

**The Effect of Classroom Performance Assessment
on EFL Students' Basic and Inferential
Reading Skills**

Abdel Salam Abdel Khalek El-Koumy

Professor of Curricula and Instruction of English as a Foreign

Language at Suez Canal University, Egypt

November 2009

Abstract

The purpose of this study was to investigate the effect of classroom performance assessment on the EFL students' basic and inferential reading skills. A pretest-posttest quasi-experimental design was employed in the study. The subjects of the study consisted of 64 first-year secondary school students in Menouf Secondary School for Boys at Menoufya Directorate of Education (Egypt) during the academic year 2006/2007. These subjects were divided into an experimental group and a control group. Both groups were pretested to measure their basic and inferential reading skills before conducting the experiment. During treatment, students in the experimental group used the KWL chart and the self-assessment checklist for assessing their own reading strategies and comprehension in each reading session. The KWL chart and the self-assessment checklist were then compiled in a portfolio for each student. This portfolio was read by the teacher every week to provide both 'feedback' and 'feedforward' for improving each student's reading strategies and comprehension. Students in the control group answered a traditional discrete item test at the very end of each lesson and unit. This traditional test focused mainly on the phonological, lexical and grammatical elements of the reading skill, and students were judged on the basis of how well they achieved as compared to each other. The experiment lasted for six months. After treatment, the same pretests were readministered to both groups. The collected data were analyzed using the t-test. The pre-test data analysis revealed that there were no significant differences in the basic and inferential reading skills between the experimental group and the control group ($t=0.48$, $p > 0.05$; $t=-0.46$, $p > 0.05$, respectively). However, the post-test data analysis showed that there was a statistically significant difference between the two groups of the study in the basic reading skills in favor of the control group and in the inferential reading skills in favor of the experimental group ($t=-2.61$, $p=0.01$; $t=7.75$, $p=0.000$, respectively). These findings suggest that classroom performance assessment is less effective in improving secondary school EFL students' basic reading skills, but more effective in developing their inferential reading skills than traditional assessment. In light of these findings, the researcher recommends that a multi-dimensional comprehensive approach to classroom assessment is more likely to improve both the basic and inferential reading skills of intermediate-level EFL students.

Background of the Problem

In the global information-based society, English reading comprehension has become essential for sharing ideas with others and obtaining up-to-date information in all fields of life because “90% of all information in the world's electronic retrieval systems,” as Hasman (2000) states, “is stored in English” (p. 2). Despite the importance of English reading comprehension, Egyptian secondary school students cannot understand what is between the lines when they read in English. More specifically, they lack the inferential reading comprehension skills although they possess a large size of English vocabulary and a lot of English grammar (El-Koumy, 2006). The researcher claims that the first reason for the lack of these skills in Egyptian secondary school students is that their teachers spend most of the instruction time testing bits of reading rather than teaching and assessing higher-order comprehension skills because they mould reading instruction around both the content and format of the traditional high-stakes test, neglect the materials that this test excludes, and depend largely on drills with old test items. In line with this reason, many studies all over the world (e.g., Shepard, 1991; Smith, Edelsky, Draper, Rottenberg and Cherland, 1991; Smith and Rottenberg, 1991; Levinson, 2000; Pendulla et al., 2003) showed that teachers aligned instruction with the content and format of the

traditional high-stakes test.

A second reason for Egyptian secondary school EFL students' lack of inferential reading skills is that their teachers devote more class time to teaching test-taking strategies rather than reading comprehension strategies to prepare them to pass the traditional high-stakes test. They, for example, instruct them about how to guess the answer when in doubt since students are not penalized for incorrect responses. Therefore, they lack the strategies they need to read effectively. In line with this reason, Taylor, Shepard, Kinner and Rosenthal (2003) found that traditional high-stakes testing forced teachers to emphasize test-taking strategies and did not provide students with strategies that could help them achieve true understanding and become successful lifelong learners. Along with the same reason, research studies (e.g., Ward and Traweek, 1993) showed that students who had difficulty in comprehension were not aware of reading strategies and did not have the ability to regulate or monitor their own comprehension.

A third reason for the low level of English reading comprehension among Egyptian secondary school students is that traditional testing encourages

them to focus on memorization, recognition and regurgitation of decontextualized bits of reading at the expense of higher level comprehension skills. This reason is in line with Haertel and Mullis' (1996) view that "overreliance on multiple-choice and similar item formats has led to curricula and instructional methods that encourage learning isolated bits of information and mechanically applying isolated skills, at the expense of more complex reasoning and meaningful problem solving" (p. 287).

A fourth reason for the low reading comprehension level of Egyptian secondary school EFL students is that their teachers hold the misconception that assessment should focus only on ranking and sorting students rather than improving their learning. These teachers view classroom assessment as an activity tacked on the end of a lesson or unit for grading purposes. They cannot interpret assessment information, or use it to decide where the students are in their learning, where they need to go, and how best to get there. More specifically, they don't use reading tests for planning the next steps in response to students' needs, or for diagnosing each student's reading difficulties and individualizing instruction to improve these difficulties. They focus only on scores which are poor predictors of how well students understand what they read. They also ask low-level cognitive

questions, and resort to ready-made tests in commercially prepared books that are fully aligned to the high-stakes test. Furthermore, they are not aware of performance assessment formats or how to use them in classrooms. In line with this reasoning, much research worldwide (e.g., Stiggins, 1993, 2002; Black, Harrison, Lee, Marshall and Wiliam, 2003; Popham, 2004; Stiggins, Arter, Chappuis and Chappuis, 2004; Burke, 2005) suggests that teachers in general are not proficient in classroom assessment practices.

A final reason for Egyptian secondary school students' low level of English reading comprehension is their high-anxiety level generated by traditional reading tests. This high-level of test anxiety, in turn, increases their problems with constructing meaning from the text and negatively affects their reading comprehension. In line with this reasoning, the results of Kellaghan, Madaus and Airasian's (1982) study on test anxiety revealed that traditional tests were a source of emotional discomfort for students. Along the same reasoning, some studies found that test anxiety negatively correlated with both reading strategy use and reading comprehension. Calvo and Eysenck (1996), for example, found that "high-anxiety subjects produced overt articulation more frequently than low-anxiety subjects,"

and that "anxious subjects showed poorer comprehension than non-anxious subjects" (p. 291). Stallworth-Clark, Cochran, Nolen, Tuggle and Scott (2000), for another example, found that students who scored higher on measures of test anxiety scored lower on reading competency tests than students whose anxiety scores were lower.

It appears from the foregoing that all the probable reasons for Egyptian secondary school students' low level of English reading comprehension are closely related to traditional testing because Egyptian EFL teachers do not teach reading comprehension, but test it all the time using traditional formats that resemble the high-stakes test. In other words, how EFL reading comprehension is tested is how it gets taught in Egyptian secondary schools.

Despite its negative effects on both teaching and learning in general and reading comprehension in particular, traditional testing remains the predominant form of assessment all over the world, including Egypt. This may be due to the fact that this form of assessment is "easy to administer and grade" (Johnson, 1989, p. 57). It may also be due to its low cost as compared to alternative forms of assessment. As Koretz and Hamilton

(2006) with reference to the U.S. General Accounting Office put it:

The problem of high costs [of alternative assessment formats] may be the most important factor contributing to states' reliance on multiple-choice testing. The magnitude of the problem is evident in a study by the U.S. General Accounting Office (2003), which estimated states' costs for implementing large-scale testing. The total estimated cost for states using only multiple-choice tests was approximately \$1.9 billion, whereas the cost if states also included a small number of hand-scored open response items such as essays was estimated to be about \$5.3 billion. The magnitude of this difference suggests that many states may remain reluctant to abandon extensive reliance on the multiple choice format, at least until the alternative testing technologies become less expensive. (p. 536)

In addition, French (2003), with reference to others, notes that the dominance of traditional high-stakes testing, particularly in the USA context, lies in the fact that this type of testing is a multi-billion-dollar industry from which testing companies gain substantial amounts of money every year; and therefore, they exert so much effort to maintain it in use.

He writes:

What we have, then, is a high stakes testing movement being fueled, not by those who best know and care about the students in our middle schools, but by others outside of our public schools who have varied interests at heart. In 1999, NCS Pearson, a large testing company, reported more than \$620 million in revenues, up 30% from the previous year. McGraw-Hill, another large testing company, also owns programs such as Open Court and Reading Mastery, two direct instruction programs that are being purchased in large numbers by districts striving to drive up their standardized test scores (Kohn, 2002). State spending on testing has increased multifold. It is estimated that the K-12 standardized testing industry is as much as \$1.5 billion per year (Kohn, 2002; Gluckman, 2002). Business leaders and legislators have lined up behind this industry as a quick fix to the dilemmas of educating a diverse student population. (p. 8)

Over and above, the worldwide prevalence of using traditional testing in teaching and assessing reading comprehension is due to the tragic fact that

teachers find it hard to move outside their comfort zone. They feel comfortable with the traditional formats they are accustomed to because they themselves were taught and assessed by them. They find it hard to try the innovative formats of classroom performance assessment because they lack the pre-service and in-service training that enable them to use these new formats.

