

Ethical Assessment: Implications for EFL Educators in Korea

Finch, A. E. (2002). Ethical assessment: Implications for EFL educators in Korea. *English Teaching*, 57(3)

Language test design is often a matter of balancing conflicting elements such as ministerial policy, pedagogic principles, ethical concerns, and administrative (logistic) requirements. Recently, at least two of these components have shown some agreement in Korea, as top level policy statements have matched performance-assessment research findings, by identifying holistic and humanistic goals for language education - goals which (according to the relevant literature) can best be approached through criterion-referenced tests and authentic assessment methods (portfolios, projects, learner journals, self- and peer-assessment, learning conversations, etc.). Other more practical elements of the test-design jigsaw (e.g. testing practices and local administrative requirements), however, lag some way behind these theoretical and pedagogical advances, with norm-referenced methodologies and restrictive grading policies effectively prohibiting any attempts at authentic assessment and succeeding only in demotivating language teachers and students. This paper therefore aims to discuss the gap between ethical principles and prescribed assessment parameters in universities in Korea, and concludes that ministerial policies and validated research findings can no longer be put aside for the sake of logistical convenience. (172 words)

I. INTRODUCTION

When designing a language test, irrespective of high/low stakes¹⁾, language level, and testing purpose, it is normal to consider the needs and preferences of the various test-consumers. These can range from the test-takers and their families, to the raters, administrators and curriculum designers, and further on to local and national policy-makers. The field of test-design is thus known for the weighing and balancing of interests which must occur before the first test item takes shape. This paper suggests that administrative concerns have, to date, received undue emphasis in this process, and that there is a need, in Korea and elsewhere, for ethical and pedagogic criteria to be given more attention, so that the introduction of holistic and humanistic testing methods can have beneficial results for teachers, test-takers, and for society as a whole. In particular, this paper focuses on the practice of using bell-curves to prescribe (rather than describe) student grades, and advocates alternative methods of assessment which involve the

¹⁾ I.e. the potential impact on the students' future.

students as active participants in the assessment process, thus training them to be critical thinkers and responsible problem-solvers when they enter society as adults.

II. SOME CURRENT ISSUES

1. Ethical testing

In a paper presented recently to a Language Assessment Ethics Conference in Pasadena, CA., Alderson (2002) makes the point that language testers and test designers “must develop, proclaim, and follow professional standards, we must identify professional ethics, and we must acknowledge the importance of “higher” principles, principles of morality and humanity, that should govern our thinking and our practice” (p. 1).

At first sight, such a statement seems far from controversial, since no language educator would advocate (or admit to) following unethical test procedures. However, the fact that this paper was presented at a conference dealing solely with the ethics of testing, indicates that there is cause for considerable professional concern among test-designers and researchers, and also that there is widespread evidence that important (high-stakes) tests, which can positively or negatively affect career possibilities for large numbers of test-takers, are lacking in terms of the “higher” principles which should govern their design and implementation.

Anderson develops his point by referring to national testing practices in Hungary. However, it is important to acknowledge that the situation in Korean universities also merits examination in terms of ethical and pedagogic principles. Despite the humanistic policy statements of the Ministry of education (Ministry of Education and Human Resources Development, 2001), and the overwhelming body of literature supporting the use of formative, criterion-referenced testing (for example), it is notable that summative, norm-referenced tests remain the preferred mode of assessing foreign-language performance in most faculties, and that grading criteria focus on administrative, rather than educational, logistics (“number-crunching”²). This situation is not only ethically undesirable, but it is educationally unsound, while being demotivating for language teachers and students. Such concerns are frequently voiced in the academic community, but, as Alderson (2002) points out, rarely heard in conferences or journals.

If we are to proclaim professional ethics, and develop standards for professional practice, then we must identify, expose, and explain the opposite – unethical practice, unprofessional behaviour³. Yet in language testing, it is extremely rare to come across accounts of such practice. (p. 2)

² This statement is based on communications with TEFL colleagues in Korea over 10 years. Email correspondence with language testers in other countries suggests that the statement is also true of many countries in which English is learned as a foreign language.

