TITLE PAGE

Title

The Learning Effectiveness Survey (LES): An Instrument for Evaluating and Improving the Effectiveness of Learning Interventions

Type

Research Paper

Topic Area

Curriculum Research and Development

Keywords

Measurement, evaluation, learning effectiveness, learning outcomes

Author Names

Daniel L. Moody, Guttorm Sindre (GS will present the paper, so communication directly related to the conference should therefore be directed to him)

Mailing Address

Until May 2003: Department of Management Science and Information Systems, University of Auckland, Old Choral Hall, 7 Symonds St, Private Bag 92019, Auckland, NEW ZEALAND

After May 2003: Department of Computer and Information Science, Norwegian University of Science and Technology (NTNU), N-7491 Trondheim, NORWAY

Email Address

dmoody@idi.ntnu.no; g.sindre@auckland.ac.nz

Phone Numbers

Business: +64 9 373 7599 ext 2778
Mobile: +64 21 038 7118

Fax Number

(47) +64 9 373 7430

The Learning Effectiveness Survey (LES): An Instrument for Evaluating and Improving the Effectiveness of Learning Interventions

Daniel L. Moody[1,2]
Guttorm Sindre[3,4]

Abstract. Currently, there is no standard instrument for evaluating the effectiveness of learning interventions. This makes it difficult to assess whether a change to teaching methods has been successful or not. 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 in using them in this way. This paper describes a survey instrument, called the Learning Effectiveness Survey (LES), 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). Unlike instruments previously proposed in the literature, the instrument was developed based on an explicit theoretical model of the learning process rather than ad hoc development of survey items. A case study is described where the instrument is used to evaluate the use of peer review as a learning activity in an Information Systems course. The instrument was found to have relatively high validity, but reliability was below acceptable levels for three of the five constructs. Some interesting results were also found on the determinants of short-term and long-term learning. In particular, attitude was found to have no significant effect on short term learning, but was found to be the primary determinant of long-term learning.

Keywords: education, measurement, evaluation, knowledge, skill, attitude, learning effectiveness, learning objectives, learning outcomes

1. INTRODUCTION

The question of how to evaluate the effectiveness of learning interventions is a problematic one. 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 educator may seek out evidence that confirms what they want to believe, while ignoring contrary evidence or downgrading its importance (Neuman 2000). This suggests a need for more systematic way of evaluating learning effectiveness.

---

1 Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia
2 School of Business Systems, Monash University, Melbourne, Australia 3800
3 Department of Computer and Information Science, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
4 Department of Management Science and Information Systems, University of Auckland, New Zealand

Theoretically, the best way to evaluate learning effectiveness is to measure improvement in performance on achievement tests (Cashin 1995), e.g., comparing performance on final examinations. However there are theoretical and practical problems in using examinations to measure the effectiveness of a particular intervention. Comparison in achievement between one year and the next means that the student groups and exam questions must be different (or, if the same exam is used, the last students will gain an advantage just from that), and besides, the teacher may have improved simply by having one extra year of experience. Hence there are several possible causes for any improvement, beside the intervention itself. Comparison in achievement between two randomly selected groups within the year is, on the other hand, a true experimental design. But again, there are problems. For instance, it is hard to isolate the students group completely from each other, and there may be other confounding variables such as teaching times or locales that explain the differences. Moreover, should it indeed be the case that one treatment gave better learning than the other, there are also issues of fairness – as some students may get more knowledge and better grades than others simply because they were lucky to get the best treatment. Finally, the need to run two parallel streams of instruction doubles the teaching resources required, which raises practical issues of increased costs and increased teaching loads.