The many criticisms leveled at traditional testing, as well as the widespread dissatisfaction with students' levels of achievement in general and reading comprehension in particular, made many educational reformers and reading-assessment specialists (e.g., Wood, 1988; Linn, Baker and Dunbar, 1991; Lavande, 1993; Hart, 1994; Valencia, 1994; Hammond, Ancess and Falk, 1995; O'Malley and Valdez Pierce, 1996; Gagliano and Swiatek, 1999; Geocaris and Ross, 1999; McNamara, 2000; Stiggins, 2002; El-Koumy, 2004) call for a shift to alternative formats often called performance or authentic assessment. They claim that these alternative formats have many advantages over traditional assessment. These advantages include building students' self-confidence, reducing their test anxiety, enhancing their self-esteem, motivating them to excel because of involving them in meaningful activities which are needed in the real-world, emphasizing and promoting

higher-order thinking skills, developing strategic learners who have knowledge of their own learning processes, providing the teacher with valuable information about how each student learns and what strategies she/he uses, equipping students to function effectively in the world beyond the school doors, allowing teachers to adjust instruction in response to each student's needs, helping students learn how to learn, and allowing them to become independent learners. However, opponents of performance assessment claim that this type of assessment neglects the basic skills; whereas its advocates hold that the whole is more than the sum of its elements and that the basics taught in isolation from context are not likely to become functional. Opponents of performance assessment maintain that this type of assessment is invalid and unreliable; whereas its supporters hold that it is valid in terms of consequences, impartiality, transference, authenticity, cognitive complexity, significance, and efficiency; and contend that its reliability can be obtained by using a variety of formats for data collection, appropriate rubrics for scoring, and more than a single observer, interviewer or reader.

It appears from the foregoing that both traditional and performance assessments have advantages and disadvantages and the advantages of one

of them imply the disadvantages of the other. With respect to reading, traditional testing allows for domain coverage by focusing on the building blocks of this skill, yet it ignores higher-order comprehension skills. On the other hand, performance assessment elicits higher-order reading skills and measures both the process and product of comprehension, yet it does not offer the potential for domain coverage and may produce students who are clueless about the basics. Therefore, it is important to investigate the relative effects of the two types of assessment to find out which one outweighs the other with respect to improving both basic and inferential reading skills.

Problem and Purpose of the Study

The problem of this study was that Egyptian first-year secondary school EFL students exhibited deficiencies in higher-order reading comprehension skills. In an attempt to find a solution to this problem, this study aimed at investigating the effects of classroom performance assessment as compared to classroom traditional assessment on their basic and inferential reading skills.

Hypotheses of the Study

This study aimed at testing the following hypotheses:

- 1. There would be a statistically significant difference ($\alpha \leq 0.05$) in the first-year secondary school EFL students' basic reading skills between the experimental group exposed to classroom performance assessment and the control group exposed to classroom traditional assessment in favor of the latter group.**
- 2. There would be a statistically significant difference ($\alpha \leq 0.05$) in the first-year secondary school EFL students' inferential reading skills between the experimental group exposed to classroom performance assessment and the control group exposed to classroom traditional assessment in favor of the former group.**

Significance of the Study

The significance of this study lies in the testing of classroom assessment which can ultimately improve teaching and learning in schools because assessment in and of itself may have a negative or positive impact on students' learning. This is because teachers worldwide instinctively teach to the test. As Herman (2004) puts it: “The instinct to simply 'teach to the test'

may in part be a survival instinct” (p. 159). Therefore, many educators all over the world consider testing as the key lever for change in education and classroom assessment as the true path to educational reform. As Womer (1984) states: “Lay persons and legislators who control education see testing-assessment as a panacea for solving our concerns about excellence in education” (p. 3). In England and Wales, for example, the secretary of state for education set up a Task Group on Assessment and Testing (TGAT) to advise on the assessment policy for the new national curriculum in 1987. The TGAT report (1988) views assessment as the core of educational reform as follows:

Promoting children’s learning is a principal aim of schools. Assessment lies at the heart of this process. It can provide a framework in which educational objectives may be set, and pupil’s progress charted and expressed. It can yield a basis for planning the next educational steps in response to children’s needs ... it should be an integral part of the educational process, continually providing both ‘feedback’ and ‘feedforward’. It therefore needs to be incorporated systematically into teaching strategies and practices at all levels. (DES/WO, 1988, paras. 3-4)

In the USA, for another example, the idea of using assessment as a lever for educational reform is stated by Linn and Herman (1997) in the following way:

Assessments play a pivotal role in standards-led reform, by: communicating the goals....providing targets..., and shaping the performance of educators and students. Coupled with appropriate incentives and/or sanctions—external or self-directed—assessments can motivate students to learn better, teachers to teach better, and schools to be more educationally effective. (p. iii)

In view of the above, educational reformers claim that schools are in need of tests 'worth teaching to' to achieve their goals. They believe that assessment should model the kinds of learning that we expect students to achieve. They also feel that traditional tests cannot meet the demands required of students in the real world; and therefore, they should be replaced by tests that embody the higher-order cognitive skills we want students to learn. As Resnick and Resnick (1992) put it: "[I]f we put debates, discussions, essays and problem solving into the testing system, children will spend time practicing those activities" (p. 59). Shepard et al. (1995), with reference to

others, make the same point in the following way:

[I]t is natural for teachers to work hard to prepare students to do well on examinations that matter. Rather than forbid 'teaching to the test,' which is impossible, it is preferable to create measures that will result in good instruction even when teachers do what is natural. (pp. 1-2)

Alongside and parallel to the earlier view, performance assessment has been considered as 'worth teaching to' because it, as its supporters claim, develops students' higher-order thinking skills which are the ultimate goals of the 21st century education. Linn and Baker (1996) with reference to Resnick and Resnick put this idea in the following way:

The desire for major reform of the curriculum provides a second major motivation for the introduction of performance-based assessments. A widely held belief is that you get what you assess and conversely that you do not get what you do not assess. A major concern about standardized tests is that they drive instruction in undesirable ways by focusing on the accumulation of facts and decontextualized skills (Resnick and Resnick, 1992). A curriculum that focuses on decomposed bits

and pieces presented without context is incompatible with the type of curriculum reform advocated by groups such as the National Council on Education....Performance-based assessment seem to be more compatible with curriculum reforms that emphasize the identification and solution of real-world problems, reasoning, and higher-order thinking skills. Indeed, performance-based assessments are considered an integral part of curriculum reform. (p. 86)

Liskin-Gasparro (1997) also views performance assessment as a means of educational reform to improve leaning and instruction as follows:

With assessment that is performance-oriented, the thinking goes, with assessment that aims to measure not only the correctness of a response, but also the thought processes involved in arriving at the response, and that encourages students to reflect on their own learning in both depth and breadth, the belief is that instruction will be pushed into a more thoughtful, more reflexive, richer mode as well. Teachers who teach to these kinds of alternative assessments will naturally teach in ways that emphasize reflection, critical thinking, and

personal investment in one's own learning.

As indicated earlier, reform in testing represents the most pressing issue in education today and performance assessment is being hailed as the true path to this reform. However, the effect of this type of assessment on students' learning has not been studied thoroughly, particularly in Egypt. Therefore, this study is urgently needed before money and effort are expended to apply it in a large scale in Egypt. It is hoped that this study will help in building an empirical knowledge base to inform Egyptian policymakers when they make decisions with respect to adopting this new type of assessment, and Egyptian teachers and students when they implement it into the heart of the teaching and learning process.

It is also hoped that a shift toward the application of classroom performance assessment will help remedy the ills which have become inherent with the emphasis on traditional assessment, and will improve students' inferential reading skills to enable them to get to the heart of things and the deeper meanings of what they read.

Operational Definition of Terms

The terms below, wherever seen in this study, have the following definitions:

Classroom performance assessment, also known as alternative or authentic classroom assessment, is a form of assessment that requires students to construct responses rather than select among preexisting options. It centers not only on the product of learning, but also on the process students go through to create that product to provide ongoing feedback and feedforward for improving each student's performance relative, not to others, but to the student herself/himself. It also occurs within the natural context of students' learning environment and calls for students to learn while they are being assessed by themselves or others. This form of assessment includes a variety of formats such as dialogue journals, verbal reports, conferences, learning logs, KWL charts, self-assessment checklists and portfolios. The present study is confined to the last three formats.

Traditional assessment is a form of assessment that requires students to select an answer from ready-made options. It focuses mainly on decontextualized fragments and gives more attention to grading and assigning students to levels rather than giving feedback about how teaching

and learning can be improved. This form of assessment includes a variety of formats such as multiple-choice, true-false, matching and fill-in-the-blank.

Basic reading skills: For the present study, this term refers to the discrete subskills of reading including word decoding, phonological awareness, vocabulary and grammatical knowledge.

Inferential reading skills: For the present study, this term refers to inference skills such as identifying the author's purpose, tone, point of view and bias, identifying the implied main idea, recognizing causal relations in the reading text, comparing and contrasting ideas across the text, drawing logical conclusions from the text, etc.

