

Authentic Assessment: Implications for EFL Performance Testing in Korea

Dr. Andrew E. Finch
(Kyungpook National University)

Finch, A. E. (2002). Authentic assessment: Implications for EFL performance testing in Korea. *Secondary Education Research, 49, 89 - 122.*

This paper examines current theory and practice regarding the assessment of foreign language oral performance, and discusses implications for curriculum designers and teachers in tertiary learning institutions in the republic of Korea. In addition to suggesting that norm-referenced assessment be replaced with criteria-referenced, "authentic" assessment in these establishments, the first part of the paper also concludes that the current grading of students in National Universities according to a prescribed bell-curve is inappropriate for language assessment, being intrinsically demotivating for students and teachers. It is therefore argued that tertiary English programs should act on recent research findings and government policy statements by promoting positive affect (attitudes, beliefs, confidence, motivation, etc.) in non-threatening learning environments, using criterion-referenced, authentic assessment.

The second part of this paper recognizes two modes of authentic assessment (self- and peer-assessment) as reliable and valid methods of evaluation, particularly suitable for assessment of oral skills at tertiary level. These reflective models encourage students to become involved in their learning, and promote positive attitude change in the fostering of life-long learning skills and socially responsible citizens. It is suggested, therefore, that when employed in a student-centered, holistic setting, self- and peer-assessment are practical and effective evaluation tools for tertiary language education.

I. INTRODUCTION

A systematic testing component is an essential part of every language program and of most language classrooms (despite the fact that many teachers feel intimidated by the terminology and use of statistical concepts, Brown, 1995, p. 12), being used to measure language aptitude, proficiency, placement, diagnosis, progress, and achievement, and providing feedback for the program evaluator(s), washback information for teachers and students, and motivational washforward implications for all concerned. However, the field of language testing in general and of performance testing in particular, is fraught with problems of theory and practice. Before discussing appropriate evaluation models for university English programs in Korea, therefore, it is appropriate at this point to view a brief survey of language assessment research.

II. ORAL TESTING: HISTORY OF RESEARCH

1. Beginnings

Defining a test as "a systematic method of eliciting performance which is intended to be the basis for some sort of decision making" (Skehan, 1998, p. 153), Skehan notes the tendency of testers to

place an emphasis on "care and standardization in assessment in the belief that such methods of examining performance will have more to contribute to reliable measurement than informal assessment by people who may be very familiar with particular language users" (Skehan, 1998, p. 153). This attitude follows on from the assumption that "there are knowable best ways of learning and that these can be discovered using a scientific method which has long been discarded by contemporary philosophers (Popper), scientists (Medawar) and physicists (Heisenberg)" (Harri-Augstein & Thomas, 1991, p. 7), and has been at the heart of language testing from its "pre-scientific" stage (Spolsky, 1975, p. 148), to its psychometric-structuralist "scientific" stage (when discrete-point testing represented the accepted behaviorist truth). According to this view, language can be learned by studying its parts in isolation, acquisition of these parts can be tested and will successfully predict performance levels, and the learner will somehow reconstruct the parts in meaningful situations when necessary. This view continued into the "psycholinguistic-sociolinguistic" stage (the 1970's), when integrative testing (e.g. cloze tests and dictation) claimed to come from a sounder theoretical base (Oller, 1979) but was shown by commentators such as Alderson (1981), Morrow (1979) and Carroll (1981, p. 9) to be still concerned with *usage* rather than *use*, therefore being only indirect tests of potential efficiency. Kelly (1978, pp. 245-246) also pointed out that it is possible to develop proficiency in the integrative test itself, and that indirect tests cannot diagnose specific areas of difficulty in relation to the authentic task. Such tests can only supply information on a candidate's linguistic competence, and have nothing to offer in terms of performance ability (Weir, 1998).

A consensus that "knowledge of the elements of a language in fact counts for nothing unless the user is able to combine them in new and appropriate ways to meet the linguistic demands of the situation in which he wishes to use the language" (Morrow, 1979, p. 145), and an acknowledgement that the easily quantifiable, reliable, and efficient data obtained from discrete (and cloze) testing implies that proficiency is neatly quantifiable in such a fashion (Oller, 1979, p. 212), led to a perception that the ability to perform should be tested in a specified socio-linguistic setting. Based on work by Hymes (1972), Canale & Swain (1980), and Morrow (1979), the emphasis shifted from linguistic accuracy to the ability to function effectively through language in particular contexts of situation (a demonstration of competence and of the ability to use this competence), and communicative testing was adopted as a means of assessing language acquisition (though with some lack of initial agreement or direction, cf. McClean 1995, p. 137; Benson, 1991).

The components of communicative language ability to be tested were variously described at this time (Canale & Swain, 1980; Canale, 1983), and early frameworks for testing communicative competence were proposed. However, these were neither practical, systematic, nor comprehensive (cf. Cziko, 1984), and were unable to advance prediction and generalization in any substantial way. This problem was addressed by Bachman (1990) through the application of categories to real contexts, and resulted in a model of oral testing which was: i) more detailed in its specification of component language competences; ii) more precise in the interrelationships between the different component competences; iii) more grounded in contemporary linguistic theory; and iv) more empirically based, allowing a more effective mapping of components of competence on to language use situations, and more principled comparisons of those components. Despite these improvements, however, Bachman's model still lacked a "rationale founded in psycholinguistic mechanisms and processes (and research findings) which can enable [it to] make functional statements about the nature of performance and the way it is grounded in competence" (Skehan, 1998, p. 164). Skehan (1988) articulated the dilemma of communicative language testing at the end of the 1980s:

