions.

The resulting survey instrument consists of six parts, corresponding to the six constructs in the theoretical model. The first five parts relate to *evaluation*, and result in numeric scores for each construct. The final part relates to *improvement*, and results in qualitative data about how the intervention could be improved.

5. Case Study: Empirical Testing Of The Instrument

5.1 The Intervention

In this study, peer review was introduced into an undergraduate requirements analysis course as a way of giving students an appreciation of the role of quality assurance in systems development. Quality assurance of deliverables is an important part of the software development process (van Vliet, 2000), but is rarely included in undergraduate courses. In this course, students were required to evaluate each others' work as a learning exercise. Peer review also provides a way of achieving deeper understanding of the subject matter. According to Bloom's (1984) taxonomy of learning objectives, evaluation activities represent the highest level of thinking. The Learning Effectiveness Survey was used to evaluate the effect of this exercise on student learning[†].

In this course, students were required to submit three major assignments: a data model, a process model and a requirements specification. In addition, they were also required to submit six quality reviews of other students' data and process models. They were assessed for all assignments on a pass-fail basis. Each participant was required to evaluate three different models, and each model was evaluated by three different reviewers. Multiple reviewers were used to enable analysis of the reliability of the evaluation process. Models were randomly assigned to reviewers, and reviews were anonymous (reviewers did not know whose models they were evaluating, and modellers did not know who their reviewers were). Each participant was assigned cases which were different to each other and to the one they had modelled themselves (to avoid possible bias). The review exercises began in the week following completion of each of the modelling exercises and participants were allowed one week to complete their allotted reviews. Results of their reviews were recorded using a web-based evaluation system.

5.2 Operationalisation of the Model

Learning Goals (Knowledge, Skill, Attitude) Items

The learning goals of the course were defined as follows:

- Knowledge:
 - K1: Understand the concepts of the modelling languages taught
 - K2: Understand the concepts of the conceptual modelling quality framework
- Skills:
 - S1: Be able to use the modelling languages to develop conceptual models
 - S2: Be able to interpret conceptual models
 - S3: Be able to evaluate the quality of conceptual models
- Attitudes:
 - A1: Understand the importance of quality assurance in conceptual modelling
 - A2: Understand the importance of conceptual modelling in systems development

[†] Authors' note: Our original intention was simply to use a standard instrument to evaluate the educational effect of the intervention, as the authors' expertise lies in the conceptual modelling field, not education. However in the (somewhat surprising) absence of such an instrument, we were forced to develop our own. The development of the instrument was thus a purely serendipitous outcome of the educational intervention.

These were used to develop items for the Knowledge, Skill and Attitude constructs.

Learning Effectiveness Items

The two standard items plus two intervention-specific items were used to measure Learning Effectiveness. The intervention specific items evaluated the separate effects of reviewing vs being reviewed.

Process Improvement Items

A number of questions were formulated to obtain feedback about how the intervention could be improved. There were five closed questions relating to how the review process was conducted and an open ended question which asked for suggestions as to how the process could be improved.

Figure 2 summarises the operationalisation of the evaluation model for the case study. In the diagram:

- Circles represent theoretical constructs (*latent variables*)
- Rectangles represent survey items (*observed variables*) used to measure the underlying theoretical constructs.
- Dotted lines represent *measurement relationships*.

