

Language Performance Assessment: Current Trends in Theory and Research

Abdel Salam Abdel Khalek El-Koumy
Full Professor of Teaching English as a Foreign Language
School of Education in Suez
Suez Canal University

Language Performance Assessment: Current Trends in Theory and Research

First published 2003 by Dar An-Nahda Al-Arabya, 32 Abdel-Khalek Tharwat St., Cairo, Egypt.

Deposit No.: 4779/2003

ISBN: 977-04-4099-x

Dedication

This paper is dedicated to my ever caring parents, and to my wife and children, for their help and encouragement. Without their support, this work would not be a reality. I would like to express my deep appreciation to all of them.

Contents

Overview

Chapter One: Background Information

- 1.0 Introduction**
- 1.1 Definition of Performance Assessment**
- 1.2 Theoretical Assumptions Underlying Performance Assessment**
- 1.3 Purposes of Performance Assessment**
- 1.4 Performance Assessment Procedures**
- 1.5 Merits and Demerits of Performance Assessment**

Chapter Two: Performance Assessment Formats

- 2.0 Introduction**
- 2.1 Naturally Occurring Assessment**
- 2.2 On-Demand Assessment**
 - 2.2.1 Oral Interviews**
 - 2.2.2 Individual or Group Projects**
 - 2.2.3 Portfolios**
 - 2.2.4 Dialogue Journals**
 - 2.2.5 Story Retelling**
 - 2.2.6 Oral Reading**
 - 2.2.7 Group Discussions**
 - 2.2.8 Role Playing**
 - 2.2.9 Teacher-Student Conferences**
 - 2.2.10 Verbal Reports**
- 2.3 Criteria for Selecting Performance Assessment Formats**

Chapter Three: Alternative Groupings for Performance Assessment

- 3.0 Introduction**
- 3.1 Self-Assessment**
- 3.2 Peer Assessment**
- 3.3 Group Assessment**

Chapter Four: Performance Assessment via Computers

4.1 Theoretical Background

4.2 Research on Performance Assessment via Computers

Chapter Five: Reliability and Validity of Performance Assessment

5.1 Theoretical Background

5.2 Research on the Validity and Reliability of Language Performance Assessment

Chapter Six: Summary and Conclusions

References

Overview

The current interest in performance assessment grew out of the widespread dissatisfaction with standardized tests, and of the widespread belief that schools do not develop productive citizens. The purpose of this paper is to review the theoretical and empirical literature relevant to recent trends in language performance assessment. Following a definition of performance assessment, this paper considers: (1) theoretical assumptions underlying performance assessment; (2) purposes of performance assessment; (3) performance assessment procedures; (4) merits and demerits of performance assessment; (5) language performance assessment formats and research relevant to each format; (6) criteria for selecting performance assessment formats; (7) alternative groupings for assessing student performance; (8) performance assessment via computers and research related to this area; (9) reliability and validity of performance assessment and research related to this area; and (10) conclusions drawn from the literature reviewed in this paper.

Chapter One

Background Information

1.0 Introduction

The current interest in performance assessment grew out of the widespread dissatisfaction with standardized tests (Bachman, 2000), and of the widespread belief that schools do not develop productive citizens (Roeber, 1995). The purpose of this paper is to review the literature relevant to this type of assessment over the last ten years.

1.1 Definition of Performance Assessment

As defined by Nitko (2001) performance assessment is the type of assessment that “(1) presents a hands-on task to a student and (2) uses clearly defined criteria to evaluate how well the student achieved the application specified by the learning target” (p. 240). Nitko goes on to state that “[t]here are two aspects of a student’s performance that can be assessed: the product the student produces and the process a student uses to complete the product” (p. 242).

In their *Dictionary of Language Testing*, Davies et al. (1999) define performance assessment as “a test in which the ability of candidates to perform particular tasks, usually associated with job or study requirements, is assessed” (p. 144). They maintain that this performance test “uses ‘real life’ performance as a criterion and characterizes measurement procedures in such a way as to approximate non-test language performance” (loc. cit.).

Kunnan (1998) states that performance assessment is “concerned with language assessment in context along with all the skills and not in discrete-point items presented in a decontextualized manner” (p. 707). He adds that in this type of assessment “test takers are assessed on what they can do in situations similar to ‘real life’” (loc. cit.).

Thurlow (1995) states that performance assessment “require[s] students to create an answer or product that demonstrates their knowledge and skills” (p. 1). Similarly, Pierce and O'Malley (1992) define performance assessment as “an exercise in which a student demonstrates specific skills and competencies in relation to a continuum of agreed upon standards of proficiency or excellence” (p. 2).

As indicated--from the aforementioned definitions--performance assessment focuses on the following:

- (a) application of knowledge and skills in realistic situations,
- (b) open-ended thinking,
- (c) wholeness of language, and
- (d) processes of learning as well as the products of these processes.

1.2 Theoretical Assumptions Underlying Performance Assessment

Performance assessment is consistent with modern learning theories. It reflects the cognitive learning theory which suggests that students must acquire both content and procedural knowledge. Since particular types of procedural knowledge are not assessable via traditional tests, cognitivists call for performance to assess this type of knowledge (Popham, 1999).

Performance assessment is also compatible with Howard Gardner's (1993, 1999) theory of multiple intelligences because this type of assessment has the potential of permitting students' achievements to be demonstrated and evaluated in several different ways. In an interview with Checkley (1997), Gardner himself expresses this idea in the following way:

The current emphasis on performance assessment is well supported by the theory of multiple intelligences. . . . [L]et's not look at things through the filter of a short-answer test. Let's look instead at the performance that we value, whether it is linguistic, logical, aesthetic, or social performance let's never pin our assessments of understanding on just one particular measure, but let's always allow students to show their understanding in a variety of ways.

Furthermore, performance assessment is consistent with the constructivist theory of learning which views learners as active participants in the construction and evaluation of their learning processes and products. Based on this theory, performance assessment involves students in the process of assessing their own performance (Shepard, 2000).

1.3 Purposes of Performance Assessment

A large portion of performance assessment literature (e.g., Arter et al., 1995; Katz, 1997; Khattri et al., 1998; Tunstall and Gipps, 1996) indicates that performance assessment serves the following purposes:

- (a) documenting students' progress over time,
- (b) helping teachers improve their instruction,
- (c) improving students' motivation and increasing their self-esteem,
- (d) helping students improve their own learning processes and products,

- (e) making placement or certification decisions, and
- (f) providing parents and community members with directly observable products concerning students' performance.

1.4 Performance Assessment Procedures

The major stages of performance assessment, synthesized from a number sources (Cheng, 2000; Gallagher, 1998; Martinez, 1998; Nitko, 2001; Palomba and Banta, 1999; Shaklee et al. 1997; Stiggins, 1994; Wiggins, 1993), are the following:

(1) Deciding What to Assess and How to Assess It

At this stage, the teacher should become quite clear about what he/she will assess. He/she should also become quite clear about how he/she will assess students' performance. More specifically, the teacher at this stage needs to address questions such as the following:

- (a) Which learning target(s) will I assess?
- (b) Will the test task(s) assess the processes, or the products of students' learning, or both?
- (c) Should my students be involved in assessing their own performance?
- (d) Should I use holistic or analytic assessment rubrics?

(2) Developing Assessment Tasks and Performance Rubrics

In light of the answers to stage-one questions, the teacher develops the assessment tasks and performance rubrics. In doing so, he/she must make sure that students will understand what he/she expects them to do. After developing the assessment tasks and performance rubrics, the teacher should pilot them on subjects that represent the target test population to identify problems and remove them.

(3) Assessing Students' Performance

In light of the performance rubrics--created at the second stage--the teacher scores students' performance. The questions the teacher might answer at this stage include:

- (a) What are the strengths in the student's performance?
- (b) What are the weaknesses in the student's performance?
- (c) What evidence of self-, peer- or group assessment appears in the student's performance?

(4) Interpreting and Reporting Students' Results

At this stage, the teacher analyses and discusses students' results in light of the teaching strategies he/she used as well the learning strategies students employed. In light of these results, the teacher also suggests ways to develop

his/her teaching strategies and to improve students' performance. As Wiggins (1993) puts it:

Assessment should improve performance, not just audit it....
Assessment done properly should begin conversations about performance not end them....If the testing we do in the name of accountability is still one event, year-end testing, we will never obtain valid and fair information. (pp. 5, 13, 267)

At this stage, the teacher should also create a performance-based report card. This card should focus on reporting the strengths and weaknesses of the student performance instead of numerical grades (Fleurquin, 1998; Stix, 1997). Simply, the teacher must address the following questions at this stage:

- (a) What do these results tell me about the effectiveness of the instructional program?
- (b) What kind of evidence will be useful to me and to my students?
- (c) How can I report my students' results?

1.5 Merits and Demerits of Performance Assessment

The advantages of performance assessment include its potential to assess 'doing,' its consistency with modern learning theories, its potential to assess processes as well as products, its potential to be linked with teaching and learning activities, and its potential to assess language as communication (Brualdi, 1998; Linn and GronLound, 1995; Mehrens, 1992; Stiggins, 1994). Although performance assessment offers these advantages over traditional assessment, it also has some distinct disadvantages. The first disadvantage is that performance assessment tasks take a lot of time to complete (Oosterhof, 1994). If such tasks are not part the instructional procedures, this means either administering fewer tasks (thereby reducing the reliability of the results), or reducing the amount of instructional time (Nitko, 2001). The second disadvantage is that performance assessment tasks do not assess all learning targets well, particularly in the situations where some learning targets focus on bits and pieces of information (Bailey, 1998; Soodak, 2000). The third disadvantage is that the scoring of performance tasks takes a lot of time (Rudner and Boston, 1994). The fourth disadvantage is that scores from performance tasks may have lower scorer reliability (Fuchs, 1995; Hutchinson, 1995; Koretz et al., 1994; Miller and Legg, 1993). The fifth and final disadvantage is that performance tasks may be discouraging to less able students (Gomez, 2000; Meisles et al., 1995).

Chapter Two

Performance Assessment Formats

2.0 Introduction

Assessment specialists (e.g., Cohen, 1994; Genesee and Upshur, 1996; Nitko, 2001; Popham, 1999; Stiggins, 1994) have proposed a wide range of alternatives for assessing students' performance. Such alternatives fall into two major categories: (I) naturally occurring assessment, and (II) on-demand assessment. Each of these categories is described below.

2.1 Naturally Occurring Assessment

This type of assessment refers to observing students' normally occurring performance in naturalistic environments without intervening or structuring the situation, and without informing the students that they are being assessed (Fisher, 1995; Stiggins, 1994; Tompkins, 2000). The major advantage of this type of assessment is that it provides a realistic view of a student's language performance (Norris and Hoffman, 1993). Another advantage of this type of assessment is that it is not a source of anxiety and psychological tensions for the students (Antonacci, 1993). However, this type of assessment does not seem practically feasible because of the following shortcomings (Nitko, 2001):

- (a) it is difficult and time-consuming to use with large numbers of students,
- (b) it is inadequate on its own because it cannot provide the teacher with all the data he/she needs to thoroughly assess students' performance, and
- (c) the teacher cannot ensure that all students will perform the same tasks under similar conditions.

Research on Naturally Occurring Assessment

A survey of research on naturally occurring assessment indicated that whereas several studies used this type of assessment as a research tool in addition to standardized testing (e.g., Brooks, 1995; Lemons, 1996; Mauerman, 1995; Santos, 1998; Wright, 1995), no studies investigated its effect on students' performance. However, indirect support for this type of assessment comes from studies which found that test anxiety negatively

affected students' language performance (e.g., Dugan, 1994; Ross, 1995; Teemant, 1997).

2.2 On-Demand Assessment

Because of the shortcomings of the naturally occurring assessment, many assessment formats were developed to elicit certain types of performance. These formats include oral interviews, individual or group projects, portfolios, dialogue journals, story retelling, oral reading, group discussions, role playing, teacher-student conferences, retrospective and introspective verbal reports, etc. This section describes the on-demand formats that are well suited for assessing language performance.

2.2.1 Oral Interviews

Oral interviews are the simplest and most frequently employed format for assessing students' oral performance and learning processes (Fordham et al., 1995; McNamara, 1997a; Thurlow, 1995). This format can take different forms: the teacher interviewing the students, the students interviewing each other, or the students interviewing the teacher (Graves, 2000).

Chalhoub-Deville (1995) claims that oral interviews offer a realistic means of assessing students' oral language performance. However, opponents of this format argue that such interviews are artificial because students are not placed in natural, real-life speech situations, and are thus susceptible to psychological tensions and to constraints of style and register (Antonacci, 1993). They also argue that a face-to-face interview is time-consuming because it cannot be conducted simultaneously with more than one student by a single interviewer (Weir, 1993).

Stansfield (1992) suggests that oral interviews should progress through the following four stages:

- (a) **Warm-up:** At this stage the interviewer puts the interviewee at ease and makes a very tentative estimate of his/her level of proficiency.
- (b) **Level checks:** During this stage, the interviewer guides the conversation through a number of topics to verify the tentative estimate arrived at during the previous stage.
- (c) **Probes:** During this stage the interviewer raises the level of the conversation to determine the limitations in the interviewee proficiency or to demonstrate that the interviewee can communicate effectively at a higher level of language.

(d) Wind-down: At this stage the interviewer puts the interviewee at ease by returning to a level of conversation that the interviewee can handle comfortably.

To effectively integrate oral interviews with language learning, Tompkins and Hoskisson (1995) suggest that students can conduct interviews with each other or with other members of the community. They further suggest that students should record such interviews and submit the tapes to the teacher for assessment (*ibid.*).

To make interviewing intimately tied to teaching, Maden and Taylor (2001) suggest that the interviewer, usually the teacher, should enter into interaction with students for both teaching and assessment purposes.

To make interviewing intimately tied to the ultimate goals of assessment, the interviewer should use interview sheets (Lumley and Brown, 1996). Such sheets usually contain the questions the interviewer will ask and blank spaces to record the student's responses (*ibid.*). Additionally, audio and video cassettes can be made of oral interviews for later analysis and evaluation (Tannenbaum, 1996).

Stansfield and Kenyon (1996) suggest using a tape-recorded format as an alternative to face-to-face interviews. They claim that such a tape-recorded format can be administered to many students within a short span of time, and that this format can help assessors to control the quality of the questions as well as the elicitation procedures (*ibid.*).

Alderson (2000) suggests that oral interviews can be extremely helpful in assessing students' reading strategies and attitudes towards reading. In such a case, students can be asked about the texts they have read, how they liked them, what they did not understand, what they did about this, and so on (*ibid.*).

Research on Oral Interviews

A survey of recent research on oral interviews indicated that whereas several studies used this format as a research tool for assessing students' oral performance (e.g., Berwick and Ross, 1996; Careen, 1997; Fleming, and Walls, 1998; Kiany, 1998; Lazaraton, 1996), and for exploring students' reading strategies (e.g., Harmon, 1996; Vandergrift, 1997), no studies used it as an on-going technique for both assessment and instructional purposes.

