

## Evaluating business writing training programs: Just how adaptable is the Context Adaptive Model?

Sheryl Ward<sup>1</sup>

### Abstract

The paper outlines the process of evaluating a 3-day technical report writing program undertaken in Hong Kong applying the Context Adaptive Model (CAM) developed by Lynch (1996). It considers both the effectiveness of the training program and the appropriateness of the CAM as a tool for evaluation in an organizational climate where resources are restricted and expectations may differ from those which apply in the educational contexts for which the model was developed. It is concluded that, overall, the training program was effective in meeting stakeholders' needs but that the CAM in its current form, while providing a useful framework for investigating what goes on in a business training program, has some serious limitations which restrict its applicability in a context of this kind.

### 1. Introduction

Despite the fact that many organisations spend large sums on improving their staff's skills through short, intensive training programs, evaluation of such programs is often limited to the use of end of program questionnaires that elicit only the participants' subjective reactions.<sup>2</sup> This failure to evaluate beyond the level of participant reactions is somewhat surprising given that the main purpose of such training programs is usually to enhance workplace performance. This paper assesses the effectiveness of using the *Context Adaptive Model* (Lynch, 1996) for evaluating a business writing training program. Although this model is more commonly used for evaluating large-scale language programs it was selected as a framework for evaluating a 3-day training program in report writing skills for senior Chinese engineers to determine if it is as

---

<sup>1</sup>For further information about the research contact  
sherylward@bigpond.com.au

<sup>2</sup> See Brandenburg, (1988) for a review of the extent of use of end-of-program participant questionnaires in training programs.

---

adaptable as its name implies. This paper thus has two main purposes:

- (i) to describe the evaluation of a report-writing training program using Lynch's CAM;
- (ii) to evaluate CAM's effectiveness as an evaluation framework for such programs.

Evaluation is defined as:

*a disciplined inquiry to gather facts and other evidence that allow an evaluator to make assertions about the quality, effectiveness, or value of a program, a set of materials or some other object of the evaluation in order to support decision making (Cummings 1998: 57).*

## 2. Language program evaluation

The focus of language program evaluation over the past four decades has changed dramatically. Lynch summarises this change in the following way:

*The history of program evaluation in applied linguistics can be seen, thus far, as a move away from a concern with tightly controlled experiments focusing on the analysis of product, or student achievement, to a concern for describing and analysing the process of programs as well (Lynch 1996:39).*

From the late 1960s until the early 1980s, the focus of language program evaluation in the United States was on summative, product-oriented evaluations (Keating 1963; Scherer and Wertheimer 1964). Despite severe criticism (by Brumfit, 1980 and others) it was not until the 1980s that other approaches to evaluation such as needs assessment, implementation studies and formative evaluation were considered. Long (1983), in his seminal 'black box' article discussed the need for a clear understanding of what was happening in the program in order to help explain the outcomes. He maintained that it was unwise to assume that the outcomes of a program were necessarily the result of the program itself.

This increasing concern with process led to the exploration of more subjective data gathering methods such as observation and case studies. Because the case study concentrates on distilling an in-depth picture of the situation under study it often uses a range of methods for collecting and analysing data. This utilisation of various data gathering methods encouraged a more pragmatic approach to research to develop which was favoured among others, by Lynch. Lynch's Context Adaptive Model (1996) was partly a response to the rigidity of earlier models that relied only on experimental research designs and statistical techniques for measuring validity. Table 1 summarises the main steps in Lynch's model and the questions that are appropriate at each step.

---

#### Lynch's Context Adaptive Model (CAM)

---

<b>Step 1</b>	Who is requesting the evaluation?
Audience & Goals	Who will be affected by the evaluation? Why is the evaluation being conducted? What information is being requested and why?
<b>Step 2</b>	Availability of comparison groups
Context Inventory	Availability of reliable and valid test measures Availability of evaluation expertise Timing Selection process Characteristics of participants Characteristics of the trainers Characteristics of the program Instructional materials and resources Perspective and purpose of the program Social and political climate.
<b>Step 3</b>	Where should the evaluation begin?
Preliminary Thematic Framework	What aspects of the program should the evaluator investigate in detail?
<b>Step 4</b>	What type of data will be gathered?
Data Collection Design	What will be the best methods for gathering the data?