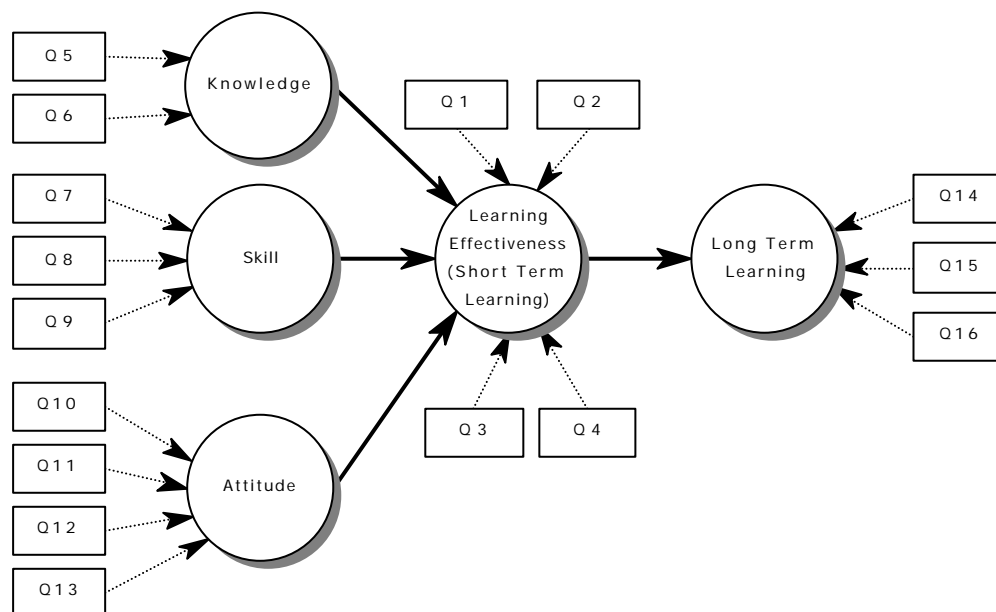


Figure 2. Operationalisation of Evaluation Model for Case Study

5.3 Validation of the Measurement Instrument

Construct Validity

Factor analysis is the preferred technique among researchers for evaluating construct validity. However in this case, the sample size was too small (61), so *inter-item correlation analysis*

was carried out instead. All items showed positive results as a result of this analysis, suggesting that the items are valid measures of the underlying constructs.

Reliability

Reliability analysis was conducted on the survey items used to measure each construct. As shown in Table 1, two of the constructs (Learning Effectiveness and Long Term Learning) had high levels of reliability ($> .7$), while all of the constructs associated with the learning goals had lower than acceptable levels.

Table 1. Item Reliabilities

CONSTRUCT	CRONBACH'S α
Knowledge	.432
Skill	.640
Attitude	.642
Learning Effectiveness	.773
Long Term Learning	.855

This suggests that more care needs to be taken in formulating learning goals clearly. The learning goals for this course were defined at quite a high level, leaving them open to interpretation. Another issue is that perhaps multiple items need to be developed for each learning goal rather than at the level of *goal types* (Knowledge, Skill and Attitude). A particular learning activity may affect different learning goals in different ways, which should not adversely impact reliability. This suggests that Knowledge, Skill and Attitude may be *classifications* of constructs rather than constructs in their own right.

5.4 Overall Results

Table 2 shows the summary statistics for each construct. Overall, students found the review exercises to be moderately effective in improving their knowledge, skills and attitude and their learning in the course, but only between slightly and moderately effective for long term learning. While these results are encouraging, there is clearly room for improvement—this represents a “lukewarm” rather than an enthusiastic endorsement of the intervention.

Table 2. Summary of Construct Values

CONSTRUCT	MEAN	STDEV	RESULT
Skill	3.28	.68	Moderate
Knowledge	2.93	.72	Moderate
Attitude	3.01	.63	Moderate
Learning Effectiveness	3.13	.61	Moderate
Long Term Learning	2.54	.85	Slight-Moderate

The reason why the responses weren't more positive can be explained by a number of contextual factors. Firstly, the course was a compulsory unit in the degree, and was relatively unpopular with students (presumably because of its non-technical nature). Research shows that students are more likely to give higher ratings when they have a strong interest in the subject matter or when they are taking the course as an elective (Marsh, 1987; Cashin, 1995). In addition, adding extra assessment to a course (in this case, six additional items of assessment) is never likely to be popular with students, so would have lowered their responses regardless of the learning value of the exercises.

5.5 Analysis of Individual Survey Responses

Table 3 summarises the responses to the individual survey items (from most positive to least positive). The mean responses to all items except one (Q14) were in the Moderate range (2.5—3.5). The most positive responses were for enthusiasm for the course, understanding the quality framework and relative effectiveness compared to alternative learning activities. The least positive responses were the value of the exercise in preparation for future courses and working life and learning from the review feedback. This suggests that:

- Students did not see a great benefit from the review exercise beyond the course itself. This suggests that more effort should have been spent explaining to students the significance of quality assurance in software engineering practice.
- Students found the process of reviewing others' models more useful than the process of being reviewed. This is partly because many of the reviews were poorly done and because of negative expertise differentials between reviewers and reviewees in many cases.