2.2.2 Individual or Group Projects

Many educators and assessment specialists (e.g., Greenwald and Hand, 1997; Gutwirth, 1997; Katz and Chard, 1998; Ngeow and Kong, 2001;

Sokolik and Tillyer, 1992) suggest assessing students' language performance with group or individual projects. Such projects are in-depth investigations of topics worth learning more about. Such investigations focus on finding answers to questions about a topic posed either by the students, the teacher, or the teacher working with students (Gutwirth, 1997).

The advantages of using projects for both instructional and assessment purposes include helping students bridge the gap between language study and language use, integrating the four language skills, increasing students' motivation to learn, taking the classroom experience out into the community, using the language in real life situations, allowing teachers to assess students' performance in a relatively non-threatening atmosphere, and deepening personal relationships between the teacher and students and among the students themselves (Fried-Booth, 1997; Katz, 1997; Katz and Chard, 1998; Warschauer et al., 2000). However, project work may take a long time and require human and material sources that are not easily accessible in the students' environment.

- Katz and Chard (1998) suggest that a project topic is appropriate if
- (a) it is directly observable in the students' environment,
 - (b) it is within most students' experiences,
 - (c) direct investigation is feasible and not potentially dangerous,
 - (d) local resources (e.g., field sites and experts) are readily accessible,
 - (e) it has good potential for representation in a variety of media (e.g., role play, writing),
 - (f) parental participation and contributions are likely,
 - (g) it is sensitive to the local culture and culturally appropriate in general,
 - (h) it is potentially interesting to students, or represents an interest that teachers consider worthy of developing in students,
 - (i) it is related to curriculum goals,
 - (j) it provides ample opportunity to apply basic skills (depending on the age of the students), and
 - (k) it is optimally specific—neither too narrow nor too broad.

Project topics are usually investigated by a small group of students within a class, sometimes by a whole class, and occasionally by an individual student (Greenwald and Hand, 1997). During project work students engage in many activities including reading, writing, interviewing, recording observations, etc. (Ngeow and Kong, 2001).

Fried-Booth (1997) suggests that a project should move through three stages: project planning, carrying out the project, and reviewing and evaluating the work. She further suggests that at each of these three stages, the teacher should work with the students as a counselor and consultant (*ibid.*). Similarly, Katz (1994) suggests the following three stages for project work:

- (a) selecting the project topic,
- (b) direct investigation of the project, and
- (c) culminating and debriefing events.

Stoller (1997) proposes the following ten-step procedure for integrating project work into content-based language classrooms:

- (a) students and instructor agree on a theme for the project,
- (b) students and instructor determine the final outcome,
- (c) students and instructor structure the project,
- (d) instructor prepares students for information gathering,
- (e) students gather information,
- (f) instructor prepares students for compiling and analyzing data,
- (g) students compile and analyze data,
- (h) instructor prepares students for the culminating activity,
- (i) students present the final product,
- (j) students evaluate the product.

Recently, new technology has made it possible to implement projects on the computer if students have the Internet access. For information about how this can be done see, Warschauer (1995) and Warschauer et al. (2000).

Research on Language Projects

A review of research on language projects revealed that only two studies were conducted in this area in the last decade. In one of them, Hunter and Bagley (1995) explored the potential of the global telecommunication projects. Results indicated that such projects developed students' literacy skills, their personal and interpersonal skills, as well as their global awareness. In the other study, Smithson (1995) found that the on-going assessment of writing through projects improved students' writing.

2.2.3 Portfolios

Portfolios are purposeful collections of a student's work which exhibit his/her performance in one or more areas (Arter et al., 1995; Barton and Coley, 1994; Graves, 2000). In language arts, there is a spreading emphasis on this format as an alternative type of assessment (Gomez, 2000; Jones and Vanteirsburg, 1992; Newman and Smolen, 1993; Pierce and O'Malley, 1992). Many advantages have been claimed for this type of assessment. The first advantage is that this alternative links assessment to

teaching and learning (Hirvela and Pierson, 2000; Porter and Cleland, 1995; Shackelford, 1996). The second advantage of this alternative is that it gives students a voice in assessment and helps them diagnose their own strengths and weaknesses (Courts and McInerney, 1993). The third advantage is that this alternative can be tailored to the student's needs, interests, and abilities (Ediger, 2000). Additional advantages of this alternative are stated by Arter et al. (1995) in the following way:

The perceived benefits [of portfolios as an assessment format] are that the collection of multiple samples of student work over time enables us to (a) get a broader, more in-depth look at what the students know and can do; (b) base assessment on more "authentic" work; (c) have a supplement or alternative to report cards and standardized tests; and (d) have a better way to communicate student progress to parents. (p. 2)

However, as with all performance assessment formats, it is quite difficult to come up with consistent evaluations of different students' portfolios (Dudley, 2001; Hewitt, 2001). Another problem with portfolio assessment is that it takes time to be carried out properly (Koretz, 1994; Ruskin-Mayher, 1999). In spite of these demerits, portfolio assessment is growing in use because its merits outweigh its demerits.

Tannenbaum (1996) suggests that the following types of materials can be included in a portfolio:

- (a) audio-and videotaped recordings of readings or oral presentations,
- (b) writing samples such as dialogue journal entries and book reports,
- (c) writing assignments (drafts or final copies),
- (d) reading log entries,
- (e) conference or interview notes and anecdotal records,
- (f) checklists (by teacher, peers, or student), and
- (g) tests and quizzes.

To gain multiple perspectives on students' language development, Tannenbaum (1996) further suggests that students should include more than one type of materials in the portfolio. More specifically, Farr and Tone (1994) suggest that the best guides for selecting work to include in a language arts portfolio are these two questions: "What do these materials tell me about the student?" and "Will the information obtained from these materials add to what is already known?" However, May (1994) contends that teachers should let students decide what they want to include in a portfolio because this makes them feel they own their

portfolios and that this feeling of ownership leads to caring about portfolios and to greater effort and learning.

Tenbrink (1999) suggests that using portfolios for assessing students' performance requires the following:

- (a) deciding on the portfolio's purpose,
- (b) deciding who will determine the portfolio's content,
- (c) establishing criteria for determining what to include in the portfolio,
- (d) determining how the portfolio will be organized and how the entries will be presented,
- (e) determining when and how the portfolio will be evaluated, and
- (f) determining how the evaluations of the portfolio and its contents will be used.

To be an effective assessment format, portfolios must be consistent with the goals of the curriculum and the teaching activities (Arter and Spandel, 1992; Tenbrink, 1999). That is, they should focus on the same targets emphasized in the curriculum as well as the daily instruction activities. As Tenbrink (1999) puts it, "Portfolios can be a very powerful tool if they are fully integrated into the total instructional process, not just a tag-on at the end of instruction" (p. 332).

To make portfolios more useful as an assessment format, some educators (e.g., Farr, 1994; Grace, 1992; Wiener and Cohen, 1997) suggest that the teacher should occasionally schedule and conduct portfolio conferences. Through such conferences, students share what they know and gain insights into how they operate as readers and writers. Although such conferences may take time, they are pivotal in making sure that portfolio assessment fulfills its potential (Ediger, 1999). In order to make such conferences time efficient, Farr (1994) suggests that the teacher should encourage students to prepare for them and to come up with personal appraisals of their own work.

Since current technology allows for the storage of information in the form of text, graphics, sound, and video, many assessment specialists (e.g., Barret, 1994; Chang, 2001; Hetterscheidt et al., 1992; Wall, 2000) suggest that students should save their portfolios on a floppy disk or on a website. Such assessment specialists claim that this makes students' portfolios available for review and judgment by others. Other advantages of electronic portfolios are stated by Lankes (1995) this way:

The implementation of computer-based portfolios for student assessment is an exciting educational innovation. This method of assessment not only offers an authentic demonstration of accomplishments, but also allows students to take responsibility

for the work they have done. In turn, this motivates them to accomplish more in the future. A computer-based portfolio system offers many advantages for both the education and the business communities and should continue to be a popular assessment tool in the “information age.” (p. 3)

Research on Language Portfolios

A survey of recent research on portfolios revealed that many investigators used this format as a research tool for assessing students’ writing (e.g., Camp, 1993; Condon and Hamp-Lyons, 1993; Hamp-Lyons, 1996). A second body of research indicated that portfolio assessment, that was situated in the context of language teaching and learning, improved the quality and quantity of students’ writing (e.g., Horvath, 1997; Moening and Bhavnagri, 1996), enabled learning disabled students to diagnose their own strengths and weaknesses (e.g., Boerum, 2000; Holmes and Morrison, 1995), and had a positive effect on teachers’ understanding of assessment and on students’ understanding of themselves as learners and writers (Ponte, 2000; Tanner, 2000; Wolfe, 1996). A third body of research investigated teachers’ or students’ perceptions of portfolios after their involvement in portfolio assessment. In this respect, Lylis (1993) found that teachers felt that portfolio assessment helped them document students’ development as writers and offered them a greater potential in understanding and supporting their students’ literacy development. Additionally, Anselmo (1998) found that students, who assessed their own portfolios, felt that their motivation increased.

2.2.4 Dialogue Journals

Dialogue journals--where students write freely and regularly about their activities, experiences, and plans--can be a rich source of information about students’ reading and writing performance (Bello, 1997; Borich, 2001; Peyton and Staton, 1993; Schwarzer, 2000). Such dialogues are also a powerful tool with which teachers can collect information on students’ reading and writing processes (Garcia, 1998; Graves, 2000).

The advantages of using dialogue journals for both instructional and assessment purposes include individualizing language teaching, making students feel that their writing has a value, promoting students’ reflection and autonomous learning, increasing students’ confidence in their own ability to learn, helping the instructor adapt instruction to better meet students’ needs, providing a forum for sharing ideas and assessing students’ literacy skills, using writing and reading for genuine communication, and increasing opportunities for interaction between students and teachers (Bromley, 1993; Burniske, 1994; Cobine, 1995a; Courts and McInerney, 1993; Garcia, 1998; Garmon, 2000; Graves, 2000;

Smith, 2000). However, dialogue journals require a lot of time from the teacher to read and respond to student entries (Worthington, 1997).

Peyton (1993) offers the following suggestions for responding to students' dialogue writing:

- (a) commenting only on the content of the student's entry,
- (b) asking open-ended questions and answering student questions,
- (c) requesting and giving clarification, and
- (d) offering opinions.

However, Routman (2000) cautions that responding only to the content of the dialogue journals may lead students to get accustomed to sloppy writing and bad spelling as the norm for writing.

Both Reid (1993) and Worthington (1997) agree that the dialogue journal partner does not have to be the teacher and that students can write journals to other students in the same class or in another class. They claim that this reduces the teacher's workload and makes students feel comfortable in asking for advice about personal problems. In such a case, Worthington (ibid.) suggests that the teacher can put a box and a sign-in notebook in his/her office to help him/her monitor the journal exchanges between pairs.

With access to computer networks, many educators and assessment specialists (e.g., Hackett, 1996; Knight, 1994; LeLoup and Ponterio, 1995; Peyton, 1993) suggest that students can keep electronic dialogue journals with the teacher or other students in different parts of the world.

Research on Dialogue Journal Writing

A review of recent dialogue journal studies indicated that keeping a dialogue journal improved students' writing (e.g., Cook, 1993; Hannon, 1999; Ho, 1992; Song, 1997; Worthington, 1997), and increased their self-confidence (e.g., Baudrand, 1992; Dyck, 1993; Hall, 1997). It is worth noting here that although dialogue journals were used in these studies as an instructional technique, the procedure of this technique actually involved an assessment stage at which teachers responded to the content of students' entries.

Regarding the effect of computer-mediated journals on students' writing performance, the writer found that three studies were conducted in this area in the last ten years. In one of them, Ghaleb (1994) found that the quantity of writing in the networked class far exceeded that of the traditional class, and that the percentage of errors in the networked class dropped more than that of the traditional class. Based on these results, Ghaleb concluded that "computer-mediated communication . . . can provide a positive writing

environment for ESL students, and as such could be an alternative to the laborious and time-engulfing method of the traditional approach to teaching writing" (p. 2865). In the second study, MacArthur (1998) found that writing dialogue journals using the word processor had a strong positive effect on the writing of students with learning disabilities. In the third study, Gonzalez-Bueno and Perez (2000) found that electronic dialogue journals had a positive effect on the amount of language generated by learners of Spanish as a second language, and on their attitudes towards learning Spanish, but did not have a significant effect on lexical or grammatical accuracy.

2.2.5 Story Retelling

Story retelling is a highly popularized format of performance assessment (Kaiser, 1997; Pederson, 1995). It is an effective way to integrate oral and written language skills for both learning and assessment (May, 1994). Students who have just read or listened to a story can be asked to retell this story orally or in writing (Callison, 1998; Pierce and O'Malley 1992). This format can be also used for assessing students' reading comprehension. As Kaiser (1997) puts it, "Story retelling can play an important role in performance-based assessment of reading comprehension" (p. 2).

The advantages of this format as an instructional and assessment technique include allowing students to share the cultural heritage of other people; enriching students' awareness of intonation and non-verbal communication; relieving students from the classroom routine; establishing a relaxed, happy relationship between the storyteller and listeners; allowing the teacher to assess students in a relatively non-threatening atmosphere; and allowing the students to assess one another (Grainger, 1995; Hines, 1995; Kaiser, 1997; Malkina, 1995; Stockdale, 1995).

Wilhelm and Wilhelm (1999) suggest that when choosing tales for retelling, language difficulty, content appropriateness, and instructional objectives should be considered. They also suggest that after story retelling, the teacher should encourage students to evaluate their own retellings (*ibid.*).

Kaiser (1997) suggests that students need to be aware of the structural elements of a story before asking them to retell stories. She further suggests that this can be achieved through instruction and practice in story structure using a story map. However, Pederson (1995) suggests that storytelling lies within the storyteller and that storytellers must go beyond the rules and develop their own unique styles.

The story retelling techniques include oral or written presentations, role playing, and pantomiming (Biegler, 1998). Students can retell the story in whatever way they prefer (ibid.).

During retelling, Grainger (1995) suggests that the student should maintain eye contact, use gestures that come naturally, vary his/her voice, and give different tones to different characters. She further suggests that teachers may divide the class into small groups so that more students can retell stories at one time. In such a case, audio and video recording can be used to help in assessing students' performance. Antonacci (1993) suggests that the teacher can help the student during retelling by clearing up misconceptions. To force students to listen attentively to the stories which their classmates tell, Gibson (1999) suggests that students should know in advance that one of them will tell the story again.

After retelling, Pederson (1995) suggests using the following activities for assessing students' performance:

- (a) analyzing and comparing characters,
- (b) discussing topics taken from the story theme,
- (c) summarizing or paraphrasing the story,
- (d) writing an extension of the story,
- (e) dramatizing the story, and
- (f) drawing pictures of the characters.

Tompkins and Hoskisson (1995) suggest that "teachers can assess both the process students use to retell the story and the quality of the products they produce" (p. 131). They further suggest that assessing "the process of developing interpretations is far more important than the quality of the product" (loc. cit.).