What we need is a theory which guides and predicts how an underlying communicative competence is manifested in actual performance; how situations are related to one another, how competence can be assessed by examples of performance on actual tests; what components communicative competence actually has; and how these interrelate. Since such definitive theories do not exist, testers have to do the best they can with such theories as are available. (Skehan, 1988, cited in Weir, 1998, p. 7)

2. Task-based oral testing

Bachman's (1990) model used familiar empirical research methods in which data was perceived in terms of the underlying structural model, to infer abilities, via a static picture of proficiency, based on the assumption that there are competence-oriented underlying abilities made up of different interacting components (cf. Canale & Swain, 1980). However, cognitive theory shows that second language performers, faced with a developing inter-language and performance pressures such as fluency, accuracy and complexity, do not draw upon "a generalized and stable underlying competence", (Skehan, 1998, p. 169) but allocate limited processing attention in appropriate ways, drawing on parallel coding systems for efficiency of real-time communication. Skehan therefore proposed a construct of "ability for use", which would allow a processing competence to operate and to be assessed, and advocated the use of tasks as a central unit within a testing context (Skehan, 1998, p. 169). In contrast to performance evaluations which use unreliable analytic scales (in areas such as grammar, vocabulary, fluency, appropriateness, and pronunciation) but which do not allow for affect and for competing demands on attention, a processing approach in a task-based framework allows generalizations to be made on the three basic language-sampling issues of: i) fluency; ii) breadth/complexity of language used; and iii) accuracy (Skehan, 1998, p. 177), though these criteria compete for processing resources in the performer, and the score may be influenced by whichever processing goals are emphasized by him/her.

While advocating tasks as the basic unit of oral testing (c.f. Lee, 1991), Skehan notes that "we need to know more about the way tasks themselves influence (and constrain) performance" (Skehan, 1998, p. 169), and that tasks also need to be rated in terms of planning, time pressure, modality, stakes, opportunity for control, manufactured surprise, and degree of support, since these factors will also affect the outcome. Task performance conditions and the way these affect performance represent "a fertile area for research" (Skehan, 1998, p. 177).

3. Authentic assessment

Kohonen (1999) extended Skehan's task-based framework, proposing "authentic assessment" as a process-oriented means of evaluating communicative competence, cognitive abilities and affective learning (Kohonen, 1999, p. 284; cf. Hart, 1994, p. 9; O'Malley & Pierce, 1996, pp. x-6), using reflective forms of assessment in instructionally-relevant classroom activities (communicative performance assessment, language portfolios and self-assessment), and focusing on curriculum goals, enhancement of individual competence and integration of instruction and assessment. In this two-way process, "the essentially interactive nature of learning is extended to the process of assessment" (Williams & Burden, 1997, p. 42), examining what learners can do with their language, through real-life language use tasks (cf. Weir, 1998, p. 9). For the learner this means developing reflective

awareness through self-assessment and peer assessment, learning "how to manage learning , rather than just managing to learn" (Williams & Burden, 1997, p. 291). For the teacher (whose professional judgement and commitment to enhancing student learning is an important part of this process), authentic assessment means collecting information about learner progress and the social learning environment in the class, along with a re-assessment of classroom roles and responsibilities. Such a teacher becomes a:

... tool-maker and provider, observer and joint interpreter of the evolving conversational experiment in which both subject and [teacher] are full but different participants. ... Only the subject/learner can tap his or her personal *experience*, but the experimenter can observe *behaviour* and recruit methodological skills to drive the experiment forward. (Harri-Augstein & Thomas, 1991, p. 6)

Kohonen (1999) offers a list of 13 ways in which authentic assessment can enhance learning, and summarizes how this approach contrasts with standardized testing (Table I, below):

TABLE 1
Comparison of standardized testing and authentic assessment (Kohonen, 1999, p. 285).

	Standardized testing	Authentic testing
1	Testing and instruction are regarded as separate activities.	Assessment is an integral part of instruction.
2	Students are treated in a uniform way.	Each learner is treated as a unique person.
3	Decisions are based on single sets of data (test scores).	Provides multiple sources of data; a more informative view.
4	Emphasis on weakness/failures: what students cannot do.	Emphasis on strengths/progress: what learners can do.
5	One-shot exams	Ongoing assessment
6	Cultural/socio-economic status bias	More culture-fair
7	Focus on one 'right answer'	Possibility of several perspectives
8	Judgement without suggestions for improvement.	Useful information for improving/guiding learning
9	Pressures teachers to narrow teaching to what is tested).	Allows teachers to develop meaningful curricula.
10	Focus on lower-order knowledge and skills.	Emphasis on higher-order learning outcomes and thinking skills.
11	Forbids students to interact promotes comparisons between students (norm- referencing).	Encourages collaborative learning compares learners to their own past performances and the aims.
12	Intrinsic learning for a grade.	Extrinsic learning for its own sake.

As can be seen from this list, authentic assessment is a learning tool, providing evaluative information to both learners and teachers. Its focus on student-centered and student-managed ongoing assessment also reflects educational thought in other areas of language acquisition: collaborative learning (Vygotsky, 1978); individual learning styles and preferences (Bickley, 1989; Keefe, [Ed.],

1979; Reid, 1987); the importance of affect (Arnold, [Ed.], 1999); and the process syllabus (Breen, 1984). The authentic assessment model is thus particularly suitable for tertiary students, since it encourages them to gradually assume responsibility for their own learning and for the assessment of that learning, as the cycle of intention, action and reflection becomes a personal endeavor, facilitated by: portfolios, projects, self- and peer-assessment, learning conversations, and reflective journals. In addition, an institutional trust and respect for the learner (which must be implicit in this approach), sees him/her as an active and socially responsible agent, fully capable of needs analysis, goal setting, and assessment of achievement.