Table 3. Summary of Item Responses

ITEM	CONSTRUCT	MEAN	SD
Q10: Enthusiasm for the course	Attitude	3.36	.89
Q6: Understanding of quality framework	Knowledge	3.33	.85
Q4: Relative effectiveness of review exercise	Learning Effectiveness	3.33	1.04
Q9: Ability to evaluate quality of conceptual models	Skill	3.30	.82
Q11: Enjoyment of the course	Attitude	3.32	.71
Q5: Understanding of modelling languages	Knowledge	3.23	.86
Q12: Awareness of importance of quality assurance in conceptual modelling	Attitude	3.17	1.03
Q2: Learning from reviewing others' models	Learning Effectiveness	2.91	.89
Q8: Ability to interpret conceptual models	Skill	2.86	.88
Q7: Ability to develop quality models	Skill	2.85	.77
Q1: Overall learning from review exercises	Learning Effectiveness	2.85	.89
Q13: Awareness of importance of modelling in IS development	Attitude	2.67	.91
Q15: Preparation for project work	Long term Learning	2.65	.98
Q3: Learning from reviews by others	Learning Effectiveness	2.64	.94
Q16: Preparation for working life	Long term Learning	2.56	.96
Q14: Preparation for future courses	Long term Learning	2.39	.96

5.6 Process Improvement

Table 4 summarises the responses to the closed process improvement questions. Overall, participants wanted reviews to remain anonymous, but would have liked to collaborate with other reviewers and have the ability to respond to reviewer comments.

Qualitative responses to the open-ended question were also analysed. The most common suggestions for improving the review process were:

- Improve the web-based evaluation system (12%): this is perhaps a predictable response from IS students.
- Requirements to pass the review exercises should be stricter (9%): This comment typically came from students who were dissatisfied because they got poor feedback on their own models. This was partly because reviews were assessed on a pass-fail basis, so many students did the minimum work required to meet the requirements.
- Should have reviewed the same case as we modelled (8%): Students were given other cases to review than the one they had modelled to avoid possible bias, but this would clearly reduce their ability to conduct a thorough review.

- Should have had more iterations (4%), i.e. modelling, getting reviews, improving the model based on reviews, getting reviewed again. This is a useful suggestion, and this would be closer to how the process takes place in practice.
- Lectures should have been more relevant (4%), from students who felt that the lectures were not closely enough related to the review task. Again, this is a fair criticism and emphasises that in adding a new learning activity to a course, it is important to make sure that it is fully integrated into the curriculum, rather than simply added on at the end.

Table 4. Responses to Process Questions

QUESTION	YES%	SIG.	RESULT
As a Reviewer			
Q17: Know identity of reviewee	5%	.000	N O
Q18: Ability to correspond with reviewee	58%	(.268)	Undecided
Q19: Collaboration with other reviewers	67%	.010	Y E S
As a Reviewee			
Q20: Know identity of reviewer	20%	.000	N O
Q22: Ability to respond to reviewer	74%	.000	Y E S

5.7 Determinants of Learning

A number of causal relationships were hypothesised between the constructs in the theoretical model:

Knowledge + Skill + Attitude → Learning Effectiveness

Learning Effectiveness → Long Term Learning

These represent hypotheses about the determinants of learning. We used regression analysis to evaluate the validity of these relationships.

Determinants of Short Term Learning

In this analysis, Knowledge, Skill and Attitude were used as independent variables and Learning Effectiveness (Short Term Learning) as the dependent variable. The results showed that both Knowledge and Skill had a significant effect on Learning Effectiveness while Attitude did not.

Relationship Between Short and Long Term Learning

In this analysis, Learning Effectiveness (short term learning) was used as the independent variable and Long Term Learning as the dependent variable. This relationship was confirmed with $\alpha < .01$.

Determinants of Long Term Learning

To test for possible effects of learning goals on Long Term Learning, a regression analysis was also carried out using Knowledge, Skill and Attitude as independent variables and Long Term Learning as the dependent variable. The regression was found to be highly significant ($\alpha < .01$) but with higher predictive power ($r^2 = .4$ compared to $.25$) than the previous analysis. Further analysis showed that Skill and Attitude have a significant effect on Long Term

Learning while Knowledge does not. This may reflect the fact that knowledge is often forgotten soon after the final exam, whereas skills and attitudes are more likely to be retained. Surprisingly, Attitude had the strongest effect of all learning goals, whereas it had a non-significant effect on short term learning.