Research on Story Retelling

To date there has been no research on story retelling as an on-going assessment format. However, indirect support for the use of this format comes from several studies which used storytelling as an instructional technique. The results of these studies revealed that this technique improved (a) reading comprehension (e.g., Biegler, 1998; Trostle and Hicks, 1998), (b) narrative writing (e.g., Gerbracht, 1994), (c) oral skills (e.g., Cary, 1998), and (d) self-esteem (e.g., Carroll, 1999; Lie, 1994). Indirect support for the use of this format also comes from Brenner's study (1997). In this study, she (Brenner) analyzed the elements of story structure used in written and oral retellings. Results indicated that written and oral retellings were of significant value in assessing students' comprehension. Based on these results, she concluded that "monitoring students' use of story structure elements provides a holistic method for the assessment of comprehension" (p. 4599).

2.2.6 Oral Reading

Listening to students reading aloud from an appropriate text can provide teachers with information on how students handle the cueing systems of language (semantic, syntactic, and phonemic) and on how they process information in their heads and in the text to construct meaning (Farrell, 1993; Manning and Manning, 1996). During oral reading the teacher can code students' miscues (Wallace, 1992). Through an analysis of such miscues, the teacher becomes aware of each student's reading strategies as well as his/her reading difficulties (May, 1994; Pierce and O'Malley, 1992; Pike and Salend, 1995). Miscues are often analyzed in terms of their syntactic and semantic acceptability. The following four questions are often asked in this procedure (Watson and Henson, 1991):

- (a) Is the sentence, as finally read by the student, syntactically acceptable within the context?
- (b) Is the sentence, as finally read by the student, semantically acceptable within the entire context?
- (c) Does the sentence, as finally read by the student, change the meaning of the text? (This question is coded only if questions 1 and 2 are coded yes.)
- (d) How much does the miscue look like the text item?

Once the miscue analysis is completed, the teacher should use the individual conferences to inform each student of his/her strengths and weaknesses and to suggest possible remedies for problems (May, 1994).

During oral reading, the teacher can also observe a student's performance by using anecdotal records (Rhodes and Nathenson-Mejia, 1992). The open-ended nature of these records allows the teacher to describe students' performance, to integrate present observations with other available information, and to identify instructional approaches that may be appropriate (ibid.).

Since there is insufficient time for each student in a large class to present his/her oral reading to the teacher, students can record their oral readings and submit the tapes to the teacher to analyze and evaluate them at leisure (Tannenbaum, 1996).

Research on Oral Reading

A survey of research on oral reading revealed that three studies were conducted in this area in the last ten years. One of them (Kitao and Kitao, 1996) used oral reading as a research tool for testing EFL students' speaking performance. The second study addressed oral reading as an assessment and instructional technique. In this study, Adamson (1998) found that the use of oral reading as an on-going assessment and

instructional technique provided the classroom teacher and parents with critical information about students' literacy development. The third study addressed oral reading as an instructional technique. In this study, Atyah (2000) found that oral reading improved EFL students oral performance.

2.2.7 Group Discussions

Group discussions are a powerful format with which teachers can collect information on students' oral and literacy performance (Graves, 2000; Butler and Stevens, 1997). This format engages students in discussing what they have just read, listened to, or written, or any topics of interest to them. The advantages of this format as an instructional and assessment technique include encouraging students to express their own opinions, allowing students to hear different points of view, increasing students' involvement in the learning process, developing students' critical thinking skills, allowing the teacher to assess students' performance in a relatively non-threatening atmosphere, and raising students' motivation level (Greenleaf et al., 1997; Kahler, 1993; McNeill and Payne, 1996). However, the teacher may not have the time to observe all discussion groups in large classes (Auerbach, 1994). To overcome this difficulty, Kahler (1993) suggests that the teacher should videotape discussion sessions for later analysis and evaluation.

May (1994) notes that the organization of groups and the choice of discussion topics play important roles in promoting successful assessment with group discussions. He further notes that students should be grouped in a way to have something to offer each other, and that the discussion topics have to be of a problematic nature and relevant to the needs and interests of the students (ibid.).

Kahler (1993) suggests that the main role of the teacher during group discussions is to act as language consultant to resolve communicative blocks, and to make notes of students' strengths and weaknesses. The teacher can also use observational checklists for recording data about students' performance (Secord et al., 1994).

To promote group discussions, Zoya and Morse (2002) suggest that the teacher should:

- (a) choose an interesting topic,
- (b) give students some materials on the topic and time limits to read and discuss,
- (c) praise every student for sharing any ideas,
- (d) let the students organize groups according to their friendship, and
- (e) invite specialists to participate in group discussions.

Nowadays, audio and video conferencing programs, such as CUSeeMe and MS NetMeeting are available options for engaging students in voice conversation. Through such computer programs, students can talk directly to their key pals in any place of the world. They can also see and be seen by the key pals they are addressing. Meanwhile, teachers can observe their discussions and progress, and make comments to individual students (Higgins, 1993; Kamhi-Stein, 2000; LeLoup and Ponterio, 2000; Muller-Hartmann, 2000; Sussex and White, 1996).

Research on Group Discussions

A survey of recent research on group discussions revealed that no studies used this format as an on-going assessment tool. However, indirect support for the use of this format comes from four studies that used group discussions as an instructional technique. These studies indicated that group discussions served to develop students' literacy (Troyer, 1992); enriched students' literacy understandings (Allen, 1994); encouraged students to introduce themes relevant to their age, interests, and personal circumstances and to develop these themes according to their own frame of reference (McCormack, 1995); and improved students' overall performance both at the individual and group levels (Mintz, 2001).

2.2.8 Role Playing

Role playing can be used not only as an activity to help students improve their language performance, but also as an assessment format to help the teacher assess students' language performance (Davies et al., 1999; Tannenbaum, 1996).

The advantages of role playing as an instructional and assessment technique include developing students' verbal and non-verbal communication skills, increasing students' motivation to learn, promoting students' self-confidence, integrating language skills, developing students' social skills, allowing the students to know and assess one another, and allowing the teacher to know and assess students in a relatively non-threatening setting (Haozhang, 1997; Krish, 2001; Maxwell, 1997; Tompkins, 1998).

Before role playing, the students are given fictitious names to encourage them to act out the roles assigned to them (Tompkins, 1998). Additionally, McNamara (1996) suggests that each student should be given a card on which there are a few sentences describing what kind of a person he or she is. However, Kaplan (1997) argues against role plays that focus solely on role cards as they do not capture the spontaneous, real-life flow of conversation.

During role playing, the assessor, usually the teacher, can take a minor role in order to be able to control the role play (Tompkins and Hoskisson, 1995). He/she can also support or guide students to perform their roles (ibid.). This “scaffolded assessment,” as Barnes (1999) notes, has a “learning potential” (p. 255). However, Weir (1993) claims that role playing will be more successful if it is done with small groups with the assessor as observer. He further states that if the assessor “is not involved in the interaction he has more time to consider such factors as pronunciation and intonation” (p. 62).

At the end of role playing, Krish (2001) suggests that the teacher should get feedback from the role players on their participation during the preparation stage and the presentation stage.

Since there is insufficient time for each group in a large class to present their role plays to the whole class, Haozhang (1997) suggests that each group should record their role play and submit the tape signed with their names to the teacher for assessment.

To make role playing more effective as an instructional and assessment technique, Burns and Gentry (1998) suggest that teachers should choose role plays that match the language level of the students.

To capture students’ interest, Al-Sadat and Afifi (1997) suggest that role-playing “must be varied in content, style, and technique” (p. 45). They add that role plays may be “comic, sarcastic, persuasive, or narrative” (loc. cit.).

Research on Role Playing

A survey of recent research on role playing indicated that only one study (Kormos, 1999) used this format as a research tool for assessing students’ speaking performance. Furthermore, two other studies investigated students’ perceptions of role playing as an instructional and assessment technique. In one of these studies, Kaplan (1997) found that students learning French as a foreign language felt that role playing boosted their confidence in speaking French. In the other study, Krish (2001) found that EFL students felt that role playing improved their English and developed their confidence to take part in this activity in the future.

2.2.9 Teacher-Student Conferences

Teacher-student conferences are another format for assessing students’ language performance (Ediger, 1999; Newkirk, 1995). Teachers often hold such conferences to talk with students about their work, to help them solve a problem related to what they are learning, and to assess their language performance (Fisher, 1995).

The advantages of teacher-student conferences as an instructional and assessment technique include allowing teachers to determine students' strengths and weaknesses in language performance; providing an avenue for students to talk about their real problems with language; allowing the teacher to discover students' needs, interests, and attitudes; and integrating language skills (McIver and Wolf, 1998; Patthey and Ferris, 1997; Sythes, 1999).

Teacher-student conferences may be formal or informal (Fisher, 1995). They may be also held with individuals or groups (May, 1994). Individual conferences are superior to group conferences in allowing the teacher to diagnose the strengths and weaknesses of each student (*ibid.*). However, it may be difficult to hold individual conferences in large classes (Tarone and Yule, 1996).

During conferring with the student, Fisher (1995) suggests that the teacher should fill in a conference form. In such a form he/she should record the date, the conference topic, the students' strengths and weaknesses, and what the student will do to overcome his/her learning difficulties (*ibid.*). Furthermore, Tompkins and Hoskisson (1995) suggest that during the conference, the teacher's role should be just a listener or guide as this role allows him/her to know a great deal about students and their learning. However, Hansen (1992, p. 100; cited in May, 1994, p. 397) suggests that the teacher should ask the following questions during the conference:

- (a) What have you learned recently in writing?
- (b) What would you like to learn next to become a better writer?
- (c) How do you intend to do that?
- (d) What have you learned recently in reading?
- (e) What would you like to learn next to become a better reader?
- (f) How do you intend to do that?

After the conference, the teacher should keep the conference form--filled during the conference--in a folder along with other evaluation forms to help him/her keep track of the student's progress in language performance (Fisher, 1995).

With access to modern technology, some educators (e.g., Freitas and Ramos, 1998; Marsh, 1997) suggest that teacher-student conferences can be mediated through the computer.

Research on Teacher-Student Conferences

A survey of recent research on teacher-student conferences revealed that whereas several studies analyzed students' and teachers' behaviors during

such conferences (e.g., Boreen, 1995; Boudreaux, 1998; Forsyth, 1996; Gill, 2000; Keebleer, 1995; Nickel, 1997), no studies investigated the effect of this format as an on-going assessment technique on students' language performance.

2.2.10 Verbal Reports

Verbal reports refer to learners' descriptions of what they do while performing a language task or immediately after completing it. Such descriptions develop students' metacognitive awareness and make teachers aware of their students' learning processes (Anderson, 1999; Matsumoto, 1993). Such an awareness can help students make conscious decisions about what they can do to improve their learning (Benson, 2001; Ericsson and Simon, 1993). It can also help teachers assist students who need improvement in their learning processes (Chamot and Rubin, 1994; May, 1994). However, students may change their actual learning processes when teachers ask them to report on these processes (O'Malley and Chamot, 1995).

Verbal reports may be introspective or retrospective. Introspective reports are collected as the student is engaged in the task. This type of reports has been criticized for interfering with the processes of task performance (Gass and Mackey, 2000). Retrospective reports are collected after the student completes the task. This type of reports has been criticized because students may forget or inaccurately recall the mental processes they employed while doing the task (Smagorinsky, 1995).

To help students produce useful and accurate verbal reports, Anderson and Vandergrift (1996) suggest that the teacher should:

- (a) provide training for students in reporting their learning processes,
- (b) elicit verbal reports as close to the students' completion of the task as possible, or even better, during the language task,
- (c) provide students with some contextual information to help them remember the strategies used during doing the task if the report is retrospective,
- (d) videotape students while doing the task, and
- (e) allow students to use either L1 or L2 to produce their verbal reports.

There are different opinions with respect to the validity and reliability of verbal reports. However, many assessment specialists (Alderson, 2000; Storey, 1997; Wu, 1998) agree that verbal reports can be valuable sources of information about students' cognitive processes when they are elicited with care and interpreted with full understanding of the conditions under which they were obtained.

Research on Verbal Reports

A survey of research on introspective and retrospective verbal reports indicated that several studies used this format as a research tool for investigating the processes test-takers employ in responding to test tasks (e.g., Gibson, 1997; Storey, 1997; Wijgh, 1996; Wu, 1998), and for exploring students' learning processes (e.g., El-Mortaji, 2001; Feng, 2001; Kasper, 1997; Lynch, 1997; Robbins, 1996; Sens, 1993).

In addition to the above studies, two other studies were conducted in this area. In one of them, Allan (1995) investigated whether students can effectively report their thoughts. Results indicated that many students were not highly verbal and found it difficult to report their thought processes. In the other study, Anderson and Vandergrift (1996) investigated the effect of verbal reports as an on-going assessment tool on students' awareness of their reading processes and their reading performance. Results indicated that students' use of verbal reports as a classroom activity helped them become more aware of their reading strategies and improved their reading performance.

Additional formats/instruments for evaluating students' performance include essay writing, dramatization, demonstrations, experiments, etc.

2.3 Criteria for Selecting Performance Assessment Formats

In selecting from the previously-mentioned formats, four general considerations should be kept in mind. First, selection should be guided primarily by its match to the teaching/learning targets as a mismatch between the assessment format and these targets will lower the validity of the results (Nitko, 2001). The second consideration in selecting among performance assessment formats is the area of assessment. Some of the previously-mentioned formats are compatible with reading and writing while others are compatible with listening and speaking; some are suitable for assessing language products while others are suitable for assessing learning processes. The third consideration in selecting among performance assessment formats is that no single format is sufficient to evaluate a student's performance (Shepard, 2000). In other words, multiple assessment formats are necessary to provide a more complete picture of a student's performance. The final consideration in determining the specific assessment format is that performance is best assessed if the selected format is used as a teaching or learning technique rather than as a formal or informal test (Cheng, 2000; McLaughlin and Warran, 1995; O'Malley, 1996).

Chapter Three

Alternative Groupings for Performance Assessment

3.0 Introduction

Many educators and assessment specialists (e.g., Barnes, 1999; Campbell et al., 2000; O'Neil, 1992; Santos, 1997) claim that students themselves need to be involved in the process of assessing their own performance. This can be done through self-assessment, peer-assessment, and collaborative group assessment. Each of these alternatives is discussed below.

3.1 Self-Assessment

Self-assessment has been offered as one of the alternatives to teacher assessment. Kramp and Humphreys (1995) define this alternative as “a complex, multidimensional activity in which students observe and judge their own performances in ways that influence and inform learning and performance” (p. 10). Many educators claim that this type of assessment has several advantages. The first of these advantages is that it promotes students' autonomy (Ekbatani, 2000; Graham, 1997; Williams and Burden, 1997; Yancey, 1998). The second advantage is that the involvement of students in assessing their own learning improves their metacognition which can, in turn, lead to better thinking and better learning (Andrade, 1999; O'Malley and Pierce, 1996; Steadman and Svinicki, 1998). The third advantage of this type of assessment is that it enhances students' motivation which can, in turn, increase their involvement in learning and thinking

(Angelo, 1995; Coombe and Kinney, 1999; Todd, 2002). The fourth advantage of this type of assessment is that it fosters students' self-esteem and self-confidence, which can, in turn, encourage them to see the gaps in their own performance and to quickly begin filling these gaps (Smolen et al., 1995; Statman, 1993; Wood, 1993). The fifth and final advantage of self-assessment is that it alleviates the teacher's assessment burden (Cram, 1995).