4. Validity/reliability

Before concluding this brief overview of oral performance testing, it is appropriate to mention that language testing has traditionally been limited by considerations of validity (whether tests actually measure what they are supposed to measure [Thrasher, 1984]), reliability (whether they produce similar results on more than one occasion), and efficiency (logistics of test administration) (Weir, 1998, p. 1). Validity (seen by Spolsky, 1975, and Messick, 1988, as the major problem in foreign language testing) includes content validity (the test is a representative sample of the language skills and structures it is meant to test), criterion-related validity, construct-validity (the extent to which the test matches a theoretical construct. Bachman, 1990), face-validity (the test looks reasonable to the test-taker), predictive validity (the predictive force of the test), concurrent-validity (the test and the criterion are administered at the same time. Davies, 1990), and educational validity (the relationship between positive test effects and students study habits. Thrasher, 1984). Nakamura (1995, p. 135) argues that predictive validity, educational validity, construct validity, concurrent validity, face validity and content validity should be analysed in tests of speaking ability, and Kohonen (1999, p. 291) also stresses validity in communicative evaluation.

Williams and Burden (1997), however, argue that the energy spent by test constructors on strengthening the reliability and validity of their tests so that they can be standardized, is largely misspent, since this assumes that the test is measuring a relatively fixed characteristic, rather than a hypothetical construct (the researcher's best attempt to define what is involved). In fact, individual- and affect-related traits are variable, and often context specific, such that "a test should be expected to produce different results on different occasions" (Williams & Burden, 1997, p. 90). As Kelly mentions (1955, p. 77), when a subject fails to meet the experimenter's expectations, all that can be said is that he/she has not conformed to those expectations or to the experimenter's definition of learning: "The point is that it is extremely difficult to construct a test which is truly valid in that it really measures what it is supposed to measure" (Kelly, 1955, p. 77). Weir (1998, p. 7) also points out that the validity of "communicative" tests is dependant on the test-constructor's understanding and definition of the term, and Van Lier (1996) goes deeper still into "accountability", noting that tests can only measure that which is measurable:

It is quite possible that the deepest, most satisfying aspects of achievement, and the most profound effects of education, both in positive and negative terms, are entirely unmeasurable. ... What if we held educators accountable for the quality of the memories they gave to their students, rather than for averages on national tests? (Van Lier, 1996, p. 120)

5. Summary

Oral language testing has evolved in a short time from a "physical science" approach (in which language learners are impersonal data) to a "personal science" (in which people explain themselves to themselves), and more recently, to a "conversational science" approach, based on the premise that the unique attribute of humans is that they converse. Testing which continues to concentrate on the "target-like appearance of forms" (Larsen-Freeman, 1997, p. 155) ignores the fact that "we have no mechanism for deciding which of the phenomena described or reported to be carried out by the learner are in fact those that lead to language acquisition" (Seliger, 1984, p. 37), as well as the fact that the learner's internal grammar is not a steady commodity and often deteriorates prior to internalizing new content. Even if we could identify and measure all of the factors in second language acquisition, complexity theory tells us that "we would still be unable to predict the outcome of their combination" (Larsen-Freeman, 1997, p. 157).

Psychologists and educators still know little about how language learning occurs, and why and how some individuals are more competent than others, so that it is inappropriate to define and test discrete symptoms of the process. However, observable factors that appear to be associated with learning include construction of meaning, sharing of experiences, identification of needs and purposes, critical evaluation of performance strategies, and awareness of this process (Harri-Augstein & Thomas, 1991, p. 7). These factors can be satisfactorily examined (from the point of view of both teacher and student) using reflective, authentic assessment methods in appropriate learner-centered classroom activities. Integrated into the day-to-day curriculum, assessment can become both a means and an end, and considerations of validity, reliability and efficiency cease to be a major issue in the ongoing reflective self-examination of language performance.

III. CRITERION-REFERENCED/NORM-REFERENCED TESTING

Authentic assessment in a task-based process setting implies a focus on language mastery (criterion-referenced performance) rather than relative performance (norm-referenced performance), a focus which Ames and Archer (1988) found to be highly motivating in the classroom, fostering long-term use of learning strategies and helping students form realistic but challenging goals. When relative performance was the goal however, learners believed that ability was shown by success with little effort, and they judged their ability lower. As Darling-Hammond (1994, p. 110) points out, assessment needs to support authentic forms of teaching and learning.

Task-based process assessment involves a criterion-referenced orientation, with Criterion-Referenced Tests (CRTs) providing direct information "about what the learner can actually do with the target language." (McClellan, 1995, p. 137). Strengths and weaknesses can be isolated across the whole test population, and specific information can be gained about an individual's performance, in contrast to Norm-Related Tests (NRTs), which tend to give information only about learners at either ends of the scale (cf. McClellan, 1995, p. 146; Cartier, 1968; Cziko, 1982; Hudson & Lynch, 1984; Brown, 1988; 1989; Bachman, 1989; 1990).

Brown (1995) classifies CRTs and NRTs according to their test characteristics and logistical dimensions (Table 2, below). as can be seen from this table, CRTs are appropriate for assessment of oral foreign language performance, in that they foster learning (learning how to learn), they are classroom specific, and they are formative, being concerned with ongoing needs analysis and the

feedback of relevant data into the learning process. In contrast, NRTs are concerned with the administration of the learners, and are summative in nature, assessing whether learners have been (or are likely to be) successful (however this is defined), but unable to comment on why or how, or on what should happen next (cf. Williams & Burden, 1994, p. 22). NRTs thus differ from CRTs in focus, timing, purpose and theoretical motivation and reflect different perspectives and goals. :

TABLE 2
Differences between NRTS and CRTS (Brown, 1995, p.12).