---

---

...continued

---

Lynch's Context Adaptive Model (CAM)

---

<b>Step 5</b> Data Collection	Have the assumptions of the design and statistical models been met in a quantitative design? Have the procedures for data gathering been portrayed accurately in a qualitative design? Have alternative interpretations of the data been pursued in a qualitative design?
<b>Step 6</b> Data Analysis	
<b>Step 7</b> Evaluation Report	What is the best way of communicating findings of the evaluation honestly and successfully?

---

**Table 1: Context Adaptive Model (adapted from Lynch 1996)**

### 3. The Study

Lynch's CAM was used to evaluate a 3-day technical report writing program conducted in Hong Kong in 1997 for twelve senior Chinese engineers who worked for an international transport planning consultancy. As a well-known and highly respected international transport consultancy, it was important that all reports submitted to clients be of the highest quality. Management was thus understandably concerned when it began to receive complaints from clients regarding the quality of the reports being submitted.

I was commissioned by the management of this company to design and deliver a training program to improve the standard of reports being written by its senior engineers.<sup>3</sup>

My main purposes in evaluating this program were to:

---

<sup>3</sup>While the design and delivery of the program were shared between a former colleague and myself, the design and implementation of the evaluation study were my sole responsibility. My colleague's only part in the evaluation was to act as one of the raters for the tests and to participate in the debriefing sessions at the end of each day's training

- 
- (i) Examine the extent to which the program met the needs of the main stakeholders: management, participants and trainers.
  - (ii) Determine the strengths and limitations of the CAM as an evaluation framework for such programs.

Lynch's model provided the theoretical framework for this evaluation. Each of the seven steps is summarised in the next few pages.

### **3.1. Identify audience and goals (Step 1)**

Lynch suggests that four questions need to be answered at this initial stage of the evaluation process.

*Who is requesting the evaluation?*

*Who will be affected by the evaluation?*

*Why is the evaluation being conducted?*

*What information is being requested and why?*

Although these questions seem relatively straightforward, they were in fact quite difficult to answer. Firstly, the client did not actually request an evaluation. Management seemed to have little interest in evaluating the training program.<sup>4</sup> Their needs were limited to some evidence that would justify the expense of the training. To meet this need the trainers suggested that short written comments on each of the participants be included in a final evaluation report along with the results of the standard end of program questionnaire.

Evaluation of this program was thus primarily trainer-driven. In my dual role as trainer/researcher my purpose was two-fold. As a trainer my purpose was more formative in that I wanted evidence of what worked and what didn't work so that the program could be improved in future. As a researcher my purpose was more academic. I wanted to explore the effectiveness of CAM in a more restricted context. Therefore, to a large extent the trainers were the stakeholders most affected by this evaluation.

---

<sup>4</sup> They seemed to regard us, the language-training consultants, as experts and were prepared to trust our professional judgement.

---

### 3.2 Compile a context inventory (Step 2)

Lynch presents a checklist of eleven potentially relevant dimensions of language education programs. A full list is given in Table 1. However, as not all of these dimensions were relevant to this particular program not all will be discussed below.

#### 3.2.1 *Evaluation expertise*

A variety of evaluation expertise was utilised in this project. Recent academic studies and professional experience in evaluation projects had given me considerable insights into the concepts and methods available in the different fields of program evaluation. In addition, in-house expertise, in the form of a 'subject-expert'<sup>5</sup> was used for confirming the reliability of the pre and posttests. The pre and posttests were thus evaluated using both subject and language experts as raters.

#### 3.2.2. *Participant characteristics*

Management selected the participants for this training program according to need. All twelve participants were native Cantonese speakers from Hong Kong who were engineering graduates from either Hong Kong or English-medium overseas universities. All were fluent English speakers who had some difficulties in communicating in written English. Writing reports was an important part of the job description of each of these engineers.

