



Developing a criterion-referenced self-assessment

Greg Rouault

Akashi National College of Technology

Reference Data:

Rouault, G. (2007). Developing a criterion-referenced self-assessment. In K. Bradford-Watts (Ed.), *JALT2006 Conference Proceedings*. Tokyo: JALT.

Alternative assessment methods can support instructors in attending to affective factors facing individual identities in our communicative language classrooms. As part of authentic assessment, courses should include learner reflection on abilities and levels of awareness of and confidence in their communicative performance in practical applications of language. Through a brief overview of tests and assessment in teaching and learning, along with attention to washback, this paper outlines the rationale for criterion-referenced self-assessment. An approach for the development of criterion-referenced, “can do” statements for self-assessment in oral communication classes is introduced along with a sample of the survey tool to be used to collect these self-perception ratings of communicative proficiency as the initial phase of a broader study, to be correlated with measures of foreign language anxiety.

コミュニケーションを主体とする英語の授業において、個々の学習者のアイデンティティに向き合い、学習者の情緒面の要素を重視しようとする場合、従来の形式にとられない評価方法は、その指導の一助となる。現実に即した評価方法の一部として、授業には、学習者が自分自身のコミュニケーション能力に対して持つ自己評価を反映させることが組込まれるべきである。本論では、指導と学習におけるテストおよび評価に関する短い概要を通じ、さらに、評価からの有益な波及効果にも留意しつつ、目標標準型の自己評価の原則について概説している。オーラル・コミュニケーション・クラスにおいて、目標標準型の「～することができる」という形式の自己評価を行う取組みが、より広範な研究の第一段階として、コミュニケーション能力に関する学習者の自己評価を収集するためのアンケート・ツールのサンプルと併せて導入されている。収集された自己評価データは、外国語学習者の持つ不安感を示す基準と関連付けされる。

The growth and development of communicative language teaching has created new explorations and interpretations of learning and performance. A view of communication as essentially involving the ability and willingness to use language in interaction has driven a need for assessment options other than standardized tests in language courses. Considering the range of affective factors and cognitive abilities involved in communicative performance, EFL instructors in group classes face

significant challenges in making subjective interpretations to assess student performance, especially in the area of oral communication. A variety of alternative assessments have surfaced with the common aim of integrating learning, teaching, and evaluation. Authentic assessment can be said to employ activities and tasks for ongoing evaluation of learner progress by integrating the classroom curriculum, instructional goals, and real-life performance in a target language use situation. As an alternative form of authentic assessment, self-assessment tools draw direct learner involvement and create awareness of the learning process. This paper outlines the development of criterion-referenced, "can do" (Jones, 2002) statements for use with a course curriculum established for first and third year college level learners of English as a foreign language in Japan. This represents the first phase of a broader survey project where the self-assessments will be correlated with foreign language anxiety as gauged through the Foreign Language Classroom Anxiety Scale (FLCAS) proposed by Horwitz, Horwitz, and Cope (1986).

Background Rationale

To further examine the rationale, it is useful to first identify a working definition of both criterion-referenced and self-assessment. In looking at criterion-referencing with regards to norm-referencing, Brindley (1989) cites Docking's definition that criterion refer to "the base dimensions used in determining achievement" and are validated by well-defined domain criteria. Criterion-referenced assessments therefore provide detail on an individual's level of mastery of a task or skill with no attention given to the relative ranking with

others. The base of criterion-referenced assessments and the domain of competencies to be assessed involve the specification of objectives (Docking, cited in Brindley, 1989, p.48). In terms of self-assessment, the can do statements used in this study are seen as "performance-ability self-assessments" (Brown & Hudson, 2002, p.83) where learners read a situation and rate their own ability to respond.

The rationale for using criterion-referenced self-assessments as an alternative form of assessment includes that the survey method is very practical and easily distributed in a timely fashion. The criterion based document initially identifies and frames the objectives for the course. The data provided by the learners offers a needs analysis for the instructor to plan the curriculum and determine time to be spent in various areas. Criterion-referenced assessments have a high degree of face validity as they can be constructed around authentic, real world tasks. Such contextualized tasks can be meaningful for learners that are intrinsically motivated. For others more focused on the course or the grade and not the functional process, there is a positive impact through washback (Hughes, 2003). This beneficial contribution results since achievement is based on objectives and the self-assessment corresponds to the target abilities which are also ideally supported with direct testing. Finally, criterion-referenced self-assessments can provide a bridge for the affective factors in learning (Brown, 2004, p.274) and support further individualized goal setting and autonomy for learners prepared to take that next step.