The alternative to performance-based assessments of learning is to ask students to assess the effectiveness of their own learning—this represents a perception-based assessment. The usual way of doing this is using an end-of-semester course evaluation survey. Such instruments 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). However traditional course evaluation instruments have been criticised as being poor measures of learning effectiveness (e.g. Hinton 1993; Langbein 1994; Seymour et al. 2000; Snare 2000; Wiese et al. 1999; Wilson 1998):

- They are designed to serve a whole range of purposes and are not specifically focused on measuring learning effectiveness.
- They tend to evaluate the *teacher* and how well the students liked the course rather than the methods used or the learning outcomes (Cashin 1995; Trout 1997; Wiese et al. 1999).
- They take a "one size fits all" approach. A standard form is generally used to enable comparison between different courses and different teachers. This is bureaucratically convenient but less useful for diagnosis and improvement. An evaluation instrument should be adaptable to the specific learning goals of the course, as different learning objectives will require different teaching methods (Cashin 1990; Gagne 1965).
- Most of them lack an explicit theoretical base (Hoyt et al. 1999), consisting of individual items developed in an *ad hoc* manner. Because different items measure different underlying constructs, this makes results difficult to interpret except at the level of individual items. Ideally, a theoretical model should be developed first and items generated based on this (Dubin 1978; Pedhazur et al. 1991).
- There are very few standard instruments for course evaluation. Most institutions develop their own course evaluation instruments to suit their own purposes.

The objective of this paper is to develop an instrument to evaluate the effectiveness of learning interventions, which addresses the limitations of existing course evaluation instruments. The paper is structured as follows: Section 2 develops the theoretical basis for the instrument, drawing on education and knowledge management literature. Section 3 describes the instrument in its general form. This includes definition of the underlying theoretical model and development of a number of standard survey items. Section 4 describes a case study where the survey instrument was used to evaluate the introduction of peer review in an IS design course. Section 5 summarises the findings, contributions and further research.

2. THEORETICAL FOUNDATIONS

Learning Effectiveness

In developing any educational programme, it is important to define explicit *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). We argue that the effectiveness of any educational unit can only be sensibly assessed in the context of its learning goals. We define *learning effectiveness* as "the extent to which the learning goals of the course were achieved".

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". Knowledge goals involve memorisation and comprehension abilities (Bloom 1984).
- Skills: "what tasks participants should be able to perform". Skill goals concern how knowledge can be used to solve problems and involve application, analysis, synthesis and evaluation abilities (Bloom 1984).
- Attitudes: "what attitudes, beliefs and motivation participants should possess".

The classification of learning goals into knowledge, skills and attitudes is widely used in practice, but its theoretical foundations are unclear. However it can be understood in the context of two separate but related areas of theory: education and knowledge management (see Table 1).

Table 1. Mapping of Learning Goals

LEARNING GOALS	Learning Domains (Bloom 1984)	Types Of Knowledge (Rescher 1977)
KNOWLEDGE	Cognitive domain (recall or recognition of knowledge)	Knowledge that (theoretical knowledge)

LEARNING GOALS	Learning Domains (Bloom 1984)	Types Of Knowledge (Rescher 1977)
SKILL	Cognitive domain (intellectual abilities or skills) Psychomotor domain (physical skills)	Knowledge how (practical knowledge)
ATTITUDE	Affective domain	

In one of the most influential works on learning goals, Bloom (1984) classifies learning as occurring in three separate domains:

- Cognitive: these are objectives that deal with the recall or recognition of knowledge and the development of intellectual abilities or skills.
- Affective: changes in interest, attitude and values that a participant will experience at the end of the learning episode.
- Psychomotor: perceptual abilities, physical abilities, movements and non-discursive communication. This is relevant in areas where physical skills are required—for example, physical education, performance arts, surgery etc.

Bloom's cognitive domain encompasses both Knowledge and (intellectual) Skills. The affective domain corresponds to Attitudes, while the psychomotor domain corresponds to (physical) Skills. The distinction between Skill and Knowledge (which are combined together in Bloom's classification) can be understood in the distinction between "knowledge that" and "knowledge how" from the knowledge management field. According to Rescher (1977), human knowledge consists of two types:

- "Knowledge that": this defines statements or assertions about the world. This is also called *theoretical* or *factual* knowledge (Cohen 1984; Cohen et al. 1980; Polanyi 1967; Sveiby 1997).
- "Knowledge how": this defines ways of doing things. This is also called *practical* or *procedural* knowledge (Cohen 1984; Cohen et al. 1980; Polanyi 1967; Sveiby 1997).

Short-Term Learning vs. Long-Term Learning

Learning is an ongoing process. Usually 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. 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 positively to the student's overall learning experience? This addresses the issue of *relevance*—for example, a course may be effective in achieving its learning goals, but the learning goals may be themselves of little relevance to practice (e.g. if it teaches outdated or discredited techniques).