	CRTs	NRTs
Test Characteristics		
Underlying Purposes	Foster learning	Classify/group students
Types of Decisions	Diagnosis, progress, achievement	Aptitude, proficiency, placement
Levels of Generality	Classroom specific	Overall, global
Students'	Know content to expect	Do not know content
Expectations		
Score Interpretations	Percent	Percentile
Score Report	Tests and answers to students	Only scores go to students
Strategies		
Logistical Dimensions		
Group Size	Relatively small group	Large group
Range of Abilities	Relatively homogeneous	Wide range of abilities
Test Length	Relatively few questions	Large number of questions
Time Allocated	Relatively short time	Long (2-4 hours) administration
Cost	Teacher time & duplication	Test booklets, tapes, proctor

IV. THE KOREAN SITUATION

1. The 7th Curriculum

The advocacy of CRTs, and in particular of authentic assessment, in tertiary EFL classrooms in Korea, is especially appropriate in the light of the goals of school education as stated in section 1 ("The direction of curriculum design") of the Korean Ministry of Education 7th Curriculum document (박, 2001, p. 3). Here we find the ideal of *홍익인간* (*hongik-ingan*: contributing to the overall benefit of humankind) at the foundation of educational objectives which aim to foster "the ability to achieve an independent life and acquire the qualifications of democratic citizens, and to be able to participate in the building of a democratic state and promoting the prosperity of all humankind" (박, 2001, p. 3). The well-educated person that these goals aim to promote is further defined as:

1. A person who seeks to develop his/her own individuality on the basis of well-rounded and wholesome development

2. A person who demonstrates creative ability on the basis of a solid grounding in basic knowledge and skills
3. A person who explores career paths on the basis of broad intellectual knowledge and skills in diverse academic disciplines
4. A person who creates new values on the basis of an understanding of the national culture
5. A person who contributes to the development of the community where he/she lives on the basis of democratic citizenship. (박, 2001, p. 4)

Such a humanistic, holistic view of education provides an excellent reference-point for curriculum designers and school teachers when considering learning environments, curriculum content, and assessment models, for it is immediately apparent that the promotion of responsible, creative individuals with critical thinking skills and awareness of professional ethics (the sort of people who will contribute actively and constructively to society in the 21st century), is not to be (and has not been) achieved through the norm-referenced assessment model, which is evidently inappropriate for language learning (especially at tertiary level), since it promotes competition, exclusion and summative testing above the ethics of *hongik-ingan*. If education is to successfully foster autonomous, informed learners who are aware of their learning goals, confident of their ability to achieve them, motivated to learn, and possessing the learning skills that will enable them to take on the unpredictable learning situations of the future, then the use of collaborative, student-centered assessment is imperative. CRTs, with their focus on real-life situations, problem-solving, learning skills, and responsibility for learning, must be adopted at every level of education, and language classrooms must focus on development of student autonomy, responsibility, confidence, and motivation.

Such a process can begin most conveniently at tertiary level, since designers of language programs at this level have the expertise and freedom (within certain restrictions) to construct student-centered conversation-based courses (c.f. Finch, 2001). However, there are instances of the NRT approach at tertiary level which must be addressed in order for this process to gain momentum.

2. The bell curve

One such educational dinosaur is the prescribed application of proportional grading (here referred to as a bell-curve) in the grading of language courses. This practice began as an attempt to ensure consistency in grading across all the academic disciplines in the Freshman year of study, since professors who assigned abnormally strict or lax grades disproportionately affected the prospects of their students (entry to second-year courses is dependant on first-year grades). Student expectations regarding final grades were also reflected in class sizes. Because of this situation, National Universities in Korea introduced a system in which professors were required to assign grades according to a given bell-curve: a defined percentage of students would receive an "A" grade, another percentage of students would receive a "B" grade, and the remainder would receive "C" or below. With the advent of online recording of grades, grading software was designed to accept only grades which conformed to these requirements.

What began as an attempt to solve a difficult situation has, however, produced its own problems, especially in the conversation-English classroom. If (as the government advises) students are to be encouraged to develop communicative competence in English, along with problem-solving, creative thinking and social awareness, then the learning environment must promote the acquisition of such

qualities through responsibility, collaboration, motivation, and positive attitude change. As can be seen, however, the bell-curve fosters competition, aggression, and exclusion, encouraging students to be "successful" despite (rather than in tandem with) their peers.

If we consider (for example) a class of well motivated students, all of whom are doing their utmost (individually and together) to set linguistic goals, work towards them, and reflect upon their achievements (i.e. a class in which the teacher has successfully promoted affective, cognitive and social growth), 30% of these excellent students must receive "C" grades, and the teacher must find "reasons" for assigning them. In another class of unmotivated students, none of whom show any interest in learning, and all of whom reject the advice and counseling of the teacher, no more than 30% can be allocated a "C" grade. The remainder of the class must be assigned "A" or "B" grades, irrespective of their lack of progress relative to students in the previously-mentioned class. The situation is particularly hard on the teacher who aims to promote positive affect (motivation, confidence, reduced stress) and attitude change. If he/she manages to fire up a student with enthusiasm to learn and with love of learning *per se*, the resultant improvement in that student's grades must be matched by the demotion of two other students (if one student rises from "C" to "A", then one must also fall from "A" to "B", and another from "B" to "C"). How are teachers to motivate students in this situation? What are they to say to the highly motivated student who receives a "C" grade simply because he/she is in a class of extremely highly motivated peers? The topic of how to provide realistic assessment information on oral achievement is largely irrelevant for the student who realizes from day 1 of the semester that he/she will receive a low grade, and who therefore makes no effort to improve.