#### 3.2.3 *Trainer characteristics*

Both of the trainers were qualified teachers of English to adults from non-English speaking backgrounds who have both lived and worked in Hong Kong and so were very familiar with the target group and the cultural context.

---

<sup>5</sup> This 'subject expert' was a senior director of the company and an experienced technical writer who was seen by his colleagues as one of the better writers in the company.

### 3.2.4 *The Program*

The 3-day program in report writing skills was structured around three themes: Clear Focus, Clear Structure and Clear English. The program included presentations, discussions, and writing tasks that reinforced the main teaching points. A program manual was produced, which contained reference material and exercises, and was designed to be used both during and after the program.

### 3.2.5 *Social and political climate*

This program was conducted less than two months after the formal return of Hong Kong to China. In the years prior to the handover attitudes to English were becoming more ambivalent in Hong Kong. However, participants in this program accepted that they needed excellent English writing skills if they wanted to retain their jobs which required them to communicate in English with international clients.

## 3.3 **Establish a preliminary thematic framework (Step 3)**

Lynch argues that establishing a thematic framework at this stage allows the evaluator to focus the evaluation in the face of 'potentially overwhelming' amounts of information. Given the differing goals of the evaluation for the trainers and the client one salient question emerged: *Did the program meet the needs of the various stakeholders?*

However, it is important to recognise that *needs* can be defined in different ways. Hutchinson and Waters (1987), for example, recognise six different types of needs: needs as *necessities*, *lacks*, *wants*, *learning strategies*, *constraints*, and as *language audits*. The three main stakeholders in this program had overlapping views of need. Management recognised that a need existed to improve the standard of the reports being written by these engineers (need as *necessity*). The participants recognised their need to improve their confidence and ability particularly in the area of report structure: writing introductions, conclusions and management summaries (need as *wants*). The trainers recognised gaps in the three areas of report focus, report structure and English expression (need as *lacks*). Research indicates that the stronger the correlations that exist between the varying stakeholders' perceptions of needs (in this case

---

those of management, participants and trainers) the greater degree of success any training program is likely to have.

### 3.4. Select a design and collect data (Steps 4 & 5)

In these steps, decisions are made about what information is needed and the best ways of gathering it to answer the question posed in the thematic framework section; that is, *Did the program meet the needs of the various stakeholders?* The design and data collection methods used to answer this question for each of the three main stakeholders: management, participants and trainers, is briefly outlined.

#### 3.4.1. Management

Management required some evidence that the training program had been effective, that is, that trainees had improved their writing skills. To obtain objective information about the writing skills of the participants a quasi-experimental design using pre-post tests without a control group was used. The pre and posttest consisted of writing a management summary and an introduction to a short report. Participants were given sufficient background data on which to base these two writing tasks. They were asked to complete the pretest at the beginning of the training program and the posttest at the end of day three. The pre and posttests were initially assessed by the two trainers and later by the 'subject expert'.

#### 3.4.2 Trainees

Participant needs and reactions were gathered from two main sources:

- (i) participant needs questionnaire;
- (ii) end of program evaluation questionnaire that included a classroom observation checklist.

Participants were asked to complete a short, structured questionnaire prior to beginning the training program to gauge their perceptions of their main writing needs. This questionnaire used a



---

three-point scale for rating the two variables *importance* and *confidence*.<sup>6</sup>

At the end of the program participants were asked to complete another questionnaire that included an observation checklist adapted from Nunan (1989) in which participants were asked to rate the extent to which they agreed with a series of statements that reflected what went on in the training room.

### 3.4.3 Trainers

Trainer perceptions of the effectiveness of the training program were documented through daily journal entries and daily post-training debriefing sessions. The trainer debriefing sessions took the form of *informal conversational interviews* (Patton 1980) in which the questions and issues arose spontaneously as both trainers relaxed together at the end of each day's training.<sup>7</sup> This method worked well in this situation because of the close working relationship the two trainers had developed from years of working together. The diary entries and audiotaped debriefing accounts were coded according to emergent themes. Both the journal entries and the trainer debriefing sessions focused on issues and concerns thought important by the trainers rather than being comprehensive and descriptive field notes of the whole program.<sup>8</sup>

---

<sup>6</sup> Two main limitations with this initial questionnaire were identified. First, the use of such a short scale meant that responses tended to cluster around the mid-point. Second, the initial perceptions regarding training needs were not followed up at the end of the program. In other words participants were not asked if they felt more confident about writing reports at the end of the program.