Evaluation vs. Improvement

One of the desired objectives of any evaluation process is improvement. In the context of educational evaluation, we would like to be able to evaluate the effectiveness of a course of education as well as improve it. The distinction between these terms is:

- Evaluation: measuring how effective the learning intervention has been.
- Improvement: modifying the learning intervention to improve its 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 (Cashin 1995; Seldin 1993).

3. LEARNING EFFECTIVENESS SURVEY INSTRUMENT

The evaluation instrument proposed in this paper is called the Learning Effectiveness Survey (LES). This section defines the instrument in its general form. However the instrument can only be fully operationalised in the context of a particular intervention and learning goals.

Theoretical Model

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

- 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.

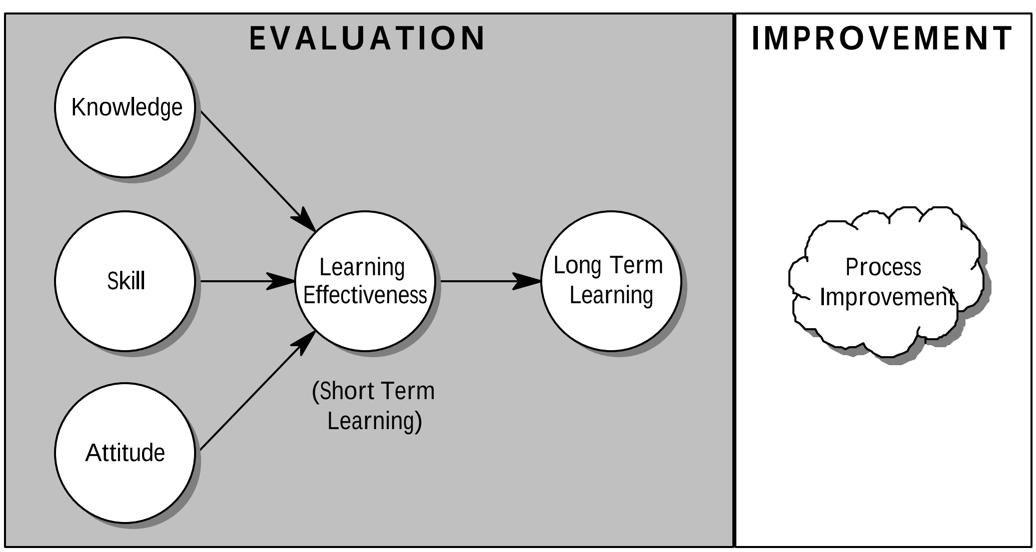

Figure 1. Theoretical Model

The definitions of the constructs are:

- Learning Effectiveness (or Short-Term Learning): how effective overall was the intervention in improving student learning in the course?
- Knowledge: how effective was the intervention in increasing knowledge?
- Skills: how effective was the intervention in improving skills?
- Attitude: how effective was the intervention in changing attitudes?
- Long-Term Learning: how valuable was the intervention in preparing students for future courses and future working life?
- Process Improvement: how could the intervention be improved 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 within a course (short term learning) is defined by how well learning goals were achieved.
- Long-Term Learning will be determined by Learning Effectiveness (Short-Term Learning). We argue that perceptions of the usefulness of the learning beyond the scope of the course will be determined by how effectively people learn during the course.

Operationalisation of the Model

To operationalise the model, survey items need to be developed to measure each of the theoretical constructs shown in Figure 1. The general operationalisation of the model is shown in Figure 2. In the diagram:

- Circles represent theoretical constructs (*latent variables*)
- Rectangles represent survey items (*observed variables*), which are used to measure the underlying theoretical constructs. These define *empirical indicators* for each construct (Dubin 1978). Multiple items are used to measure each construct, following Churchill's (1979) recommendation to use a minimum of two indicators for latent variables.
- Dotted lines represent *measurement relationships* (relationships between latent variables and their empirical indicators).

