Instituting Exit Competencies: Procedures, Problems and Future Developments

Abstract

For the past few decades the field of second language acquisition pedagogy has seen the increasing popularity of a communicative approach within the classroom. This has led to many educational institutions becoming desirous of shifting to a competency based assessment system in order to better compliment their curriculum. While advances have been made, the inherent difficulty in defining and capturing the tangible, essential constitutes of communication for the purposes of evaluation has proved incredibly difficult. This paper reports on the results of a series of trials that aimed at establishing an exit competency system of assessment at a foreign language university in Japan. Explicated will be the various challenges faced and the results encountered so as to inform other language institutions interested in implementing a similar system, as well as adding to the general discussion and debate on the issue.

Introduction and background to exit competencies

Exit Competencies, in this paper, are defined as a series of objectives that outline student’s ability to convey meaning in order to achieve a goal--an inventory of what they ‘can do’ in a communicative sense. According to Brindley (1994: 41), “competency-based approaches to curriculum and assessment are beginning to exert a major influence of teaching practice”. Competency-based approaches in the assessment of the language learner have become increasingly common in language testing around the world, with several large-scale competency-based assessment criteria in use, such as the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2007) in Europe, the ACTFL performance guidelines for K-12 learners (ACTFL, 1999) in the United States, and the ESL Scales (Curriculum Corporation, 1994) in Australia. Adopting a competency-based approach to assessment has the potential benefits of rewarding particular achievements among students, contributing to the raising of standards, working
towards ensuring consistency of benchmarks across different classes and as well as empowering students to take more responsibility for their learning by providing an ongoing assessment of achievement (Brindley, 1989; Docking, 1994). This approach to assessment is complementary with the current movement of communicative language teaching and task-based approaches to learning, which center on developing communicative language competency rather than knowledge of the language alone.

For these reasons, a system of exit competencies was developed and trialed at a university in Japan, to better reflect teaching practices and curriculum within the university. This paper aims at reporting on the results of this trial. First, the context of the university will be explained, to better understand the rationale for the development of such a system. Then, the method of the study will be explained followed by the results with interpretation.

**Context of the trial**

The university in which the trial took place is a language-specialist university in the greater Tokyo area and one which holds held a reputation for innovative directions in language education, as evidenced by awards given by the Ministry of Education. The university has long adopted a communicative and task-based approach to language education, with the curriculum at the university centered on tasks grouped under general themes of content, rather than on grammatical, functional or situational language points. In addition, students were assessed by their completion of a variety of tasks, involving work such as presentations, the scripting, acting and filming of a commercial or adapted movie, reading and writing tasks, and so forth. Despite this innovative and modern curriculum, the university continued to operate under a traditional system of evaluation based on formal knowledge of the language. This was problematic as it was subjective in that there was no formal ‘standard’ by which grades were administered, it lacked unity as there was no agreement as to what specific criteria was considered when evaluating, and, similar to other universities in Japan, progressed students from basic to advanced courses on the basis of years spent in the system rather than on the grounds of their proficiency or competency. The university investigated ways to restructure the system of
evaluation and it was decided that a competency-based assessment should be pursued for the following reasons:

1. It makes a public statement about a student’s language ability and achievement.
2. It allows students to progress through the curriculum in accordance with their demonstrated language achievement.
3. It accommodates differing levels of proficiency.
4. It provides a statement of accountability on curriculum design and efficacy.
5. It enables students and teachers to become aware of course objectives.
6. It allows students to become aware of their own strengths and weaknesses.
7. It encourages more engagement from students in classes due their realization of accountability.

(Basic English Proficiency Project Report, 2007)

In order to better develop well rounded, high-level language users, it became the aim of the English curriculum at the university for all students to attain a stated level of ability expressed in terms of:

1. A level of language **proficiency** – which was measured by a standardized proficiency test similar to TOEFL or IELTS.
2. A measure of language **competency** – which was to be measured by a series of tasks designed to demonstrate students’ competency in using the language.
3. A measure of language **achievement** – measured by performance in enrolled courses at the university.

