

AUTHOR Crabbe, David  
 TITLE Bringing Evaluation and Methodology Closer Together.  
 PUB DATE 91  
 NOTE 8p.; In: Anivan, Sarinee, Ed. Issues in Language Programme Evaluation in the 1990's. Anthology Series 27; see FL 021 869.  
 PUB TYPE Reports - Evaluative/Feasibility (142) -- Guides - Non-Classroom Use (055) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.  
 DESCRIPTORS \*Class Activities; Classroom Techniques; Evaluation Criteria; Foreign Countries; Independent Study; \*Instructional Design; Instructional Effectiveness; \*Learning Motivation; Second Language Instruction; \*Second Languages; \*Student Evaluation of Teacher Performance; Student Motivation; \*Teaching Methods

ABSTRACT

Problems inherent in student evaluation of second language teaching methodology, particularly task-based instruction, are discussed and recommendations are made. It is argued that course evaluation by learners is based on their assessment of their own communicative performance, and that they are likely to attribute a general sense of progress, at least in part, to class activities. However, if self-direction is an instructional goal, a general sense of progress is not an adequate criterion. Evaluation of methodology should be encouraged from the outset and be focused on specific tasks. If learners see value in the tasks, they are more likely to use them for independent work. If teachers were to examine instructional tasks to determine whether learners can evaluate them positively for effectiveness, they would find that the tasks are often inappropriate. Therefore, tasks should be: (1) identifiable by learners as involving a specific type of performance; (2) designed in such a way that they can be staged easily by a learner working on his own; and (3) include an element of enhanced feedback and practice to demonstrate improved performance. Incorporation of these elements in learning tasks is viewed as part of a broader strategy to develop learner self-direction. (MSE)

\*\*\*\*\*  
 \* Reproductions supplied by EDRS are the best that can be made \*  
 \* from the original document. \*  
 \*\*\*\*\*

# BRINGING EVALUATION AND METHODOLOGY CLOSER TOGETHER

David Crabbe

## 1 INTRODUCTION

In this paper I address the problem of learner evaluation of methodology and, in particular, task-based methodology. At the end of any course it is almost required practice to give a questionnaire that elicits learner evaluation of aspects of the course. This is often the only formal evaluation of the course that is carried out. I have distributed many end-of-course questionnaires, but on more than one occasion I have felt that they somehow miss the point. Firstly, there is, typically, a paucity of information that comes back from them. They only give a sense of the degree of client satisfaction or dissatisfaction which is commercially but not pedagogically informative. It may well be that the design of these questionnaires is lacking but if client satisfaction is the main thing that comes out of them, the end of the course is a little late to gauge it. Secondly the comments made are usually one-off statements that can easily be dismissed as the eccentric whim of one respondent. Sometimes one would like to be able to discuss a viewpoint with a respondent because one feels that there is a serious lack of shared expectations of the course, something that should have been ironed out in a different way. Clearly a final questionnaire does not do justice to the depth of involvement in the pedagogical decisions one would like students to demonstrate.

This paper begins with the assumption that course evaluation by learners is based on self-assessment of their own communicative performance. A general sense of progress is likely to be attributed, at least in part, to course activities. What is suggested here is that evaluation based on a general sense of progress is not good enough if self-direction is a goal. Evaluation of methodology needs to be encouraged right from the beginning and to be focussed on specific tasks. If learners see value in tasks they are more likely to use them for independent work. The paper proposes certain requirements of task design that may help this important process. Any final questionnaire should reflect how far this process has worked.

As with much teaching procedure, the proposals are somewhat speculative in character. They arise out of experience with a task-based pre-session course for overseas postgraduate students run at Victoria University, Wellington. One of the four general aims of this course is for the learners "to know what steps they might personally take for further improvement of their communication in English in an academic context." It should be added that the students on this course are highly motivated by their imminent need to survive in demanding academic contexts.

## 2 THE LEARNER AND EVALUATION OF TASKS

When people are learning a new skill, my experience is that they are informally evaluating a great deal - evaluating not primarily the effectiveness of the task for learning but their own performance in doing it. If you are learning to ski you are constantly critical of what you are doing, that you are leaning at the right angle to the slope, that your feet are appropriately placed and so on. This is a natural informal evaluative process that one would expect of any skill learning, including language learning.