However, opponents of self-assessment claim that this type of assessment is an unreliable measure of learning and thinking. They further claim that the unreliability of this type of assessment is due to two main reasons. The first reason is that students may under- or over-estimate their own performance (McNamara and Deane, 1995). The second reason is that students can cheat when they assess their own performance (Gardner and Miller, 1999). Another disadvantage of self-assessment is that a few students may engage in it (Cram, 1993).

Gipps (1994) suggests that learners need sustained training in ways of self-assessment to become competent assessors of their own performance. In support of this suggestion, Marteski (1998) found that instruction in self-rating criteria had a positive effect on students' ability to assess their writing.

Barnes (1999) makes the point that questions can encourage learners to evaluate their own performance in a more structured way. She adds that these questions should be generic such as "How are you doing and what do you need to do to improve?" Answers to such questions can help the learner decide what is exactly needed (ibid.). She (Barnes, 1996) also suggests that self-assessment can be aided through the use of a logbook or a course guide.

Arter and Spandle (1992) suggest asking students the following questions to encourage them to engage in self-assessment:

- (a) What is the process you went through to complete this assignment?
- (b) Where did you get ideas?
- (c) What are the problems you encountered?
- (d) What revision strategies did you use?
- (e) How does this activity relate to what you have learned before?
- (f) What are the strengths of your work? and
- (g) What still makes you uneasy?

Anderson (2001) suggests that teachers can help students evaluate their strategy use by asking them to respond thoughtfully to the following questions:

- (a) What are you trying to accomplish?
- (b) What strategies are you using?
- (c) How well are you using them? and

(d) What else could you do?

Furthermore, a number of instruments have been developed for encouraging students to engage in assessing their own learning processes and products. These instruments include K-W-L charts, learning logs, and self-assessment checklists. Each of these instruments is briefly described below.

(1) K-W-L Charts

The K-W-L chart (what I “Know”/what I “Want” to know/what I’ve “Learned”) is one form of self-assessment instruments (O’Malley and Chamot, 1995). The use of this chart improves students’ learning strategies, keeps learners focused and interested during learning, and gives them a sense of accomplishment when they fill in the L column after learning (Shepard, 2000).

Tannenbaum (1996) suggests that this chart can be used as a class activity or on an individual basis before and/or after learning, and that this chart can be completed in the first language for students with limited English proficiency.

Research on K-W-L Charts

A survey of recent research on the K-W-L chart revealed that only one study was conducted in this area. In this study, Burns (1995) found that this chart had a significant effect on fifth-grade students’ reading comprehension.

(2) Learning Logs

Learning logs are a self-assessment tool which students keep about what they are learning, where they feel they are making progress, and what they plan to do to continue making progress (Carlisle, 2000; Lee, 1997; Pike and Salend, 1995; Yung, 1995). At regular intervals, the students reflect on and analyze what they have written in their logs to diagnose their own strengths and weaknesses and to suggest possible remedies for problems (Castillo and Hillman, 1997; Cobine, 1995b). Additional advantages of this format as a learning and assessment technique are (Angelo and Cross, 1993; Commander and Smith, 1996; Conrad, 1995; Kerka, 1996):

- (a) encouraging students to become self reflective,**
- (b) promoting autonomous learning,**
- (c) fostering students’ self-confidence, and**
- (d) providing the teacher with assessable data on students’ metacognitive skills, and with valuable suggestions for improving students’ performance.**

However, learning logs require time and effort from students and teachers (Angelo and Cross, 1993). Moreover, unless a continuing attempt is made to focus on strengths, this format can leave students demoralized from paying too much attention to their weaknesses and failures (ibid.).

McNamara and Deane (1995) propose many activities that students might describe in their logs. These activities include listening to the radio, watching TV, speaking and writing to others, and reading newspapers. They further state that for each experience, students should record the date, the activity, the amount of time engaged in the use of English, the ease or difficulty of the activity, and the reasons for the ease or difficulty of this activity (ibid.). Cranton (1994) suggests that the learner can use one side of a page for the description of his/her activities and the other for thoughts and feelings stimulated by this description.

Since students may find it difficult to know what to write in their logs, Walden (1995) suggests that teachers should give them specific guiding questions such as “What did you learn today and how will you apply that learning?”

Paterson (1995) suggests that learning logs should be shared with the teacher. In such a case, the teacher should not grade them for writing style, grammar, or content, but they can be considered as part of the overall assessment (ibid.).

Perham (1992) and Perl (1994) agree that learning logs can be shared with other students in the class. They further suggest using a loose-leaf notebook--accessible to the whole class--in which learners can reflect on what they learn and read other students' reflections.

Research on Learning Logs

A review of recent research on learning logs revealed that studies conducted in this area were varied as briefly shown below.

Holt (1994) found that six of the ten students who kept learning logs did not find this format helpful. In light of this result, Holt concluded that either the guiding questions those students were given did not motivate reflection or they did not know how to write reflectively.

Matsumoto (1996) found that learning logs improved students' reflection.

Demolli (1997) found that learning logs along with group discussions increased students' abilities to use critical thinking skills.

Saunders et al. (1999) investigated the effects of literature logs, instructional conversations, and literature logs plus instructional conversations on ESL students' story comprehension. Results indicated that students in the literature logs group and literature logs plus instructional conversations group scored significantly higher than those in the conversations group on story comprehension.

Vann (1999) investigated the effects of students' daily learning logs as a means of assessing the progress of advanced ESL students at the college level. Results indicated that this technique enhanced both teaching and learning.

Halbach (2000) found that learning logs revealed many differences between successful and less successful students with respect to their learning strategies.

(3) Self-Assessment Checklists

A checklist consists of a list of specific behaviors and a place for checking whether each is present or absent (Tenbrink, 1999). Through the use of checklists students can evaluate their own learning processes or products (Angelo and Cross, 1993; Burt and Keenan, 1995; Harris et al., 1996). Such checklists can be developed by the teacher or the students themselves through classroom discussions (Meisles, 1993). Moreover, many examples of checklists are nowadays available for students to use for self-assessing their own learning processes (e.g., Oxford's SAS, 1993) and products (e.g., Robbins' Effective Communication Self-Evaluation, 1992). These checklists help students diagnose their own strengths and weaknesses, and help teachers adapt their teaching strategies to suit students' levels and learning style preferences (Tenbrink, 1999). However, such checklists often focus on bits and pieces of students' performance (ibid.). Furthermore, the preparation of checklists is rather time-consuming (Angelo and Cross, 1993).

Research on Self-Assessment Checklists

A survey of recent research on self-assessment checklists revealed that only one study was conducted in this area. In this study, Allan (1995) found that ready-made checklists risked skewing students' responses to those the checklist writer had thought of.

In addition to the previously mentioned instruments, other performance assessment formats (e.g., portfolios, dialogue journals) provide opportunities for self-assessment.

Additional Research on Self-Assessment

In addition to the empirical studies conducted in the areas of learning logs and self-assessment checklists, many other studies were conducted on self-assessment in the last ten years. These studies fall into two broad categories: (1) investigating students' ability to assess themselves and/or the factors that affect this ability, and (2) investigating the effects of self-assessment on student motivation and language performance. The first category includes six studies and two review articles. In one of the six studies, Moritz (1995) found that self-assessment was influenced by many factors such as language learning background, experience, and self-esteem. She (ibid.) concluded that "it seems unreasonable to employ self-assessment as a measurement tool in any situation which entails a comparison of students' abilities. It may, however, be a useful formative learning device, that is, one used throughout the course of a learning program, for feedback to both learners and teachers about the learners' progress, strengths, and weaknesses" (p. 2592). In the second study, Thomson (1995) found that learners were capable of carrying out self-assessment, but noted some variations in the levels of their self-ratings according to gender and ethnic background. In the third study, Graham (1997) found that effective language learners seemed willing and able to assess their own progress. In the fourth study, Shameem (1998) found that Indo-Fijians self-reported their oral Fiji Hindi ability at a level higher than their judged level of performance. In the fifth study, Shoemaker (1998) found that fourth-grade students with special education needs provided evidence of their ability to engage in self-assessment of literacy learning when they were asked to do so, but their self-assessments tended to reflect surface elements of reading and writing rather than reflections of strategic thinking. In the sixth study, Kruger and Dunning (1999) found that learners whose skills or knowledge bases were weak in a particular area tended to overestimate their ability in this area.

In one of the two reviews undertaken in this area, Cram (1995) found that the accuracy of self-assessment varied according to several factors, including the type of assessment, language proficiency, academic record and degree of training; and that students' willingness and ability to engage in self-assessment practices increased with training. She (ibid.) recommended that self-assessment can work best in a supportive environment in which "teachers would place high value on independent thought and action; [and] learners' opinions would be accepted non-judgmentally" (p. 295). In the other review, Oscarson (1997) concluded that learners are capable of assessing their own language proficiency under appropriate conditions.

The second category includes only two studies. In one of these studies, Smolen et al. (1995) found that self-assessment developed students' self-awareness and self-confidence and improved the quantity of their writing. In the other study, Diaz (1999) investigated the effects of self-assessment on student motivation and second language proficiency. Results indicated that

self-assessment helped to improve students' motivation as well as their oral and written proficiency in the target language.

3.2 Peer-Assessment

Many performance assessment specialists (e.g., Johnson, 1998; Norris, 1998; Van-Daalen, 1999) advocate the use of peer-assessment as an alternative to teacher assessment. The advantages of this alternative as a learning and assessment technique include (O'Donnell, 1999; King, 1998; Topping and Ehly, 1998):

- (a) helping students learn from each other,
- (b) developing students' sense of responsibility for their fellows' progress,
- (c) reinforcing students' self-esteem and self-confidence, and
- (d) saving the teacher's time.

Hughes and Large (1993) suggest that learners need training in performance assessment before asking them to assess each other. King (1998) further suggests that students should be given assessment forms to use while assessing each other. Furthermore, Mansour and Mansour (1998) propose that at the end of peer assessment the teacher should assess students' assessments.

Anderson and Vandergrift (1996) suggest that peers can be involved in assessing the strategies they employ while doing a language task.

Research on Peer-Assessment

A survey of recent research on peer-assessment revealed that studies conducted in this area focused on students' perceptions of peer-assessment, the effect of peer vs. teacher assessment on students' writing performance, and the effect of peer- vs. self-assessment on students' writing performance.

With respect to students' perceptions of peer assessment, Qiyi (1993) discovered that Chinese EFL students who used peer-assessment found themselves more interested in the writing class than before and thought that peer-assessment helped them make greater gains in writing quality than did the teacher evaluation. Similarly, Huang (1995) investigated university students' perceptions of peer-assessment in an EFL writing class. Results indicated that students had a positive perception of how they and their peers performed in peer-assessment sessions.

With respect to the effect of peer vs. teacher assessment, only one study was conducted in this area in the last ten years. In this study, Richer (1993) investigated the effect of peer directed vs. teacher based assessment on first year college students' writing proficiency. Results showed that there was a significant difference in writing proficiency in favor of the peer-assessment group.

With respect to the effect of peer- vs. self-assessment, only two studies were conducted in this area in the last ten years. In one of these studies, Mooko (1996) investigated the effect of guided peer-assessment vs. guided self-assessment on the quality of ESL students' compositions. Results revealed that guided peer-assessment was superior to guided self-assessment in enabling students to refine the opening (introduction) and closing statements (conclusion) of their compositions, and in assisting in the reduction of micro-level errors. Results also revealed that self-assessment was more effective than guided peer-assessment in improving composition content. In the other study, Al-Hazmi (1998) investigated the effect of peer-assessment vs. self-assessment on the quality of word processed ESL compositions. Results indicated that both peer-assessment and self-assessment improved the quality of EFL students' writing. However, subjects in the peer-assessment group showed slightly more improvement between drafts with respect to mechanics, grammar, vocabulary, organization, and content than those in the self-assessment group. The self-assessment subjects, nonetheless, recorded slightly higher scores in their final drafts for mechanics, language use, vocabulary, organization, content, and length.

3.3 Group Assessment

Group assessment is a further extension of peer-assessment. This type of assessment provides students with a genuine audience whose response is immediate (Barnes, 1999; Berridge and Muzamhindo, 1998). Moreover, through involvement in group assessment students become more critical of their own work (Graham, 1997). Additional advantages of collaborative group assessment are (Stahl, 1994; Webb, 1995):

- (a) developing students' sense of responsibility,
- (b) helping weak students to learn from their colleagues,
- (c) developing students' social skills, and
- (d) reducing the assessment load of the teacher.

However, compared to self- and peer-assessment, group assessment requires more preparation from the teacher to form groups. Additionally, conflict is more likely to arise among group members (Imel, 1992). Therefore, the teacher should move among groups to observe group members while assessing their own performance, and to resolve the conflicts that may arise among them.

Research on Group Assessment

A survey of recent research in the area of group assessment revealed that only one study was conducted in this area in the last ten years. In this study, Lejk, Wyvill, and Farrow (1999) found that that low-ability students performed better when having their work done and assessed in mixed-ability groups and that high-ability students obtained lower grades in heterogeneous groups than in homogeneous groups.

To sum up this chapter, the writer claims that we cannot assume that students at all levels are capable of assessing their own performance in English as a foreign language. Nor can we assume that teachers have the time to continuously assess all students' performance in large classes. Therefore, both teachers and students need to be involved in the process of assessment.

Chapter Four

Performance Assessment via Computers

4.1 Theoretical Background

In response to the widespread use of computers at schools, homes and workshops, many educators (e.g., Alderson, 1999, 2000; Bernhardt, 2000; Darling-Hammond et al., 1995; Gruba and Corbel, 1997) call for the administration of performance tests via computers. Such educators claim that advances in multimedia and web technologies offer the potential for designing and developing performance tests that are more interactional than their paper-and-pencil counterparts. They also claim that the computer lends authenticity to assessment tasks because it is connected to students' lives and to their learning experiences.

4.2 Research on Performance Assessment via Computers

The introduction of computer administered tests raised a concern about the equivalence of performance yielded via computers versus paper-and-pencil tests. As a result of this concern many studies investigated the effect of computer versus paper-and-pencil tests on students' performance. In this respect, Mead and Drasgow (1993) reported on a meta-analysis of 29 studies that computerized tests were slightly harder than paper-and-pencil tests. They concluded that the results of their meta-analysis "provide strong support for the conclusion that there is no medium effect for carefully constructed power tests" (p. 457).

In a more recent review, Sawaki (1999) also found that there was little consensus in research findings regarding whether test takers either performed better or preferred computer-based as opposed to paper-and-pencil tests of reading. However, Russell and Haney (1997) found that writing performance on the computer was substantially better for students accustomed to writing on computers than that written by hand.