³ Individual authors’ original spellings are adhered to in quotes in this paper.

2. The Bell-curve

A particular example of a questionable testing practice which occurs throughout Korea at tertiary level, is the use of a prescribed bell-curve to allocate grades. Thus, a testing tool which began as a means of describing and analyzing test-data has become an achievement predictor, requiring teachers and professors to adhere to defined grading parameters, whatever their philosophy of evaluation. In universities and colleges using this system, final grades are entered online, using software which rejects “incorrect” input. Irrespective of individual (absolute) progress and achievement, therefore, students must be allocated grades according to the defined (relative) spread, and language teachers in Korea must contribute to a system of grading (and consequently, of testing) which contradicts educational pedagogy. Concerned educators, who are successful in promoting meaningful learning, holistic development and personal reflection in their classes, are thus faced with the dilemma of having to find ways of differentiating between sincere, hard-working students (ways which will inevitably demotivate 30% of them). Irrespective of the inspirational qualities of the teacher and the excellence or indifference of the students, the relative number of students achieving high (and low) grades remains fixed.

Justifications for this situation are mostly logistic, and are shared by educational administrators around the world. When designing university entrance selection procedures, or when faced with limited access to courses of study, there is a need for a means of classifying and ranking students in terms of their suitability. In Korea, Freshman English programs are often used to satisfy this need, and are instrumental in the selection of students for popular sophomore courses. This is not the full story, however, since it appears that before the present system was initiated, grade-allocation across disciplines was inconsistent, with certain professors being known for assigning abnormally strict or lax grades and for consequently affecting the prospects of their students disproportionately. Student expectations regarding final grades were thus reflected in respective class sizes, and the current system was introduced as an attempt to provide an ethical framework for grading: in current terminology “an even playing field for all”.

What began as an attempt to solve a difficult situation has, however, produced its own problems, especially in the English-conversation classroom. If (as the government advises) students are to develop communicative competence in English, along with problem-solving skills, creative thinking and social awareness, then the learning environment must promote the development of such skills and qualities through responsibility, collaboration, motivation, and positive attitude change, rather than the competition and exclusion which results when only 20% of students in a class are allowed to receive an “A” grade. It is meaningless in this situation for the teacher to encourage team-work (a quality required by most employers) and social responsibility. Instead, students are prompted to be “successful” despite (rather than in tandem with) their peers, and teachers tend to use easily quantifiable testing methods (e.g. multiple-choice, discrete-item, grammar-based tests), which provide little information about affective or social growth.

If we take a case in point, and imagine 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), then we find that a defined proportion (e.g. 30%) of these excellent students must receive “C” grades, and that the teacher must find reasons for assigning these grades. If we next imagine a class of unmotivated students, none of whom show any interest in learning, and all of whom reject the advice and counseling of the teacher, we find that no more than 30% can be allocated a “C” grade, and that the remainder of this class must be assigned “A” or “B” grades, irrespective of their lack of progress relative to students in the previously mentioned class. This situation is particularly hard on the teacher who aims to promote positive affect (motivation, confidence, stress-management) and attitude change, for if he/she manages to fire up a non-achieving 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”).

This adherence to mediocrity, in which each class is defined as containing no more than a set proportion of excellent students, is counterproductive in terms of holistic and practical educational goals, in that students and teachers feel their efforts to be irrelevant to the final grades. How are teachers to motivate students in this situation? What is one 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 linguistic achievement is meaningless 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. It must be of particular concern that the wildly varying language levels (and attitudes to learning) in Freshman English classes (normally heterogeneous collections of students from the same departments) means that an “A” in one class bears little resemblance to an “A” in another, and that even the gate-keeping exercise is fatally flawed⁴).