Limitations of the Study

The generalization of the results of the study is limited to first-year secondary school EFL students. It is also limited to the three performance assessment formats used in the study (KWL charts, self-assessment checklists and portfolios), the operational definition of the independent and dependent variables, the length of the experiment, and the instruments used to collect data for the study.

Theoretical Framework

The theoretical framework of the present study is organized around both the behaviorist and constructivist theories of learning. Traditional assessment has its roots in the behaviorist assumption that each macro-skill includes many sub- or micro-skills that need to be mastered and measured separately and sequentially before learners can proceed to the next. This form of assessment uses closed questions with only one correct answer to discover whether the student knows the predetermined subskill to progress to the next. The following quotation from Skinner (1954) illustrates this assumption:

The whole process of becoming competent in any field must be divided into a very large number of very small steps, and reinforcement must be contingent upon the accomplishment of each step. This solution to the problem of creating a complex repertoire of behavior also solves the problem of maintaining the behavior in strength. . . . By making each successive step as small as possible, the frequency of reinforcement can be raised to a maximum, while the possibly aversive consequences of being wrong are reduced to a minimum. (p. 94)

On the other hand, performance assessment is based on the constructivist theory which views assessment as an integral part of the teaching/learning process. This theory contends that assessment should focus on students' learning processes and products rather than the accumulation of bits and pieces of information. It also contends that assessment tasks should be open-ended, authentic, meaningful and valuable beyond the classroom. In addition, according to Shepard (2000), the constructivist view of assessment includes student self-assessment and feedback as a central "part of the social processes that mediate the development of intellectual abilities, construction of knowledge, and formation of students' identities" (p. 3). Shepard (Ibid.) maintains that the constructivist view considers "assessment as a source of insight and help instead of its being the occasion for meting out rewards and punishments" (p. 53). In essence, according to Rudner and Boston (1994), "the process of [performance] assessment is itself a constructivist learning experience, requiring students to apply thinking skills, to understand the nature of high quality performance, and to provide feedback to themselves and others" (p. 7).

Review of Related Research

The effect of testing on both teaching and learning has been a subject of research for many years. In various content areas, there are several studies (e.g., Neil and Medina, 1989; Herman and Golan, 1991; Smith and Rottenberg, 1991; McNeil and Valenzuela, 2000; Amrein and Berliner, 2002; Moon, Brighton and Callahan, 2003; Neil, 2003) suggesting that traditional tests result in negative consequences on both teaching and learning; in contrast, there are several other studies (e.g., Gaynor and Millham, 1976; Glover, Zimmer and Bruning, 1979; Cizek, 2001; Fuller and Johnson, 2001; Roderick and Engel, 2001; Skrla and Scheurich, 2001) suggesting that frequent traditional tests result in improving students' learning. However, still other studies (e.g., Nungester and Duchastel, 1982; Mehrens and Kaminski, 1989; Van Horn, 1997; Vining and Bell, 2005) indicate that the higher scores obtained by students, who are frequently tested by traditional tests, are attributed to students' test wiseness and the teaching of test-taking strategies.

Similarly, there are several studies (e.g., Koretz, Stecher, Klein and McCaffrey, 1994; Khattri, Kane, and Reeve, 1995; Shepard et al., 1995; Koretz, Barron, Mitchell and Stecher, 1996; Tilton, 1996; Khattri, Reeve

and Kane, 1998; Supovitz, 2001) suggesting that performance assessment results in modest or equivocal effects on learning and instruction; in contrast, there are several other studies (e.g., Borko, Flory and Cumbo, 1993; Falk and Darling-Hammond, 1993; Koretz, Stecher, Klein, McCaffrey and Diebert, 1993; Dunne, 1996; Newmann, Marks and Gamoran, 1996; Stretcher and Mitchell, 1996; Cross, Greer, and Pearce, 1998; Rhine and Smith, 2001; Kim, 2003; Nicol and Owen, 2008) suggesting that this type of assessment results in a number of positive effects on teachers' practices and students' learning. However, still other studies (e.g., Koretz, Mitchell, Barron and Keith, 1996; Koretz and Barron, 1998) indicate that performance assessment is not immune to score inflation.

To conclude this section, it can be said that the existing evidence, with respect to drawing any conclusions about the consequences of traditional and performance assessments, is inadequate because the findings of the related studies are contradictory; and the evidence against traditional assessment is not as strong as it has been theoretically claimed. In addition, most of the studies were attached to large-scale not classroom assessment; and none of them was conducted with Egyptian students. Therefore, this study seems essential before applying performance assessment in Egyptian

schools.

Methodology

Design of the Study

A pretest-posttest quasi-experimental design was employed in the study. In this design the researcher used an experimental group and a control group. Both groups were pre-tested to measure their basic and inferential reading skills before conducting the experiment. During the experiment, the experimental group students were exposed to classroom performance assessment; whereas the control group students were exposed to classroom traditional assessment. After treatment, the two groups were post-tested to investigate any significant differences in their basic and inferential reading skills. This design is displayed in Table 1 below.

Table 1

Design of the study

Experimental Group	01	02	X1	01	02
Control Group	01	02	X2	01	02

Where:

01= Basic Reading Skills Test

02= Inferential Reading Skills Test

X1= Classroom Performance Assessment

X2= Classroom Traditional Assessment

Subjects of the Study

The subjects of the study consisted of 64 first-year secondary school students in Menouf Secondary School for Boys at Menoufya Directorate of Education (Egypt) during the academic year 2006/2007. These subjects were assigned to an experimental group and a control group. Almost all of them were 16 years old. They were also similar regarding their economic and social conditions.

Data Collection Instruments

For the purpose of collecting data for the study, the researcher developed two tests to measure students' basic and inferential reading skills (one for each) before and after conducting the experiment. The basic reading skills test consisted of 4 subtests for measuring word decoding, phonological awareness, vocabulary, and grammatical knowledge (one for each). The

word decoding subtest consisted of 80 single words of increasing difficulty. These words were equally selected from the reading passages in the student's book (five from each). Students had to read these words correctly and as quickly as possible. The score was the number of words read correctly in 1 minute.

The phonological awareness subtest consisted of three parts. Each part consisted of 10 items and each item consisted of three words that were also selected from the reading passages in the student's book. In the three parts students had to tick the word that differs from the two other words. In the first part they had to tick the word that does not rhyme with the other two. In the second and third parts they had to tick the word that differs by first or last phoneme, respectively.

The vocabulary knowledge subtest consisted of three parts (10 items for each). In the first part students had to choose from four options the word that is closest in meaning to another word. In the second part they had to choose from four options the definition that is closest in meaning to a single word. In the third part they had to match words of opposite meaning. All words were selected from the key ones introduced in the reading passages

(nearly two from each passage) to cover the vocabulary domain within the student's book

The grammatical knowledge subtest consisted of three parts. Each part consisted of 10 items. In the first part students had to construct a sentence from an unordered string of component words. In the second part they had to choose from four options the grammatical structure that completes the sentence correctly. In the third part they had to change the form of the word between brackets to fit into the sentence. All grammatical structures were selected from those introduced in the reading passages in the student's book.

The inferential reading skills test consisted of three reading comprehension passages with 15 questions (5 for each). The five inferential questions on each passage comprised: (1) inferring the implied main idea, (2) identifying the author's implicitly stated purpose for writing the text, (3) inferring the author's tone within the text, (4) inferring the relation that holds between two propositions in the text, and (5) drawing a logical conclusion from the text.

To ensure the validity of the two tests, a jury of two EFL supervisors and two university professors was consulted, and their comments were taken into consideration. To ensure their reliability, the two tests were administered to a sample of twenty first-year secondary school students out of the sample of the study and readministered thirteen days later to the same sample to investigate their stability over time. The Pearson correlation coefficients between the scores of the two administrations were 0.91 for the basic reading skills test and 0.78 for the inferential reading skills test which indicated that the two tests were stable over time.

Materials for the Study

The instructional materials for the study consisted of the sixteen reading passages involved in the Student's Book (*Hello! 6*). Students in the two groups of the study were exposed to these materials with the exception that the experimental group students were exposed to classroom performance assessment; whereas the control group ones were exposed to classroom traditional assessment.

Variables of the Study

Independent variables:

- (a) Classroom performance assessment**
- (b) Classroom traditional assessment**

Dependent variables:

- (a) Basic reading skills**
- (b) Inferential reading skills**

Procedures of the Study

The following procedures were followed for the purpose of collecting data for the study:

- (1) Getting the approval of Menoufya Directorate of Education to conduct the experiment.**
- (2) Choosing the subjects for the study from Menouf Secondary School for Boys.**
- (3) Pre-testing the experimental group and the control group on September 24, 2006, to measure their basic and inferential reading skills before conducting the experiment. The results of the analysis of the pre-test scores revealed that the t-value of the difference in the mean scores**

between the experimental group and the control group on the basic reading skills test was 0.48 and on the inferential reading skills test was - 0.46. These values are not significant at the 0.05 level which indicates that the two groups were equivalent in both their basic and inferential reading skills before conducting the experiment.

