

PEDAGOGICAL IMPLICATIONS FOR AN AUTHENTIC ORAL TESTING

BUSHRA SAADOON MOHAMMED AL-NOORI & DHEA MIZHIR KREBT

Research Scholar, University of Baghdad, College of Education Ibn-Rushed, Iraq

ABSTRACT

The present study examines current theory and practice regarding the assessment of foreign language oral performance, and discusses implications for curriculum designers and teachers in Ministry of Education in Iraq. 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 Iraqi schools according to prescribed textbooks "English for Iraq" and "Iraq Opportunities" for language assessment, being intrinsically demotivating for students and teachers. It is therefore argued that English Language programs should act on recent research findings and Ministry of Education policy statements by promoting positive affect (attitude, beliefs, confidence, motivation, etc.) in non-threatening learning environments, using criterion-referenced, and authentic assessment. On the other hand 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 universities level. It is hoped that these reflective models will 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.

KEYWORDS: Testing, Norm/Criterion Reference, Authentic

INTRODUCTION

It is necessary to note that 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, 1995:12), being used to measure language aptitude, proficiency, placement, diagnosis, progress, and achievement, and providing feedback for the program evaluator(s), wash-back information for teachers and students, and motivational wash-forward 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 intermediate and high schools English programs in Iraqi schools, therefore, it is appropriate at this point to view a brief survey of language assessment research.

HISTORY OF AUTHENTIC ORAL TEST

Defining a test as "a systematic method of eliciting performance which is intended to be the basis for some sort of decision making" 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, 1998:153). 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 and has been at the heart of language testing from its "pre-scientific" stage, 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 but was shown by commentators such as Alderson (1981), Morrow (1979) and Carroll (1981) to be still concerned with usage rather than use, therefore being only indirect tests of potential efficiency. Kelly (1978:245-246) also points 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. 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", 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, 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 (Weir, 1998:63).

TESTING THE ORAL SKILLS

Testing Listening

It can be tested alone, though very often it also involves speaking (think of oral answers to listening comprehension) and it always has a spoken (live or recorded) stimulus. (Hughes, 1989: 134)

The material here should be as authentic as possible and the recordings should be natural (with fillers and pauses) and with good quality. In order to write the items, we should keep in mind that with extended listening items should be kept sufficiently far apart in the passage and that students should be warned by key words. Next, time should not put pressure on candidates. If we just want to test oral comprehension, items and responses can be written in the native language. We should try and avoid setting questions which require the memorization of individual words in sentences. When administering the test, it is helpful if the speaker can be seen by the listeners (Oller, 1976:59).

Testing Speaking

Many testing experts and teachers coincide in mentioning the difficulties in testing the speaking skills. Madsen (1983; 148) mentions some of them: how to test fluency, how to get students to speak, how to evaluate so many things at once and, in addition, the practical problem of having to test each student individually. As we said with listening, sometimes it is neither possible nor desirable to separate the speaking skills from the listening ones. In spite of the obvious problems of scoring (highly subjective) and administration, we have to admit the necessity of its testing, especially nowadays when the ability to produce language is a requisite of the communicative trend. Apart from its importance, as

Doff (1988:63) suggests, oral tests should be given from time to time to give seriousness to this skill and also to parallel the importance given to it in class and in our methodology.

The oral test should not be improvised and we should try and make students feel at ease, including major areas and interesting topics and not talking too much ourselves (Hughes, 1989: 105-107). For beginners can use imitation exercises (repetition of sentences), directed requests, reading aloud and directed-response role-play. Paraphrase (combining speaking with either listening or reading and with the help of pictures), guided role-play (with prompts) and split dialogues are useful with intermediate students. At advanced levels can set oral interviews, speaking from tape-recorded stimuli, short talks, group discussion (especially with consensus-seeking activities) and role-playing.

Bachman's Model of Communicative Language Test

The components of communicative language ability to be tested were variously described at this time, and early frameworks for testing communicative competence were proposed. However, these were neither practical, systematic, nor comprehensive, and were unable to advance prediction and generalization in any substantial way, this problem was addressed by Bachman (1990:44) 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 (1988:155) articulates 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".

Communication Effectiveness Scoring

In a sense similar in objectives to primary- trait scoring, this of measuring the quality of prose is also concerned with the effects it has on an audience. But, operationally, the method is very different from primary- trait scoring. Hirsch and Harrington(1981:342)describe the theoretical basis for this new method and some of its advantages over traditional method of scoring. The method is also similar in some ways to recent approaches being taken by cognitive educationalists, in which the theory and structures of reading comprehension research are applied to analysis of text. Usually, an objective index of communicative effectiveness, such as reading speed or comprehension, is derived for the assessment (Bracewell and Frederiksen, 1982: 7).