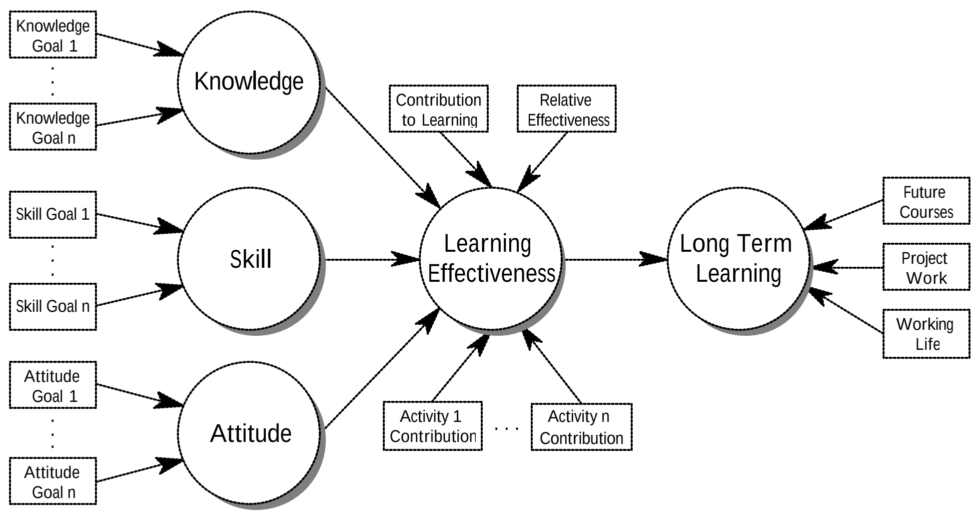

Figure 2. General Operationalisation of the LES Model

Items used to measure Knowledge, Skill and Attitude are course-specific, and must be developed based on the specific learning goals of the course. A single item is used to measure each learning goal.

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 this 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 separate contribution of specific components of the intervention to learning, where the intervention consists of multiple component activities. These items will be of the form:

- Activity i Contribution: How much did you learn from «component i» of the «learning activity»?

Clearly, this is a compromise solution: evaluating the contribution of each activity of the intervention to each learning goal would lead to a combinatorial explosion in the number of questions, which would make the survey much more complex and onerous to complete

Long-Term Learning is measured by three standard items. These 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 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 may consist of a combination of closed and open questions.

Survey Instrument

The resulting survey instrument consists of six parts, each of which corresponds to one of the 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. This gives the survey diagnostic power as well as evaluation power: the quantitative responses provide the basis for evaluating the effectiveness of the learning intervention, while the qualitative responses provide the basis for determining *why* the intervention was successful and *how* it could be improved. Note that the underlying theoretical model can only be fully operationalised in the context of a particular intervention and specific learning goals, so is not a "one size fits all" instrument. The next section describes the operationalisation of the model for a particular learning intervention (peer review) in an Information Systems course. This was the first empirical test of the instrument.

4. CASE STUDY: EMPIRICAL TESTING OF THE INSTRUMENT

Peer Review as a Learning Activity

In this study, peer review was introduced in the exercise part of the course SIF8035 Information Systems at the Norwegian University of Science and Technology. The course is taken by third year IT students pursuing the degree Master of Technology. Peer review of work products is an important part of the software development process, but is rarely included in Information Systems courses. Typically students are required to submit assignments to be evaluated by teaching staff. In this course, students were required to evaluate each other's work as a learning activity. The instrument described in the previous section was used to evaluate the effectiveness of this activity for improving 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 reviews of other students' data and process models, for which they were assessed on a pass-fail basis. Reviews were conducted using a quality framework defined by Lindland et al (1994), and the study also acted as a test of the reliability and validity of this framework (Moody et al. 2002). The review exercises began in the week following completion of each of the modelling exercises. Each participant was required to evaluate three different models, and each model was evaluated by three different reviewers. This enabled analysis of the reliability of the evaluation process. Participants were allowed one week to complete the evaluation task. 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 that were different to each other, and to the one they had modelled themselves (to avoid possible bias on the part of reviewers). The participants recorded the results of their reviews using a web-based evaluation system.

At the end of the course, after conducting six reviews of data and process models, students were asked to complete a 22-item survey about the learning effectiveness of the peer reviews and how it could be improved. The survey was based on the instrument described in Section 4. There were 66 respondents completing this questionnaire.