What I would like to suggest is that other evaluation arises largely out of this self-assessment. If you are being instructed at skiing and you are not told about angle to the slope until you have made several undignified slides downhill in a prone position, you will probably evaluate negatively the way in which your practice task has been explained to you. In the

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

This document has been reproduced as  
received from the person or organization  
originating it

Minor changes have been made to improve  
reproduction quality

Points of view or opinions stated in this docu-  
ment do not necessarily represent official  
OEI position or policy

PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

Wong Kim  
Wong

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)

same way, the learner of a language is likely to judge classroom tasks against the progress that he perceives he is making. If he perceives improvement or no improvement, he will either attribute that improvement or the lack of it to the classroom tasks, or to his personal study and use of the language, or to both.

It seems to me that it is important that the learner is able to distinguish what contribution the classroom tasks on the one hand and personal study and general use of the language on the other, have made to his language learning. The reason for this is that the classroom tasks are in the public domain and the personal study is in the private domain and in the interests of establishing self-direction, the boundaries between these two domains, usually raised by education, need to be removed. In other words, the language learner needs to be evaluating across both domains which aspects work and which do not so that he has control over them. In this way, classroom tasks become a source of information, a model, for personal study (and vice versa) and even a model for managing general use. Classroom tasks should therefore highlight and not just require efficient strategies for language learning and use.

This seems an obvious criterion for task design when self-direction is an objective. Yet it is not often a criterion which is met by classroom tasks. If we are to help a learner to evaluate a classroom task for the degree to which it enhances performance and learning and, in so doing, to help him to build up a personal arsenal of independent activities, then I think that certain requirements have to be met by classroom tasks in a course. I have listed these below and each will be discussed in turn in subsequent sections. Considerably more attention will be given to the third requirement.

- (1) Tasks should be identifiable by learners as involving a specific piece of communicative performance
- (2) Tasks need to be done in such a way that they can be easily staged by a learner working on his own.
- (3) Tasks need to include an element of enhanced feedback and practice to demonstrate improved performance and thus facilitate evaluation.

### 3 TASKS SHOULD BE IDENTIFIABLE BY LEARNERS AS INVOLVING A SPECIFIC PIECE OF PERFORMANCE

If learners are to be able to attribute improvement in language development to any particular activity, a general sense of improvement is not easily attributable in a valid way. I know if I practise at the piano simply by playing it as often as possible, and I improve, then it is difficult to say why, beyond the fact that I have played a lot. If, on the other hand, someone says to me that I need to focus on my fingering and shows me a technique to practise that aspect of playing the piano, I can say the technique was either useful or not depending on whether there is an immediate improvement in fingering. In the same way, if I am using and studying the language a lot and gradually improve, I do not know what to attribute that improvement to. It may be vocabulary study, it may be extensive listening to the radio, it may be both, it may be neither. Because I have not paid attention to any specific task and focused on that, I cannot really tell. Does it really matter, so long as we get there? Well I think it does matter. For one thing, it may be more efficient to concentrate on one bit of performance at a time on the grounds that an in-depth case study of a bit of language in use is better than trying to draw generalisations from language data spread over several bits of performance. This view draws on an information processing view of language (McLaughlin 1987, Chap. 6) rather than the comprehensible input view of Krashen. For another thing, it gives the learner a greater satisfaction with the learning process in that he can evaluate specific progress as it happens on one front rather than have a vague sense of general progress on several fronts at the same time.

This is really an argument for task-based learning in general - not any tasks but tasks that have as their goal a specific bit of performance with high face validity, that is

performance that relates to the learner's target communication. This bit of performance might be writing a particular genre of report, being interviewed for a job, giving a seminar presentation. The performance may of course be considerably guided or simplified for the level of the learner using various techniques available. (See, for example, Phillips 1983 for the principle of reality control and Widdowson 1979 for the technique of gradual approximation).