TASK-BASED ORAL TESTING

Based on above, 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. 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", but allocate limited processing attention in appropriate ways ,drawing on parallel coding systems for efficiency of real-time communication. Skehan therefore proposes 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, 1998:169).

In contrast to performance evaluations which can use reliable 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, 1998:177), 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, Skehan notes that "we need to know more about the way tasks themselves influence (and constrain) performance", 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, 1998:177).

AUTHENTIC ASSESSMENT

In this type of 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 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 "examining what students can do with their language, through real-life language use tasks. 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, 1997:291).

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 behavior

and recruit methodological skills to drive the experiment forward.

(Harri-Augstein&Thomas, 1991:6)

In this way 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

	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 students 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 suggestion 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 learningcomparesstudents to their own past performances and the aims .
12	Intrinsic learning for a grade.	Extrinsic learning for its own sake .

CRITERION-REFERENCE AND NORM-REFERENCE OF 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 (1988:33) finds a high 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, students believed that ability was shown by success with little effort, and they judged their ability lower. As Darling-Hammond (1994:110) 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.". 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 students at either rends of the scale .(Brown (1995) classifies CRTs and NRTs according to their test characteristics and logistical dimensions (Table 2), 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 students, and are summative in nature, assessing whether students have been (orare likely to be) successful (however this is defined), but unable to comment on why or how, or on what should happen next (Williams & Burden, 1994:22). 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

Test Characteristics	Crts	Nrts
Underlying Purpose	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 Strategies	Test and answers to students	Only scores go to students
Logistical Dimension		
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(2-4hours) administration
Cost	Teacher time & duplication	Test booklets, tapes, proctor

(Brown, 1995:12)

THE IRAQI SITUATION

The prescribed textbooks for English language subject in Iraqi schools are “English for Iraq” and “Iraq Opportunity”. The Schools consists of thousands full-time teachers, all non-native English speakers. This yearlong program is divided into two courses, each lasting one semester. Classes meet 50 minutes a day per week per semester.

Our students have all had 6 years of English instruction in secondary school and, thus, possess a basic understanding of the language. On the whole, their written English is better than their spoken English; they have had little experience with English as a living language. Therefore, the primary goals of our program are to teach English as a linguistic-cultural-social unit, to facilitate students’ use of the language as a tool for communication, and to “actively develop students’ ability to communicate in a socially appropriate manner” (Kurzweil, et al., 2002: 32).

The advocating of CRTs, and in particular of authentic assessment, in tertiary EFL classrooms in Iraq, is especially appropriate in the light of the goals of school education of the Iraqi Ministry of Education. Here we find the ideal of 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". The well-educated person that these goals aim to promote is further defined as :

- 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.

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. If education is to successfully foster autonomous, informed students 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 (Finch, 2001: 21). However, there are instances of the NRT approach at tertiary level which must be addressed in order for this process to gain momentum.

Implications and Issues for Iraqi Teachers

Self- and peer-assessment are thus practical and effective assessment methods in English language classes, addressing educational goals espoused in the Iraqi Curriculum. In handing over a large part of the assessment burden and responsibility to the students, 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. There is a fairly consistent overall agreement between self-assessment and external criteria. Doubts about the sincerity of the students. One reason put forward by teachers for not sharing responsibility for assessment is that students will "cheat" and produce unrealistic scores. Dickinson (1987: 150), however, points out that "cheating" (a process in which a learner seeks to obtain personal advantage by unfair means, 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".

Doubts regarding the reliability of self-assessment in formal education. Research on peer assessment 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: 85).

Reluctance of teachers to lose control of assessment. 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-centered evaluation techniques" (Oscars son, 1989: 11).

The need for students to receive training and practice in assessing their own performances. Learner training for self-assessment can help students 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, 1988:101).

CONCLUSIONS

Testing is very affective since we talk about the growth of students, this issue due to the transferring of learning and integration of that learning with the individual constructs and meaning.

The present study concludes that it is no longer acceptable to use discrete-item testing as a sample of behavior as a mean of analyze underlying skill or capability, self/peer testing represents a modern mean of dealing with testing, this can be done by making our students participating and taking part of the process learning and students should have an awareness of their improvement.

As for Iraqi students, self-assessment or self-monitoring could be consider of a great importance that leads to integrate such assessment with everyday classroom activities. So, our aim is to reach an authentic assessment in contextualized criterion-reference because this matter consider as desirable in testing English language in Iraq.

One of the most important factor of getting an authentic assessment is recognizing of rising awareness of promoting the social context of the test. The aim of education is to teach not how to think, what to think - rather to enhance minds, so, the educated student is the only one who has to learn for change.