Method

The initial phase of the study was conducted over 2 years

in 8 oral communication classes of approximately 20 low proficiency 1st and 3rd year students at a technical college in Western Japan. Several students were experiencing English study with a native English speaking instructor for the first time and most were experiencing their first English-only class. The aims of the course were a) to equip students with the practical skills needed to interact effectively in English and b) to introduce students to becoming more active learners. The course syllabus was anchored by the functional skills featured in a popular commercial EFL textbook targeted for the Japanese market. The learning objectives from each text unit were further adapted into “can do” statements (Jones, 2002), similar to what Oskarsson (1980, p.45) introduces as a “situation outline” and “task specification”. Piloting the survey confirmed that L1 support was necessary for comprehension of the language. A questionnaire containing 30 “can do” statements was then drawn up in both English and Japanese (Appendices 1 & 2). The self-assessment was administered during class time at the beginning and end of the term with responses recorded on a 5-point Likert scale. The learners were also clearly informed that the self-assessment was for research purposes only and would not form part of their class grade.

Results

Student performance on the “can do” statements was monitored and assessed by the instructor in three ways. First, participation and the level of communication demonstrated in paired and small group conversations tasks and role plays were observed. These observations followed introductions to the communicative function, key vocabulary, grammar forms,

together with listening in context and practice sessions using the target language. An activity was conducted the following week to recycle the communicative purpose in addition to review sessions for the six units covered at the mid-point and end of the 14-week term. Second, an oral test provided by the publisher in the Teachers’ Guide was used. Students selected their own partner from among their classmates and, much like the learning process and tasks in class, engaged in a question and answer and follow up dialogue on various topics from the course, requiring a demonstration of language on the performance target measures. Finally, brief interviews were conducted with a cross-section of learners who made themselves available to discuss the process and to provide feedback on the use of the criterion-referenced scale for self-assessment.

Table 1. Hi-Lo range for performance clusters by individual class

| | Class code | Class size = n | Cluster 1 students/ percentage | Cluster 2 students/ percentage | Cluster 3 students/ percentage |
|---------------|------------|----------------|--------------------------------|---------------------------------|--------------------------------|
| Year 1 | 1 - A | 21 | 3 = 14.3% | 11 = 52.3% (*2) | 5 = 23.8% |
| | 1 - B | 20 | 4 = 20.0% | 8 = 40.0% (*3) | 5 = 25.0% |
| | 3 - C | 22 | 3 = 13.6% | 9 = 40.9% (*4) | 6 = 27.3% |
| | 3 - D | 20 | 2 = 10.0% | 8 = 40.0% (*5) | 5 = 25.5% |
| Year 2 | 1 - E | 22 | 6 = 27.3% | 9 = 40.9% (*2) | 5 = 22.7% |
| | 1 - F | 24 | 3 = 12.5% | 10 = 41.7% (*5) | 6 = 25.0% |
| | 3 - G | 22 | 3 = 13.6% | 14 = 63.6% (*0) | 5 = 22.7% |
| | 3 - H | 23 | 3 = 13.0% | 10 = 43.5% (*5) | 5 = 21.4% |
| Hi - Lo Range | | | 10.0 - 27.3% | 40.0 - 63.6% *14.6% unusable | 21.4 - 27.3 % |

Although the study design lacked the rigor to examine quantitative measures with reliability, in summarizing the results more qualitatively, three distinct groups were seen to emerge (See Table 1). The first group, ranging from 10.0% - 27.3 % of the members of a given class, was generally made up of the learners with the greatest language proficiency, yet not necessarily the highest communicative competence as shown in their output or active participation performance. An interesting, possibly social phenomenon, is that generally the learners in this group did not rate their self-assessment as highly as they were rated in observations of performance recorded by the instructor. Further investigation would need to be conducted to determine if this variance results from the learner interpretation of the self-assessment scale, a general self-effacing character, or possibly factors of anxiety or unfamiliarity over autonomous subjective performance. This final factor may be significant in the Japanese context and others similar where performance has previously been routinely assessed through an objective measure of accuracy in the foreign language classes before the college level. This socio-educational context could possibly lead to a skewed orientation of performance hinging on right or wrong answers. Also, in junior and senior high school in Japan, oral output for assessment is very rarely required as it was in this Oral Communication college class.