<sup>7</sup> This type of interview "has no predetermined agenda or set of questions; instead it allows the interaction between interviewer and interviewee to establish its own dynamic" (Lynch 1996:126).

<sup>8</sup> This subjectivity highlights one of the main limitations of being an insider trying to conduct research in that it is possible that what was observed and noted was "overly influenced by prior knowledge and understanding of the setting" and the insider could have easily failed to "observe aspects that did not conform to that understanding" (Lynch 1996:121).

---

### 3.5. Analyse data (Step 6)

#### 3.5.1. Pre and posttests

The pre and posttests were rated by the two trainers who each wrote brief comments about salient features of each of the tests. One trainer then collated these two sets of comments about each of the tests and wrote a short report for inclusion in a report to management. Because these tests were not 'high stakes' tests this impressionistic method was at first felt to be justified. However, a post hoc analysis of the comments revealed that the trainers did not always focus on the same criteria in each test despite initial discussions and agreement about the criteria on which to base their assessments. However, a Chi-square analysis ( $\text{Chi}^2 = 2.671$ ,  $\text{df} = 3$ ) indicated that the trainers did not differ significantly on the criteria that they evaluated. Nevertheless, the use of a simple, quantitative rating instrument would have ensured greater consistency.

Overall, the pre and posttest analysis by the trainers shows an improvement in most participants' ability to write an introduction and management summary. Only in one of the twelve cases did the trainers disagree about whether any improvement had taken place. Unfortunately, when attempts to confirm the validity of these results was made by asking a subject expert to analyse the same tests, there was almost no correlation between the subject and language experts' evaluation of the tests ( $p=0.03$ ). Close analysis of two of the posttests each rated by a language and subject expert reveals that although both raters were focusing on similar criteria they were rating the criteria in very different ways. The raters had different ideas about what was important under each criterion. The subject expert had not been given any rater training which seems to be an important variable when raters from different professional backgrounds are used. Had rater training been provided we would expect the correlations between the ratings assigned by each rater to have been greater.

#### 3.5.2 Participants' Pre-training Needs Questionnaire

The greatest training needs were defined as those aspects of writing that participants felt least confident about but which were very important to their workplace writing. Table 2 reveals that at least

seven of the participants lacked confidence in their ability to write introductions, recommendations, and management summaries yet these skills were perceived as most important by the majority of the participants.

Writing Skills	Importance of skill for your work			Confidence in using this skill		
	<i>Very important</i>	<i>Important</i>	<i>Not very important</i>	<i>Very confident</i>	<i>Confident</i>	<i>Not very confident</i>
<i>General</i>						
Using correct grammar	9	2	1	0	8	4
Writing clearly	8	4	0	0	8	4
Writing concisely	6	6	0	0	6	6
Using an appropriate tone	6	6	0	1	3	8
Constructing correct sentences	4	8	0	0	9	3
Using linking words	1	11	0	0	7	5
Writing logical paragraphs	8	4	0	0	7	5
<i>Report writing skills</i>						
Planning and organising	9	3	0	0	7	5
Using standard formats effectively	4	8	0	0	8	4
Describing tables, graphs or charts	4	7	1	2	9	1

...continued

Writing Skills	Importance of skill for your work			Confidence in using this skill		
	<i>Very important</i>	<i>Important</i>	<i>Not very important</i>	<i>Very confident</i>	<i>Confident</i>	<i>Not very confident</i>
Writing introductions	9	3	0	0	5	7
Writing up results	8	4	0	1	7	4
Writing conclusions	10	2	0	1	6	5
Writing recommendations	9	3	0	1	4	7
Writing management summaries <sup>9</sup>	7	4	0	1	1	9

Table 2: Trainees' perceptions of importance of and confidence in specific skills

### 3.5.3 End of Program Evaluation Questionnaire

The main purpose of the end of training questionnaire is usually to elicit participant reactions to the program. As mentioned earlier this is often the only form of evaluation that takes place in short training programs. In addition to the standard questions about most/least useful aspects of the program, this particular questionnaire included an observation checklist about how participants felt about what went on in the training room. The overall result of 4.1 on a 5 point scale (*1= does not reflect what went on; 5= reflects totally what went on*) indicates that participants were on the whole satisfied with the delivery of the program.