The use of authentic testing in contextualized criterion reference is also a call for the enhancing of responsibility in the language classroom. It is no more acceptable for teachers to say that my job is only to teach English. The "English for Iraq" Curriculum highlights the qualities that must be promoted in students in every educational settings, and in every classroom. Authentic assessment, self- and peer-assessment are practicalwaysto achieve this goal, and can be used by teachers at the local level.

The results also indicate that there is a gap between theory and practice. In fact, teachers believe that students` examinations should be scored in the light criterion reference. But in practice, they neglect or put little emphasis on the authentic test. Iraqi teachers should be trained to use informal assessment techniques, namely tests and examinations.

Criterion-referenced grades are most suitable when teachers wants to test the oral skills that a student has learned through classroom teaching. Most criterion-referenced test have a cut score, which determines success or failure based on an established percentage correct.

It is clear to say that criterion-referenced grades give us how well a student performs against an objective or standard, as opposed to against another student. Learning objective in the class is 'students should be able to correctly divide fractions.' The criterion-referenced grades shows that student meets the objective successfully. One of the problems of criterion-referenced grades is that the assessment of oral skills is difficult to determine through the use of one score on an assessment. In the other hand, norm-referenced grades are useful when teachers want to compare among large numbers of students.

REFERENCES

1. Ames, C., & Archer, J. (1988). Achievement Goals in the Classroom: Students' Learning Strategies and Motivation Process. *Journal of Educational Psychology*, 80, 260-267 .
2. Bachman, L. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press .

3. Blue, G.M. (1988). Self-Assessment: The Limits of Learner Independence. In A. Brookes & P. Grundy(Eds.). Individualization and Autonomy in Language Learning. ELT Documents100-118. London: Modern English Publications and the British Council.
4. Brace well, R.J. and Frederiksen, J.D. (1982) Cognitive Processes in Composing and Comprehending. Education Psychologist Journal. Vol.17, No.3, May, pp.65-78.
5. Brown, J. D. (1995). Differences between Norm-Referenced and Criterion-Referenced Tests. In J. D .
6. Darling-Hammond, L. (1994). Performance-Based Assessment and Educational Equity. Harvard Educational Review, 64(1), 5-30 .
7. Dickinson, L. (1978). Autonomy, Self-Directed Learning and Individualization. In A. Brookes & P .Grundy (Eds.). Individualization and Autonomy in Language Learning. ELT Documents 103 28-7Modern English Publications and the British Council .
8. Dickinson, L. (1987). Self-instruction in Language Learning. Cambridge. Cambridge University Press .
9. Doff, A. (1988). Teach English: A Training Course for Teachers. Cambridge: Cambridge University Press.
10. Finch, A. E. (2001). A Formative Evaluation of a Task-Based Conversation English Program. PACJournal, 1(1), 125-146 .
11. Harri-Augstein S., & Thomas, L. (1991). Learning Conversations: The Self-Organized Learning Way to Personal and Organizational Growth. London: Rout ledge .
12. Harrington, D.P. (1981). Measuring the Communicative Effective Prose: In Writing. The Nature, Development , and Teaching of Writing Communications Journal, Vol.2.No.3,pp.90-120 C.H. Frederiksen and J. F. Dominic, eds Hillsdale, N.J., LawreceEarlbaum Associates.
13. Hughes, A. (1989). Testing for Language. Teachers. Cambridge: Cambridge University Press.
14. Kelly, R. (1978). On the Construct Validity of Comprehensive Tests: An Exercise in Applied Linguistics . University of Queensland Ph. D Thesis .
15. Kurzweil, J. et al. (2002). Communication Syllabus: Designed by Consensus. Kansai University Forum for Foreign Language Education; 31-43.
16. Madsen, H. S, (1983). Techniques in Testing. London: Longman.
17. Oller, J. (1976). Evidence of a General Language Proficiency Factor: An Expectancy Grammar". Retrieved from <http://www.DieNeureSprachen76:165-174>.
18. Oscars son, M. (1978). Approaches to Self-Assessment in Foreign Language Learning. Council of Europe, Council for Cultural Co-operation, Strasburg, France .
19. Oscars son, M. (1989). Self-Assessment of Language Proficiency: Rationale and Implications. Language Testing, 6(1), 1-13 .

20. Oscars son, 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.
21. Skehan, P. (1988). Language Testing, Part 1: State of the Art Article. Language Teaching, 21(4) 218-211
22. Skehan, P. (1998). A Cognitive Approach to Language Learning. Oxford: Oxford University Press .
23. Weir, C. J. (1998). Communicative Language Testing. Exeter: University of Exeter Press .
24. Williams, M., & Burden, R. L. (1997). Psychology for Language Teachers: A Social Constructivist Approach. Cambridge: Cambridge University Press.