Such distinctions are consistent with Brindley’s (1989: 14-15) original conceptualization of competency based assessment where three levels of competency were envisaged: (1) Level one – achievement in overall proficiency, measured by a proficiency test; (2) level two - achievement in functional proficiency, measured by performance on a series of tasks; and (3) level three – achievement in structural proficiency, measured by attainment of goals set out in a particular course. In Brindley’s conceptualization, language proficiency (level one) would continue to be measured by a proficiency test called the Kanda English Proficiency Test, which
was already in place. Functional proficiency, from here on referred to as language competency (level two), would be measured by a new system, which would be developed and trialed. Language achievement (level three), which is the most informal and subjectively evaluated of the three levels, would continue to be measured by student performance in enrolled courses, in terms of feedback on assignments and class tasks. Thus, of the three levels, a system concerned with assessing language competency (level two), had yet to be created and was the aim of the current investigation.

**Methodology**

The investigation of a competency-based assessment system was carried out as an action-based research project. According to Nunan (1994), action research is a form of research that is becoming increasingly significant in language education. He states that in the area of language education it involves research that is carried out by practitioners, or classroom teachers, rather than by outside investigators (Nunan 1994). According to Kemmis and McTaggart, who were the early advocates and formulators of action research:

> A distinctive feature of action research is that those affected by planned changes have the primary responsibility for deciding on the course of critically informed action which seem likely to lead to improvement, and for evaluating the results of strategies tried out in practice. Action research is a group activity. (1988: 6)

Action research was deemed the most appropriate methodological framework to use because it allowed the researchers to trial a number of systems within real learning environments. In addition, another benefit of conducting the project as action research was that the “practitioner as researcher” role takes advantage of knowledge of an understanding and appreciation of class dynamics and student identity that may go unnoticed by an outside researcher. Such knowledge is vital when dealing with a study involving the trialing of a system in regards to its feasibility of use in real classroom settings and its accuracy and appropriateness at measuring student competency.

The methodology used in the current research uses a typical action research framework to investigate methods of competency based assessment, through a
series of action research cycles. Each cycle consists of the following processes: system development, trial, data collection and system reform based on results of the trial (see appendix one). The research project has completed three cycles since its commencement in 2005. One trial was carried out outside of the classroom environment in December 2005. The second trial was carried out in two first year English skills classes from April to June 2006. The third trial was carried out from September to December 2006 with eight first year English skills classes.

**Results**

Each of the three trials yielded information through interviews with teachers on the feasibility of using the system in class and on its perceived validity and appropriateness, leading to the further and on-going refinement of the system itself (see appendix 2 for summary of the results of each cycle).

**Results of First Cycle**

In the first cycle, an evaluation system consisting of a number of tasks that would be used to elicit certain competencies from students was tested. In the initial development of the system, it was thought that two-assessors evaluating students’ performing a task would increase the objectivity of the evaluations (See North, 1993, for discussion). However, after consultation with the university management and administration, it was decided that a system that used multiple assessors was impractical for the university setting as it required too much manpower, and that an evaluation that occurred outside the classroom might draw students’ focus away from the in-class curriculum to that of a preoccupation with the external evaluation. Therefore, the first system to be trialed was a compromise in that a single teacher would observe students’ completion of a pre-designed, standard task. A bank of six tasks was created, designed to elicit a wide range of communicative competencies from the students. One task, where a student was required to read an article and then give a presentation on the key points raised in the material was performed in a controlled environment. Four students of varying language proficiency were chosen to take part in this trial – one advanced level student, two mid-range, and one lower level student as measured by their performance on the Kanda English
Proficiency Test. The tasks were performed in small groups with one student presenting to three students. The audience was a key element to this system of evaluation, as they became part of the evaluation process. It was thought that it would be possible to judge whether a student was able to competently present key points raised in the article by appraising the audience’s understanding of these key points after the presentation in a test. The results of this test could be used to verify whether there had been an authentic exchange of meaning, and further complement the teacher’s evaluation. This would in turn improve inter-rater reliability, which is a major concern for competency-based assessment carried out in classroom settings (Brindley, 1994). Furthermore, in order to stress the importance of communication for the presenter as well as the audience, the participants were told that the result of the test completed by the audience would form part of the presenter’s final score. The tasks were carried out and video recorded. Data was collected in the form of teacher evaluations of the students’ performance of the task, researchers’ evaluations of the students’ performance of the task after watching the video recordings, and the audience’s comprehension tests.