(4) Training the experimental group students in implementing the performance assessment formats used in the study by modeling the use of a KWL chart and a self-assessment checklist (See Appendixes A and B) to them through thinking out loud and asking them to apply both formats independently until they became quite comfortable with their use. After that, the experimental group teacher was trained on how to identify each student's strengths and weaknesses in reading strategies and comprehension while reading the KWL chart and the self-assessment checklist in each portfolio, without assigning grades to responses.

(5) Conducting the experiment from the beginning of October until the end of March during the academic year 2006/2007. During treatment, students in the experimental group used the KWL chart, and the self-assessment checklist parts sequentially and circularly (one per session), for assessing their own reading strategies and comprehension in each

reading session. The KWL chart and the self-assessment checklist part were then compiled in a portfolio for each student. This portfolio was read by the teacher every week to provide both ‘feedback’ and ‘feedforward’ for improving each student’s reading strategies and comprehension. Students in the control group answered a traditional discrete item test at the very end of each lesson and unit. This traditional test focused mainly on the phonological, lexical and grammatical elements of the reading skill, and students were judged on the basis of how well they achieved on this test as compared to each other. The teacher only told them the right answers of the items they got wrong.

(6) Post-testing the experimental group and the control group on April 2, 2007, to measure their basic and inferential reading skills after treatment.

Findings and Discussion

The t-test was used to determine the significance of the difference in the basic and inferential reading skills between the experimental group and the control group on the post-tests. The results are shown and discussed below.

(1) Analysis and interpretation of the basic reading skills post-test data

Table (2)

The t-value of the difference in the mean scores between the experimental group and the control group on the basic reading skills post-test

Group	N	M	SD	DF	T
Experimental	32	80.25	4.44	62	-2.61
Control	32	82.41	1.46		

Table (2) shows that the mean score of the experimental group on the basic reading skills post-test was 80.25 with a standard deviation of 4.44, but the mean score of the control group on the same test was 82.41 with a standard deviation of 1.46. It also shows that the difference in the mean scores between the experimental group and the control group was statistically significant ($t=-2.61$, $p=0.01$). This result shows that classroom performance assessment was less effective in improving students' basic reading skills than traditional assessment. Therefore, the first hypothesis of the study was accepted. This finding may be attributed to two reasons. First, unlike performance assessment in which students assessed their own reading

strategies and responded to whole texts, traditional assessment focused on the recall of non-contextualized, isolated pieces of reading throughout the academic year, which could in turn help the control group students memorize more of these pieces than those of the experimental group whose attention might have been shifted from these pieces due to the holistic and process-oriented nature of performance assessment. Second, traditional assessment pushed instruction toward basic reading skills and made the teacher use “drill and skill” instruction throughout the academic year; and therefore, the control group students achieved higher scores in these skills than those of the experimental group. In line with this interpretation, the control group teacher stated that he aligned instruction with the content of the traditional test and focused on lower skills in every reading session. In support of the control group teacher's behavior, from their study on the consequences of traditional testing, Smith and Rottenberg (1991) concluded that this type of testing made teachers neglect the material that testing excludes and encouraged them to use instructional methods that resemble tests. This traditional testing-driven instruction, as I argue, could improve the control group students' achievement in the basic reading skills more than that of the experimental group without an equal gain in comprehension. In line with this argument, Shepard (1989) states that

students, who are taught to the traditional test, become good test takers and their test scores go up without a commensurate gain in performance. Along with the same argument, Neil (2003) reported cases where students, who had been taught to the traditional reading test, reached the right answers to multiple-choice questions without actually understanding what they read.

(2) Analysis and interpretation of the inferential reading skills post-test data

Table (3)

The t-value of the difference in the mean scores between the experimental group and the control group on the inferential reading skills post-test

Group	N	M	SD	DF	T
Experimental	32	4.09	0.59	62	7.75
Control	32	1.75	1.61		

Table (3) shows that the mean score of the experimental group on the inferential reading skills post-test was 4.09 with a standard deviation of 0.59, but the mean score of the control group on the same test was 1.75 with a standard deviation of 1.61. It also shows that the difference in the mean

scores between the experimental group and the control group was statistically significant ($t=7.75$, $p=0.000$). This result shows that classroom performance assessment was more effective in improving students' inferential reading skills than traditional assessment. Therefore, the second hypothesis of the study was accepted. This finding may be attributed to seven reasons. First, unlike traditional assessment in which students only recalled facts, performance assessment allowed thoughtful routes—such as making predictions before and during reading and reflecting on reading strategies and comprehension after reading—for developing and assessing higher-order thinking skills. These routes could in turn foster the experimental group students' thinking skills in general and inferential reading skills in particular. In line with this interpretation, many educators (Resnick and Resnick, 1992; Wiggins, 1993; Shohamy, 1994; Fischer and King, 1995; Newmann, 1996) assert that traditional assessment does not offer opportunities for thinking and the methods used for teaching to this type of assessment are often boring and uninspiring and deemphasize higher-order thinking skills; whereas performance assessment involves a wider spectrum of opportunities for incorporating teaching, learning and assessment with higher-order thinking skills.

Second, unlike traditional assessment which increased the control group students' test anxiety by concentrating on scores rather than learning, performance assessment decreased the experimental group students' level of test anxiety and increased their comfort zone by focusing on learning without the threat of scoring. This could encourage the experimental group students to think freely and to take risks in inferring what is between the lines while reading, thereby developing their inferential reading skills. Along with this interpretation, Sadler (1989) states that the norm-referenced grading system can give students the wrong message since it is more concerned with grades than with learning. Taras (2002) also points out that grades “have serious repercussions on learning” (p. 508).

Third, the experimental group students' self-assessment of their own reading strategies, by using part of the self-assessment checklist in every reading session, developed their awareness of the processes they go through in understanding a written text and made them aware of the reading strategies that work best for them. And as a result, they became active, strategic readers who could read inferentially and use a variety of reading strategies before, while and after reading. In line with this interpretation, Tierney, Carter and Desai (1991) state that assessment practices should

involve students if we want them to develop into independent thinkers. In support of the same interpretation, some researchers (e.g., Barnett, 1988; Carrell, 1989; Schoonen, Hulstijn and Bossers, 1998) found that awareness of reading strategies significantly predicted reading comprehension. In further support of the same interpretation, Schneider, Korkel and Weinert (1989) found that 3rd, 5th and 7th grade students who were better able to use metacognitive strategies were also significantly better able to make inferences.

Fourth, self-assessment might have increased the experimental group students' self-confidence and raised their feeling of accomplishment which could in turn encourage them to take risks and read thoughtfully. This interpretation is supported by Biondi (2001) who found that self-assessment resulted in higher self-confidence, higher self-esteem and better achievement. On the other hand, the anxiety-generating nature of traditional assessment might have negatively affected the control group students' self-image and threatened their self-esteem which could in turn lead them to concentrate on passing the test rather than learning. Along with this reasoning, Paris, Lawton, Turner and Roth (1991) found that as students got older they felt "greater resentment, anxiety, cynicism, and

mistrust of standardized achievement tests" (p. 16). Smith and Rottenberg (1991) also found that teachers believed that traditional tests "cause stress, frustration, burnout, fatigue, physical illness, misbehavior and fighting, and psychological distress" (p. 10).

Fifth, the KWL chart helped the experimental group students to be active thinkers by having them relate their prior knowledge to the information in the text and reflect on what they read. With this emphasis on the learner's prior knowledge rather than the teacher's and on the active construction of knowledge rather than the passive receipt of information, the experimental group students became independent thinkers and developed their inferential reading skills, which require knowledge of the world rather than knowledge of words. In line with this interpretation, Harvey and Goudvis (2000) in their book, *Strategies that Work*, state: "When children [or adults] understand how to connect the text they read to their lives, they begin to make connections between what they read and the larger world. This nudges them into thinking about bigger, more expansive issues beyond their universe of home, school and neighborhood" (p. 68). In support of the same interpretation, Carr (1991) found that content schema activation developed the inferential reading comprehension skills of students with learning

disabilities. In contrast, students in the traditional assessment group, as I argue, accepted all what they were told as facts without activating their own prior knowledge, which could in turn stifle their thinking in general and suffocate their inference generation in particular.

Sixth, the nonthreatening interactive nature of portfolio assessment might have reduced students' reading anxiety, which could in turn encourage them to use global reading strategies, thereby thinking of what is between the lines. In contrast, the fear of being judged on the basis of scores might have increased the control group students' reading anxiety and pushed them to use local strategies, which could in turn standardize their minds and hamper their thinking. In line with this interpretation, Monteiro (1992) found that, for both reading in the L1 and L2, poor readers tended to be more local in their perception of effective reading strategies compared to better readers, and the less readers perceived local strategies as effective strategies, the higher their reading ability. Along with the same interpretation, Sellers (2000) found that highly anxious readers used more local strategies, such as focusing on vocabulary, grammar and translation; whereas less anxious readers approached the text more holistically.