This suggestion, that specific performance tasks may enhance self-assessment and evaluation of task effectiveness, implies a problem with tasks such as reordering jumbled sentences, some information transfer tasks, spotting the difference in two pictures, and many information gap activities. These kinds of tasks are stock in trade for communicative language teaching and whilst I do not wish to decry their value for fluency development, they are often tasks that are not specific in performance. What they usually aim at is general improvement in proficiency and this provides no performance focus for learners to evaluate except artificial classroom performance. While learners can evaluate their progress in these artificial tasks, such progress may be seen as trivial. More importantly, however, they provide no encouragement for the learner to relate the classroom task to the real tasks that he will face or is already facing in the real world. He is therefore, I believe, less likely to identify or consciously transfer learning strategies.

The current ESP practice, then, of simulating target communication in the classroom through tasks such as essay writing, preparing and delivering seminar presentations, listening to lectures, is valuable not only because they meet Phillips' criteria of non-triviality and authenticity (Phillips 1983) but also because at the same time they meet one requirement for training learners to meet real learning needs themselves. If a learner in the target situation has a problem with oral presentations, then will his mind turn to an information gap activity to improve his oral performance? Probably not, but if he had done a task involving individual preparation procedures for a short talk he would then have a model procedure to follow.

#### **4 TASKS NEED TO BE DONE IN SUCH A WAY THAT THEY CAN EASILY BE STAGED BY A LEARNER WORKING ON HIS OWN.**

This requirement is a practical one. I think that not infrequently on English language courses, including EAP courses, tasks are selected that involve special materials or special classroom equipment or special management (information distributed in a certain way, for example). Again, I do not wish to decry the innovative design that is behind many of these activities, but I do believe that it may mystify the teaching process by making the teacher the powerful magician. This may prevent the learners from evaluating a task as something which they can usefully stage themselves. The boundaries between the public and private domain remain intact. Students - and teachers - often believe that purpose-built materials are necessary for language learning. I am not suggesting a contrary minimalist approach, but purpose-built strategies are so much more important, and so much more difficult to provide.

On the EAP course in Wellington, naturalistic performance is emphasised as much as possible in the sense that tasks are mostly tasks that can be done either in groups or individually without any extra resources. The fashion for group work, supported as it is by work demonstrating the quality of the interaction involved in such work (Long and Porter 1985) tends to overshadow the arguments in favour of individual work in EAP where the conceptual performance is intimately bound up with the communicative performance and in the end the learner is on his own. (Crabbe 1987)

#### **5 TASKS NEED TO INCLUDE AN ELEMENT OF ENHANCED FEEDBACK AND PRACTICE TO DEMONSTRATE IMPROVED PERFORMANCE AND THUS FACILITATE EVALUATION.**

One of the biggest problems that I see with many communicative tasks, even if they concentrate on specific performance, is that they obviously provide for communication but they do not obviously provide for learning. Of course the current theory is that communication does lead to learning and the consequent principle is that the more communication you do, the more learning that will take place. This principle is not always accepted by learners. They worry about their performance - about not understanding bits of the communication, about making production mistakes. Moreover their fears about their performance means they feel they are not making progress. This is likely to have negative repercussions. The students are likely to undervalue the course as a whole and, moreover, no particular task will stand out as one they can take into the private domain as independent language learning strategy. I believe that there is an element of task design that is critical here and to illustrate this I want to describe two language learning demonstrations I experienced, one 20 years ago as a beginning student of Russian and the other 10 years ago at a seminar in Lancaster.

The Russian course I attended was a traditional grammar and translation course with a bit of audio-lingual laboratory thrown in. Once a week we had a Russian evening in a local cafe and on one night a member of the Russian diplomatic staff came along to engage in conversation. He chose to play a recording of Goldilocks and the Three Bears in Russian perhaps to avoid holding what must have been painful conversation with us. Then he had us retell the story, person by person, one sentence from each person. If the person did not get it right the turn passed to the next person. The recording was replayed between each retelling. I thought this was a marvellous way of learning as we each struggled with our sentence, not only with the form but also with the content. We had to do it several times and every time round we each got a different sentence to do.

The seminar at Lancaster was held by Celia Roberts, at that time teaching English to Hong Kong policemen. She demonstrated a piece of performance which was answering a telephone at the police station. She played the part of an irate member of the public phoning in to complain about the noise in the port area. She held an imaginary phone to her ear, made a ringing noise and pointed at a hapless member of the audience to answer. The answer was inappropriate and so she hung up and started again pointing at different participants and hanging up until a correct or appropriate response was made. Each participant learned from others mistakes although there was no explicit feedback.