Data collected in the form of feedback from the assessors of these classes indicated some problematic areas in the system. According to both teachers and verified by three researchers who watched the filmed tasks, two students were evaluated as having competently completed the task according to the criteria. However, data which tested audience understanding of the presentation indicated that the task was more successfully completed by the two presenters who did not satisfactorily meet the competency criteria. The reason for this is that audience members’ understanding of the subject matter relied on the listening comprehension and thus language proficiency of each audience member. Some audience members were unable to answer questions related to the key points of the reading, even when included in the presentation, and so could be a reflection of the audience’s competency rather than that of the presenter. In addition, more proficient audience members were often able to infer the answers to the test from the presentation, even though the presenter had failed to competently communicate these ideas. This result indicated that audience understanding was not a reliable method to assess a performer's competency as the proficiency of the audience played too large a role.
Thus, in the refinement of the system at the end of the first trial, the aspect of audience assessment was abandoned.

The video recording of performers in the first trial provided invaluable data into the nature of what competencies teachers expect students to display in order to complete a task successfully. The recording would be used in future trials during rater training sessions to ensure inter-rater agreement on the standard of performance required to pass the competency based assessment.

Results of Second Cycle

In the second cycle, the results of the first cycle and teacher feedback led to the creation of a list detailing the competencies a student should possess after completing the existing course. These were then mapped to certain in-class tasks where competencies could be observed. For example, presentation skills competencies were mapped to rubrics used to assess performance ability in presentations (appendix three). The underlying concept of this procedure was this appraisal would take place in class, the teacher would be the sole rater of competencies, and that competencies would be woven into existing tasks so as to maintain a focus on the curriculum. In addition, by linking competencies to a standardized rubric that all teachers used, it was hoped that subjectivity could be partially minimized.

Data in the form of a number of interviews throughout the trial period was collected. Some concerns emerged in this refined version of the system. Firstly, due to the nature of some of the tasks in the curriculum, teachers found it difficult to apply the standardized rubric. For example, while some presentation tasks were traditional in that the presenter desired to impart information to the audience, others were more of a performance nature, such as the 'selling' of a country to potential vacationers, or making a film adaptation of a fairy tale. These performances included more than merely imparting information, and so it became unclear to teachers whether students had fulfilled certain competencies according to the rubric. In addition, teachers became concerned that a chosen task in the curriculum might not provide sufficient opportunity for students to display their language
competency. This concern is mirrored in research into performance assessment that competency based assessment appears to be highly task dependent (see, for example, Brandt 1992). According to Brindley (1994: 50), “this [concern] underlines the necessity of using multiple tasks and multiple assessment formats”. As a consequence to the problem that a certain task may not provide sufficient opportunity for students to display certain language competencies (such as in one highly contextualized presentation) the use of exit competencies mapped to single assessment items was abandoned and a more holistic approach to competency based assessment was sought where teachers were at more liberty to observe a students’ competency over a wider range of tasks and time.

Results of Third Cycle

In the third cycle of the study, instructors utilized a competency checklist in observations of performance four times a year. Of all the previous systems tested, this was the most holistic, involving teacher judgment to evaluate many of the competencies (see appendix 4 for the competency checklist). However, in order to maintain some partial level of objectivity, teachers were asked to fill out competencies as much as possible based on real observations of student achievement in class, rather than inferences made based on subjective recollections of achievement. This system was tested with eight classes over a semester-long period and data was collected through interviews throughout the semester with each of the teachers. This was by far the biggest trial of the three systems, because it was felt that as the system was being refined, a larger trial was necessary to highlight issues of feasibility and validity of introducing the system into all classes. It was felt progress was being made in this the third trial and teachers were genuinely positive concerning the introduction of this system. Despite limited success, further refinement and testing is required. Some concerns that need to be addressed can be found below along with suggested improvements that can be made in the future development of the system.

1. Unobservable competencies.
   i) Teachers indicated difficulty in accurately assessing certain competencies. For instance, teachers found it very problematic to assess writing and reading as there was little emphasis on the
assessment of these skills in the course in which the system was tested, which focused more on students’ verbal communication. It was suggested that in future trials, material read and written for the purpose of the presentation by the student might be collected and used to form part of the writing and reading score.

ii) Teachers also mentioned that other competencies involving students’ management of learning and learner autonomy were difficult to define and therefore observe. Other modes of assessment need to be experimented with in order to better discover whether certain competencies have been achieved. For example, filming of classes, or considering student reflections when grading final projects/presentations could occur. It is recommended a portfolio of a student’s work be amassed in an attempt to evaluate intangibles such as autonomy and independence. A portfolio would also increase objectivity as a means for teachers to collaborate and be more accountable in their evaluations.