3. Summative Testing

Inflexible allocation of grades according to administratively desirable proportions impacts on language testing by promoting “snapshot” proficiency tests above ongoing, student-centered, process assessment, since the discrete-item, “right/wrong”, answers which the former provide are particularly suitable for student-selection purposes. The pressure for teachers and professors to “come-up with the numbers”, and to rank language achievement within classes, thus breeds a focus on summative memory-based testing, with its avoidance of personal, social and affective development. Hence, norm-referenced (see section III.2) mid-term and final tests in university courses can count for 60% (or more) of

⁴ The author has experience of this situation being alleviated to some extent when teachers were allowed to apply relative grading across all their classes.

the final grade. Not only is student study product-driven in this situation, but lack of attention to the process of learning promotes last minute cramming and immediate relegation of tested materials to the mental “trash-can”. However, such testing, which concentrates 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).

Were we able to put such considerations aside, it would nevertheless be important to recognize that the competitive allocation of grades and the testing culture that it spawns (and vice-versa), has been shown to be unsuccessful, according to its own criteria. As Gardner (1993) points out, even those students who are most successful (success being defined as attendance at prestigious schools, high grades, high test scores, excellent references from teachers, etc.) “typically do not display an adequate understanding of the materials and concepts with which they have been working” (p. 3). One must wonder about the ethics of high-stakes knowledge-based testing that fails to assess whether the learning content has been internalized and can be applied (or is applicable) in non-rehearsed situations.

III. PRINCIPLES OF ASSESSMENT⁵⁾

1. Authentic Assessment

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. In view of these considerations, the attention of language testers turned (in the 1980’s and 90’s) towards performance skills and communicative competence. Research on the importance of affect in language learning also had implications for test-design. Various researchers thus proposed “authentic assessment” as a process-oriented means of evaluating communicative competence, cognitive abilities and affective learning (Hart, 1994, p. 9; Kohonen, 1999, p. 284; O’Malley & Pierce, 1996, pp. 1-6), 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

⁵⁾ For a more detailed analysis of language testing issues, the reader is referred to Skehan’s “state of the art article” (1988), and Lee’s work on task-based oral testing in Korea (1991).

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 (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 judgment 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 ways in which authentic assessment can enhance learning, and summarizes how this approach contrasts with standardized testing (Table 1).

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 | Judgment 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 | Intrinsic 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 (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.

2. Criterion-referenced Testing

Authentic assessment in a task-based process setting implies a focus on language mastery (criterion-referenced performance), fostering long-term use of learning strategies and helping students to form realistic but challenging goals, in contrast to relative performance (norm-referenced performance), which encourages learners to believe that ability is shown by success with little effort (Ames & Archer, 1988).

As Darling-Hammond (1994) points out, “assessment needs to support authentic forms of teaching and learning” (p. 110), and in this context, task-based process assessment (with criterion-referenced tests - CRTs) provides direct information “about what the learner can actually do with the target language.”(McClean, 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, whereas norm-referenced tests (NRTs) tend to give information only about learners at either ends of the scale (Bachman, 1989, 1990; Brown, 1988, 1989; Cartier, 1968; Cziko, 1982; Hudson & Lynch, 1984; McClean, 1995, p. 146).

Brown (1995) classifies CRTs and NRTs according to their test characteristics and logistical dimensions (Table 2).