What are the features of this and the Russian task? Firstly the feedback is built into the task. Getting the performance right is an essential criterion for it to be completed. Sometimes there was an explicit model for comparison as in the Goldilocks example, sometimes the feedback is by listening to other speakers perform or getting correction, implicit or explicit, from the teacher.

The second feature is that there is what I call repeated performance. In other words, learners get a chance to have another crack at the same communication, not another piece of communication with some similar features. I think as we emerged from the audio-lingual era we forgot the importance of repetition in language learning, so keen were we to slough off the old paradigm. Of course unlike the audio-lingual repetition of unconnected forms, I am talking here about the repetition of connected meanings, discourse.

In the tasks on the Wellington EAP programme, there is an attempt to incorporate as much as possible the features of built-in feedback and repeated performance. An example of a ready-made task that can be used with any content available is the 4-3-2 technique (Maurice 1983). In this task a learner gives a talk for 4 minutes to a partner and then listens to the partner give a talk on the same topic, a new partner is found and the same talk is given and listened to but this time for 3 minutes. Finally a third partner is found and the talks are given again in 2 minutes. Feedback comes from listening to others give the same performance but to enhance the feedback process I would deliver a talk myself on the same topic as a native speaker model after the 4 minute and the 3 minute talks, thus providing a native-speaker model for performance comparison. The topics are usually drawn from the study theme that was currently being worked on. The activity can be used at any point for practice in oral presentation.

Writing tasks also involve repeated performance and built-in feedback as the writing is done in the form of a workshop where the learners work independently calling on a

teacher as informant when needed. On the same principle, when giving a short seminar style presentation the students are encouraged to practice the presentation several times on their own at home before they deliver it. Feedback is provided after the presentations and the learners are able to benefit from the feedback given to others before their own turn comes around.

A similar and more fully developed approach to task design, is described in Willis and Willis (1987) in which there is emphasis on rehearsal of tasks and on listening to native speaker models. There is less emphasis on the repetition of the same performance although this is not precluded.

I think that there are important reasons why built-in feedback and repeated performance are necessary components of any task and this is to do with the nature by which we learn. Built-in feedback enables the learner to critically assess how well he has performed, not in general but in specific details. The best way in which this is to be encouraged is still not clear although I favour procedures by which the learner has to discover the errors for himself (see Chaudron 1987, however, for a review of error correction by the teacher). Repeated performance enables the learner to apply the results of feedback as well as develop a degree of automaticity. After a number of repetitions, students nearly always report improvement, at least in fluency. This gives repeated performance high face validity with learners - there is hard work involved in repeating a bit of communication four or five times but the perceived return avoids any sense of tedium. Some research to support the perception of improved performance is that carried out by Brown and colleagues (Brown et al 1984) in which little performance improvement in oral tasks was noted after simple repetition but when the speakers had a chance to listen to others do the same task, there was significant improvement in subsequent performance. This research was with native speakers. Arevart (1988) looked specifically at the 4-3-2 technique with second language learners and found that their fluency increased and that "repetition also results in improvement in the accuracy of the language used in the talk. The case studies show that the learners correct grammatical errors previously committed while speaking. They set out a discourse plan formulate utterances, establish language rules and try them out." (p 80) This happened without a native speaker model for comparison.

Clearly research is needed to confirm that any gain in performance in such tasks is permanent. At this stage however, I am satisfied with the fact that immediate improvement is evident to the learners themselves. The learners are actively engaged in managing improvements and I believe this helps to break down the public/private boundary. The task is more likely to be transferred to the private study domain as a useful strategy for practice.

In the private domain there is of course the problem with feedback in self-directed productive tasks. A model will usually provide feedback of a comparative nature for learners to identify salient lexical, structural or even pragmatic information for their own personal learning. However models are not available for tasks that you do without a teacher unless you are working with specially prepared materials with models built in (See Willis and Willis 1987). But tasks can be made out of models. In other words a learner can take a piece of available native speaker performance (printed or recorded), put it aside as a model, do the performance himself and then pick up the model again. Classroom tasks may have to reflect this order of going about things.