Professors and language teachers need the freedom not only to award high grades where they are merited, but also to award low grades, thus providing students with the information that their academic efforts are inadequate. A recommendation of this paper is, therefore, that the "bell-curve" system of grade allocation be dropped forthwith. If professors and teachers are successful in motivating their students to achieve excellence, then these students deserve appropriate grades, just as the student who shows sudden improvement must not punish the others by his/her success.

V. SELF/PEER-ASSESSMENT

1. Introduction

Returning to the promotion of oral skills in a student-centered setting, with authentic assessment integrated into the curriculum, we find that self- and peer-assessment are forms of CRT-based assessment which have proved practical and effective. The majority of the research in this area has been performed on self-assessment, though the findings also apply to peer-assessment, which is justified largely by the same arguments (Tudor, 1996, p. 182).

2. Justifications

Based on work carried out since the late 1970s, various authors and researchers agree on self-assessment as a vital part of learner autonomy (Henner-Stanchina & Holec, 1985, p. 98; Dickinson, 1987, p. 16; Blanche, 1988, p. 75; Harris, 1997, p. 12), providing the opportunity for learners to assess their own progress and thus helping them to focus on their own learning. Hunt, Gow

and Barnes (1989) even claim that without learner self-evaluation and self-assessment "there can be no real autonomy" (Hunt et. al., 1989, p. 207). Rea (1981) sees self-appraisal as helping the learner become aware of his/her responsibilities in planning, executing and monitoring his/her language learning activities, and Oscarsson (1978), while agreeing on this formative prime aim, adds a more summative secondary aim of enabling the learner "to assess his total achievement at the end of a course or course unit". Dickinson points out that this does not necessarily devalue or conflict with external evaluation, which still has relevance for supplying official certification of learning (Dickinson, 1987, p. 136). Rather, as Dickinson and Carver observe:

A language course can only deal with a small fraction of the foreign language therefore one objective of language courses should be to teach learners how to carry on learning the language independently. Part of the training learners need for this purpose is training in self-assessment and self-monitoring. (Dickinson & Carver, 1980, p. 7)

The favourable correlation of self-rating scores and external test scores in research findings mostly support the use of self-assessment in second language learning, and Oscarsson's "rationale of self-assessment procedures in language learning" (1989, p. 3) serves as a framework for the various justifications for self-assessment that have been proposed: i) promotion of learning; ii) raised level of awareness; iii) improved goal orientation; iv) expansion of range of assessment; v) shared assessment burden; and vi) beneficial postcourse effects.

Harris sees self-assessment as appropriate in test-driven secondary and tertiary education, claiming that self-assessment can help learners in such environments to become more active, to locate their own strengths and weaknesses, and to realize that they have the ultimate responsibility for learning. By encouraging individual reflection, "self-assessment can begin to make students see their learning in personal terms [and] can help learners get better marks." (Harris, 1997, p. 13). Peer assessment is especially applicable to the classroom setting, aiming to encourage students to take increased responsibility for their own curricula and to become active participants in the learning process (Hill, 1994, p. 214; Miller & Ng, 1996, p. 134). Tudor adds that critical reflection on the abilities of other learners with respect to a shared goal is a practical form of learner training which helps individuals to assess their own performance, and which reduces the stress of error correction through identifying them in others (Tudor, 1996, p. 182). Thus, Assinder (1991, pp. 218-28) reports increased motivation, participation, real communication, in-depth understanding, commitment, confidence, meaningful practice and accuracy, when students prepare and deliver learning tasks for each other.

Haughton & Dickinson (1989) (cited in Miller & Ng, 1996, p. 135) set out to test nine hypotheses about peer assessment in their study of a collaborative post-writing assessment. Five of these hypotheses (items 1 to 5, below) dealt with the practicality of peer assessment, and four (6 to 9) with the benefits of the scheme:

1. Students are sincere and do not use the scheme as a means of obtaining higher grades than they themselves think they deserve.
2. Students are or become able to assess themselves at about the same level as their tutors, i.e. they can interpret the criteria in the same way.
3. Students are or become able to negotiate with tutors on the appropriate level of criteria.
4. Students are or become able to negotiate grades in a meaningful and satisfying manner.
5. The scheme does not result in a lowering of standards on the course.

6. Students perceive collaborative assessment as fairer than other (traditional) forms of assessment.
7. Students benefit in enhanced understanding of and attitude towards assessment.
8. Students become more self-directed as a result.
9. The scheme demands more thoroughly worked out criteria of assessment and hence results in fairer assessment.

This study showed "a relatively high level of agreement between the peer assessments and the marks given by the lecturers" (Miller & Ng, 1996, p. 139). Similar reliability of results was reported by Bachman and Palmer (1982) with the self-rating of communicative language ability of ESL learners (aged 17-67) in the USA. Fok (1981), looking at a group of university students in Hong Kong, also found a high degree of similarity between the students' self-assessment and past academic records for Reading and Speaking. Haughton and Dickinson (1989) claim that to a large extent the scheme worked and that the students were able to assess their own work realistically. Miller & Ng (1996, p. 141), commenting on the scheme, considered that: i) the students were sincere; ii) they demonstrated a similar level of assessment to that of the lecturers; iii) the scheme did not result in a lowering of standards; and iv) the students benefited in their understanding of and attitude towards assessment by taking part in the study:

Language students are able to make a realistic assessment of each others' oral language ability. (Miller & Ng, 1996, p. 142)

3. Implications and issues

Self- and peer-assessment are thus practical and effective assessment methods in tertiary language classes, addressing educational goals espoused in the 7th National Curriculum (박 2001). In handing over a large part of the assessment burden and responsibility to the learners, these forms of authentic assessment offer opportunities for affective (and cognitive) growth and development of social awareness - opportunities that can be monitored by the teacher, and used in conjunction with the students for reflection on issues as they arise.