Operationalisation of the Model

As discussed earlier, the evaluation instrument can only be fully operationalised in the context of the specific intervention and the learning goals of the course being evaluated. Now we show how this was done in the case of the peer review intervention of SIF8035 Information Systems. Two intervention-specific items were defined for Learning Effectiveness, which evaluated the effect of reviewing vs. being reviewed. These represent two different activities involved in the review process, and would thus instantiate the slots "Activity i contribution" in Figure 4:

Q3. How much did you learn from reviewing other people's models?
Q4. How much did you learn from the reviews you received from other people?

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 modelling in the systems development process

These learning goals were used to develop items for the Knowledge, Skill and Attitude constructs as shown in Figure 4—one item was used to measure each learning goal. In addition, two general items were defined to measure Attitude, which evaluated the effect of the peer review exercise on participants' enthusiasm / motivation and their enjoyment of the course.

Q12. How much did the review exercises contribute to your enthusiasm and motivation for the course?
Q13. Did the review exercises contribute to your enjoyment of the course?

There were also some questions for process improvement, consisting of 5 yes/no questions and one open question asking for improvement ideas. The process improvement part is considered so specific to the particular course and intervention that they will not be discussed further in this paper.

Validation of the Measurement Instrument

To evaluate the results for the underlying theoretical constructs, it is necessary first to evaluate the validity and reliability of their empirical indicators. This tests the *measurement part* of the model. The sample size was too small for factor analysis, so *inter-item correlation analysis* was carried out instead to establish construct validity. All items showed positive results, with convergent validity > discriminant validity, cf. Campbell and Fiske (1959), suggesting that the items are valid measures of the underlying constructs.

Reliability analysis was conducted on the survey items used to measure each construct. In the literature, alphas of 0.7 or above are considered to be acceptable (Nunally 1978), with alphas as low as 0.5 considered acceptable in some circumstances (Caplan et al. 1984). As shown in Table 2, 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 2. Item Reliabilities

CONSTRUCT	CRONBACH'S a
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 to ensure that they are clearly defined. The learning goals for this course were defined at quite a high level, leaving them open to interpretation and possible measurement error.

Evaluation of Learning Effectiveness

Table 3 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 response. If we take 3 as the break-even point, the only item which is significantly greater than 3 is Skill ($\alpha < .05$).

Table 3. 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 response wasn't more positive can be explained by a number of environmental factors. Firstly, the course was a compulsory unit in the degree. Students are more likely to give high ratings when they have a strong interest in the subject matter, taking the course as an elective (Cashin 1995; Marsh 1987). In fact, student motivation has a higher correlation with student ratings than any other variable (Cashin 1990). This would also explain the lower value found for long-term learning, as many students felt that the content of the course had little relevance to their future work. Finally, 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.

Determinants of Learning

A number of causal relationships were hypothesised between the constructs in the theoretical model, as shown in Figure 3:

- Knowledge + Skill + Attitude → Learning Effectiveness
- Learning Effectiveness → Long-Term Learning

Both these were confirmed by regression analysis. The first with $\alpha < .01$, the adjusted r^2 statistic showing that together the learning goals accounted for 44% of the variance in Learning Effectiveness. The second with $\alpha < .01$, the r^2 statistic showing that Learning Effectiveness accounted for almost 25% of the variance in Long Term Learning.

While the hypotheses were thus confirmed, and the r^2 values are statistically good, we were still suspicious that there was such a high percentage of the learning not accounted for by the assumed construct. Hence we made exploratory investigations of the individual effects of Knowledge, Skill and Attitude on both Learning Effectiveness and Long-Term Learning. This showed that Attitude was the only variable which had a significant effect on Long Term Learning, while Knowledge and Skill (and not Attitude) had a significant effect on Learning Effectiveness (which again did not have a significant effect on Long-Term Learning compared to Attitude). This yields the revised model is shown in Figure 3. 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 – although it is too early to say if this result is only valid for the particular course or more generally.