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

| | CRTs | NRTs |
|------------------------------|---|---|
| Test Characteristics | | |
| Underlying Purposes | <ul style="list-style-type: none"> • Foster learning • Diagnosis, progress, achievement | <ul style="list-style-type: none"> • Classify/group students • Aptitude, proficiency, placement |
| Types of Decisions | <ul style="list-style-type: none"> • Classroom specific | <ul style="list-style-type: none"> • Overall, global |
| Levels of Generality | <ul style="list-style-type: none"> • Know content to expect | <ul style="list-style-type: none"> • Do not know content |
| Students’ Expectations | <ul style="list-style-type: none"> • Percent | <ul style="list-style-type: none"> • Percentile |
| Score Interpretations | <ul style="list-style-type: none"> • Tests and answers to students | <ul style="list-style-type: none"> • Only scores go to students |
| Score Report Strategies | | |
| Logistical Dimensions | | |
| Group Size | <ul style="list-style-type: none"> • Relatively small group | <ul style="list-style-type: none"> • Large group |
| Range of Abilities | <ul style="list-style-type: none"> • Relatively homogeneous | <ul style="list-style-type: none"> • Wide range of abilities |
| Test Length | <ul style="list-style-type: none"> • Relatively few questions | <ul style="list-style-type: none"> • Large number of questions |
| Time Allocated | <ul style="list-style-type: none"> • Relatively short time | <ul style="list-style-type: none"> • Long (2-4 hours) administration |
| Cost | <ul style="list-style-type: none"> • Teacher time & duplication | <ul style="list-style-type: none"> • Test booklets, tapes, proctor |

3. Humanistic Policy-making

The advocating of CRTs and 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’s 7th Curriculum document (Ministry of Education and Human Resources Development, 2001). Here, the ideal of “Hongik-Ingan” (contributing to the overall benefit of humankind) underlies 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” (p. 3). The well-educated person that these goals aim to promote is further defined as follows.

- A person who seeks to develop his/her own individuality on the basis of well-rounded and wholesome development
- A person who demonstrates creative ability on the basis of a solid grounding in basic knowledge and skills
- A person who explores career paths on the basis of broad intellectual knowledge and skills in diverse academic disciplines
- A person who creates new values on the basis of an understanding of the national culture
- A person who contributes to the development of the community where he/she lives on the basis of democratic citizenship. (Ministry of Education and Human Resources Development, 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 (or the proportional grading model), which 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.

IV. ETHICAL APPLICATIONS

When looking for viable and practical alternatives to a norm-referenced, summative model of language testing and grading, it is important to remember that 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 et al, 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. Alternative assessment (e.g. portfolios, self- and peer-assessment, learning journals, learning conversations) are also well-suited to the humanistic view of education advocated for some time by politicians, researchers, philosophers and educational practitioners (Bruner, 1960; Dewey, 1916; Krishnamurti, 1981; Ministry of Education and Human Resources Development, 2001). Let us therefore put aside the logistic requirements of administrators for the moment, and examine the characteristics of a learning environment in which testing is an integral part of the process of affective, cognitive and social growth.

1. The Non-threatening Learning Environment

The first point to make is that teachers can fashion such an environment by “doing ordinary things well” (O’Neill, 1991, pp. 300-301), and by putting them in a humanistic context. Underhill (1989) considers the act of “really listening to the student and to the content of what he or she says” (p. 256) as having the most dramatic effect on the learning atmosphere, since “our students don’t necessarily need reassurance, what they need is to be heard” (p. 256). A second focus of attention is on the creation of a non-threatening learning environment (Finch, 2001) in which teachers:

1. develop a stress-free climate;
2. develop peer-support networks;
3. promote self-confidence without focusing on competence or performance;
4. examine hidden agendas;
5. offer unconditional trust, inspiring confidence, motivation and independence;
6. reflect on personal assumptions;
7. reflect on counseling skills and management of affect;
8. reflect a holistic, affective, student-centered view of language learning;
9. act as learning resources. (p. 138)

The American Association for Supervision and Curriculum Development (ASCD, 1964) offers further advice on how to help students “expand the self”.