The results of this analysis leads to ambiguity in determining Long Term Learning—the second analysis showed that it is determined by Short Term Learning, while the third analysis showed that it is jointly determined by Skill and Attitude. To resolve this ambiguity, a regression analysis was carried out with Long Term Learning as the dependent variable and all other variables as independent variables (this may seem like a “brute force” approach but is defensible given the exploratory nature of the research). The results of this analysis showed that Attitude was the only variable which had a significant effect on Long Term Learning. This is a surprising result, and suggests that motivating students plays a much more important role in whether knowledge is retained and used beyond the course than originally thought.

Model Revision

It is common practice to drop any variable or relationship from the model that is found to be non-significant (Pedhazur & Schmelkin, 1991). This means that the relationships between Attitude and Learning Effectiveness and between Learning Effectiveness and Long Term Learning should be removed. In addition, the relationship between Attitude and Long Term Learning which was discovered through exploratory analysis, should be added to the model. The revised model is shown in Figure 3. This gives a very different picture of the causal relationships between constructs compared to the model originally proposed in Figure 1. It suggests that different types of learning objectives have different impacts on short term and long term learning. In particular, it suggests that attention to attitude goals is needed for long term retention and application of knowledge. However these are only preliminary findings and need to be verified in different courses and using different sample populations.

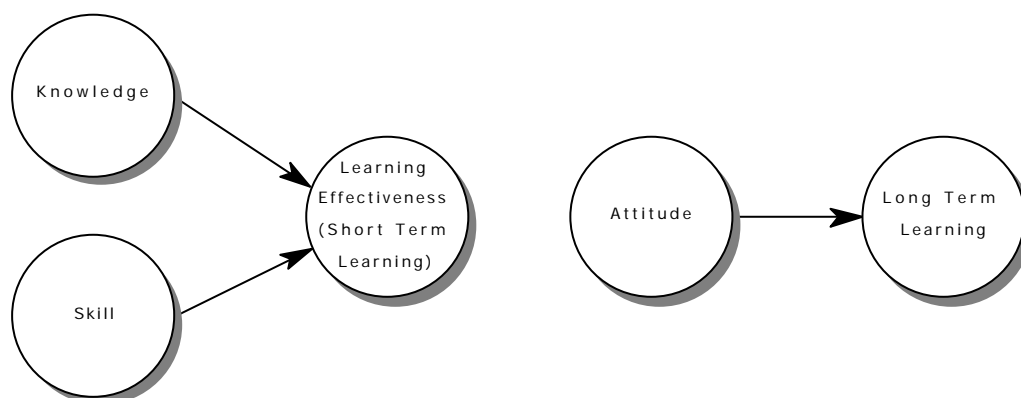


Figure 3. Revised Theoretical Model

6. Conclusion

6.1 Summary

This paper has described an instrument for evaluating and improving the effectiveness of learning interventions. It can be adapted to evaluate learning interventions of any kind, and in any domain. The instrument could be used for *formative assessment* while the intervention is in progress, or for *summative assessment* (as in the case study described in this paper) to evaluate its value at the end of the learning episode. The paper describes the first empirical test of the instrument, which showed that the instrument had relatively high validity, but reliability was below acceptable levels. As well as providing feedback on the specific intervention applied, it also resulted in some more general findings about the determinants of learning. We believe the instrument provides the basis for developing a standard instrument for measuring learning effectiveness where currently none exists.

6.2 Strengths of the Instrument

The proposed instrument provides a number of advantages over traditional course evaluation instruments:

- It is based on an explicit theoretical model of the learning process rather than *ad hoc* development of survey items.
- It is tailored to the specific learning goals of the course rather than taking a “one size fits approach.”
- It addresses the issue of evaluation as well as process improvement. A combination of quantitative and qualitative data is collected to facilitate this.

6.3 Weaknesses of the Instrument

The major limitations of the instrument are:

- It relies on perception-based data. This represents a threat to validity in that students may *think* they have learned effectively when they have not. However student perceptions of their learning have been found to be highly correlated with scores on achievement tests (Cohen, 1981; 1986; Marsh, 1987; Feldman, 1989; Seymour et al, 2000).
- Reliability was found to be low for three of the five constructs in the model.
- It does not consist of a standard set of questions, but must be tailored to the learning goals of the course and the intervention being evaluated. This makes it less bureaucratically convenient than traditional course evaluation instruments, as it is more difficult to make comparisons across courses. It also requires more effort on the part of teachers to explicitly define learning goals and incorporate them into the instrument.