The responses to the question on the most useful part of the program are shown in Table 3 and indicate that the participants perceived that the report structure section of the program was the most beneficial with five of the eleven responses referring to this.

<sup>9</sup>One respondent said this aspect of writing was not applicable so their results do not appear in this section of the chart.

Most useful aspects of program	Frequency	Reasons Given
Clear Structure	5	<ul style="list-style-type: none"> <li>• It helps to organise materials into sections</li> <li>• It's the foundation of producing an appropriate report</li> <li>• The exercise of reformatting the company report is very useful.</li> <li>• It is because I can use the skills learnt in the course in a real situation.</li> <li>• All arguments are requested to group under different headings such as background, purpose and scope etc. This makes the report very clear and easy to understand.</li> </ul>
Clear Focus	1	<ul style="list-style-type: none"> <li>• This directs me in report writing.</li> </ul>
Clear English	2	<ul style="list-style-type: none"> <li>• Plain English sentence strategies are useful for trainees to write clear and precise English for an ideal and/or logical sequence.</li> <li>• Many of the (editing) errors have appeared on my previous reports. Hopefully the errors can be reduced in future.</li> </ul>
Pre and Post Tests	3	<ul style="list-style-type: none"> <li>• so we can see the difference</li> <li>• I can evaluate my understanding as to what I've learned in the course.</li> <li>• this gave good indication of how one progressed through the course</li> </ul>

**Table 3 : Most useful section of training program**

#### 3.5.4 Trainer debriefing sessions/journal entries

A summary of the main issues that emerged from this more impressionistic data collected by the trainers follows:

##### Content

While the program seemed to address the needs of trainees, many comments related to the *quantity* of material provided. On Day 1 in particular, the amount of material to cover seemed excessive given that trainees found some sections confusing and needed extra time for clarification. A number of comments related to what to omit in any

---

future program, for example, grammar test, particular exercises. Other comments related to what to include in any future program, for example, a vocabulary section was mentioned. Comments were also made on the sequencing of material with some sections being "a bit disjointed".

#### Classroom organisation

One comment mentioned the "teacher-centredness" of sections of Day 1 with trainees being fairly passive "not contributing very much, but listening". By Day 2 there was more "lively chatter" and "eager participation" in activities as participants relaxed.

#### Quality of training

Generally trainers felt that co-teaching was a real strength of the program as it provided a "change of focus, pace and voice". This allowed each trainer to focus on areas of particular strengths or interests.

#### Overall impressions

Some comments related to the success of program objectives, "I felt the afternoon session met the needs of the trainees. They realised that their sentences are long and difficult to understand".

Some areas of 'cognitive dissonance' emerged with the program providing "a lot for them to think about" in that the training was encouraging the participants to rethink their approach to report writing. This raised the question about how much trainers should act as agents of change.

### **3.6. Write Evaluation Report (Step 7)**

In preparing the final evaluation report, the trainers were aware of the sensitive nature of some of the comments about individual participants. Thus, the comments were worded carefully and as positively as possible while still communicating the findings of the evaluation "honestly and successfully" (Lynch 1996: 9).

## **4. Discussion**

The two main purposes in evaluating this program were to:

- (i) Examine the extent to which the program met the needs of the main stakeholders: management, participants and trainers.

- 
- (ii) Determine the strengths and limitations of the CAM as an evaluation framework for such programs.