Chapter Five

Reliability and Validity of Performance Assessment

5.1 Theoretical Background

As opposed to standardized forms of testing, performance-based assessment does not have clear-cut right or wrong answers. However, advocates of performance assessment claim that there are ways to make performance assessment valid and reliable. The first way is to use performance assessment rubrics (Boyles, 1998; Linn, 1993). Such rubrics, as Elliot (1995) suggests, should be developed jointly by the teacher and students. In support of developing the assessment rubrics in this way, Graves (2000) found that by assessing students' performance with rubrics created jointly by the teacher and students, there was "much less cause for complaint, whining, accusations of unfairness, or claims of ignorance" (p. 229). Furthermore, allowing "students to assist in the creation of rubrics may be a good learning experience for them" (Brualdi, 1998, p. 3). Additional advantages of the development of assessment rubrics with students are:

- (a) allowing students to know how their own performance will be evaluated and what is expected from them, and
- (b) promoting students' awareness of the criteria they should use in self-assessing their own performance.

The assessor can use either holistic or analytic assessment rubrics for the evaluation of students' performance. However, many performance assessment specialists (e.g., Hyslop, 1996; Moss, 1997; Pierce and O'Malley, 1992; Wiig, 2000) strongly advocate the use of the holistic rubrics for the assessment of students' performance. Such assessment specialists contend that these rubrics focus on the communicative nature of the language. As Pierce and O'Malley

(1992) put it, “Scoring criteria should be holistic with a focus on the student’s ability to receive and convey meaning. Holistic scoring procedures evaluate performance as a whole rather than by its separate linguistic or grammatical features” (p. 4). However, the use of such rubrics may result in wide discrepancies among raters (Davies et al., 1999). Therefore, the second way to make performance assessment valid and reliable is to have a student’s performance assessed by two or more raters (McNamara, 1997b; Mehrens, 1992). These raters should agree upon the assessment criteria and obtain similar scores on some performance samples prior to scoring (Ruth, 1998).

The third way to make performance assessment valid and reliable is to use multiple assessment formats for assessing the same learning objective. Shepard (2000) expresses this idea in the following way:

Variety in assessment techniques is a virtue, not just because different learning goals are amenable to assessment by different devices, but because the mode of assessment interacts in complex ways with the very nature of what is being assessed. For example, the ability to retell a story after reading it might be fundamentally a different learning construct than being able to answer comprehension questions about the story: both might be important instructionally. Therefore, even for the same learning objective, there are compelling reasons to assess in more than one way, both to ensure sound measurement and to support development of flexible and robust understandings. (p. 48)

It is worth noting here that some performance assessment specialists (e.g., Bachman and Palmer, 1997; Kunnan, 1999; Moss, 1994, 1996) argue against a reliance on the traditional, fragmented approach to reliability and validity as sole or best means of achieving fairness and equity in evaluating students’ performance. Kunnan (1999), for example, gives primacy to test fairness and argues that “if a test is not fair there is little or no value in it being valid and reliable or even authentic and interactive” (p 10). He further proposes that fairness in language assessment can be achieved through the following:

- (a) equity in constructing the test in terms of culture, academic discipline, gender, etc.,
- (b) equity in treatment in the testing process (e.g., equal testing conditions, equal opportunity to be familiar with testing formats and materials), and
- (c) equity in the social consequences of the test (e.g., access to university, promotion).

Beyond the concern with traditional reliability and validity, Bachman and Palmer (1997) also propose that the most important consideration in designing a performance test is its usefulness. They add that the usefulness of a language test can be defined in terms of the following six qualities:

- (a) consistency of measurement,
- (b) meaningfulness and appropriateness of the interpretations that we make on the basis of the test scores,
- (c) authenticity of the test tasks—that is, the correspondence between the characteristics of the target language use tasks and those of the test tasks,
- (d) interactiveness of the test tasks—that is, the capacity of the test tasks to engage the test taker in performing cognitive and metacognitive aspects of language,
- (e) impact of the test on the society and educational system, and on the individuals within this system, and
- (f) availability of the resources required for the design, development, and use of the test.

Moss (1994, 1996) also argues that the traditional approach to reliability and validity is “inadequate to represent social phenomena” (ibid., 1996, p. 21). She further proposes a unified approach to reliability and validity, the hermeneutic approach, which requires the inclusion of teachers’ voices in the context of assessment and a dialogue among judges about the specific performance being evaluated.

Furthermore, Baker and her colleagues (1993) suggest that, beyond the fragmented approach to reliability and validity, there are five characteristics that performance assessment should exhibit. These characteristics are:

- (a) meaning for students and teachers,
- (b) current standards of language performance,
- (c) demonstration of complex cognition which is applicable to important problem areas,
- (d) explicit criteria for judgment, and
- (e) minimizing the effects of ancillary skills that are irrelevant to the focus of assessment.

5.2 Research on the Validity and Reliability of Language Performance Assessment

Empirical evidence in support of the claims concerning the reliability and validity of language performance assessment has in general been lacking. In contrast, several studies found differences in language performance due to rater characteristics (e.g., background, experience) both in the assessment of speaking (e.g., Brown, 1995; Chalhoub-Deville, 1996; McNamara, 1996) and of writing (e.g., Lukmani, 1996; Schoonen et al., 1997; Weigle, 1998; Wolfe, 1995). Moreover, some studies found that rater differences survived training (Lumley and McNamara, 1995; McNamara and Adams, 1994; Tyndall and Kenyon, 1995). Lumley and McNamara (1995), for example, examined the stability of speaking performance ratings by a group of raters on three occasions over a period of 20 months. Such raters participated in a training session followed by

rating of a series of audiotaped recordings of speaking performance to establish their reliability. The results of the study indicated that rater differences survived this training. Based on these results, the researchers concluded that

One point that emerges consistently and very strongly from all of these analyses is the substantial variation in rater harshness, which training has by no means eliminated, nor even reduced to a level which would permit reporting of raw scores for candidate performance. (p. 69)

In another line of research, some investigators found differences in students' performance across different types of speaking performance tasks (e.g., McNamara and Lumley, 1997; Shohamy, 1994; Upshur and Turner, 1999) and of reading performance tasks (e.g., Riley and Lee, 1996).

The results of the previously-mentioned studies indicate that the reliability and validity of performance assessment remain a major obstacle in the implementation of this type of assessment and that assessment specialists need to exert so much effort to refine the criteria as well as the procedures by which teachers can establish the reliability and validity of this type of assessment.

Chapter Six

Summary and Conclusions

The last ten years have seen a growth of interest in performance assessment. This interest has led to the development of many alternatives which teachers or assessors can use to elicit and assess students' performance. These alternative assessment techniques highlight the assessment of language as communication, and integrate assessment with learning and instruction. However, for the time being, such techniques remain difficult and costly to use for high-stakes assessment (Wrigley, 2001). Furthermore, assessment specialists are still refining the criteria and procedures by which teachers can establish the reliability and validity of these alternatives (Van-Duzer, 2002). Therefore, my own view is that we should utilize both quantitative and qualitative assessment tools in a complementary fashion. In other words, it seems reasonable to employ performance assessment as a formative learning device throughout the course of the curriculum for feedback to both teachers and learners, and quantitative measures at the end of the curriculum for the comparison of students' abilities. This conclusion is supported by Nitko (2001) in the following way:

If your evaluations are based only on one type of assessment format (e.g., if you rely only on performance tasks), you are likely to have an incomplete picture of each student learning. *You increase the validity of your assessment results by using information gathered from multiple assessment formats: short-answer items, objective items, and a variety of long-term and short-term performance tasks.* (p. 244, emphasis in original)

Before the implementation of performance assessment in our context, there is a need for teachers and students to understand performance assessment alternatives and their limitations. There is also a need for the development of performance standards, adopting performance-based instruction, and

supplying schools with all types of resources (e.g., tape and video recorders, computers, references).

We cannot assume that students at all levels are capable of assessing their own performance in English as a foreign language. Nor can we assume that the teacher have the time to continuously assess all students' performance in large classes. Therefore, I agree with assessment specialists who suggest that the teacher should share the responsibility for assessment with his/her students.

Finally, to ensure the success of performance assessment, I strongly agree with the educators (e.g., Brualdi, 1998; Elliott, 1995; Pachler and Field, 1997) who suggest that performance assessment should be an integral part of teaching and learning because this will save the time for both teachers and students.

About the Author

Abdel-Salam A. El-Koumy is a Full Professor of TEFL and Vice-Dean for Postgraduate Studies and Research at the School of Education in Suez, Suez Canal University, Egypt. E-Mail: Koumi_5000@Yahoo.Com.

References

- Adamson, P. (1998). An investigation into the effects of using oral reading for assessment and practice in the grade one level: Resource/classroom teacher collaboration. *DAI-A*, 37(2), 423.
- Alderson, J. (1999). Reading constructs and reading assessments. In M. Chalhoub-Deville (Ed.), *Issues in Computer-Adaptive Testing of Reading Proficiency (49-70)*. Cambridge: University of Cambridge.
- Alderson, J. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.
- Al-Hazmi, S. (1998). The effect of peer-feedback and self-assessment on the quality of word processed ESL compositions. *DAI-A*, 60(1), 16.
- Allan, A. (1995). Begging the questionnaire: Instrument effect on readers' responses to a self-report checklist. *Language Testing*, 12(2), 133-156.
- Allen, S. (1994). Literature group discussions in elementary classrooms: Cases studies of three teachers and their students. *DAI-A*, 55(6), 1494.
- Al-Sadat, A. and Afifi, E. (1997). Role-playing for inhibited students in paternal communities. *English Teaching forum*, 35(3), 43-48.
- Anderson, N. (1999). *Exploring Second Language Reading: Issues and Strategies*. Boston: Heinle and Heinle.
- Anderson, N. (2001). *The Role of Metacognition in Second Language Teaching and Learning (On-Line)*. Available at : eric@cal.org.
- Anderson, N. and Vandergrift, L. (1996). Increasing metacognitive awareness in the L2 classroom by using think-aloud protocols and other verbal report formats. In Rebecca L. Oxford (Ed.), *Language Learning Strategies*

- Around the World: Cross-Cultural Perspectives* (pp. 3-18). Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.
- Andrade, H. (1999). *Student Self-Assessment: At the Intersection of Metacognition and Authentic Assessment*. ERIC Document No. ED 431 030.
- Angelo, T. (1995). Improving classroom assessment to improve learning: Guidelines from research and practice. *Assessment Update*, 7, 1-2, 12-13.
- Angelo, T. and Cross, K. (1993). *Classroom Assessment Techniques: A Handbook for College Teachers*. San Francisco, CA: Jossey-Bass.
- Anselmo, C. (1998). Experiences students encounter with portfolio assessment. *DAI-A*, 58(11), 4237.
- Antonacci, P. (1993). Natural assessment in whole language classroom. In Angela Carrasquillo and Carolyn Hedley (Eds.), *Whole Language and the Bilingual Learner* (pp. 116-131). Norwood, NJ: Ablex Publishing Company.
- Arter, J. and Spandel, V. (1992). NCME instructional module: Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice*, 11(1), 36-44.
- Arter, J., Spandel, V., and Culham, R. (1995). *Portfolios for Assessment and Instruction*. ERIC Digest No. ED 388890.
- Atyah, H. (2000). The impact of reading on oral language skills of Saudi Arabian children with language learning disability. *DAI-A*, 61(3), 819.
- Auerbach, E. (1994). *Making Meaning, Making Change: Participatory Curriculum Development for Adult ESL Literacy*. ERIC Document No. ED 356688.
- Bachman, L. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17(1), 1-42.
- Bachman, L. and Palmer, A. (1997). *Language Testing Practice*. Oxford: Oxford University Press.
- Bailey, K. (1998). *Learning about Language Assessment: Dilemmas, Decisions, and Directions*. Boston: Heinle and Heinle.
- Baker, E. O'Neill, H., Jr., and Linn, R. (1993). Policy and validity prospects for performance-based assessments. *American Psychologist*, 48, 1210-1218.
- Barnes, A. (1996). Getting them off to a good start: The lead-up to that first A level class. In A. Brien (Ed.), *German Teaching 14* (pp. 23-28). Rugby: Association of Language Learning.
- Barnes, A. (1999). Assessment. In Norbert Pachler (Ed.), *Teaching Modern Foreign Languages at Advanced Level* (pp. 251-281). London: Routledge.
- Barret, H. (1994). Technology supported assessment portfolios. *Computing Teacher*, 21(6), 9-12.
- Barton, P. and Coley, R. (1994). *Testing in America's Schools*. Princeton, NJ: Educational Testing Service Policy Information Center.
- Baudrand, A. (1992). Dialogue journal writing in a foreign language classroom: Assessing communicative competence and proficiency. *DAI-A*, 54(2), 444.

- Bello, T. (1997). *Improving ESL Learners' Writing Skills*. ERIC Document No. ED 409746.
- Benson, P. (2001). *Teaching and Researching Autonomy in Language Learning*. Harlow: Pearson Education Limited.
- Bernhardt, E. (2000). If reading is reader-based, can there be a computer-adaptive test of reading. In M. Chalhoub-Deville (Ed.), *Issues in Computer-Adaptive Testing of Reading Proficiency (1-10)*. Cambridge: University of Cambridge.
- Berridge, R. and Muzamhindo, J. (1998). Group oral tests. In J. D. Brown (Ed.), *New Ways of Classroom Assessment* (pp. 137-139). Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc.
- Berwick, R. and Ross, S. (1996). Cross-cultural pragmatics in oral proficiency interview strategies. In M. Milanovic, and N. Saville (Eds.), *Performance Testing, Cognition and Assessment (43-54)*. Cambridge: University of Cambridge.
- Biegler, L. (1998). *Implementing Dramatization as an Effective Storytelling Method To Increase Comprehension*. ERIC Document No. ED 417377.
- Boerum, L. (2000). Developing portfolios with learning disabled students. *Reading and Writing Quarterly: Overcoming Learning Difficulties*, 16(3), 211-238.
- Boreen, J. (1995). The language of reading conferences. *DAI-A*, 56(10), 3860.
- Borich, J. (2001). Learning through dialogue journal writing: A cultural thematic unit. *Learning Languages*, 6 (3), 4-19.
- Boudreaux, M. (1998). Toward awareness: A study of nonverbal behavior in the writing conference. *DAI-A*, 59(4), 1145.
- Boyles, P. (1998). Scoring rubrics: Changing the way we grade. *Learning Languages*, 4(1), 29-32.
- Brenner, S. (1997). Storytelling and story schema: A bridge to listening comprehension through retelling. *DAI-A*, 58(12), 4599.
- Bromley, K. (1993). *Journaling: Engagements in Reading, Writing, and Thinking*. New York: Scholastic, Inc.
- Brooks, H. (1995). Uphill all the way: Case study of a non-vocal child's acquisition of literacy. *DAI-A*, 56(6), 2196.
- Brown, A. (1995). The effect of rater variables in the development of an occupation-specific language performance test. *Language Testing*, 12(1), 1-15.
- Brown, J. (Ed.). (1998). *New Ways of Classroom Assessment*. Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc.
- Brualdi, A. (1998). *Implementing Performance Assessment in the Classroom*. ERIC Digest No. ED 423312.
- Burniske, R. (1994). Creating dialogue: Teacher response to journal writing. *English Journal*, 83(4), 94-87.
- Burns, A. and Gentry, J. (1998). Motivating students to engage in experiential learning: A tension-to-learn theory. *Simulation and Gaming*, 29, 133-151.