Such issues (e.g. plagiarism, peer-pressure, and unrealistic expectations) have been cited as disadvantages of self-assessment. However, they have all received adequate responses ("•", below) in research literature:

1. Doubts on the reliability and feasibility of learners assessing their own self-directed learning and carrying out individual needs analysis (Dickinson, 1987, p. 150).
 - There is evidence that learners can make satisfactorily accurate self-assessments (Blanche, 1988, p. 85; Blue, 1988, p. 100) and that there is a fairly consistent overall agreement between self-assessment and external criteria (Dickinson 1987, p. 150).
2. Doubts about the sincerity of the learners (Dickinson 1987, pp. 150-151).
 - One reason put forward by teachers for not sharing responsibility for assessment is that students will "cheat" and produce unrealistic scores. Dickinson (1987), however, points out that "cheating" (a process in which a learner seeks to obtain personal advantage by unfair means [Dickinson, 1987, p. 150]), is not about learning but about demonstrating the results of learning to someone else, usually in situations which value scores and rank over actual

- learning: "Where the learner is concerned with real learning objectives, and where self-assessment is mainly used, cheating offers no advantages" (Dickinson, 1987, p. 151).
3. Doubts regarding the reliability of self-assessment in formal education (Blue, 1988, p. 100; Janssen-van Dieten, 1989, p. 31; Pierce, Swain & Hart, 1993, p. 38).
 - Research on peer assessment (cf. Miller & Ng, 1996) has shown that peer- self-assessment has an important place in formal education, and that it focuses attention on communicative competence levels in the classroom (Blanche, 1988, p. 85).
 4. Reluctance of teachers to lose control of assessment (Blue, 1988, p. 100).
 - Teachers need to be aware of the rationale behind self-assessment as well as the means of promoting it: "Relevant training of teachers may actually constitute a prerequisite for the effective realization of student-centred evaluation techniques" (Oscarsson, 1989, p. 11).
 5. The need for learners to receive training and practice in assessing their own performances (Nunan, 1996, p. 22).
 - Learner training for self-assessment can help learners successfully identify their needs. This not only enhances learning, but also frees the teacher to concentrate on developing learning materials and giving help in other parts of the learning process (Blue, 1988, p. 101).
 6. Conflict of the need for students to be in control of aspects of evaluation, and demands of external imperatives (Dickinson, 1978). The question of whether self-assessment is both formative and summative, or whether it should only be seen as a process-oriented, integrative, and ongoing (formative) activity (Oscarsson, 1997).
 - Self-assessment for formative self-monitoring is "both possible and desirable" (Dickinson, 1987, p. 151). It is also feasible for placement testing, diagnostic testing, and self-evaluation of "total achievement at the end of a course or course unit" (Oscarsson, 1978, p. 3).

VI. CONCLUSION

In the shift from transmission of knowledge to transformation of knowledge and to integration of knowledge with existing personal constructs and meanings (Kohonen, 1999, p. 280), evaluation is taking on new affective goals in which the personal growth of the learner is becoming increasingly important (Ranson, 1994, p. 116). Thus it is no longer defensible to use discrete-item testing of dubious constructs or to sample performance as a means of inferring underlying competence or abilities, if assessment is really concerned with providing information on learning.

Self- and peer-assessment offer ways of addressing this situation by encouraging the student to become part of the whole process of language learning, and to be aware of his/her progress. Of particular significance for students in Korea, Harris (1997, p. 19) sees self-assessment as a practical tool that should be integrated into everyday classroom activities, and Blanche suggests that it "would be particularly helpful in the case of false beginners" (Blanche, 1988, p. 85).

It would be possible to close at this point, concluding that authentic assessment in a CRT context is both feasible and desirable for the tertiary language classroom in Korea, promoting as it does the qualities of "the educated person" as defined by the 7th National Curriculum, and also fostering cognitive and affective development in the students. However, it is important to take a step further, and to recognize the importance of the promotion of social awareness, which is also a feature of authentic assessment. Even a very brief survey of quotations made by educators and others on this topic shows a consensus that the role of education goes beyond the preparation of an individual for a

prospective career (the following quotations, apart from the one by Rogers, are from "The quotations page": <http://www.quotationspage.com/>):

Education is not merely a means for earning a living or an instrument for the acquisition of wealth. It is an initiation into life of spirit, a training of the human soul in the pursuit of truth and the practice of virtue. (Vijaya Lakshmi Pandit)

Education is not the filling of a pail, but the lighting of a fire. (W. B. Yeats)

The object of education is to prepare the young to educate themselves throughout their lives. (R. M. Hutchins)

The aim of education should be to teach us rather how to think, than what to think - rather to improve our minds, so as to enable us to think for ourselves, than to load the memory with thoughts of other men. (Bill Beattie)

The only person who is educated is the one who has learned how to learn and how to change (Rogers, 1969, p. 120)

Putting such quotations into context, it can be claimed that the immediate goal of mass education when it began (in Korea and elsewhere) was the need for a society of knowledgeable citizens who could contribute to the economic growth of an emerging nation, and that intense intellectualization of learning was therefore necessary and excusable. Since that time, however, industrialism and consumerism have shown destructive potential, and education is currently seen by many educators not simply as a means of improving society, but of preventing its collapse:

Human history becomes more and more a race between education and catastrophe. (H. G. Wells: <http://www.quotationspage.com/>)

Thus Harri-Augstein and Thomas (1991, p. 7) call for attention to "our capacity for learning " to "provide us with the resource to negotiate change, to prevent man-made catastrophes, to achieve success and to attain new standards of excellence and quality in our various human endeavors." In such a light, a call for the use of authentic assessment in CRT contexts is also a call for the fostering of social responsibility in the language classroom. It has never been justifiable to state that "I am just a language teacher. I teach language", but such an attitude is even more indefensible in a contemporary society crying out for creative, problem-solving, critical-thinking citizens. The 7th National Curriculum highlights the qualities that must be promoted in students in every educational establishment, and in every classroom. Authentic assessment, self- and peer-assessment are practical means towards this goal, and can be adopted by teachers at the local level, while waiting for them to be adopted globally by educational policy makers. Promotion of *홍익인간* in the language classroom requires program designers to incorporate authentic oral testing into their programs, and university administrators to drop the "bell-curve" requirement for class grades, trusting professors and teachers to assess students according to their professional judgement and expertise.