#### **4.1 Effectiveness of the program in meeting stakeholders' needs**

Overall, the results from the data analysis indicate that each of the stakeholder's needs was met. The evaluation report produced for management appeared to satisfy their need for some hard 'evidence' as to the success of the program. This report included comments about the progress of each participant based on both observation and the results of the pre and posttests. The report also included a summary of the end of program questionnaire which rated most aspects of the program quite highly. Management's satisfaction with the program was also confirmed in discussions with the training manager who stated that "they are learning, I have no doubt about that".

The simple instrument used to establish the participants' initial perceptions of what the most important writing skills were and their confidence in these areas revealed that some participants were not very confident in some of the skills they felt were very important. This was particularly the case with the skills of writing introductions, management summaries and conclusions. Knowing this information allowed the trainers to focus on these aspects in more depth. The end of program questionnaire also confirmed that the participants appreciated this focus with five of the eleven respondents stating that the Clear Structure section of the program was the most useful. In addition, the enthusiastic comments made by the training manager at the end of the program regarding improvements in the structure of a report written by one of the participants also indicated that the focus on report structure was appropriate.

Finally, the results of the observation checklist, the debriefing discussions and diary entries all provided valuable insights for the trainers into the program practices that would require future attention. In addition, the conceptual and methodological difficulties encountered in designing, conducting and rating the pre and posttests raised some interesting evaluation issues that have been documented elsewhere (Ward 1997).

---

## 4.2 Context Adaptive Model

The second purpose of this study was to determine the strengths and limitations of CAM for evaluating this type of program. CAM was developed in response to a need for a more flexible approach to evaluation of language programs than had previously been used in the applied linguistics field. This model is most commonly used to evaluate large-scale program evaluations, so using it as a framework for evaluating a short, language-training program tested its flexibility and adaptability. In fact the model worked well in this context and CAM seems to be an extremely useful framework around which to organise an evaluation. One of the main strengths of this model is its comprehensiveness. By working iteratively through the model's seven main steps a comprehensive evaluation of the program was conducted. This approach to evaluation goes far beyond the usual participant reaction questionnaire distributed at the end of most training programs.

Another strength of CAM is its flexibility in allowing the use of a variety of data sources and methods thereby avoiding the biases inherent in any one particular data source or method.<sup>10</sup> The CAM is a "flexible, adaptable heuristic - a starting point for inquiry into language education programs that will constantly reshape and redefine itself, depending on the context of the program and the evaluation" (Lynch 1996: 3).

However, one related limitation of CAM is that although the model allows for both qualitative and quantitative methods to be used in a mixed design these can be extremely time-consuming and expensive to use.<sup>11</sup> As such they are not always very practical for evaluating language-training programs in the business world. Nevertheless, using a variety of methods overcomes the limitations of any one method and allows for increased reliability when the

---

<sup>10</sup> However, not all agree with this pragmatic approach. For example, Smith and Heshusius (1986) argue that such an approach arises from a confusion between the two paradigms and obscures important differences. Additionally both critical theory and constructivism ("fourth generation evaluation") recognise fundamental differences between the paradigms.

<sup>11</sup> For example, the evaluation of authentic writing samples to assist in the needs analysis, the pre and posttest analysis, the coding of the journal entries and the transcription of the debriefing interviews between trainers all took an inordinate amount of time.



---

same data is collected from different sources or by different methods.

Perhaps a more serious limitation of CAM is that although it provides a very comprehensive framework for evaluating what goes on in a language-training program, it did not provide any guidance for evaluating what happens *after* the program has finished. Research from the business and industry field of program evaluation suggests that what happens *after* the training program is vital for measuring the effectiveness of any program.<sup>12</sup> Therefore, the emphasis on evaluation of program processes may be inappropriate as the quality of the actual training intervention is only one of many factors that determine whether what is learned will be applied on the job. Of at least equal importance to the successful transfer of training appear to be the extent of management commitment to, and involvement in, the training process and a supportive organisational climate.

In this study no data were collected on the extent to which workplace performance improved after this training program. Nor were any data collected on which factors may have contributed to or hindered this transfer of training. Most of the data collected for this evaluation related to program objectives and processes. The pre and posttests were the only form of outcome data collected. While Lynch recognises the need for both process and outcome data when he states that "in order to accomplish an effective evaluation, program objectives, program processes, and program outcomes must all be investigated" (Lynch 1996), there is no mention of evaluating the effects of longer-term programs.