- Burt, M. and Keenan, F. (1995). *Adult ESL Assessment: Purposes and Tools*. ERIC Digest No. ED 386962.
- Burns, P. (1995). The effect of the K-W-L strategy on fifth graders' reading comprehension and reading attitude. *DAI-A*, 55(1), 149.
- Butler, F. and Stevens, R. (1997). Oral language assessment in the classroom. *Theory into Practice*, 36(4), 214-219.
- Callison, D. (1998). Authentic assessment. *School Library Media Activities*, 14(5), 42-43, 50.
- Camp, R. (1993). Changing the model for the direct assessment of writing. In M. Williamson and B. Huot (Eds.), *Validating Holistic Scoring for Writing Assessment*. Creskill, NJ: Hampton Press.
- Campbell, D., Melenzer, B., Nettles, D., and Wyman, R. Jr. (2000). *Portfolio and Performance Assessment in Teacher Education*. Boston: Allyn and Bacon.
- Careen, K. (1997). A study of the effect of cooperative learning activities on the oral comprehension and oral proficiency of grade 6 core French students. *DAI-A*, 36(4), 889.
- Carlisle, A. (2000). Reading logs: An application of reader-response theory in ELT. *ELT Journal*, 54(1), 12-19.
- Carroll, S. (1999). *Storytelling for Literacy*. ERIC Document No. ED 430234.
- Cary, S. (1998). The effectiveness of a contextualized storytelling approach for second language acquisition. *DAI-A*, 59(3), 0758.
- Castillo, R. and Hillman, G. (1997). Ten ideas for creative writing in the EFL classroom. *English Teaching Forum*, 33(4), 30-31.
- Chalhoub-Deville, M. (1995). Deriving oral assessment scales across different tests and rater groups. *Language Testing*, 12(1), 16-33.
- Chalhoub-Deville, M. (1996). Performance assessment and the components of the oral construct across different tests and rater groups. In M. Milanovic, and N. Saville (Eds.), *Performance Testing, Cognition and Assessment* (55-73). Cambridge: University of Cambridge.
- Chamot, A. and Rubin, J. (1994). Comments on Janie Rees-Miller's "A critical appraisal of learner training: Theoretical bases and teaching implications": Two readers react. *TESOL Quarterly*, 28(4), 771-776.
- Chang, C. (2001). Construction and evaluation of a web-based learning portfolio system: An electronic assessment tool. *Innovation in Education and Teaching International*, 38(2), 144-155.
- Checkley, K. (1997). The first seven . . . and the eighth: A conversation with Howard Gardner. *Educational Leadership*, 55(1), 8-13.
- Cheng, L. (2000). *Washback or Backwash: A Review of the Impact of Testing on Teaching and Learning*. ERIC Document No. ED 442280.
- Cobine, G. (1995a). *Effective Use of Student Journal Writing*. ERIC Digest No. ED 378587.
- Cobine, G. (1995b). *Writing as a Response to Reading*. ERIC Digest No. ED 386734.

- Cockrum, W. and Castillo, M. (1991). Whole language assessment and evaluation strategies. In Bill Harp (Ed.), *Assessment and Evaluation in Whole Language Programs* (pp. 73-86). Norwood, MA: Christopher-Gordon Publishers, Inc.
- Cohen, A. (1994). *Assessing Language Ability in the Classroom*. Boston, MA: Heinle and Heinle.
- Commander, N. and Smith, B. (1996). Learning logs: A tool for cognitive monitoring. *Journal of Adolescent and Adult Literacy*, 39(6), 446-453.
- Condon, W. and Hamp-Lyons, L. (1993). Questioning the assumptions about portfolio-based assessment. *College Composition and Communication*, 44, 167-190.
- Conrad, L. (1995). *Assessing Thoughtfulness: A New Paradigm*. ERIC Document No. ED 409352.
- Cook, S. (1993). Dialogue journals in the junior high school. *DAI-A*, 32(5), 1258.
- Coombe, C. and Kinney, J. (1999). Learner-centered listening assessment. *English Teaching Forum*, 37, 21-24.
- Courts, P. and McInerney, K. (1993). *Assessment in Higher Education: Politics, Pedagogy and Portfolios*. New York: Praeger.
- Cram, B. (1993). Training learners for self-Assessment. *TESOL in Context*, 2, 30-33.
- Cram, B. (1995). Self-assessment: From theory to practice. In G. Brindley (Ed.), *Language Assessment in Action* (pp. 271-306). Sydney: Macquarie University.
- Cranton, P. (1994). *Understanding and Promoting Transformative Learning*. San Francisco: Jossey-Bass.
- Darling-Hammond, L. Acness, J. and Falk, B. (1995). *Authentic Assessment in Action*. New York: Teachers College Press.
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T. and McNamara, T. (1999). *Dictionary of Language Testing*. Cambridge: University of Cambridge Local Examinations Syndicate.
- Davies, F. (1995). *Introducing Reading*. London: Penguin.
- Demolli, R. (1997). *Improving High School Students' Critical Thinking Skills*. ERIC Document No. ED 420600.
- Diaz, C. (1999). *Authentic Assessment: Strategies for Maximizing Performance in the Dual-Language Classroom*. ERIC Document No. ED 435171.
- Dudley, M. (2001). Portfolio assessment: When bad things happen to good ideas. *English Journal*, 90(6), 19-20.
- Dugan, S. (1994). Test anxiety and test achievement: A cross cultural study. *DAI-A*, 55(11), 3432.
- Dyck, B. (1993). Dialogue journals in content classroom: An interdisciplinary approach to improve language, learning, and thinking. *DAI-A*, 33(3), 704.
- Ediger, M. (1999). Evaluation of reading progress. *Reading Improvement*, 36(2), 50-56.

- Ediger, M. (2000). *Reading, Portfolios, and the Pupil*. ERIC Document No. ED 442078.
- Ekbatani, G. (2000). Moving toward Learner-Directed Assessment. In G. Ekbatani and H. Pierson (Eds.), *Learner-Directed Assessment in ESL* (pp. 1-11). New Jersey: Mahwah.
- Elliot, S. (1995). *Creating Meaningful Performance Assessments*. ERIC Digest No. ED 381985.
- El-Mortaji, L. (2001). Writing ability and strategies in two discourse types: A cognitive study of multilingual Moroccan university students writing in Arabic (L1) and English (L2). *DAI-A*, 62(4), 499.
- Ericsson, K. and Simon, H. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.
- Farr, R. (1994). *Portfolio Conferences: What They Are and Why They Are Important*. Los Angeles: IOX Assessment Associates.
- Farr, R. and Tone, B. (1994). *Portfolio and Performance Assessment: Helping Students Evaluate Their Progress as Readers and Writers*. Orlando, FL: Harcourt Brace and Company.
- Farrell, M. (1993). Three-part oral reading practice for large classes. *English Teaching Forum*, 31(2), 43-45.
- Feng, H. (2001). Writing an academic paper in English: An exploratory study of six Taiwanese graduate students. *DAI-A*, 62(9), 3033.
- Fisher, B. (1995). *Thinking and Learning Together: Curriculum and Community in a Primary Classroom*. Portsmouth, NH; Heinemann.
- Fleming, F. and Walls, G. (1998). What pupils do: The role of Strategic planning in modern foreign language learning. *Language Learning Journal*, 18, 14-21.
- Fleurquin, F. (1998). Seeking authentic changes: New performance-based report cards. *English Teaching Forum*, 36(2), 45-47.
- Fordham, P., Holland, D., and Millican, J. (1995). *Adult Literacy*. Oxford: Oxfam/Voluntary Service Overseas.
- Forsyth, S. (1996). Individual reading conferences: An exploration of opportunities for learning from a socio-cognitive perspective. *DAI-A*, 57(12), 5054.
- Freitas, C. and Ramos, A. (1998). *Using Technologies and Cooperative Work To Improve oral, writing, and Thinking Skills*. ERIC Document No. ED 423835.
- Fried-Booth, D. (1997). *Project Work*. Oxford: Oxford University Press.
- Fuchs, L. (1995). *Connecting Performance Assessment to Instruction: Comparison of Behavioral Assessment, Mastery Learning, Curriculum-Based Measurement, and Performance Assessment*. ERIC Digest No. ED 381984.
- Gallagher, J. (1998). *Classroom Assessment for Teachers*. Columbus, Ohio: Prentice-Hall, Inc.
- Garcia, J. (1998). Don't talk about it, write it down. In J. D. Brown (Ed.), *New Ways of Classroom Assessment* (pp. 36-37). Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc.

- Gardner, D. and Miller, L. (1999). *Establishing Self-Access: From Theory to Practice*. Cambridge: Cambridge University Press.
- Gardner, H. (1993). *Frames of Mind: The Theory of Multiple Intelligences*. New York: Basic Books.
- Gardner, H. (1999). Are there additional intelligences? The case for naturalist, spiritual, and existential intelligences. In J. Hane (Ed.), *Education, Information, and Transformation* (pp. 111-131). Englewood Cliffs, NJ: Prentice Hall.
- Garmon, M. (2001). The benefits of dialogue journal: What prospective teachers say. *Teacher Education Quarterly*, 28(4), 37-50.
- Gass, S. and Mackey, A. (2000). *Stimulated Recall Methodology in Second Language Research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Genesee, F. and Upshur, J. (1996). *Classroom-Based Evaluation in Second Language Education*. New York: Cambridge University Press.
- Gerbracht, G. (1994). The effect of storytelling on the narrative writing of third-grade students. *DAI-A*, 55(12), 7341.
- Ghaleb, M. L. (1994). Computer networking in a university freshman ESL writing class: A descriptive study of the quantity and quality of writing in networking and traditional writing classes. *DAI-A*, 54 (8), 2865.
- Gibson, B. (1997). *Talking the Test: Using Verbal Report Data in Looking at the Processing of Cloze Tests*. ERIC Document No. ED 409713.
- Gibson, B. (1999). A story-telling and re-telling activity. *The Internet TESL Journal*, 5(9), 1-2.
- Gill, S. (2000). Reading with Amy: Teaching and learning through reading conferences. *Reading Teacher*, 53(6), 500-509.
- Gipps, C. (1994). *Beyond Testing: Towards a Theory of Educational Assessment*. London: Falmer Press.
- Gomez, E. (2000). *Assessment Portfolios: Including English Language Learners in Large-Scale Assessments*. ERIC Digest No. ED 447725.
- Gonzalez-Bueno, M., and Perez, L. (2000). Electronic mail in foreign language writing: A study of grammatical and lexical accuracy, and quantity of language. *Foreign Language Annals*, 33(2), 189-198.
- Grace, C. (1992). *The Portfolio and Its Use: Developmentally Appropriate Assessment of Young Children*. ERIC Digest No. ED 351150.
- Graham, S. (1997). *Effective Language Learning*. Clevedon: Multilingual Matters.
- Grainger, T. (1995). *Traditional Storytelling in the Primary Classroom*. Warwickshire: Scholastic Ltd.
- Graves, K. (2000). *Designing Language Courses: A Guide for Teachers*. Boston: Heinle and Heinle.
- Greenleaf, C. Gee, M. and Ballinger, R. (1997). *Authentic Assessment: Getting Started*. ERIC Document No. ED 411474.
- Greenwald, C. and Hand, J. (1997). The project approach in inclusive preschool classrooms. *Dimensions of Early Childhood*, 25(4), 35-39.

- Gruba, P. and Corbel, C. (1997). Computer-based testing. In C. Clapham and D. Corson (Eds.), *Language Testing and Assessment* (pp. 141-149). Dordrecht: Kluwer.
- Gutwirth, V. (1997). A multicultural family project for primary students. *Young Children*, 52(2), 72-78.
- Hackett, L. (1996). The Internet and e-mail: Useful tools for foreign language teaching and learning. *On-Call*, 10(1), 15-20.
- Halbach, A. (2000). Finding out about students' learning strategies by looking at their diaries: A case study. *System*, 28(1), 58-69.
- Hall, S. (1997). Factors in teachers usage of journal writing in first grade classrooms. *DAI-A*, 58(10), 3865.
- Hamp-Lyons, L. (1996). Applying ethical standards to portfolio assessment of writing English as a second language. In M. Milanovic, and N. Saville (Eds.), *Performance Testing, Cognition and Assessment* (323-233). Cambridge: University of Cambridge.
- Hannon, J. (1999). Talking back: Kindergarten dialogue journals. *Reading Teacher*, 53(3), 200-203.
- Haozhang, X. (1997). Tape recorders, role-plays, and turn-taking in large EFL listening and speaking classes. *English Teaching Forum*, 35(3), 33-41.
- Harmon, J. (1996). Constructing word meaning: Independent strategies and learning opportunities of middle school students in a literature-based reading program. *DAI-A*, 57(7), 2943.
- Harris, D., Carr, J., Flynn, T., Petit, M., and Rigney, S. (1996). *How to Use Standards in the Classroom*. Alexandria, Virginia: Association for Supervision and Curriculum Development.
- Hetterscheidt, J. and Others (1992). Using the computer as a reading portfolio. *Educational Leadership*, 49(8), 73-81.
- Hewitt, G. (2001). The writing portfolio: Assessment starts with A. *Clearing House*, 74(4), 187-190.
- Higgins, C. (1993). *Computer-Assisted Language Learning: Current Programs and Projects*. ERIC Digest No. ED 355835.
- Hines, M. (1995). Story theater. *English Teaching Forum*, 33(2), 6-11.
- Hirvela, A. and Pierson, H. (2000). Portfolios: Vehicles for Authentic Self-Assessment. In G. Ekbatani and H. Pierson (Eds.), *Learner-Directed Assessment in ESL* (pp. 105-126). New Jersey: Mahwah.
- Ho, B. (1992). Journal writing as a tool for reflective learning: Why students like it. *English Teaching Forum*, 30(4), 40-42.
- Holmes, J. and Morrison, N. (1995). *Research Findings on the Use of Portfolio Assessment in a Literature Based Reading Program*. ERIC Document No. ED 380794.
- Holt, S. (1994). *Reflective Journal Writing and Its Effects on Teaching Adults*. ERIC Document No. ED 375302.
- Horvath, L. (1997). Portfolio assessment of writing, development and innovation in the secondary setting: A case study. *DAI-A*, 58(8), 2985.