1. Observing and listening to learners with care and concern.
2. Achieving openness in pupil-teacher relationships, to permit improved response and interaction.
3. Helping learners toward the objective of personal relevance.
4. Recognizing and accepting different ways of responding, according to learners' individualized styles and needs.
5. Stimulating creation and recreation of self-image that encourages further development.
6. Questioning, probing, and responding in ways that lead learners to assume responsibility.
7. Standing aside judiciously to let the learner discover and exercise his own resources.
8. Making development of the learner the chief goal in teaching subject matter.
9. Achieving free affective responses and seeing its relevance to intellectual development.
10. Achieving free and constructive communication with learners.
11. Helping learners sense the living dynamics of man's creations, as revealed by history and the current scene.
12. Clearing the way, by whatever means, for stretching learners' minds and abilities in creative, self-fulfilling endeavor. (pp. 161-162)

2. Alternative assessment

When endeavoring to incorporate such concepts into a student-centered, process-oriented, humanistic learning environment, alternative assessment models offer practical and effective assessment methods for tertiary language classes. Self- and peer-assessment in particular have been well researched, and found to be both valid and reliable. Harris (1997) sees them 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." (p. 13). Peer assessment, a form of self assessment (Tudor 1996, p. 182) which is justified largely by the same arguments, 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; Miller & Ng, 1996). Tudor (1996) 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 (p. 182). Thus Assinder (1991) 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. In handing over a large part of the assessment burden and responsibility to the

learners, these forms of humanistic 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.

Portfolios (and associated project work) offer a practical approach to assembling student work, interpreting evidence of student performance, and assessing student performance relative to instructional objectives. If we examine Rogers' description of learning (appendix A), we find that portfolios can indeed act as agents of behavioral change (cf. items 1 and 4), they are concerned with personal meanings and ideas (items 2, 3 and 8), they can be cooperative and collaborative experiences (item 5), they grow over time and reflect the personal nature of learning (items 6, 7 and 11), and they allow the learner to present his/her own formative learning experiences (items 12 and 13). There is both process and product in this exploration of learning and assessment, and consequently infinite scope for development of ideas in the direction of self-directed project work. By containing examples of work in progress, formal products, and "ratings or other evidence of student knowledge relative to specific objectives or purposes" (Valencia, 1990), portfolios can reflect the current status of the learner's educational journey.

Along with other alternative assessment methods such as learner-journals (c.f. Finch, 1998) and learning conversations (Harri-Augstein et al, 1991, p.6), such humanistic assessment can focus on developing linguistic skills in a socially-aware learner, fostering autonomy at the same time as collaboration.

3. Logistics

Returning to the issue of how to select students for future academic courses, we have now a framework of humanistic (and ethical) testing methods which promotes the personal and social qualities described in the policy statement of the 7th Curriculum, and which trains students in the process of assessing themselves (individually and collectively) according to cognitive, affective and social criteria, through self- and peer-assessment, portfolios, presentations, learner journals and learner conversations.

Rather than handcuff language teachers to the issue of selection, we would do well to remember that language tests are notoriously unsuccessful as predictors of academic success (Kerstjens & Nery, 2000, p. 85). If the goal of relative assessment is to infer academic success in future unrelated courses, therefore, it will not be achieved. If, on the other hand, language tests (and the resultant grades) are serving the function of gate-keeping for future courses of study and for career-related opportunities, then factors to be promoted should be those that are effective and beneficial in such situations (e.g. collaboration, critical thinking, problem-solving, etc.). If we take the broader view that education is about preparing the individual for his/her place in society, then a method of holistic and humanistic testing is indispensable.

4. Social Awareness

It would be possible to close the discussion at this point, concluding that relative assessment and its associated Gaussian grading system should be replaced by authentic assessment in a CRT context, since this would promote the qualities of “the educated person” as defined by the 7th National Curriculum, while 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 task-based 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 career⁶.

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)

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

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>)

Now the whole question of the educator’s role in dealing with planetary crisis becomes prominent. (O’Sullivan, 2001, p. 46)

Some type of holistic, or participating consciousness and a corresponding sociopolitical formation have to emerge if we are to survive as a species. (Berman, 1981, p. 23)

Thus Harri-Augstein et al (1991) call for attention to “our capacity for learning ... to

⁶ The quotations on this page, apart from the one by Rogers, are from “The quotations page”: <http://www.quotationspage.com/>

“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” (p. 7). In such a light, a call for the use of ethical assessment in CRT contexts is also a call for the fostering of social responsibility in the language classroom. It has never been justifiable to say “I am just a language teacher. I produce the grades I am told to produce,” but such an attitude is even more indefensible in a contemporary society crying out for creative, problem-solving, critical-thinking citizens.