While the model does not *exclude* collecting information of this nature it does not include any explicit steps to encourage evaluators to consider this important follow-on stage of evaluation. This is in contrast with models from the field of industry and business program evaluation that do explicitly include steps for considering the longer term impact of training. (See Kirkpatrick 1994; Brinkerhoff 1989, 1991 and 1995.)

---

<sup>12</sup> See Tannenbaum and Yukl (1992) for a comprehensive review of the research in this field.

---

Adopting a model of evaluation that considers both implementation processes and immediate and longer-term outcomes would seem to be worthwhile. Such models could provide both trainers and clients with a better understanding of the overall worth of training programs and indicate where improvements were required. Those who are involved in language program evaluation may find it useful to consider evaluation beyond the confines of their traditional fields and examine the research being produced in other disciplines that use different paradigms.

## 5. References

- Beretta, A. (1986) Toward a methodology of ESL program evaluation. *TESOL Quarterly* 20, 1: 144-55.
- Brandenburg, D. C. (1988) *The status of training evaluation: An update*. Paper presented at the National Society for Performance and Instruction Conference, Washington, DC.
- Brinkerhoff, R. O. (1989) Evaluating training programs in business and industry. *New Directions for Program Evaluation* 44: 5-19. San Francisco: Jossey Bass.
- Brinkerhoff, R. O. (1991) *Achieving results from training*. San Francisco: Jossey Bass.
- Brinkerhoff, R. O. and M. Montesino (1995) Partnerships in training transfer: Lessons from a corporate study. *Human Resource Development Quarterly* 6, 30: 263-274.
- Brinkerhoff, R. O. (1998) Clarifying and directing impact evaluation. In S. M. Brown and C. Seidner (Eds.), *Evaluating corporate training*. Boston, USA: Kluwer Academic Press.
- Brumfit, C. J. (1980) *Problems and principles in English teaching*. New York: Pergamon Press.
- Cummings, C. W. (1998) What stakeholders want to know. In S. M. Brown and C. Seidner (Eds.), *Evaluating corporate training*. Boston, USA: Kluwer Academic Press.

- 
- Keating, R. F. (1963) *A study of the effectiveness of language laboratories: A preliminary evaluation in twenty-one school systems of the Metropolitan School Study Council*. New York: The Institute of Administrative Research, Teachers College, Columbia University.
- Kirkpatrick, D. L. (1994) *Evaluating training programs: The four levels*. San Francisco: Don Berrett-Koehler.
- Hutchinson, T. and A. Waters (1987) *English for Specific Purposes*. Cambridge: Cambridge University Press.
- Long, M. H. (1983) Inside the black box: Methodological issues in classroom research on language learning. In H. W. Seliger and M. H. Long (Eds.), *Classroom oriented research in second language acquisition*. Rowley, Massachusetts: Newbury House.
- Lynch, B. K. (1996) *Language program evaluation: Theory and practice*. USA: Cambridge University Press.
- Nunan, D. (1989) *Understanding language classrooms: A guide to teacher-initiated action*. Englewood Cliffs, NJ: Prentice-Hall.
- Patton, M. Q. (1980) *Qualitative evaluation methods*. Beverly Hills, CAL: Sage.
- Scherer, G. A. C. and M. Wertheimer (1962) *Extended classroom experimentation with varied sequencing of the four skills in German instruction*. Final report, Contract number SAE 8823. Washington, DC: US Office of Education.
- Smith, J. K. and L. Heshusius (1986) Closing down the conversation: The end of the quantitative-qualitative debate among educational inquirers. *Educational Researcher* 15, 1: 4-12.
- Tannenbaum, S. and G. Yukl (1992) Training and development in work organisations. *Annual Review of Psychology*, 43: 399-441.
- \* \* \* \* \*

---

Ward, S. (1997) *Inter-rater reliability in an ESP context*.  
Unpublished paper, The Department of Linguistics and  
Applied Linguistics, The University of Melbourne, Melbourne.