6.4 Further Research

Based on the results of this first empirical study, the instrument needs to be refined and subjected to further testing. A clear direction for future research is to improve the reliability of the measurement instrument. However the survey could also be enhanced to collect more qualitative data to improve its diagnostic power. In particular, the survey could be expanded to allow students to write comments for each item rather than restricting them to one “catch-

all” item at the end. This would serve to increase the specificity of written remarks and their utility for process improvement (Seldin, 1993).

References

1. BLOOM, B. (1984): *Taxonomy of Educational Objectives*, Longman, New York.
2. CASHIN, W.E. (1990): *Student Ratings of Teaching: Recommendations for Use*, Idea Paper No. 22, Centre for Faculty Education and Development, Kansas State University.
3. CASHIN, W.E. (1995): *Student Ratings of Teaching: The Research Revisited*, Idea Paper No. 32, Centre for Faculty Education and Development, Kansas State University.
4. CHURCHILL, G.A. (1979): “A Paradigm for Developing Better Measures of Marketing Constructs”, *Journal of Marketing Research*, 16, 1, pp. 64-73, February.
5. COHEN, P.A. (1981): “Student Ratings of Instruction and Student Achievement: A Meta-Analysis of Multi-Section Validity Studies”, *Review of Educational Research*, 51, pp. 281-309.
6. COHEN, P.A. (1986): “Un Updated and Expanded Meta-Analysis of Multisection Student Rating Validity Studies”, *Annual Meeting of the American Educational Research Association*, San Francisco, USA, April.
7. COMRIE, A. and LIM, H. (2001): “Opening Pandora's Box? Are we really listening to what students have to say about their learning experience?”, *Staff and Educational Development Association (SEDA) Conference*, University of Glasgow, Glasgow, Scotland, 2 - 3 April.
8. FELDMAN, K.A. (1989): “The Association Between Student Ratings of Specific Instructional Dimensions and Student Achievement: Refining and Extending the Synthesis of Data from Multisection Validity Studies”, *Research in Higher Education*, 30, pp. 583-645.
9. GAGNE, R.M., BRIGGS, L.J. and WAGER, W.W. (1992): *Principles of instructional design*, Harcourt Brace Jovanovich, Fort Worth.
10. HINTON, H. (1993): “Reliability and Validity of Student Evaluations: Testing Models versus Survey Research Models”, *Political Science and Politics*, 26, September, pp. 562-69.
11. LANGBEIN, L. (1994): “The Validity of Student Evaluations of Teaching”, *Political Science and Politics*, 27, 3, pp. 545-53.
12. MARSH, H.W. (1987): *Students' Evaluations of University Teaching: Research Findings, Methodological Issues and Directions for Future Research*, Pergamon Press, Elmsford, NY.
13. NEUMAN, W.L. (2000): *Social Research Methods - Qualitative and Quantitative Approaches* (4th edition), Allyn and Bacon, Needham Heights, MA.
14. PEDHAZUR, E.J. and SCHMELKIN, L.P. (1991): *Measurement, Design and Analysis: An Integrated Approach*, Lawrence Erlbaum Associates Inc, Hillsdale, NJ.
15. SELDIN, P. (1993): “How Colleges Evaluate Professors 1983 vs 1993”, *American Association for Higher Education (AAHE) Bulletin*, 46, 2.
16. SEYMOUR, E., WEISE, D.J., HUNTER, A.-B. and DAFFINRUD, S.M. (2000): “Using Real-World Questions to Promote Active Learning”, *Proceedings of the National Meeting of the American Chemical Society Symposium*, San Francisco, March 27.
17. SNARE, C.E. (2000): “An Alternative End-of-Semester Questionnaire”, *Political Science Online*, December.
18. VAN VLIET, H. (2000): *Software Engineering: Principles and Practice* (2nd edition), John Wiley & Sons.

19. WIESE, D., SEYMOUR, E. and HUNTER, A.-B. (1999): Report on a panel testing of the student assessment of their learning gains instrument by instructors using modular methods to teach undergraduate chemistry, Bureau of Sociological Research, University of Colorado.
20. WILSON, R. (1998): "New Research Casts Doubt on Value of Student Evaluations of Professors", *The Chronicle of Higher Education*, pp. A12-A14, January 16.