- Huang, S. (1995). *ESL University Students Perceptions of Their Performance in Peer Response Sessions*. ERIC Document No. ED 399766.
- Hughes, I. and Large, B. (1993). Staff and peer-group assessment of oral communication skills. *Studies in Higher Education*, 18(3), 379-385.
- Hunter, B. and Bagley, C. (1995). *Global Telecommunications Projects: Reading and Writing with the World*. ERIC Document No. ED 387081.
- Hutchinson, N. (1995). *Performance Assessment of Career Development*. ERIC Document No. ED 414518.
- Hyslop, N. (1996). *Using Grading Guides to Shape and Evaluate Business Writing*. ERIC Digest No. ED 399562.
- Imel, S. (1992). *Small Groups in Adult Literacy and Basic Education*. ERIC Digest No. ED 350490.
- Johnson, J. (1998). So, how did you like my presentation. In J. D. Brown (Ed.), *New Ways of Classroom Assessment* (pp. 67-69). Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc.
- Jones, J. and Vanteirsburg, P. (1992). *What Teachers Have Been Telling Us about Literacy Portfolios*. Dekalb, IL: Northern Illinois University.
- Kahler, W. (1993). Organizing and implementing group discussions. *English Teaching Forum*, 31(1), 48-50.
- Kaiser, E. (1997). *Story Retelling: Linking Assessment to the Teaching-Learning Cycle*. Available at: http://www.wcer.wisc.edu/ccvi/p.../Story_Retelling_V2_No1.htm.
- Kamhi-Stein, L. (2000). Looking to the future of TESOL teacher education: Web-based bulletin board discussions in a methods course. *TESOL Quarterly*, 34, 423-456.
- Kaplan, M. (1997). Learning to converse in a foreign language: The reception game. *Simulation and Gaming*, 28, 149-163.
- Kasper, L. (1997). Assessing the metacognitive growth of ESL student writers. *TESL. EJ*, 3(1), 1-23.
- Katz, L. (1994). *The Project Approach* (On-Line). Available at: <http://ericece.org/>
- Katz, L. (1997). *A Developmental Approach to Assessment of Young Children*. Available at: <http://ericece.org/>
- Katz, L. and Chard, S. (1998). *Issues in Selecting Topics for Projects* (On-Line). Available at: <http://ericece.org/>
- Keebler, R. (1995). "So, there are some things I don't quite understand...": An analysis of written conferences in a culturally diverse second-grade classroom. *DAI-A*, 56(5), 1644.
- Kerka, S. (1995). *Techniques for Authentic Assessment*. Available at: http://www.ed.gov/databases/ERIC_Digests/
- Kerka, S. (1996). *Journal Writing and Adult Learning*. ERIC Document No. ED 399413.
- Khattari, N. Reeve, A. and Kane, M. (1998). *Principles and Practices of Performance Assessment*. Hillsdale, NJ: Erlbaum.

- Kiany, G. (1998). Extraversion and pedagogical setting as sources of variation in different aspects of English proficiency. *DAI-A*, 59(3), 488.
- King, K. (1998). Teachers and students assessing oral presentations. In J. D. Brown (Ed.), *New Ways of Classroom Assessment* (pp. 70-72). Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc.
- Kitao, S. and Kitao, K. (1996). *Testing Speaking*. ERIC Document No. ED 398261.
- Knight, S. (1994). Making authentic cultural and linguistic connections. *Hispania*, 77, 289-294.
- Koretz, D. (1994). *The Evolution of Portfolio Program: The Impact and Quality of the Vermont Portfolio Program in Its Second Year (1992-1993)*. ERIC Document No. ED 379301.
- Koretz, D., Stecher, G., Klein, S., and McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice*, 13(3), 5-16.
- Kormos, J. (1999). Simulating conversations in oral-proficiency assessment: A conversation analysis of role plays and non-scripted interviews in language exams. *Language Testing*, 16(2), 163-188.
- Kramp, M. and Humphreys, W. (1995). Narrative, self-assessment, and the habit of reflection. *Assessment Update*, 7 (1), 10-13.
- Krish, P. (2001). A role play activity with distance learners in an English language classroom. *The Internet TESL Journal*, 4(8), 1-7. Available at: <http://iteslj.org/>
- Kruger, J. and Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessment. *Journal of Personality and Social Psychology*, 77, 1121-1134.
- Kunnan, A. (1998). Language testing: Fundamentals. In A. Kunnan (Ed.), *Validation in Language Assessment: Selected Papers from the 17th Language Testing Research Colloquium* (pp.707-711). Mahwah, NJ: Lawrence Erlbaum.
- Kunnan, A. (1999). Fairness and justice for all. In A. Kunnan (Ed.), *Fairness and Validation in Language Assessment* (pp. 1-14). Cambridge: Cambridge University Press.
- Lankes, A. (1995). *Electronic Portfolios: A New Idea in Assessment*. ERIC Digest No. ED 390377.
- Lazaraton, A. (1996). Interlocutor support in oral proficiency interviews: The case of CASE. *Language Testing*, 13, 151-172.
- Lee, E. (1997). The learning response log: An assessment tool. *English Journal*, 86(1), 41-44.
- Lejk, M., Wyvill, M., and Farrow, S. (1999). Group assessment in systems analysis and design: A comparison of the performance of streamed and mixed-ability groups. *Assessment and Evaluation in Higher Education*, 24(1), 5-14.

- LeLoup, J. and Ponterio, R. (1995). Networking with foreign language colleagues: Professional development on the Internet. *Northeast Conference Newsletter*, 37, 6-10.
- LeLoup, J. and Ponterio, R. (2000). *Enhancing Authentic Language Experiences through Internet Technology*. ERIC Digest No. ED 442277.
- Lemons, A. (1996). The effects of reading strategies on the comprehension development of the ESL learner. *DAI-A*, 57(7), 2945.
- Lie, A. (1994). *Paired Storytelling: An Integrated Approach for Bilingual and English as a Second Language Students*. ERIC Document No. ED 372601.
- Linn, R. (1993). Educational assessment: Expanded expectations and challenges. *Educational Evaluation and Policy Analysis*, 15, 1-16.
- Linn, R. and Gronlund, N. (1995). *Measurement and Assessment in Teaching*. Englewood Cliffs NJ: Prentice Hall.
- Lukmani, Y. (1996). Linguistic accuracy versus coherence in academic degree programs: A report on stage 1 of the project. In M. Milanovic, and N. Saville (Eds.), *Performance Testing, Cognition and Assessment* (130-150). Cambridge: University of Cambridge.
- Lumley, T. and Brown, A. (1996). Specific-purpose language performance tests: Task and interaction. *Australian Review of Applied Linguistics*, 13, 105-136.
- Lumley, T. and McNamara, T. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing*, 12(1), 54-71.
- Lylis, S. (1993). A descriptive study and analysis of two first grade teachers' development and implementation of writing-portfolio assessments. *DAI-A*, 54(2), 420.
- Lynch, W. (1997). An investigation of writing strategies used by high-ability seventh graders responding to a state-mandated explanatory writing assessment task. *DAI-A*, 58(6), 2056.
- MacArthur, C. (1998). Word processing with speech synthesis and word prediction: Effects on dialogue journal writing of students with learning disabilities. *Learning Disability Quarterly*, 21(2), 151-166.
- Maden, J. and Taylor, M. (2001). *Developing and Implementing Authentic Oral Assessment Instruments*. ERIC Document No. ED 548803.
- Malkina, N. (1995). Storytelling in early language teaching. *English Teaching Forum*, 33(1), 38-39.
- Manning, M. and Manning, G. (1996). Teaching reading and writing, tried and true practices. *Teaching PreK-8*, 26(8), 106-109.
- Mansour, S. and Mansour, W. (1998). Assess the assessors. In J. D. Brown (Ed.), *New Ways of Classroom Assessment* (pp. 78-83). Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc.
- Marsh, D. (1997). Computer conferencing: Taking the loneliness out of independent learning. *Language Learning Journal*, 15, 21-25.
- Marteski, F. (1998). Developing student ability to self-assess. *DAI-A*, 59(3), 714.
- Martinez, R. (1998). *Assessment Development: A Guidebook for Teachers of English Language Learners*. ERIC Document No. ED 419244.

- Matsumoto, K. (1993). Verbal-report data and introspective methods in second language research: State of the art. *RELC Journal: A Journal of Language Teaching and Research in Southeast Asia*, 24(1), 32-60.
- Matsumoto, K. (1996). Helping L2 learners reflect on classroom learning. *ELT Journal*, 50(2), 143-148.
- Mauerman, D. (1995). The effects of a literature intervention on the language development of second grade children. *DAI-A*, 34(1), 39.
- Maxwell, C. (1997). *Role Play and Foreign Language Learning*. ERIC Document No. ED 146688.
- May, F. (1994). *Reading as Communication*. New York: Macmillan Publishing Company.
- Mccormack, R. (1995). Learning to talk about books: Second graders' participation in peer-led literature discussion groups. *DAI-A*, 56(4), 1244.
- McIver, M. and Wolf, S. (1998). *Writing Conferences: Powerful Tools for Writing Instruction*. ERIC Document No. ED 428126.
- McLaughlin, M. and Warran, S. (1995). *Using Performance Assessment in Outcomes-Based Accountability Systems*. ERIC Digest No. ED 381987.
- McNamara, M. and Deane, D. (1995). Self-assessment activities: Toward autonomy in language learning. *TESOL Journal*, 5, 17-21.
- McNamara, T. (1996). *Measuring Second Language Performance*. London: Longman.
- McNamara, T. (1997a). "Interaction" in Second language performance assessment: Whose performance. *Applied Linguistics*, 18(4), 446-466.
- McNamara, T. (1997b). *Measuring Second Language Performance*. London: Longman.
- McNamara, T. and Adams, R. (1994). *Exploring Rater Characteristics with Rasch Techniques*. ERIC Document No. ED 345498.
- McNamara, T. and Lumley, T. (1997). The effect of interlocutor and assessment mode variables in overseas assessments of speaking skills in occupational settings. *Language Testing*, 14(2), 140-156.
- McNeill, J. and Payne, P. (1996). *Cooperative Learning Groups at the College Level: Applicable Learning*. ERIC Document No. ED 404920.
- Mead, A. and Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin*, 114(3), 449-458.
- Mehrens, W. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice*, 11(1), 3-9, 20.
- Meisles, S. (1993). Remaking classroom assessment with the work sampling system. *Young Children*, 48(5), 34-40.
- Meisles, S., Dorfman, A., and Steele, D. (1995). Equity and excellence in group-administered and performance-based assessments. In M. Nettles and A. Nettles (Eds.), *Equity and Excellence in Educational Testing* (pp. 243-261). Boston, Kluwer: Academic Publishers.
- Miller, M. and Legg, S. (1993). Alternative assessment in a high-stakes environment. *Educational Measurement: Issues and Practice*, 12(2), 9-15.

- Mintz, C. (2001). The effects of group discussions session attendance on student test performance in the self-paced, personalized, interactive, networked system of instruction. *MAI*, 40(1), 239.
- Moening, A. and Bhavnagri, N. (1996). Effects of the showcase writing portfolio process on first graders' writing. *Early Education and Development*, 7(2), 179-199.
- Mooko, T. (1996). An investigation into the impact of guided peer feedback and guided self-assessment on the quality of compositions written by secondary school students in Botswana. *DAI-A*, 58(4), 1155.
- Moritz, C. (1995). Self-assessment of foreign language proficiency: A critical analysis of issues and a study of cognitive orientations of French learners. *DAI-A*, 56(7), 2592.
- Moss, B. (1997). A qualitative assessment of first graders' retelling of expository text. *Reading Research and Instruction*, 37(1), 1-13.
- Moss, P. (1994). Can there be validity without reliability? *Educational Researcher*, 23(2), 5-12.
- Moss, P. (1996). Enlarging the dialogue in Educational measurement: Voices from interpretive research traditions. *Educational Researcher*, 25(1), 20-28.
- Muller-Hartmann, A. (2000). The role of tasks in promoting intercultural learning in electronic learning networks. *Language Learning and Technology*, 4(2), 129-147.
- Newkirk, T. (1995). The writing conference as performance. *Research in the Teaching of English*, 29(2), 193-215.
- Newman, C. and Smolen, L. (1993). Portfolio assessment in our schools: Implementation, advantages, and concerns. *Mid-Western Educational Researcher*, 6, 28-32.
- Ngeow, K. and Kong, Y. (2001). *Learning to Learn: Preparing Teachers and Students for Problem-Based Learning*. ERIC Digest No. ED 457524.
- Nickel, J. (1997). An analysis of teacher-student interaction in the writing conference. *DAI-A*, 37(1), 39.
- Nitko, A. (2001). *Educational Assessment of Students*. New Jersey: Merrill.
- Norris, J. (1998). The audio mirror: Reflecting on students' speaking ability. In J. D. Brown (Ed.), *New Ways of Classroom Assessment* (pp. 164-167). Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc.
- Norris, J. and Hoffman, P. (1993). *Whole language Intervention for School-Age Children*. San Diego, CA: Singular Publishing Group, Inc.
- O'Donnell, A. (1999). Structuring dyadic interaction through scripted cooperation. In Angela O'Donnell and Alison King (Eds.), *Cognitive Perspectives on Peer Learning* (pp. 179-196). Mahwah, NJ: Lawrence Erlbaum Associates.
- O'Malley, J. (1996). *Using Authentic Assessment in ESL Classrooms*. Glenview, Ill: Scott Foresman.

- O'Malley, J. and Chamot, A. (1995). *Learning Strategies in Second Language Acquisition*. New York: Cambridge University Press.
- O'Malley, J. and Pierce, L. (1996). *Authentic Assessment for English Language Learners: Practical Approaches for Teachers*. Reading, MA: Addison-Wesley.
- O'Neil, J. (1992) Putting performance assessment to the test. *Educational Leadership*, 49, 14-19.
- Oosterhof, A. (1994). *Classroom Applications of Educational Measurement*. Englewood Cliffs, NJ: Merrill/Prentice Hall.
- Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham and D. Corson (Eds.), *Language Testing and Assessment* (pp. 175-187). Dordrecht: Kluwer.
- Oxford, R. (1993). *Style Analysis Survey (SAS): Assessing Your Own Learning and Working Styles*. Tuscaloosa, AL: Dept. of Curriculum, University of Alabama.
- Pachler, N. and Field, K. (1997). *Learning to Teach Modern Foreign Languages in the Secondary school*. London: Routledge.
- Palomba, C. and Banta, T. (1999). *Assessment Essentials: Planning, Implementing, and Improving Assessment in Higher Education*. San Francisco: Jossey-Bass.
- Paterson, B. (1995). Developing and maintaining reflection in clinical journals. *Nurse Education Today*, 15(3), 211-220.
- Patthey, C. and Ferris, D. (1997). Writing conferences and the weaving of multi-voiced texts in college composition. *Research in the Teaching of English*, 31(1), 51-90.
- Pederson, M. (1995). Storytelling and the art of teaching. *English Teaching Forum*, 33(1), 2-11.
- Perham, A. (1992). *Collaborative Journals*. ERIC Document No. ED 355555.
- Perl, S. (1994). Composing texts, composing lives. *Harvard Educational Review*, 54(4), 427-449.
- Peyton, J. (1993). *Dialogue journals: Interactive Writing to Develop Language and Literacy*. ERIC Digest No. ED 354789.
- Peyton, J. and Staton, J. (1993). *Dialogue Journals in the Multilingual Classroom: Building Language Fluency and Writing Skills through written Interaction*. Norwood, NJ: Ablex.
- Pierce, L. and O'Malley, J. (1992). *Performance and Portfolio Assessment for Language Minority Students* [On-Line]. Available: <http://www.ncbe.gwu.edu/ncbepubs/pigs/pig9.htm>.
- Pike, K. and Salend, S. (1995). Authentic assessment strategies: Alternatives for norm-referenced testing. *Teaching Exceptional Children*, 28(1), 15-20.
- Ponte, E. (2000). Negotiating portfolio assessment in a foreign language classroom. *DAI-A*, 61(7), 2581.
- Popham, W. (1999). *Classroom Assessment: What Teachers Need to Know*. Boston: Allyn and Bacon.