VI. CONCLUSION

The situation regarding relative grading of language classes in Korean universities is reminiscent of Rea’s comment on testing in the late 1970s:

Although we would agree that language is a complex behaviour and that we would generally accept a definition of overall language proficiency as the ability to function in a natural language situation, we still insist on, or let others impose on us, testing measures which assess language as an abstract array of discrete items, to be manipulated only in a mechanistic way. Such tests yield artificial, sterile and irrelevant types of items which have no relationship to the use of language in real life situations. (Rea, cited in Weir 1998, p. 3)

Thus, university evaluation, which should be concerned with the personal growth of the learner (Ranson, 1994, p. 116) is often sacrificed to logistical demands of administrators, with any attempts to move towards humanistic, process-oriented, and student-centered assessment, being required to submit to unsound and unethical grading practices.

In response to this situation, this paper points to recent ministry policy statements, contemporary literature on language testing, and pressing demands for creative and responsible citizens in society in general, and concludes that language testing (in Korea and across the EFL community) must give these concerns greater weight in the test-design equation than heretofore. Given that high-stakes tests control the learning and teaching that occurs in the language classroom, it is proposed that holistic and humanistic testing methods (based on the “authentic” model of CRTs) incorporating “alternative” assessment procedures (self/peer-assessment, etc.) would promote the ideals of “Hongik-Ingan” and individual responsibility for learning. Further, it is recommended that program designers and professors employ ethical assessment practices (student-centered, pedagogically sound testing methods) in their programs, and that university administrators drop the “bell-curve” requirement for student grades.

If educators are successful in motivating their students to achieve excellence, then these students deserve appropriate grades, just as the student who shows sudden cognitive, affective and social improvement must not be made to punish his/her peers by his/her success. The 7th National Curriculum highlights the qualities that must be promoted in

every educational establishment and in every classroom. Authentic assessment, CRTs, and process-oriented alternative assessment methods are practical means towards this goal, and at the very least, can be adopted by teachers at the local level, while waiting for them to become an educational reality for all. Finally, this paper reminds the reader of W. B. Yeats' comment that education is not the filling of a pail, but the lighting of a fire (<http://www.quotationspage.com>). In order to address pressing issues in education and in society as a whole, educators must be encouraged and empowered to spark fires and to fan the resultant flames.

REFERENCES

- Alderson, C. J. (May, 2002). Ethics and politics, or politics as unethical behaviour? Paper presented at *The Language Assessment Ethics Conference*. Pasadena, S. California.
- ASCD 1964 yearbook committee. (1964). *Individualizing Instruction*. Washington: Association for Supervision and Curriculum Development.
- 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), 218-229.
- 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.
- Berman, M (1981). *The reenchantment of the world*. Ithaca, NY: Cornell University press.
- Bickley, V. (Ed.). (1989). *Language teaching and learning styles within and across cultures*. Hong Kong: Institute of Language in Education, Education Department.
- 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.
- Bruner, J. (1960). *The Process of Education*. Cambridge, Mass.: Harvard University 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.
- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64(1), 5-30.
- Dewey, J. (1916). *Democracy and Education*. New York: Macmillan.
- Finch, A. E. (1998). Designing and using a learner journal for false beginners: self-assessment and organization of learning. In *Self-Access Language Learning*, pp. 23-35. Hong Kong: Hong Kong Association of Self-Access Learning and Development.
- Finch, A. E. (2001). The non-threatening learning environment. *The Korea TESOL Journal*, 4(1), 133-158.
- Gardner, H. (1993). *The Unschooled mind: How children think and how schools should teach*. London: Fontana Press
- 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.
- 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.
- Keefe, J. W. (Ed.). (1979). *Student learning styles: Diagnosing and prescribing programs*. Reston, VA: National Association of Secondary School Principals.
- Kerstjens, M. and Nery, C. (2000). Predicative validity in the IELTS test. *IELTS Research Reports 2000 (vol 3)*, p. 85-108. IELTS Australia.
- Kohonen, V. (1999). Authentic assessment in affective foreign language education. In J. Arnold, (Ed.). *Affect in language learning*. Cambridge: Cambridge University Press. (pp. 279-294).
- Krishnamurti, J. (1981). *Education and the Significance of Life*. New York: Harper Collins
- Larson-Freeman, D. (1997). Chaos/complexity science and second language acquisition. *Applied Linguistics*, 18(2), 141-165.
- Lee, W. (1991). A task-based approach to oral communication testing of English as a foreign language. Ph.D. thesis, Manchester University. 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.
- 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.