## 6 THE WIDER CONTEXT

Designing tasks that assist evaluation of effectiveness, do not of course represent the whole picture of how students are encouraged to evaluate - either their own performance or the effectiveness of the programme. Evaluation, to be effective, involves the gathering of information, the coding of it in a way that is sensible and usable and then applied. All this is hard enough for a teacher to do of his own performance. It is extremely difficult to encourage the learner to do on their own behalf. Yet unless we take this on as part of our job and particularly so in the case of EAP courses where the learners will soon be fending for

themselves without formal language instruction, there is little hope for any evaluation of the course being very meaningful except as a measure of client satisfaction. What we want those final questionnaires to reveal is not that the learners liked the course, although that is important enough, but that they were so involved in the learning experience that they knew what was going on.

The Wellington EAP course referred to here used a wide range of strategies, embodied in minimum standards for self-direction, to encourage learner evaluation. These strategies included student record booklets distributed at the beginning of the course and completed by the students as the course progressed, personal interviews, explicit discussion of tasks, an introductory study theme on how people learn languages, a self-access centre with advisors and, in all of that, an attempt to foster metacognitive awareness of learning. Even with all that effort one cannot be sure that self-direction is developing. The affective aspect is another factor in the process, perhaps the biggest factor of all and although that can be addressed through continuous monitoring of individuals, there are always limitations.

## 7 SUMMARY OF MAIN POINTS

In this paper, I have made the following claims

- 7.1 That final course evaluation questionnaires do not reveal a great deal of information about the course.
- 7.2 Part of the reason for this is that the learners do not usually have much basis for attributing improvement to any particular aspect of the course.
- 7.3 That learners should therefore be given such a basis as it helps them to be self-directed and to transfer tasks from the classwork to personal study
- 7.4 That, to achieve this basis for evaluation, three requirements of task design are that
  - (i) tasks should involve one specific piece of performance so that improvement in that performance is attributable to that task
  - (ii) class tasks need to be stageable by learners on their own. All group tasks should also be performable as individual tasks.
  - (iii) tasks need to include repeated performance in order to enable learners to evaluate progress in specific performance and thus to increase their management of learning.
- 7.5 That these requirements of task design constitute one aspect of a broader strategy to develop self-direction.

## 8 CONCLUSIONS

I have said that if we evaluate tasks for whether learners can evaluate them positively for effectiveness, I think it would be surprising how often the tasks do not measure up to this criterion. In the end, of course, we should as teachers know more than the learners about learning and a traditional view would be that the learner should trust us. I think it would not be too difficult however to make our wisdom about learning, such as it is, more transparent and accountable so that learners can take it on and not need us when we are not available.

This suggests that in addition to research questions designed to evaluate which aspects of tasks seem to contribute to performance improvement, we need parallel research questions to evaluate how visible this performance improvement appears to learners and whether high visibility leads to transfer of strategies.

#### REFERENCES

- Arevart, Supot 1988. The effects of repetition on spoken fluency and accuracy. Unpublished M.A. thesis, Victoria University of Wellington.*
- Brown, G., A. Anderson, R. Shillcock and G. Yule 1984. Teaching talk: strategies for production and assessment. Cambridge University Press.*
- Chaudron, Craig 1986 The role of error correction in second language teaching. In B.K. Das (ed) Patterns of classroom interaction in Southeast Asia. pp 17 - 50 SEAMEO Regional Language Centre (Anthology Series 17)*
- Crabbe, David 1987. Developing conceptual performance in a foreign language. Paper presented at RELC Regional Seminar.*
- Long, Michael and Patricia Porter 1985. Group work, interlanguage talk and second language acquisition. TESOL Quarterly 19,2.*
- MacLaughlin, B. 1987. Theories of second-language learning. Edward Arnold.*
- Maurice, K. 1983. The fluency workshop. TESOL Newsletter, 17,4*
- Phillips, Martin K. 1983 Towards a theory of LSP methodology. In J.Wilson and P. Horey (eds) ELC Occasional Papers No 1. King Abdulaziz University.*
- Widdowson, H.G. 1979. Gradual approximation. In Explorations in Applied Linguistics. Oxford University Press.*
- Willis, David and Jane Willis 1987. Varied activities for variable language. ELT Journal 41,1*