

# Using Self-assessment for Evaluation

**S**ELF-ASSESSMENT APPEARED TO COME OF AGE IN 1980 WITH THE publication of a Council of Europe text on the topic (Oskarsson 1980). Since then, more and more programmes around the world have attempted to integrate self-assessment into the learning and evaluation<sup>1</sup> process, with varying degrees of success. The usefulness of self-assessment for learning purposes seems to be widely accepted, as illustrated by the widespread use of learner diaries. Self-assessment for evaluation purposes, however, is far less common, and many teachers actively resist its implementation. This situation is due, in part, to the ways in which self-assessment is frequently conducted. In this paper, I will argue that learners can conduct reliable, global self-assessment, and I will suggest three ways in which such data-driven self-assessment can be done.

1. I am using *assessment* as a broad term for all attempts to gain information concerning learners' performance and ability, regardless of the purpose. *Evaluation*, in contrast, is used for those attempts that produce quantitative data that are used to generate scores measuring the learners' performance and ability.

## Purposes of self-assessment

Several reasons for using self-assessment have been suggested including:

- Self-assessment is a prerequisite for a self-directed learner. If a goal of learning is for learners to be self-sufficient and independent in language use, then training and experience in self-assessment are needed.
- Self-assessment can raise learners' awareness of language, effective ways of learning, and their own performance and needs.
- Self-assessment increases motivation and goal orientation in learning.
- Some aspects of language learning, such as effort and learner beliefs, can only be assessed through self-assessment.
- Self-assessment can reduce the teacher's workload.

The first four reasons clearly suggest that self-assessment can be integrated into courses for learning purposes.

Less clear, however, is whether these reasons imply that self-assessment should be used as part of the input in generating a learner's score for a course. This depends on the objectives of the course. For final evaluations of learners' performance on a course to be valid, the evaluations should match the course objectives. If the objectives include increased motivation, positive attitudes towards English, and greater independence and awareness, for example, then self-assessment should be seriously considered as a potential part of the overall evaluation for a course. Most teachers, however, strongly resist such a move, arguing that self-assessment is subjective, unreliable, open to cheating, and more reflective of the learner's self-image than actual performance and ability. Such an attitude is at least partially due to the nature and characteristics of existing self-assessment instruments.

## Self-assessment instruments

Some self-assessment instruments, while powerful when used for learning purposes, are inappropriate for evaluation purposes. These include learner diaries; the task-based self-assessment instruments of Tudor (1996) by which learners are encouraged to analyse various aspects of their learning, such as their difficulties in completing a task; and the critical incidents in learning of Singh (1998). These instruments are very subjective—indeed, sub-

jectivity is the *raison d'être* of critical incidents—and produce qualitative information that cannot be converted into scores for evaluation purposes.

Self-assessment instruments that produce quantitative information that can be used for evaluation purposes fall into two categories: global self-assessments and self-marking instruments.

### *Global self-assessment*

It is in the area of global self-assessment that Oskarsson's (1980) work is most influential. Oskarsson suggested that global self-assessments could be conducted through rating scales and checklists. However, both of these, as Oskarsson suggests, are very problematic. To illustrate this, here are example questions used to measure learners' speaking ability:

- Give yourself a rating for your speaking skills on a scale of 0 to 10, where 10 means *I am completely fluent in English* and 0 means *I cannot speak English at all*.
- Can you ask someone to help you to arrange an appointment with a doctor?
- Can you express sympathy using phrases like *I am sorry to hear that*?

At face value, these questions may seem fairly straightforward. But if we were to apply Oskarsson's question format to teaching, we would produce questions such as:

- Give yourself a rating for your classroom management skills as a teacher on a scale of 0 to 10, where 10 is a perfect classroom manager and 0 is a complete incompetent.
- Can you explain the meaning of *behave*?
- Can you give clear instructions for a jigsaw reading activity?

As a teacher, your reaction to the first of these is likely to be a complete lack of confidence in your answer. Maybe, like me, you just plumped for a number in the middle of the range. For all of the items, you also probably feel that your answer depends on the teaching situation. I'd have no problem explaining *behave* to my postgraduate students, but I wouldn't even attempt an explanation with a class of undisciplined kids on a Monday evening. In fact, it may even seem unfair to ask items like these.

Yet these are exactly the sort of items learners are faced with in Oskarsson's questions promoting self-assessment of speaking and in other instruments for global self-assessment.

This type of self-assessment instrument lacks specificity and is divorced from reality. Instead of rating any real-world language performance, learners are asked to rate their own beliefs and perceptions with little or no evidence on which to base their assessments. Such self-assessment, although valuable for learning, is grist for the mill for teachers who argue that self-assessment is too subjective to be used for evaluation purposes.

### *Self-marking instruments*

Self-marking involves learners in giving themselves a score for a piece of work. Where the task is objective, such as a multiple-choice exercise, an answer key can be provided, and learners can mark their own work easily. This reduces the teacher's marking load and provides a reliable score (with a little cross-checking to discourage cheating) that can be used for evaluation purposes. The learning benefits of this approach are, however, negligible.

For more open-ended tasks, where there may be a very large number of possible answers, self-marking is more problematic. One example of how self-marking may be conducted is given by Gardner and Miller (1999). In their example, the task is to skim a newspaper and then listen to the news on the radio. Learners then give themselves two marks: one for their understanding of the main ideas of the news, and one for their understanding of details. This self-assessment task serves a useful learning purpose by highlighting areas in which learners need to do further work, but the marks from the self-assessment are hardly reliable enough to persuade most teachers to include them as part of the final score for a course.