REFERENCES

- 박주영, (2001). *The school curriculum of the republic of Korea*. 7th Curriculum-3.doc, http://www.moe.go.kr/eng_26/
- Alderson, C. J. (1981). Report of the discussion on communicative language testing. In C. J. Alderson & A. Hughes (Eds.), *Issues in language testing. ELT document 111*. The British Council.
- Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation process. *Journal of Educational Psychology*, 80, 260-267.
- Arnold, J. (Ed.). (1999). *Affect in language learning*. Cambridge: Cambridge University Press.
- Assinder, W. (1991). Peer teaching, peer learning. *ELT Journal*, 45(3).
- Bachman, L. F. (1989). The development and use of criterion-referenced tests of language proficiency in language program evaluation. In K. Johnson, (Ed.), *Program design and evaluation in language teaching*. Cambridge: Cambridge University Press.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L. & Palmer, A. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly* 16, 449-465.
- Benson, M. (1991). Attitudes and motivation towards English: A survey of Japanese freshmen. *RELC Journal*, 22(1), 34-48.
- Bickley, V. (Ed.). (1989). *Language teaching and learning styles within and across cultures*. Hong Kong: Institute of Language in Education, Education Department.
- Blanche, P. (1988). Self-assessment of foreign language skills: Implications for teachers and researchers. *RELC Journal*, 19(1), 75-93.
- Blue, G.M. (1988). Self-assessment: the limits of learner independence. In A. Brookes & P. Grundy (Eds.). *Individualization and Autonomy in Language Learning. ELT Documents, 131*. (pp. 100-118). London: Modern English Publications and the British Council, .
- Breen, M. P. (1984). Process syllabuses for the language classroom. In C. J. Brumfit (Ed.). *General English Syllabus Design. ELT Documents No. 118*. (pp. 47-60). London: Pergamon Press & The British Council. .
- Brown, J. D. (1988). *Understanding research in second language learning*. New York: Cambridge University Press.
- Brown, J. D. (1989). Criterion-referenced test reliability. University of Hawai'i Working Papers in ESL, 1, 79-113.
- Brown, J. D. (1995). Differences between norm-referenced and criterion-referenced tests. In J. D. Brown, & S. O. Yamashita (Eds.), *Language testing in Japan* (pp. 12-19). Tokyo: The Japan Association for Language Teaching.
- Canale, M. (1983). On some dimensions of language proficiency. In J. W. Oller, Jr. (Ed.). *Issues in language testing research*. Rowley, Mass.: Newbury House.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Carroll, B. (1981). *Testing communicative performance*. Oxford: Pergamon Press.
- Cartier, F. A. (1968). Criterion-referenced testing of language skills. *TESOL Quarterly*, 2(1), 27-32.
- Cziko, G. A. (1982). Improving the psychometric, criterion-referenced and practical qualities of integrative language tests. *TESOL Quarterly*, 16(3), 367-379.
- Cziko, G. A. (1984). Some problems with empirically-based models of communicative competence.

- Applied Linguistics*, 5, 23-38.
- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64(1), 5-30.
- Davies, A. (1990). *Principles of language testing*. Oxford: Basil Blackwell.
- Dickinson, L. (1978). Autonomy, self-directed learning and individualization. In A. Brookes & P. Grundy (Eds.). *Individualization and autonomy in language learning. ELT Documents 103* (pp. 7-28). Modern English Publications and the British Council.
- Dickinson, L. (1987). *Self-instruction in language learning*. Cambridge. Cambridge University Press.
- Dickinson, L., & Carver, D. J. (1980). Learning how to learn: Steps towards self-direction in foreign language learning in schools, *English Language Teaching Journal*, 35, 1-7.
- Finch, A. E. (2001). A formative evaluation of a task-based conversation English program. *PAC Journal*, 1(1), 125-146.
- Fok, A. C. Y. Y. (1981). *Reliability of student self-assessment*. Hong Kong: H.K.U. Language Centre.
- Harri-Augstein S., & Thomas, L. (1991). *Learning conversations: The self-organised learning way to personal and organisational growth*. London: Routledge.
- Harris, M. (1997). Self-assessment of language learning in formal settings. *English Language Teaching Journal*, 51(1), 12-20.
- Hart, D. (1994). *Authentic assessment: A handbook for educators*. New York: Addison Wesley.
- Houghton, G., & Dickinson L. (1989). Collaborative assessment by masters' candidates in a tutor based system. *Language Testing* 5(2): 233-46.
- Henner-Stanchina, C., & Holec, H. (1985). Evaluation in an autonomous learning scheme. In P. Riley (Ed.), *Discourse and Learning*. London: Longman.
- Hill, B. (1994). Self-managed learning: State of the art. *Language Teaching* 27, 213-223.
- Hudson, T., & Lynch, B. (1984). A criterion-referenced approach to ESL achievement testing. *Language Testing*, 1(2), 171-200.
- Hunt, J., Gow, L., & Barnes, P. (1989). Learner self-evaluation and assessment - a tool for autonomy in the language learning classroom, in V. Bickley, (Ed.), (pp. 207-217). Hong Kong: Institute of Language in Education.
- Hymes, D., Cazden, C. B., & John, V. P. [Eds.]. (1972). *Functions of language in the classroom*. New York: Teachers College Press.
- Janssen-van Dielen, A. M. (1992). Self-assessment in second language learning. Ph.D. dissertation, Kotholieke Universiteit, Nijmegen, The Netherlands.
- Keefe, J. W. (Ed.). (1979). *Student learning styles: Diagnosing and prescribing programs*. Reston, VA: National Association of Secondary School Principals.
- Kelly, G. (1955). *The psychology of personal constructs*. New York: Norton.
- Kelly, R. (1978). On the construct validity of comprehensive tests: An exercise in applied linguistics. University of Queensland PhD thesis.
- Kohonen, V. (1999). Authentic assessment in affective foreign language education. In J. Arnold, (Ed.). (pp. 279-294).
- Larson-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, 18(2), 141-165.
- Lee, Wan-ki, (1991). A task-based approach to oral communication testing of English as a foreign language. Ph.D. thesis, Manchester University, U.K.. Seoul: Hanshin Publishing Co.
- McClellan, J. M. (1995). Negotiating a spoken-English scheme with Japanese university students. In J. D. Brown, & S. O. Yamashita (Eds.), *Language testing in Japan* (pp. 136-147). Tokyo: The