2. **Time constraints.**
   1) Teachers found it difficult to assess all the students in their class during the performance of certain tasks. For example, some tasks such as discussions, would be taking place simultaneously and thus it was impossible for the teacher to assess all students during the performance of that task. Furthermore, even in cases of presentations, teachers commented it was difficult to judge achievement of certain language competencies from some of the more unusual presentations that occur in the curriculum. Despite these comments, teachers went on to say these concerns were slightly mitigated as the evaluation took place over a semester, and granted leeway to teacher judgment and inference. It is thought that for future trials, the current competency rubric can be further modified to better reflect the existing curriculum.

3. **Subjectivity.**
   1) Interviews revealed it was difficult for teachers to agree on what constituted a competent student, i.e., students who were evaluated as competent by some teachers may not have been evaluated as competent by others. It is the hope that this problem can be circumvented with increased teacher norming of teachers prior to future trials.

**Conclusion and implications for future research**

A system of competency based assessment could have many benefits for the English course at the Kanda university and other educational institutions in Japan. Although there are logistical difficulties, it has potential advantage in terms of positive learning outcomes for its students and in the long term by developing a reputation for producing students of a higher standard. Furthermore, through the
trial of various systems, teachers and students involved have become more aware of desired outcomes of assessment tasks and of the English curriculum in totality. This system of outcomes has helped indicate tasks that were hindering students from displaying their language competencies that has led to the refinement and general improvement of the curriculum. In terms of the systems trialed, the third and more holistic approach was shown to be most successful in this particular institution, however other systems may be more applicable to other learning contexts. If the more holistic approach is pursued, other measures of accountability need to exist, one possibility being the creation of student learning portfolios.

While impediments do exist, the establishment of an accurate, reliable communicative assessment system is achievable and of benefit to all levels to the educational institution interested. Further development can bring about this future where a curriculum is conducted and is evaluated in, a communicative fashion. It is a hope that future research can build upon the findings of these initial trials to bring a reliable system of competency-based assessment closer to realization.
References


Appendix 1: Research methodology

1. System Development
2. Trial
3. Data Collection
4. System Reform
## Appendix 2: A summary of each system on trial in each of the cycles

<table>
<thead>
<tr>
<th>Cycle</th>
<th>System Development</th>
<th>Trial</th>
<th>Data Collection</th>
<th>System reform</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYCLE 1</td>
<td>A system was developed to evaluate students' competency through a series of tasks to be completed outside of class. Competency was measured through peer understanding and response to the tasks in conjunction with teachers' assessment of the student's performance.</td>
<td>The system was trialed with six groups of students from differing levels as defined by the KEPT test. They were given the task of reading an article on a given topic and asked to prepare a short presentation that highlighted the key information in the article.</td>
<td>Three different raters assessed each presentation. The presentation was videoed and analysed so that a clearer understanding of the target competencies was obtained.</td>
<td>The trial suggested that as a system involving multiple raters was a strain on resources and would be impossible to implement outside of the classroom setting, one that was adapted to a regular class setting would be of more benefit.</td>
</tr>
<tr>
<td>CYCLE 2</td>
<td>Target competencies were developed from the previous cycle’s results and through teacher interviews and questionnaires. A system was developed in which competencies would be assessed during student performance of tasks in their language classroom. For example, some exit competencies for spoken communication were evaluated during students’ in-class presentations.</td>
<td>The system was trialed in two classes of average proficiency level over a semester. During this time, teachers assessed students’ language competency by observing in-class discussions and presentations utilizing similar criteria for assessment. [This criteria was derived from data and interviews from instructors on criteria they felt students should possess upon graduation from the class.]</td>
<td>Teachers were interviewed 3 times throughout the semester to gain feedback on the system, especially in regards to the feasibility of using such a system within the current curriculum.</td>
<td>Interviews with teachers of each of the classes highlighted flaws in the usability of the system. Among these were that not all competencies were observable in the language task. Also, the subjectivity concerning what denoted ‘competent’ was noted as a problem.</td>
</tr>
<tr>
<td>CYCLE 3</td>
<td>As the second trial suggested concerns of subjectivity, operational definitions and examples of competencies were added to assessment sheets. Also, to further enhance usability, a more holistic approach to observing competencies was developed, where teachers evaluated competencies over a wide range of tasks over a given time.</td>
<td>The system was trialed in 8 classes (2 from each proficiency tier) over a whole semester. Teachers in these classes assessed students in a more holistic manner through observations of regular performance in class activities and assessment tasks.</td>
<td>Teachers were interviewed three times throughout the semester to get feedback on the system.</td>
<td>Interviews with teachers suggested usability of the system had been improved, however, issues of subjectivity were still a concern.</td>
</tr>
</tbody>
</table>
Appendix 3: An example of language competencies mapped to an assessment of a presentation task, used in the second cycle.