To increase reliability, self-assessment on open-ended tasks needs to be clearly guided by detailed scoring criteria. The easiest way to generate such criteria is to break down the task into smaller components. For example, for a letter-writing task, the finished product could be self-marked for how well it follows the standard letter-writing conventions, such as introducing the purpose of the letter in the first paragraph, assigning each topic to a separate paragraph, and so on. The close guidance of scoring criteria such as these is likely to increase the reliability of the learner's self-assessment, making it more palatable for inclusion in the final score for a course.

### **Data-driven global assessments**

The use of objective tasks and detailed scoring criteria for self-marking, as described above, are restricted to self-assessment of particular tasks. Such self-assessment may be included in the overall score for a course. However, it can provide only a snapshot of a learner's performance. To obtain a measurement of a learner's development throughout a course, self-assessment at a more global level is needed.

The usual approaches to global self-assessment, such as those of Oskarsson, are, as we have seen, fraught with problems. Divorced from any real-world performance, they end up as very subjective and unreliable guesses that are unsuitable as components of a final score. What is needed is some way to directly relate a learner's performance to his or her global self-assessment to make that self-assessment more reliable and more reflective of actual performance and ability.

A key question in designing global self-assessments, therefore, is how they can be directly related to learners' experiences. At the task level, the process of completing the task and the finished product provide a clear focus for and input into self-assessment. For global self-assessments at, say, the level of course, what things could be used as data driving the self-assessment?

### *Portfolios*

The most obvious and widely-used learning instrument that could be used as input for self-assessment is the portfolio. A portfolio is "a purposeful collection of students' work that demonstrates to students and others their efforts, progress, and achievements in given areas" (Genesee and Upshur 1996:99). Since the portfolio is evidence to learners of their own efforts, progress, and achievements, it is suitable for self-assessment. To use a portfolio as self-assessment for evaluation purposes, questions to guide the self-assessment must be provided. Sample questions could include the following:

- To what extent did you achieve your goals in learning during this course?
- To what extent did you improve your reading? List some of the problems you faced while reading and how you solved those problems.
- To what extent has your knowledge of vocabulary improved? List the new words you have learnt from your portfolio.

- To what extent has your confidence in using English improved?

By referring to their portfolios in answering these questions, learners have concrete evidence of their performance and are not forced to rely on their intuition and possible bias about their performance or ability.

#### *Pre- and post-course writing*

A second way of conducting data-driven, rather than intuition-driven, global self-assessments is to use the time-honoured research technique of pre- and post-tests. Learners can be asked to write two essays about their attitudes towards learning English, one at the start and another at the end of the course. Comparing the two, learners are able to see the extent of their development through the course. With guiding questions, learners' perceptions of their own development based on the pre- and post-course writing can provide self-assessment that can be used for evaluation purposes. The two pieces of writing can also be self-marked for certain language points. Whereas self-marking instruments applied to a given task provide a snapshot of the learner's performance at a given moment in a course, a comparison of self-marking on pre- and post-course writing can give a clear indication of the learner's development and improvement throughout the course.

#### *Learner contracts*

A third potential instrument for global self-assessment is the learner contract (e.g., Dickinson 1987). At the start of a course, learners identify two or three goals they want to achieve in the course, tasks and materials that can be used to reach these goals, and ways of measuring the extent to which the goals have been reached. For example, a learner may decide to increase his or her speed in reading. The learner can then identify some texts with comprehension questions to be used as practice and set a target level of achievement, such as an increase in reading speed of 50 words per minute while retaining a minimum of 70% for comprehension questions answered correctly. A learner contract, then, provides an organised series of tasks throughout a course

and makes attaining specific goals an integral part of the learning process. The choice of goals in learner contracts can be left to the learner or can be controlled by the teacher to match the objectives of the course. In the latter case, self-assessment in learner contracts can be used as a valid part of the overall evaluation of learners in the course.

#### **Conclusion**

At present, self-assessment is a valuable tool in the teacher's repertoire of techniques that enhance learning. If its uses are to be extended to include evaluation, self-assessment needs to be set up in such a way as to overcome the resistance of teachers. In this article, I have suggested that this can be done by basing self-assessment on concrete evidence of the learner's performance and by giving guidelines on how to conduct the self-assessment. In these ways, self-assessment can become more reliable and fulfil an important role in providing learner input into evaluation for a course.

#### **References**

- Dickinson, L. 1987. *Self-instruction in language learning*. Cambridge: Cambridge University Press.
- Gardner, D. and L. Miller. 1999. *Establishing self-access: From theory to practice*. Cambridge: Cambridge University Press.
- Genesee, F. and J. A. Upshur. 1996. *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- Oskarsson, M. 1980. *Approaches to self-assessment in foreign language learning*. Oxford: Pergamon, for the Council of Europe.
- Singh, K. 1998. *Using critical incidents in self-assessment*. Paper presented at the 3rd Annual Thai TESOL English for Specific Purposes Special Interest Group Conference ESP at the Cutting Edge: New Challenges, New Solutions, Bangkok.
- Tudor, I. 1996 *Learner-centredness as language education*. Cambridge: Cambridge University Press. ✎

---

RICHARD WATSON TODD is an associate professor in applied linguistics at King Mongkut's University of Technology Thonburi in Bangkok, Thailand.