Finally, unlike traditional assessment which did not offer opportunities of discovery into what learners did when they were reading or what problems they faced when they were failing to understand, performance assessment opened windows of discovery into what learners did when they were reading and where their reading strategies were strong or weak, and then provided feedback and support for improving the weak ones. This could in turn help the experimental group students read strategically and compensate for their linguistic inadequacies. In support of this reasoning, Carrell, Pharis and Liberto (1989) found that good second language readers compensated for a lack of language proficiency by using reading strategies during reading to make sense of the reading text.

In summery, the results of this study agree with what Lauren Resnick (cited in Wiggins, 1990, p. 5) says: “What you assess is what you get; if you don’t test it, you won’t get it. To improve student performance we must recognize that essential intellectual abilities are falling through the cracks of conventional testing.” They are also in line with Shepard et al.'s (1995) conclusion that "performance assessments are a key element in instructional reform, but they are not by themselves an easy-cure all" (p. 27).

Implications for Assessment, Teaching and Learning

This study provides direct evidence that traditional assessment does help first-year secondary school EFL students improve their basic reading skills, and performance assessment does help them develop their inferential reading skills. This indicates that both types of assessment are complementary and that one type cannot significantly improve both basic and inferential reading skills, nor can it be responsive to individual differences. Therefore, a multi-dimensional comprehensive approach, that encompasses both traditional and performance assessments, is more likely to improve intermediate-level EFL students' basic and inferential reading skills. This implication is in line with Smith and Levin's (1996) contention that "no single type of assessment can always meet all purposes, in all situations," therefore, the solution, as they argue, is to "make the best use possible of various assessment strategies in order to meet the diverse criteria of and purposes for the overall assessment" (p. 111). The same implication is consistent with Lane and Stone's (2006) notion that performance and traditional assessments should be combined to capitalize on the advantages of each type as follows:

Performance assessment tasks ... [should be] combined with multiple-choice items in assessments to capitalize on the

advantages of each type of approach. Performance assessment tasks, for example, offer the potential for more direct assessment, more complex items and more response information. Multiple-choice items, for example, offer the potential for more domain coverage, thus yielding higher reliability and more precise individual-level scores. An assessment that combines these different item formats offers the potential for more direct assessment, more complex items, more response information, and at the same time adequate domain coverage and high reliability for individual-level scores. (p. 417)

With respect to reading, the results of the study indicate that inferential reading comprehension is not simply a decoding activity, but an interactive process between the reader's background content knowledge and the text. Therefore, it requires activation of prior content knowledge and a transaction between the reader and information in the text through employing a wide range of strategies before, while and after reading. When this occurs, the reader can draw successful inferences related to the text. In line with this implication, Anderson, Reynolds, Schallert and Goetz (1977) state that “every act of comprehension involves one's knowledge of the

world” (p. 369). Along with the same implication, Aebersold and Field (1997) state: “If the topic...is outside [students’] experience or base of knowledge, they are adrift to an unknown sea” (p. 41).

The results of the study also suggest that focusing on basic skills out of context does not lead to inferential reading comprehension improvement because such isolated skills remain in isolation and cannot compensate for students' lack of content knowledge. In other words, the basic blocks of reading are not enough for constructing meaning from the text and inferring what is between the lines because readers create meaning and make inferences depending on their prior content knowledge and on the strategies they employ to activate and connect this knowledge to the text they are reading. Therefore, one cannot expect students to think inferentially if they do not have enough prior content knowledge to base their thinking on. In support of this implication, some researchers found that content schema was more important for reading comprehension than formal and linguistic schemata. Freebody and Anderson (1983), for example, found that familiar text content aided comprehension more than familiar vocabulary. Nunan (1985), for a second example, found that the text which was linguistically easier but with unfamiliar content seemed to

be significantly more difficult to comprehend than the text that was linguistically more difficult but with more familiar content. Taft and Leslie (1985), for a third example, found that third grade children with high prior content knowledge could comprehend up to 75% of the texts that were at a 5th-6th grade readability level and concluded that readers with high background content knowledge can not only read better, but also comprehend beyond what is considered their normal reading level. Carrell (1987), for a fourth example, found that unfamiliar content schema negatively affected reading comprehension to a greater extent than unfamiliar formal schema and that reading familiar content even in an unfamiliar rhetorical form was relatively easier than reading unfamiliar content in a familiar rhetorical form. Moreover, of particular importance for foreign language students, Keshavarz, Atai and Ahmadi (2007) found that content schema had a greater effect than linguistic simplification on both reading comprehension and recall.

The results of the studies mentioned above are in line with the implication that prior content knowledge plays a more significant role in reading comprehension than linguistic knowledge because readers can compensate for their linguistic deficiencies by guessing the general meaning according

to assumptions derived from their content schema, but not the reverse. However, this does not mean that linguistic knowledge is not necessary for reading comprehension, but it is not enough for achieving a higher level of comprehension. The experimental group students in the present study reached a higher level of reading comprehension than the control group ones not only because they activated their own content schema and responded to whole texts, but also because they had a threshold level of foundational reading skills before the beginning of the study. This in turn enabled them to use global reading strategies to read strategically and inferentially. The implication here is that a certain amount of linguistic competence is needed before applying performance assessment particularly in the initial stage of learning a foreign language. In line with this implication Takahashi and Beebe (1987, cited in Ellis, 1994, p. 181) state that "learners may need to reach a threshold level of linguistic proficiency before pragmatic transfer can take place." In support of the same implication, Smith et al. (1997) reported from their study that nearly two thirds of teachers believed that pupils "need to master basic skills before they can progress to higher order thinking and problem solving" (p. 41). Also, in Feinberg's (1990) opinion, it is important that students acquire a foundation of basic skills on which to build their thinking skills. However,

this, as I argue, does not mean delaying higher order reading skills until students master an advanced level of basic skills, but a minimum level of these skills can serve the purpose of developing higher-order comprehension skills because thinking activities can simultaneously develop the basic reading skills, at least to some extent; but the basic skills drills do hamper the development of thinking skills in general. In support of this implication, subordinate data analysis of the pretest and posttest scores of the present study, using the paired samples t-test, showed that the mean scores of the experimental group on the basic reading skills post-test were higher than those of the pre-test, though statistically insignificant ($t= 1.75$, $df= 31$, $p= 0.09$); and the mean scores of the control group on the inferential reading skills posttest remained nearly the same as those of the pretest ($t= 0.37$, $df= 31$, $p= 0.71$). The same implication is supported by Rodgers, Paredes and Mangino's (1991) study, in which they looked at the effects of the Texas Educational Assessment of Minimum Skills (TEAMS) on high school students' basic and higher-order thinking skills. The study took place over five years, using 12,404 eleventh grade students from Austin Independent School District. The test focused on language arts and math. Rodgers et al. found that the basic skills, as measured on the Tests of Achievement and Proficiency (TAP), increased as a result of the minimum

competency test, but higher-order thinking skills remained the same. They concluded that districts should be cautious about narrowing the curriculum and letting higher order skills suffer for the sake of improving test scores. In further support of the same implication, Amrein and Berliner (2002) examined data from 18 states of America, that implemented traditional high-stakes testing, to assess whether students gained any knowledge that they could apply elsewhere other than learning the necessary facts for doing a state's high-stakes test. From the data analysis they concluded: "[I]f the intended goal of high stakes testing policy is to increase student learning, then that policy is not working. While a state's high stakes test may show increased scores, there is little support ... that such increases are anything but the result of test preparation and/or the exclusion of [low proficient] students from the testing process" (p. 2).

Recommendations

In light of the results of the study, the researcher recommends a comprehensive classroom assessment approach, which encompasses students' learning processes and products and treats assessment as part of the teaching/learning process, to provide both teachers and students with ongoing information to adjust teaching and learning accordingly. This

approach should utilize different kinds of traditional and performance assessment formats—such as multiple-choice items, open-ended response questions, reading conferences, classroom discussions, role playing, interviews, KWL charts, self-assessment checklists, and portfolios—to improve students basic and higher-order thinking skills and to support validity and increase reliability.

Just as we need a link among teaching, learning and assessment, so too, do we need a link between classroom formative assessment and external summative assessment. None of them should constitute the sole basis for assessing students' learning particularly when making critically important decisions for grade-level promotion and graduation. In this respect, the researcher recommends that the portfolio, in which the teacher keeps the student's classroom assessments throughout the academic year, should make up 50% of the final grade. This portfolio should be read by the class teacher every week to diagnose each student's strengths and weaknesses and suggest remedies for her/his weaknesses, and by a jury of raters in the end of the academic year to score it blindly in terms of standardized rubrics. These rubrics should be developed by assessment specialists in collaboration with teachers and students to be uniformly used by raters all

over the country to make sure that scoring is reliable and fair. In addition, external summative assessment, which is still necessary to ensure uniformity of content and complete coverage of all domains within the curriculum, should make up the other 50% of the final grade and be reformed to include higher-order thinking tasks.

EFL teachers should receive training in classroom performance assessment as a necessary prerequisite for the use of this type of assessment in schools. They should be informed of the purposes and advantages of this new type of assessment to shift their mindset from ‘assessment of learning’ to ‘assessment for learning.’ They also need practical training in the development and implementation of the various formats of this type of assessment as well as the ways in which to give and take feedback based on classroom assessment information, without assigning grades to responses so as not to lead students to concentrate on passing the test rather than learning. They should also be provided with training on the use of standardized rubrics for scoring students performance in all language arts.