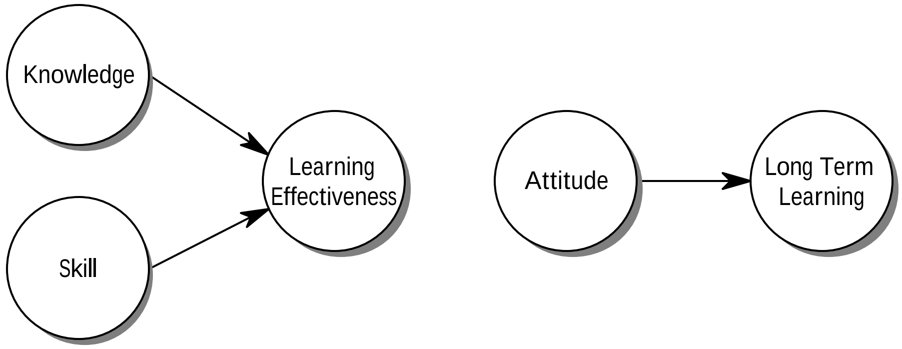

Figure 3. Revised Theoretical Model

To test this revised model, two regression analyses were carried out. First, a regression analysis was carried out using Knowledge and Skill as independent variables and Learning Effectiveness as the dependent variable (i.e. removing Attitude from the original analysis). When Attitude is removed as a determinant of Learning Effectiveness, the r^2 is reduced to .43, which is a loss of only 1% in predictive power (compared to Equation 1). However it also simplifies the model (thus increasing its parsimony), and the relationships between variables are now all highly significant. Secondly, a regression analysis was carried out using Attitude as the independent variable and Long Term Learning as the dependent variable. The regression was found to be highly significant, with an r^2 value of .32. This means that all relationships in the revised model are strongly confirmed.

5. CONCLUSION

This paper has described an instrument, called the Learning Effectiveness Survey (LES), for evaluating and improving the effectiveness of learning interventions. Unlike instruments previously proposed in the literature, the instrument was developed based on an explicit theoretical model of the learning process rather than *ad hoc* development of survey items. It can be adapted to evaluate learning interventions of any kind, and is applicable to educational programmes in any discipline. More research is needed, but it may provide the foundation for developing a standard instrument for measuring learning effectiveness.

A weakness is of course that the instrument relies on self-reported data. However while student ratings are often dismissed by teachers (Felder 1992), the research evidence shows that they are relatively valid against a variety of indicators of effective teaching (Cashin 1995; Cohen 1990). Hence, the instrument can be used to supplement other evaluation instruments such as final exams (which are designed to evaluate performance of students) and course evaluation surveys (which are designed to evaluate performance of teachers). Student ratings are only one source of data about teaching effectiveness, and should be used in combination with other sources of data (Cashin 1995). There are many aspects of teaching effectiveness that students are not qualified to judge (e.g. grading practices, currency of course content) (Hoyt et al. 1999).

Our empirical test of the instrument showed that it had relatively high validity, but reliability was below acceptable levels for three of the five constructs. As well as providing feedback on the intervention applied (the use of peer review), it also resulted in some more general findings about the determinants of short term and long-term learning. The empirical study also provided valuable feedback for improving the instrument. Based on these results, the instrument needs to be refined and subjected to further empirical testing in a variety of settings. We expect that the instrument will evolve significantly as a result of continuous testing and subsequent refinement.

REFERENCES

1. BLOOM, B. (1984): *Taxonomy of Educational Objectives* Longman, New York.
2. CAMPBELL, D.T., and FISKE, D.W. (1959): "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," *Psychological Bulletin* **56**, pp 81-105.
3. CAMPBELL, D.T., and STANLEY, J.C. (1966): *Experimental and Quasi-Experimental Designs for Research* Houghton Mifflin College.
4. CAPLAN, R.D., NAIDU, R.K., and TRIPATHI, R.C. (1984): "Coping and Defense: Constellations vs. Components," *Journal of Health and Social Behaviour* **25**, pp 303-320.
5. CASHIN, W.E. (1990): *Student Ratings of Teaching: Recommendations for Use* Idea Paper No. 22, Centre for Faculty Education and Development, Kansas State University.
6. CASHIN, W.E. (1995): *Student Ratings of Teaching: The Research Revisited* Idea Paper No. 32, Centre for Faculty Education and Development, Kansas State University.
7. CHURCHILL, G.A. (1979): "A Paradigm for Developing Better Measures of Marketing Constructs," *Journal of Marketing Research* **16**:1, February, pp 64-73.
8. COHEN, N.J. (1984): "Preserved Learning Capacity in Amnesia: Evidence for Multiple Memory Systems," in: *Neuropsychology of Memory*, L. Squire and N. Butters (eds.), Guilford Press, New York.
9. COHEN, N.J., and SQUIRE, L.R. (1980): "Preserved Learning and Retention of Pattern Analysing Skill in Amnesia Using Perceptual Learning," *Cortex* **17**, pp 273-278.
10. 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.
11. 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.
12. COHEN, P.A. (1990): "Bring Research Into Practice," in: *Student Ratings of Instruction: Issues for Improving Practice*, M. Theall and J. Franklin (eds.), Jossey-Bass, San Francisco.
13. DUBIN, R. (1978): *Theory Building (revised edition)* The Free Press, New York.
14. FELDER, R.M. (1992): "What Do They Know, Anyway?," *Chemical Engineering Education* **26**:3, Summer, pp 134-135.