- O'Malley, M., & Pierce, L. V. (1996). *Authentic assessment for English language learners*. New York: Addison Wesley.
- O'Neill, R. (1991). The plausible myth of learner-centredness: or the importance of doing ordinary things well. *English Language Teaching Journal*, 45(4), 293-304.
- O'Sullivan, E. (2001). *Transformative learning: Educational vision for the 21st century*. Canada: University of Toronto Press.
- Ministry of Education and Human Resources Development, (2001). *The school curriculum of the republic of Korea*. http://www.moe.go.kr/eng_26/
- Ranson, S. (1994). *Towards the learning society*. London: Cassell.
- Reid, J. (1987). The learning style preferences of ESL students. *TESOL Quarterly*, 21(1), 87-111.
- Rogers, C. R. (1951). *On becoming a person*. Boston: Houghton Mifflin.
- 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.), pp. 35-47. *Universals of second language acquisition*. Rowley, MA.: Newbury House.
- Skehan, P. (1988). State of the art article: Language testing, part 1. *Language Teaching*, 21(4), 211-218
- The quotations page*: <http://www.quotationspage.com/>
- Tudor, I. (1996). *Learner-centredness as language education*. Cambridge: Cambridge University Press.
- Underhill, N. (1989). Process in humanistic education. *English Language Teaching Journal*, 43(4), 250-260.
- Valencia, S. W. (1990). A portfolio approach to classroom reading assessment: The whys, whats, and hows. *The Reading Teacher*, 43, 338-340
- 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. (1994). The role of evaluation in ELT project design. *English Language Teaching Journal*, 48(1), 22-27.
- Williams, M., & Burden, R. L. (1997). *Psychology for language teachers: A social constructivist approach*. Cambridge: Cambridge University Press.

APPENDIX A

The learning process (adapted from Rogers, 1951)

1. Learning is the process of changing behavior in a positive direction.
2. Learning is an experience that occurs inside the learner and is activated by the learner.
3. Learning is the discovery of the personal meaning and relevance of ideas.
4. Learning (behavioral change) is a consequence of experience.
5. Learning is a cooperative and collaborative process.

6. Learning is an evolutionary process.
7. Learning is sometimes a painful process.
8. One of the richest resources for learning is the learner himself.
9. The process of learning is emotional as well as intellectual.
10. Learning fuses work and play.
11. Learning is a 'religious' experience.
12. The learner is a free and responsible agent.
13. The processes of problem solving and learning are highly unique and individual.
14. Teaching is learning. (Rogers 1951, p. 115)

Applicable levels: secondary, tertiary, adult

Key words: alternative assessment, authentic assessment, curriculum, criterion-referenced testing, humanism.

Andrew Edward Finch

경북대학교 사범대학 영어교육과

720-710 대구광역시 북구 산격동 1370

Tel: (053) 950-5832

Fax: (053) 384-1345

Email: aef@knu.ac.kr

Revised version received in July, 2002