Before the implementation of performance assessment, school administrators should provide school libraries with books that assist

teachers in the development of performance tasks and other books that excite students to read extensively to build their own background knowledge. They should also offer facilities that aid performance assessment such as tape recorders, videos, computers, and the Internet. In addition, they should develop cooperative structures that lead teachers to work cooperatively to achieve the goal of performance assessment.

Curriculum developers should take performance assessment formats into their own consideration during the process of developing English language curricula. Lessons should involve activities that are amenable to classroom performance assessment such as project-based learning, role-playing, journal writing, and classroom discussion. They should also know that performance assessment requires authentic materials and authentic methods of learning and instruction; and that learning, instruction and assessment should occur simultaneously.

In order to help students with higher-order reading comprehension difficulties, teachers should know the problems these students encounter during their reading process and help them overcome these problems by modeling the effective reading strategies for them, including inference-

making strategies. Teachers should also make poor readers aware of the local strategies they employ while reading and encourage them to use global reading strategies instead of them. In addition, they should build their students' background content knowledge and invite them to use strategies that activate this knowledge before, during and after reading.

The public needs to be informed of the benefits of performance assessment to obtain their support for the inclusion of this new type of assessment in large-scale testing and to make them abandon their traditional notions about testing. Lastly, policy makers should bear in mind that classroom performance assessment is a must in the information age if we want to be exporters of inventions rather than importers.

Suggestions for Future Research

Researchers are invited to investigate the effect of a multi-dimensional comprehensive approach, that encompasses both traditional and performance assessments, on students' higher-order reading comprehension skills at various proficiency levels. They are also invited to investigate the effect of other classroom performance assessment formats than those used in the present study (e.g., dialogue journals, interviews,

conferences and learning logs) on students' higher-order reading comprehension skills, and to replicate the present study with different types and levels of students for longer periods of time. Finally, research is needed to investigate the interrelationships among reading strategy awareness, reading strategy use, and reading comprehension below and above the normal linguistic level.

References

- Aebersold, J., and Field, M. (1997). *From reader to reading teacher*. Cambridge: Cambridge University Press.
- Amrein, A. L., and Berliner, D. C. (2002). *The impact of high-stakes tests on student academic performance: An analysis of NAEP results in states with high-stakes tests and ACT, SAT, and AP test results in states with high school graduation exams*. Retrieved February 3, 2007, from <http://www.asu.edu/educ/epsl/EPRU/documents/EPsl-0211-126EPRU.pdf>.
- Anderson, R. C., Reynolds, R. E., Schallert, D. L., and Goetz, E. T. (1977). Frameworks for comprehending discourse. *American Educational Research Journal*, 14, 367-381.
- Barnett, M. A. (1988). Reading through context: How real and perceived strategy use affects L2 comprehension. *Modern Language Journal*, 72, 150-162.
- Biondi, L. A. (2001). *Authentic assessment strategies in fourth grade*. Available at: <http://www.ERIC.ed.gov>. [ED460165]
- Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Maidenhead, Berkshire, UK: Open University Press.

- Borko, H., Flory, M., and Cumbo, K. (1993). *Teachers' ideas and practices about assessment and instruction: A case study of the effects of alternative assessment in instruction, student learning, and accountability practice* (CSE Technical Report 366). Los Angeles: Center for Research on Evaluation, Standards, and Student Testing.
- Burke, K. (2005). *The mindful school: How to assess thoughtful outcomes* (4th Ed.). Thousand Oaks: Corwin Press.
- Calvo, M. G., and Eysenck, M. W. (1996). Phonological working memory and reading in test anxiety. *Memory*, 4(3), 289-307.
- Carr, S. (1991). *The effects of prior knowledge and schema activation strategies on the inferential reading comprehension performance of learning disabled and nonlearning disabled children*. Available at: <http://www.ERIC.ed.gov>. [ED329924]
- Carrell, P. L. (1987). Content and formal schemata in ESL reading. *TESOL Quarterly*, 21, 461-481.
- Carrell, P. L. (1989). Metacognitive awareness and second language reading. *Modern Language Journal*, 73, 121-134.
- Carrell, P. L., and Eisterhold, J. C. (1983). Schema theory and ESL reading pedagogy. *TESOL Quarterly*, 17, 553-569.
- Carrell, P. L., Pharis, B. G., and Liberto, J. C. (1989). Metacognitive

- strategy training for ESL reading. *TESOL Quarterly*, 23, 647-678.
- Cizek, G. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 20(4), 19-27.
- Cross, B., Greer, T., and Pearce, M. (1998). *Improving student reading comprehension skills through the use of authentic assessment*. Chicago, Illinois: Saint Xavier University, IRI, Skylight Publishing Inc.
- DES/WO (1988). *National Curriculum Task Group on Assessment and Testing: A report*. London: DES/WO.
- Dunne, M. K. (1996). *Methods for measuring student growth in reading and writing*. Available at: <http://www.ERIC.ed.gov>. [ED398541]
- El-Koumy, A. A. (2004). *Language performance assessment: Current trends in theory and research*. Available at: <http://www.ERIC.ed.gov>. [ED490574]
- El-Koumy, A. A. (2006). *The effects of the directed reading-thinking activity on EFL students' referential and inferential comprehension*. Available at: <http://www.ERIC.ed.gov>. [ED502645]
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Fabrikant, W., Siekierski, N., and Williams, C. (1999). *Improving students' inferential and literal reading comprehension*. Masters Action

Research Project. Chicago, IL: Saint Xavier University and IRI/Skylight.

Falk, B., and Darling-Hammond, L. (1993). *The primary language record at P.S. 261: How assessment transforms teaching and learning.* New York: National Center for Restructuring Education, Schools, and Teaching.

Feinberg, L. (1990). Multiple choice and its critics: Are the alternatives any better? *The College Board Review*, 157, 13-31.

Fischer, D. F., and King, R. M. (1995). *Authentic assessment: A guide to implementation.* Thousand Oaks: Corwin Press.

Freebody, P., and Anderson, R. C. (1983). Effects on text comprehension of differing proportions and locations of difficult vocabulary. *Journal of Reading Behavior*, 15, 19-39.

French, D. (2003). A new vision of authentic assessment to overcome the flaws in high stakes testing. *Middle School Journal*, 35(1), 1-13.

Fuller, E. J., and Johnson, J. F., Jr. (2001). Can state accountability systems drive improvements in school performance for children of color and children from low income homes? *Education and Urban Society*, 33(3), 260-283.

Gagliano, K., and Swiatek, L. (1999). *Improving student assessment through*

the implementation of portfolios in language arts. Chicago, Illinois:
Saint Xavier University and IRI Skylight Publishing Inc.

Gaynor, J., and Millham, J. (1976). Student performance and evaluation under variant teaching and testing methods in a large college course. *Journal of Educational Psychology*, 68, 312-317.

Geocaris, C., and Ross, M. (1999). A test worth taking. *Educational Leadership*, 57(1), 29-33.

Glover, J. A., Zimmer, J. W., and Bruning, R. H. (1979). Utility of the Nelson-Denny as a predictor of structure and themanticity in memory for prose. *Psychological Reports*, 45, 44-46.

Haertel, E., and Mullis, I. (1996). The evolution of the national assessment of educational progress: Coherence with best practice. In Joan B. Baron and Dennie P. Wolf (Eds.), *Performance-based student assessment: Challenges and possibilities* (pp. 287-304). Chicago, Illinois: The University of Chicago Press.

Hammond, L. D., Aness, J., and Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work.* New York: Teachers College Press.

Hart, D. (1994). *Authentic assessment: A handbook for educators.* Menlo Park, CA: Addison-Wesley.

- Harvey, S., and Goudvis, A. (2000). *Strategies that work: Teaching comprehension to enhance understanding*. Portland, ME: Stenhouse.
- Hasman, M. A. (2000). The role of English in the 21st century. *English Teaching Forum Online*, 38(1), 1-2. Available at: <http://exchanges.state.gov/forum/vols/vol38/no1/index.htm>.
- Herman, J. (2004). The effects of testing on instruction. In Susan H. Fuhrman and Richard F. Elmore (Eds.), *Redesigning accountability systems for education* (pp. 141-166). New York: Teachers College Press.
- Herman, J., and Golan, S. (1991). *Effects of standardized testing on teachers and learning—another look* (CSE Technical Report 334). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing, University of California.
- Johnson, J. (1989)...Or none of the above. *The Science Teacher*, 56 (4), 57-61.
- Kellaghan, T., Madaus, G., Airasian, P. (1982). Effects of Testing on pupils. In T. Kellaghan (Ed.), *The effects of standardized testing* (pp. 131-150). Boston: Kluwer-Kijhoff Publishing.
- Keshavarz, M. H., Atai, M. R., and Ahmadi, H. (2007). Content schemata, linguistic simplification, and EFL readers' comprehension and recall.

Reading in a Foreign Language, 19, 19-33.