The second group contained the bulk of the learners (40-63.6%). Generally this group was made up of students showing a moderate level of proficiency both in class participation and in the oral examination. There were observable cases where in the comfort of the classroom the learner demonstrated a higher level of proficiency than

demonstrated in the oral exam. Again, the issue of possible examination anxiety and stress is of interest for further study. Some of these cases also corresponded to higher self-assessment ratings, completed outside of situations requiring use of the target language. Therefore the ability seems to exist to understand the metalanguage of the “can do” performance statements and have some cognitive awareness of potential performance criteria. As noticed in the walk-arounds during the completion of the self-assessments, some learners were even capable of softly verbalizing possible language forms for the task. The nature of the transfer of this cognitive capacity into performance as it relates to the confidence to try, the willingness to communicate, and overcoming foreign language anxiety is beyond the scope of this paper but remains a most salient point for investigation. Although names were asked to be included on the self-assessment forms to potentially allow for pre- and post-comparisons, an unexpected percentage of these learners (14.6%) provided unusable self-assessments. While other issues concerning limitations and constraints will be discussed below, upon further follow up with some of the learners directly, this factor could only be attributed to the immaturity of the responders.

The third group was generally made up of the lowest proficiency students and, over time, observation pointed to them as the least motivated learners. Potentially negative past experiences in the foreign language classroom jaded some of these individuals, or a mismatch in their learning styles and the typical transmission method for memory work caused them to be labeled as low performers. Although class activities included a wide range more suited to various

learning styles, one 14-week term is very little time to counter the influence of their past experiences. Although the future work goals and dreams of several of these learners seemed to rival those of their more motivated peers, they generally believed that English would not play a significant role in their future. Whether this was convenient rationale or blind optimism would involve a deeper level interview that neither time nor professional credentials would allow as quite clearly maturity and issues beyond academics were involved in some of these cases. The self-assessment scores for these learners tracked in the mid-upper range for some of the easier and potentially previously covered statements of performance and generally mid-low range for the new items covered during the course. The lower percentage of unusable surveys than the second group above was initially a hopeful sign. However, observation of the individual ratings varied little in pre and post self-assessments so it is possible that little conscious thought was actually being given. The performances in class and in the oral examination of the 21.4-27.7% in this cluster support the claims in the research that motivation and need are critical factors for learning (Dörnyei, 2001). In addition, some of the classes contained a returnee student or a foreign student with already highly developed English skills who were required to take the course for credit but were not considered for the purpose of this study.

Discussion

In the push to more learner-directed methods, a tension naturally exists between involving learners in assessment and the reliability of such assessments. Much of the nature

of assessment focused on reliability relates to what Brindley (1989) refers to as “objective externally administered instruments” (p.61). Such methods of assessment, often in the form of tests, require a) knowledge of learning goals, b) an understanding of linguistic and pragmatic output to realize these goals, and c) criteria and methods for evaluation (see Tudor, 1996, p.164). Executing such professional skills with reliability is beyond many learners as certainly is the case in this context. However the purpose of the self-assessment in this case is to improve awareness through a set of criteria that the learner can define internally as part of their learning objectives. Self-assessment by nature involves a shift in the role of the instructor and the student, but providing that the technique is taken as part of the learning process and not for certification, demanding objectivity, then criterion-referenced self-assessments can be a useful tool.

Institutional factors can of course become constraints, such as the changes to the curriculum which occurred between year 1 and 2 of this study in effect eliminating the possibility of following up with the 1st year students in their 3rd year. Limited term work contracts also did not permit ongoing assessments with the progressive curriculum of one commercial text series together with target language recycling.

Future Directions

In closing, it can be seen that additional training in self-assessment and the use of the survey will improve the input from the learners. For improved awareness and consistency in interpretation of the scale, a rubric could be developed for general course objectives. For the specific unit objectives, it

may be more sensible to simply use a 3-point scale to clarify the meaning of *can* among students of various personalities, self-image, or cultures. In the future, the 30 item survey could again be conducted at the start of the course as part of a needs analysis and level check. However, to improve on the formative nature of the assessments, it is recommended to conduct the assessment weekly for the two-four specific learning objectives in each unit. If resources and student access would allow for it, this could also be conducted online for improved ease of processing the data. Lastly, since academic institutions need results to grade their students it is important to have additional forms of authentic assessment of learner performance. These results should be triangulated with instructor observations of performance and learner feedback on the criterion-referenced self-assessment for ongoing improvement.

References

- Brindley, G. (1989). *Assessing achievement in the learner-centred curriculum*. Sydney, AUS: National Centre for English Language Teaching and Research.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.
- Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge: CUP.
- Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge: CUP.
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal*, 70(2), 125-132.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: CUP.
- Jones, N. (2002). Relating the ALTE Framework to the Common European Framework of Reference. In Council of Europe (Eds.), *Case studies on the use of the Common European Framework of Reference*. Retrieved April 6, 2007, from <www.alte.org/can_do/index.php>.
- Oskarsson, M. (1980). *Approaches to self-assessment in foreign language learning*. Oxford: Pergamon.
- Tudor, I. (1996). *Learner-centredness as language education*. Cambridge: CUP.