15. 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.

16. GAGNE, R.M. (1965): *The Conditions of Learning* Holt, Rinehart and Winston, Inc., New York.

17. GAGNE, R.M., BRIGGS, L.J., and WAGER, W.W. (1992): *Principles of instructional design* Harcourt Brace Jovanovich, Fort Worth.

18. HINTON, H. (1993): "Reliability and Validity of Student Evaluations: Testing Models versus Survey Research Models," *Political Science and Politics* **26**:September, pp 562-569.

19. HOYT, D.P., and PALLETT, W.H. (1999): *Appraising Teaching Effectiveness: Beyond Student Ratings* Idea Paper No. 36, Centre for Faculty Education and Development, Kansas State University.

20. LANGBEIN, L. (1994): "The Validity of Student Evaluations of Teaching," *Political Science and Politics* **27**:3, pp 545-553.

21. LINDLAND, O.I., SINDRE, G., and SØLVBERG, A. (1994): "Understanding Quality In Conceptual Modelling," *IEEE Software* **11**:2, March, pp 42-49.

22. MARSH, H.W. (1987): *Students' Evaluations of University Teaching: Research Findings, Methodological Issues and Directions for Future Research* Pergamon Press, Elmsford, NY.

23. MILLER, R.I. (1987): *Evaluating Faculty for Promotion and Tenure* Jossey-Bass, San Francisco.

24. MOODY, D.L., SINDRE, G., BRASETHVIK, T., and SØLVBERG, A. (2002): "Evaluating the Quality of Process Models: Empirical Analysis of a Quality Framework," 21st International Conference on Conceptual Modeling (ER'2002), Tampere, Finland.

25. NEUMAN, W.L. (2000): *Social Research Methods - Qualitative and Quantitative Approaches (4th edition)* Allyn and Bacon, Needham Heights, MA.

26. NUNALLY, J. (1978): *Psychometric Theory (2nd edition)* McGraw Hill, New York.

27. PEDHAZUR, E.J., and SCHMELKIN, L.P. (1991): *Measurement, Design and Analysis: An Integrated Approach* Lawrence Erlbaum Associates Inc, Hillsdale, NJ.

28. POLANYI, M. (1967): *The Tacit Dimension* Routledge & Kegan-Paul, London.

29. RESCHER, N. (1977): *Methodological Pragmatism: Systems-Theoretic Approach to the Theory of Knowledge* Basil Blackwell, Oxford.

30. SELDIN, P. (1993): "How Colleges Evaluate Professors 1983 vs 1993," *American Association for Higher Education (AAHE) Bulletin* **46**:2.

31. 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.

32. SNARE, C.E. (2000): "An Alternative End-of-Semester Questionnaire," *Political Science Online*:December.

33. SVEIBY, K.-E. (1997): *The New Organisational Wealth: Managing and Measuring Knowledge-Based Assets* Berret-Koehler Publishers, San Francisco.

34. TROUT, P. (1997): "How to Improve Your Teaching Evaluation Scores without Improving Your Teaching!," *The Montana Professor* **7**:3, pp 17-22.

35. 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.

36. WILSON, R. (1998): "New Research Casts Doubt on Value of Student Evaluations of Professors," *The Chronicle of Higher Education*, January 16, pp A12-A14.