- Porter, C. and Cleland, J. (1995). *The Portfolio as a Learning Strategy*. Portsmouth, NH: Heinemann.
- Qiyi, L. (1993). Peer editing in my writing class. *English Teaching Forum*, 31(3), 30-31.
- Reid, J. (1993). *Teaching ESL Writing*. Englewood Cliffs, NJ: Prentice Hall.
- Rhodes, L. and Nathenson-Mejia, S. (1992). Anecdotal records: A powerful tool for ongoing literacy. *The Reading Teacher*, 45, 502-509.
- Richer, D. (1993). The effects of two feedback systems on first year college students' writing proficiency. *DAI-A*, 53 (8), 2722.
- Riley, G. and Lee, J. (1996). The discourse accommodation in oral proficiency interviews. *Studies in Second Language Acquisition*, 14, 159-176.
- Robbins, H. (1992). *How To Speak and Listen Effectively*. New York: American Management Association.
- Robbins, J. (1996). Between 'hello' and 'see you later': Development of strategies for interpersonal communication in English by Japanese EFL students. *DAI-A*, 57(7), 3000.
- Roeber, E. (1995). *Emerging Student Assessment Systems for School Reform*. ERIC Document No. ED 389959.
- Ross, D. (1995). Effects of time, test anxiety, language preference, and cultural involvement on Navajo PPST essay scores. *DAI-A*, 56(8), 3032.
- Routman, R. (2000). *Conversations: Strategies for Teaching, Learning, and Evaluating*. Portsmouth, NH: Heinemann.
- Rudner, L. and Boston. C. (1994). Performance Assessment. *The ERIC Review*, 3(1), 3-12.
- Ruskin-Mayher, S. (1999). Whose portfolio is it anyway? Dilemmas of professional portfolio building. *English Education*, 32(1), 6-15.
- Russell, M. and Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis Archives*, 5(3), 1-23.
- Ruth, N. (1998). Portfolio assessment and validation: The development and validation of a model to reliably assess writing portfolios at the middle school level. *DAI-A*, 58(11), 4247.
- Santos, M. (1997). Portfolio assessment and the role of learner reflection. *English Teaching Forum*, 35, 10-15.
- Santos, S. (1998). Exploring relationships between young children's phonological awareness and their home environments: Comparing young children identified at risk for learning disabilities and young children not at risk. *DAI-A*, 59(7), 2446.
- Saunders, W., Goldenbeg, C., and Rennie, J. (1999). *The Effects of Instructional Conversations and Literature Logs on the Story comprehension and Thematic Understanding of English Proficient and Limited English Proficient Students*. ERIC Document No. ED 426631.
- Sawaki, Y. (1999). Reading in Print and from Computer Screens: Comparability of Reading Tests Administered as Paper-and-Pencil and

- Computer-Based Tests. Unpublished Manuscript, Department of Applied Linguistics and TESL, University of California, Los Angeles.
- Schoonen, R., Vergeer, M., and Eiting, M. (1997). The assessment of writing ability: Expert readers versus lay readers. *Language Testing*, 14, 157-184.
- Schwarzer, D. (2001). Whole language in a foreign language class: From theory to practice. *Foreign Language Annals*, 34(1), 52-59.
- Searle, J. (1998). De-coding reading at work: Workplace reading competencies. *Literacy Broadsheet*, 48, 8-11.
- Secord, W., Wigg, E., Damico, J., and Goodin, G. (1994). *Classroom Communication and Language Assessment*. Chicago, IL: Riverside.
- Sens, C. (1993). The cognition of self-evaluation: A descriptive model of the writing process. *DAI-A*, 31(4), 1489.
- Shackelford, R. (1996). Student portfolios: A process/product learning and assessment strategy. *Technology Teacher*, 55(8), 31-36.
- Shaklee, B., Barbour, N., Ambrose, R., and Hansford, S. (1997). *Designing and Using Portfolios*. Boston: Allyn and Bacon.
- Shameem, N. (1998). Validating self-reported language proficiency by testing performance in an immigrant community: The Wellington Indo-Fijians. *Language Testing*, 15(1), 86-108.
- Shepard, L. (2000). *The Role of Classroom Assessment in Teaching and Learning*. CSE Technical Report, No. 517. CRESST, University of California, Los Angeles, and CREDE, University of California, Santa Cruz.
- Shoemaker, S. (1998). Literacy self-assessment of students with special education needs. *DAI-A*, 59(2), 0410.
- Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing*, 11(2), 99-123.
- Smagorinsky, P. (1995). *Speaking about Writing: Reflections on Research Methodology*. Thousand Oaks, CA: Sage.
- Smith, C. (2000). *Writing Instruction: Current Practices in the Classroom*. ERIC Digest No. ED 446338.
- Smithson, I. (1995). *Assessment of Writing across the Curriculum Projects*. ERIC Document No. ED 382994.
- Smolen, L., Newman, C., Wathen, T., and Lee, D. (1995). Developing Student Self-Assessment Strategies. *TESOL Journal*, 5(1), 22-27.
- Sokolik, M. and Tillyer, A. (1992). Beyond portfolios: Looking at students' projects as teaching and evaluation devices. *College ESL*, 2(2), 47-51.
- Song, M. (1997). The effect of dialogue journal writing on overall writing quality, reading comprehension, and writing apprehension of EFL college freshmen. *DAI-A*, 58(3), 781.
- Soodak, L. (2000). Performance assessments and students with learning problems: Promoting practice or reform rhetoric? *Reading and Writing Quarterly: Overcoming Learning Difficulties*, 16(3), 257-280.
- Stahl, R. (1994). *The Essential Elements of Cooperative Learning in the Classroom*. ERIC Digest No. ED 370881.

- Stansfield, C. (1992). *ACTFL Speaking Proficiency Guidelines*. ERIC Digest No. ED 347852.
- Stansfield, C. and Kenyon, K. (1996). *Stimulated Oral Proficiency Interviews: An Update*. ERIC Digest No. ED 395501.
- Statman, D. (1993). Self-assessment, self-esteem and self-acceptance. *Journal of Moral Education*, 22, 55-62.
- Steadman, M. and Svinicki, M. (1998). CATs: A student's gateway to better learning. In T. Angelo (Ed.), *Classroom Assessment and Research: An Update on Uses, Approaches, and Research Findings* (pp. 13-20). New Directions for Teaching and Learning No. 75. San Francisco: Jossey-Bass.
- Stiggins, R. (1994). *Student-Centered Classroom Assessment*. Englewood Cliffs, NJ: Merrill/Prentice Hall.
- Stix, A. (1997). *Empowering Students through Negotiable Contracting*. ERIC Document No. ED 411274.
- Stockdale, J. (1995). Storytelling. *English Teaching Forum*, 33(1), 22-27.
- Stoller, F. (1997). Project work: A means to promote language content. *English Teaching Forum*, 35 (4), 2-9.
- Storey, P. (1997). Examining the test-taking process: A cognitive perspective on the discourse cloze test. *Language Testing*, 14(2), 214-231.
- Sussex, R. and White, P. (1996). Electronic Networking. *Annual Review of Applied Linguistics*, 16, 200-225.
- Sythes, L. (1999). *Making Sense of Conflict: A "Maternal" Framework for the Writing Conference*. ERIC Document No. ED 439427.
- Tannenbaum, J. (1996). *Practical Ideas on Alternative Assessment for ESL Students*. ERIC Digest No. ED 395500.
- Tanner, P. (2000). Embedded assessment and writing: Potentials of portfolio-based testing as a response to mandated assessment in higher education. *DAI-A*, 61(7), 2694.
- Tarone, E. and Yule, G. (1996). *Focus on the Language Learner*. Oxford: Oxford University Press.
- Teemant, A. (1997). The role of language proficiency, test anxiety, and testing preferences in ESL students' test performance in content-area courses. *DAI-A*, 58(5), 1674.
- Tenbrink, T. (1999). Assessment. In James M. Cooper et al. (Eds.), *Classroom Teaching Skills* (pp. 310-384). Boston, MA: Houghton Mifflin Company.
- Thomson, C. (1995). Self-assessment in self-directed learning: Issues of learner diversity. In R. Pemberton et al. (Eds.), *Taking Control: Autonomy in Language Learning* (pp. 77-91). Hong Kong: Hong Kong University Press.
- Thurlow, M. (1995). *National and State Perspectives on Performance Assessment*. ERIC Digest No. ED 381986.
- Todd, R. (2002). Using self assessment for evaluation. *English Teaching Forum*, 40(1), 16-19.
- Tompkins, G. (2000). *Teaching Writing: Balancing Process and Product*. New Jersey: Merrill.

- Tompkins, G. and Hoskisson, K. (1995). *Language Arts: Content and Teaching Strategies*. Englewood Cliffs, NJ: Prentice Hall.
- Tompkins, P. (1998). Role playing/simulation. *The Internet TESL Journal*, 7(7), 1-5. Available at: <http://iteslj.org/>
- Topping, K. and Ehly, S. (1998). Introduction to peer-assisted learning. In K. Topping and S. Ehly (Eds.), *Peer-Assisted Learning*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Trostle, S and Hicks, S. (1998). The effects of storytelling versus reading on comprehension and vocabulary of British primary school children. *Reading Improvement*, 35(3), 127-136.
- Troyer, C. (1992). Language in first-grade: A study of reading, writing, and talking in school. *DAI-A*, 53(6), 1855.
- Tunstall, P. and Gipps, C. (1996). Teacher Feedback to young children in formative assessment: A typology. *British Education Research Journal*, 22(4), 389-404.
- Tyndall, B. and Kenyon, D. (1995). Validation of a new holistic rating scale using Rasch multifaceted analysis. In A. Cumming and R. Berwick (Eds.), *Validation in Language Testing*. Clevedon, Avon: Multilingual Matters.
- Upshur, J. and Turner, C. (1999). Systematic effects in the rating of second language speaking ability: Test method and learner discourse. *Language Testing*, 16, 82-111.
- Van-Daalen, M. (1999). Test usefulness in alternative assessment. *Dialog on Language Instruction*, 13(1-2), 1-26.
- Vandergrift, L. (1997). The Cinderella of communicative strategies: Reception strategies in interactive listening. *Modern Language Journal*, 81(4), 494-505.
- Van-Duzer, C. (2002). *Issues in Accountability and Assessment for Adult ESL Instruction*. Available at email: ncle@cal.org; web: <http://www.cal.org/ncle/DIGESTS>.
- Vann, S. (1999). *Assessment of Advanced ESL Students Through Daily Action Logs*. ERIC Document No. ED 435187.
- Walden, P. (1995). Journal writing: A tool women developing as knowers. *New Directions for Adult and Continuing Education*, 65, 13-20.
- Wall, J. (2000). *Technology-Delivered Assessment: Diamonds or Rocks?* ERIC Digest No. ED 446325.
- Wallace, C. (1992). *Reading*. Oxford: Oxford University Press.
- Warschauer, M (Ed.). (1995). *Virtual Connections: Online Activities and Projects for Networking Language Learners*. Honolulu, HI: University of Hawaii Press.
- Warschauer, M., Shetzer, H., and Meloni, C. (2000). *Internet for English Teaching*. Alexandria, VA: Teachers of English to Speakers of Other Languages, Inc.
- Watson, D. and Henson, J. (1991). Reading Evaluation—Miscue Analysis. In Bill Harp (Ed.), *Assessment and Evaluation for Whole Language Programs* (pp. 51-72). Norwood, MA: Christopher-Gordon Publishers, Inc.

- Webb, N. (1995). Group collaboration in assessment: Multiple objectives, processes and outcomes. *Educational Evaluation and Policy Analysis*, 17(2), 239-261.
- Weigle, S. (1998). Using FACETS to model rater training effects. *Language Testing*, 15(2), 263-267.
- Weir, C. (1993). *Understanding and Developing Language Tests*. Trowbridge, Wiltshire: Prentice-Hall.
- Wiener, R. and Cohen, J. (1997). *Literacy Portfolios: Using Assessment to Guide Instruction*. ERIC Document No. ED 412532.
- Wiggins, G. (1993). *Assessing Student Performance: Exploring the Purpose and Limits of Testing*. San Francisco: Jossey-Bass Publishers.
- Wiig, E. (2000). Authentic and other assessments of Language disabilities: When is fair fair? *Reading and Writing Quarterly*, 16(3), 179-211.
- Wijgh, I. (1996). A communicative test in analysis: Strategies in reading authentic texts. In A. Cumming and R. Berwick (Eds.), *Validation in Language Testing* (pp. 145-170). Clevedon: Multilingual Matters.
- Wilhelm, K. and Wilhelm, T. (1999). Oh, the tales you'll tell! *English Teaching Forum*, 37(2), 25-28.
- Williams, M. and Burden, R. (1997). *Psychology for Language Teachers: A Social Constructivist Approach*. Cambridge: Cambridge University Press.
- Wolfe, E. (1995). A study of expertise in essay scoring (performance-assessment). *DAI-A*, 57(3), 1023
- Wolfe, E. (1996). *Student Reflect in Portfolio Assessment*. ERIC Document No. ED 396004.
- Wood, N. (1993). Self-corrections and rewriting of student compositions. *English Teaching Forum*, 31(3), 38-40.
- Worthington, L. (1997). Let's not show the teacher: EFL students' secret exchange journals. *English Teaching Forum*, 35(3), 2-7.
- Wright, S. (1995). Literacy behaviors of two first-grade children with down syndrome. *DAI-A*, 57(1), 156.
- Wrigley, H. (2001). Assessment and accountability: A modest proposal. *Field Notes*, 10(3), 1, 4-7.
- Wu, Y. (1998). What do tests of listening comprehension test? A retrospection study of EFL test-takers performing a multiple-choice task. *Language Testing*, 15(1), 21-44.
- Yancey, K. (1998). Getting beyond exhaustion, reflection, self-assessment, and learning. *Clearing House*, 72, 13-17.
- Yung, V. (1995). Using reading logs for business English. *English Teaching Forum*, 33(2), 40-41.
- Zoya G. and Morse, K. (2002). Debate as a tool of English language learning. Paper presented at the 36th TESOL Conference, Salt Lake City, Utah, USA, April 9-13, 2002.