- Japan Association for Language Teaching.
- Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer, & H.I. Braun, (Eds.). *Test validity* (pp. 33-45). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Miller, L., & Ng, R. (1996). Autonomy in the classroom: peer assessment. In R. Pemberton, S. L. Edward, W. W. F. Or, and H. D. Pierson (Eds.). *Taking control: Autonomy in language learning* (pp. 133-146). Hong Kong: Hong Kong University Press.
- Morrow, K. (1979). Communicative language testing: Revolution or evolution. In C. J. Brumfit, & K. Johnson (Eds.). (pp. 143-158).
- Nakamura, Y. (1995). Making speaking tests valid: Practical considerations in a classroom setting. In J. D. Brown, & S. O. Yamashita (Eds.), *Language testing in Japan* (pp. 126-135). Tokyo: The Japan Association for Language Teaching.
- Nunan, D. (1996). Towards autonomous learning: some theoretical, empirical and practical issues. In R. Pemberton, S. L. Edward, W. W. F. Or, and H. D. Pierson (Eds.). *Taking Control: Autonomy in Language Learning*. (pp. 13-26). Hong Kong: Hong Kong University Press.
- Oller, J. W. (1979). *Language tests in schools*. London: Longman.
- O'Malley, M., & Pierce, L. V. (1996). *Authentic assessment for English language learners*. New York: Addison Wesley.
- Oscarsson, M. (1978). *Approaches to self-assessment in foreign language learning*. Council of Europe, Council for Cultural Co-operation, Strasbourg, France.
- Oscarsson, M. (1989). Self-assessment of language proficiency: rationale and implications. *Language Testing*, 6(1), 1-13.
- Oscarsson, M. (1997). Self-assessment of foreign and second language proficiency. In *Encyclopedia of Language and Education. Vol 7: Language testing and assessment*. (pp. 175-187). Dordrecht: Kluwer Academic, .
- Pierce, B. N., Swain, M., & Hart, D. (1993). Self-assessment, French immersion, and locus of control. *Applied Linguistics*, 14, 25-42.
- Ranson, S. (1994). *Towards the learning society*. London: Cassell.
- Rea, P. R. (1981). Formative assessment of student performance: the role of self-appraisal. *Indian Journal of Applied Linguistics*. 7, 66-68.
- Reid, J. (1987). The learning style preferences of ESL students. *TESOL Quarterly*, 21/1, pp. 87-111.
- Rogers, C. R. (1969). *Freedom to Learn*. Columbus, OH: Merrill Publishing Co.
- Seliger, H. (1984). Processing universals in second language acquisition. In F. Eckman, L. Bell, & D. Nelson (Eds.). *Universals of second language acquisition*. Rowley, MA.: Newbury House.
- Skehan, P. (1988). Language testing, part 1: State of the art article. *Language Teaching*, 21(4), 211-218
- Skehan, P. (1998). *A Cognitive approach to language learning*. Oxford: Oxford University Press.
- Spolsky, B. (1975). Language testing: The problem of validation. In L. Palmer & B. Spolsky (Eds.). *Papers on language testing 1967-1974* (pp. 147-53). Washington, D.FC.: TESOL.
- Thrasher, R. H. (1984). Educational validity. Annual reports, International Christian University, 9, 67-84.
- The quotations page: <http://www.quotationspage.com/>
- Tudor, I. (1996). *Learner-centredness as language education*. Cambridge University Press: Cambridge.
- Van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy, and authenticity*.

London: Longman.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, Mass.: MIT Press.

Weir, C. J. (1998). *Communicative language testing*. Exeter: University of Exeter Press.

Williams, M., & Burden, R. L. (1997). *Psychology for language teachers: A social constructivist approach*. Cambridge: Cambridge University Press.

Applicable levels: secondary, tertiary, adult

Key words: alternative assessment, authentic assessment, communicative language testing, curriculum, criterion-referenced testing

Dr. Andrew Edward Finch

경북대학교 사범대학 영어교육과

720-710 대구광역시 북구 상격동 1370

Tel: (053) 950-5832

Fax: (053) 384-1345

Email: aef@knu.ac.kr