<table>
<thead>
<tr>
<th>Exit Competencies (Check all that apply)</th>
<th>F</th>
<th>C</th>
<th>B</th>
<th>A</th>
<th>A+</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content</strong></td>
<td>Has not provided enough information to satisfactorily complete the task.</td>
<td>Has given enough content to satisfactorily finish the task. Could have dealt with topic in more depth / with better information.</td>
<td>Appropriate content on topic. Could have been improved with further research of relevant information from different sources.</td>
<td>Relevant, well-researched and well-selected content from a number of sources. Has dealt with topic in depth.</td>
<td>Shows excellent research skills, selecting relevant information from a variety of sources to discuss topic with originality and depth.</td>
</tr>
<tr>
<td><strong>Presentation Skills</strong></td>
<td>Very little eye contact. Shows no awareness of voice, eye contact and body language in oral presentations.</td>
<td>Has shown some awareness of voice, eye contact and body language in oral presentations, but needs to improve these skills.</td>
<td>Uses appropriate eye contact. Shows awareness of voice and body language in oral communication, but somewhat unnatural delivery.</td>
<td>Makes eye contact and uses body language to communicate. Good voice / delivery of information, but may not always be perfectly executed.</td>
<td>Has mastered presentation skills in terms of voice, natural delivery of information and body language.</td>
</tr>
<tr>
<td><strong>Language Competence</strong></td>
<td>Grammar and lexical errors frequently interfere with meaning. Lexical and grammatical knowledge insufficient to complete task satisfactorily.</td>
<td>Frequent grammatical and lexical errors, especially in late-acquired grammar. Language is competent enough to satisfactorily complete the task.</td>
<td>Grammatical and lexical errors may be present, but does not interfere with meaning. Pronunciation and intonation not always clearly executed.</td>
<td>Speaks with few grammatical and lexical errors. Clear pronunciation and intonation. Somewhat complex language used.</td>
<td>Speaks with very few grammatical and lexical errors. Complex language used. Clear pronunciation and intonation. Negotiates meaning.</td>
</tr>
<tr>
<td><strong>Organization</strong></td>
<td>Little evidence of preparation and practice, resulting in a disorganised and ineffective presentation.</td>
<td>Information is organised into key areas / concepts. Presentation needed more preparation and practice to really be effective.</td>
<td>Information is well organised according to presentation conventions. Presentation was well practiced and prepared, but some aspects could have been improved.</td>
<td>Information is well organised, leading to relevant conclusion. Evidence of a good deal of preparation and practice resulting in a well-organised and presented presentation.</td>
<td>Information is masterfully organized, leading to a relevant and original conclusion. Evidence of extensive preparation and practice.</td>
</tr>
<tr>
<td></td>
<td>Can interpret, organize, summarize/paraphrase information from a number of sources</td>
<td>Can respond to questions on their presentation</td>
<td>Can make persuasive, informative presentations with natural delivery. Can use voice to stress importance, interest and to clarify information.</td>
<td>Can make effective eye contact and gestures to communicate information</td>
<td>Can speak fluently without long pauses Can speak with clear pronunciation Speaks with few grammatical/lexical errors. Can negotiate meaning</td>
</tr>
<tr>
<td></td>
<td>Can research to find relevant information for a task. Displays a concrete knowledge of the organization of a presentation. Shows co-operative learning in group presentations</td>
<td>Speaks within time limits.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 4: The more holistic language competency checklist, used in the third cycle.