Khattri, N., Kane, M. B., and Reeve, A. L. (1995). How performance assessments affect teaching and learning. *Educational Leadership*, 53(3), 80-83.

Khattri, N., Reeve, A., and Kane, M. (1998). *Principles and practices of performance assessment*. Mahwah, NJ: Lawrence Erlbaum.

Kim, S. (2003). The effect of authentic assessment strategy on students achievement in a constructivist classroom. In A. Rossett (Ed.), *Proceedings of world conference on E-Learning in corporate, government, healthcare, and higher education* (pp. 257-260). Chesapeake, VA: AACE.

Koretz, D., and Barron, S. (1998). *The validity of gains on the Kentucky Instructional Results Information System (KIRIS)*. Santa Monica: CA, RAND.

Koretz, D., Barron, S., Mitchell, K., and Stecher, B. (1996). *Perceived effects of the Kentucky Instructional Results Information System (KIRIS)*. Washington, DC: Institute on Education and Training, RAND.

Koretz, D., and Hamilton, L. (2006). Testing for accountability in K-12. In Robert L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 531-578). Westport, CT: American Council on Education Praeger.

- Koretz, D., Linn, R. L., Dunbar, S. B., and Shepard, L. (1991, April). *The effects of high-stakes testing on achievement: Preliminary findings about generalization across tests*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- Koretz, D., Mitchell, K., Barron, S., and Keith, S. (1996). *The perceived effects of the Maryland school performance assessment program* (CSE Technical Report 409). Los Angeles: University of California, Center for the Study of Evaluation.
- Koretz, D., Stecher, B., Klein, S., and McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice*, 13(3), 5-16.
- Koretz, D., Stecher, B., Klein, S., McCaffrey, D., and Diebert, E. (1993). *Can portfolios assess student performance and influence instruction? The 1991–1992 Vermont Experience* (CSE Technical Report 371). Los Angeles: CRESST.
- Lane, S., and Stone, C. (2006). Performance assessment. In Robert L. Brennan (Ed.), *Educational measurement*. (4th ed., pp. 387-432). Westport, CT: Praeger Publishers.
- Lavande, D. (1993). Standardized reading tests: Concerns, limitations and alternatives. *Reading Improvement*, 30(2), 125-127.

- Levinson, C. Y. (2000). Student assessment in eight countries. *Educational Leadership*, 57(5), 58-61.
- Linn, R., and Baker, E. (1996). Can performance-based assessments be psychometrically sound? In Joan B. Baron and Dennie P. Wolf (Eds.), *Performance-based student assessment: Challenges and possibilities* (pp. 84-103). Chicago, Illinois: The University of Chicago Press.
- Linn, R., Baker, E., and Dunbar, S. (1991). Complex performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15-21.
- Linn, R., and Herman, J. (1997). *A policymaker's guide to standards-led assessment*. Denver: ECS Distribution Centre.
- Liskin-Gasparro, J. (1997). *Testing in an age of assessment: Theoretical and practical considerations*. Plenary address. University of Texas, Spanish Second Language Acquisition Symposium, Austin, TX.
- McNamara, T. F. (2000). *Language testing*. Oxford: Oxford University Press.
- McNeil, L. M., and Valenzuela, A. (2000). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. In G. Orfield and M. Kornaber (Eds.), *Raising standards or raising barriers? Inequality and high stakes testing in public education* (pp.

127-150). New York: Century Foundation.

Mehrens W. A., and Kaminski, J. (1989). Methods for improving standardized test scores: Fruitful, fruitless, or fraudulent? *Educational Measurement: Issues and Practice*, 8(1), 14-22.

Monteiro, S. Q. (1992). *A contrastive investigation of 'reading strategy awareness' and 'reading strategy use' by adolescents reading in the first language (Portuguese) and in the foreign language (English)*. Unpublished Doctoral Dissertation, University of Essex.

Moon, T. R., Brighton, C. M., and Callahan, C. M. (2003). State standardized testing programs: Friend or foe of gifted education? *Roeper Review*, 25(2), 49-60.

Neil, M. (2003). High stakes, high risk: The dangerous consequences of high-stakes testing. *American School Board Journal*, 190(2), 18-21.

Neil, M., and Medina, N. J. (1989). Standardized testing: Harmful to educational health. *Phi Delta Kappan*, 70, 688-697.

Newmann, F. M. (1996). *Authentic achievement: Restructuring schools for intellectual quality*. San Francisco, CA: Jossey-Bass Inc.

Newmann, F. M., Marks, H. M., and Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education*, 104(4), 280-312.

- Nicol, D., and Owen, C. (2008). *Formative assessment and feedback as drivers for transformational change: Evidence of learning and workload gains*. Paper presented at the 33rd International Conference of Improving University Teaching, July 29-August 1, 2008. University of Strathclyde, Glasgow, Scotland.
- Nunan, D. (1985). Content familiarity and the perception of textual relationships in second language reading. *RELC Journal*, 16, 43-50.
- Nungester, R., and Duchastel, P. (1982). Testing vs. review: Effects on retention. *Journal of Educational Psychology*, 74, 18-22.
- O'Malley, J., and Valdez Pierce, L. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. Boston, MA: Addison-Wesley Publishing Company.
- Paris, S., Lawton, T., Turner, J., and Roth, J. (1991). A developmental perspective on standardized achievement testing. *Educational Researcher*, 20(5), 12-20.
- Pendulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., and Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers*. Boston: National Board on Educational Testing and Public Policy.
- Popham, W. J. (2004). *Classroom assessment: What teachers need to know*

(4th ed.). Boston, MA: Allyn and Bacon.

Resnick, L. B., and Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford and M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37-75). Boston: Kluwer Academic Publishers.

Rhine, S., and Smith, E. (2001). *Appropriate assessment of primary grade students*. Available at: <http://www.ERIC.ed.gov>. [ED456128]

Roderick, M., and Engel, M. (2001). The grasshopper and the ant: Motivational responses of low achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis*, 23(3), 197-228.

Rodgers, N., Paredes, V., and Mangino, E. (1991). *High stakes minimum skills tests: Is their use increasing achievement?* Paper presented at the annual meeting of the American Education Research Association, Chicago. (ERIC Document Reproduction Service No. ED336422).

Rudner, L., and Boston, C. (1994). Performance assessment. *ERIC Review*, 3(1), 2-12.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144.

Schneider, W. Korkel, J., and Weinert, F. (1989). Domain-specific

- knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology*, 81(3), 116-127.
- Schoonen, R., Hulstijn, J., and Bossers, B. (1998). Metacognitive and language-specific knowledge in native and foreign language reading comprehension: An empirical study among Dutch students in grades 6, 8 and 10. *Language Learning*, 48(1), 71-106.
- Sellers, V. D. (2000). Anxiety and reading comprehension in Spanish as a foreign language. *Foreign Language Annals*, 33, 512-521.
- Shepard, L. A. (1989). Why we need better assessments. *Educational Leadership*, 46, 4-9.
- Shepard, L. A. (1991). Psychometricians' beliefs about learning. *Educational Researcher*, 20(6), 2-16.
- Shepard, L. A. (1992). *Will national tests improve student learning?* (CSE Technical Report 342). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.
- Shepard, L. A. (2000). *The role of classroom assessment in teaching and learning* (CSE Technical Report 517). Los Angeles: CRESST/University of Colorado at Boulder.
- Shepard, L. A., and Dougherty, K. (1991, April). *Effects of high-stakes*

testing on instruction. Paper presented at the annual meeting of the American Education Research Association and the National Council on Measurement in Education, Chicago.

Shepard, L. A., Flexer, R. J., Hiebert, E. H., Marion, S. F., Mayfield, V., and Weston T. J. (1995). *Effects of introducing classroom performance assessments on student learning*. Los Angeles, CA: CRESST, University of California.

Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing*, 11(2), 99–123.

Skinner, B. (1954). The science of learning and the art of teaching. *Harvard Educational Review*, 24, 86-97.

Skrla, L., and Scheurich, J. (2001). Displacing deficit thinking in school district leadership. *Education and Urban Society*, 33(3), 235-259.

Smith, M., Edelsky, C., Draper, K., Rottenberg, C., and Cherland, M. (1991). *The role of testing in elementary schools*. Los Angeles: University of California.

Smith, M., and Levin, J. (1996). Coherence, assessment, and challenging content. In Joan B. Baron and Dennie P. Wolf (Eds.), *Performance-based student assessment: Challenges and possibilities* (pp. 104-124). Chicago, Illinois: The University of Chicago Press.

- Smith, M., Noble, A., Heinecke, W., Seek, M., Parish, C., Cabay, M., Junker, S., Haag, S., Tayler, K., Safran, Y., Penley, Y., and Bradshaw, A. (1997). *Reforming schools by reforming assessment: Consequences of the Arizona student assessment program* (CSE Technical Report No. 425). Los Angeles: University of California.**
- Smith, M., and Rottenberg, C. (1991). Unintended consequences of external testing in elementary schools. *Educational Measurement: Issues and Practice*, 10(4), 7-11.**
- Smith, M., and Shepard, L. (1989). Flunking grades: A recapitulation. In M. Smith and L. Shepard (Eds.), *Flunking grades: Research and policies on retention* (pp. 214-236). New York: Falmer Press.**
- Stallworth-Clark, R., Cochran, J., Nolen, M., Tuggle, D., and Scott, J. (2000). Test anxiety and performance on reading competency tests. *Research and Teaching in Developmental Education*, 17(1), 93-47.**
- Stecher, B. M., and Mitchell, K. J. (1995, April). *Portfolio-driven reform: Vermont teachers' understanding of mathematical problem solving and related changes in classroom practice* (CSE Technical report 400). National Center for Research on Evaluation, Standards, and Student Testing (CRESST), Graduate School of Education and Information Studies, University of California, Los Angeles.**

- Stiggins, R. J. (1993). Teacher training in assessment: Overcoming the neglect. In S. L. Wise (Ed.), *Teacher training in assessment and measurement skills* (pp. 27-40). Lincoln, NE: Buros Institute of Mental Measurements.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment *FOR* learning. *Phi Delta Kappan*, 83(10), 758-765.
- Stiggins, R. J., Arter, J., Chappuis, J., and Chappuis, S. (2004). *Classroom assessment for student learning: Doing it right—using it well*. Portland: Assessment Training Institute.
- Supovitz, J. A. (2001). Translating teaching practice into improved student achievement. In S.H. Fuhrman (Ed.). *From the capitol to the classroom: Standards-based reform in the states* (pp. 81-98). Chicago: The University of Chicago Press.
- Taft, M., and Leslie, L. (1985). The effects of prior knowledge and oral reading accuracy on miscues and comprehension. *Journal of Reading Behavior*, 17(2), 163-179.
- Takahashi, T., and Beebe, L. (1987). The development of pragmatic competence by Japanese learners of English. *JALT Journal*, 8, 131-155.
- Taras, M. (2002). Using assessment for learning and learning from

- assessment. *Assessment and Evaluation in Higher Education*, 27(6), 501-510.
- Taylor, G., Shepard, L., Kinner, F., and Rosenthal, J. (2003). *A survey of teachers' perspectives on high-stakes testing in Colorado: What gets taught, what gets lost*. Los Angeles: University of California.
- Tierney, R., Carter, M., and Desai, L. (1991). *Portfolio assessment in the reading-writing classroom*. Norwood, MA: Christopher-Gordon Publishers.
- Tilton, P. (1996). Relationship between teacher commitment to performance-based assessment and student achievement on the optional reading and writing measures in fourth-grade English-language arts of the California Learning Assessment System (CLAS). *DAI-A*, 57(3), p. 1110.
- Valencia, S. W. (1994). *Authentic reading assessment: Practices and possibilities*. DE: International Reading Association.
- Van Horn, R. (1997). Improving standardized test scores. *Phi Delta Kappan*, 78(7), 584-590.
- Vining, A., and Bell, S. (2005). The impact of teaching multiple-choice strategies on test scores of eighth grade students: A collaborative action research project. *TAMS Journal*, 32, 3-11.

- Ward, L., and Traweek, D. (1993). Application of a metacognitive strategy to assessment, intervention, and consultation: A think-aloud technique. *Journal of Psychology*, 31, 469-485.
- Wiggins, G. (1990). *The case for authentic assessment*. Available at: <http://www.ERIC.ed.gov>. [ED328611]
- Wiggins, G. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, California: Jossey-Base Publishers.
- Womer, F. (1984). Where's the action? *Educational Measurement: Issues and Practice*, 3(3), 3.
- Wood, N. (1988). Standardized reading tests and the postsecondary reading curriculum. *Journal of Reading*, 32(3), 224-230.

Appendix A

The KWL Chart

What I Know about the Topic of the Text	What I Want to Know	What I Learned from Reading the Text

Appendix B

The Self-Assessment Checklist for Assessing Reading Strategies

Directions: The purpose of this self-assessment checklist is to help you identify the reading strategies that work best for you. It makes you aware of the strategies you employ as well as their effects on your comprehension. It also invites you to experiment with other strategies until you find the ones that work best for you.

This checklist consists of three parts. The first part aims to help you recognize and assess pre-reading strategies. The second and third parts aim to help you recognize and assess while and after reading strategies, respectively. For time restriction, you should use the three parts sequentially and circularly, one per session.

Part I

The Self-Assessment Checklist for Assessing Pre-Reading Strategies

Student Name:-----.

Lesson: -----.

Date: / /

Directions: The purpose of this part of the self-assessment checklist is to help you identify the strategies that work best for you before reading. Put a tick in the box to the left of each strategy you employed before reading in the present session and in the box that indicates the extent to which the strategies you employed helped you understand what you read. In light of your self-assessment, experiment with other strategies in the next sessions until you find the ones that work best for you.

1. Before reading,

- I read the title out loud to myself.
- I analyzed the wording of the title.

- I translated the title word by word to my mother tongue.**
- I looked at the length of the text to estimate the time I will take to finish reading it.**
- I looked at the outer text organization structure.**
- I visualized the title in my mind.**
- I predicted what the content would be in reaction to the title.**
- I looked over the pictures and diagrams in the text.**
- I skimmed the text quickly to get its gist.**
- I read the first and last paragraphs of the text.**
- I activated my background knowledge related to the title by filling in the "K" and "W" columns on the KWL chart.**
- I asked myself questions that can be answered by the text.**

2. I think the pre-reading strategies I employed in this session helped me in understanding what I read -----.

- To a very little extent**
- To a little extent**
- To a moderate extent**
- To a great extent**
- To a very great extent**

Part II

The Self-Assessment Checklist for Assessing While-Reading Strategies

Student Name:-----.

Lesson: -----.

Date: / /

Directions: The purpose of this part of the self-assessment checklist is to help you identify the strategies that work best for you while reading. Put a tick in the box to the left of each strategy you employed during reading in the present session and in the box that indicates the extent to which the strategies you employed helped you understand what you read. In light of your self-assessment, experiment with other strategies in the next sessions until you find the ones that work best for you.

1. While reading,

I read word by word.

- I perceived more than one word at a time.**
- I used context clues to help me understand unfamiliar words.**
- I analyzed unfamiliar words into roots, prefixes and suffixes to determine their meanings.**
- I used a bilingual dictionary to get the Arabic meaning of each word.**
- I guessed the meaning of the unknown words from the context.**
- I dissected sentences into parts to understand their meanings.**
- I answered the questions I generated prior to reading.**
- I made up additional questions and looked for answers to them.**
- I checked and revised the predictions I formulated prior to reading.**
- I created graphic organizers to help me collect thoughts from the text.**
- I created semantic maps to help me identify the relationships among ideas in the text.**
- I transformed what I read into a graphic organizer to make connections among ideas.**
- I focused on the logical sequence of information in the text.**
- I related new information to visual concepts in my memory.**
- I inferred implicit ideas based on my prior knowledge.**
- I made inferences about implicit details based on my prior knowledge.**
- I drew meanings from pictures and other visuals in the text.**

- I formulated mental images of the ideas in the text.**
- I took notes on the margin.**
- I summarized the text in my own words.**
- I visualized what I read.**
- I focused on the overall meaning of the text.**
- I highlighted important ideas in the text with colors.**
- I underlined important ideas in the text.**
- I anticipated what would come next.**
- I questioned the text and argued with it.**
- I made an inference about the author's purpose (persuade, inform, or entertain) based on evidence from the text.**
- I made an inference about the author's tone (neutral, irritated, amused, surprised, disgusted, sad, or suspicious) based on evidence from the text.**
- I skipped the parts I did not understand.**
- I reread the parts I did not understand.**
- I reread the parts that came before and after the problematic ones I did not understand.**
- I made a connection between information in the text and my prior knowledge when the meaning was lost.**

2. I think the while-reading strategies I employed in this session helped me in understanding what I read -----.

- To a very little extent**
- To a little extent**
- To a moderate extent**
- To a great extent**
- To a very great extent**

Part III

The Self-Assessment Checklist for Assessing Post-Reading Strategies

Student Name:-----.

Lesson: -----.

Date: / /

Directions: The purpose of this part of the self-assessment checklist is to help you identify the strategies that work best for you after reading the text. Put a tick in the box to the left of each strategy you employed after reading in the present session and in the box that indicates the extent to which the strategies you employed helped you in complimenting and deepening your understanding of what you read. In light of your self-assessment, experiment with other strategies in the next sessions until you find the ones that work best for you.

1. After reading,

I made a list of the key words I learned from the text to fix them in my

memory.

- I discussed the text information with classmates to confirm my comprehension.**
- I thought about what else I need to know about the topic of the text.**
- I filled in the “L” column of the KWL chart to consolidate information learned from the text.**
- I summarized the overall meaning of the text orally or in written form.**
- I expanded what I read in writing.**
- I evaluated the author’s tone/attitude in the text.**
- I evaluated the underlying message of the text.**
- I recited text information aloud to myself to fix it in my memory.**
- I discussed the author’s line of reasoning with colleagues.**
- I judged the author’s word choice and how it advanced the theme of the text.**
- I compared and contrasted different points of view in the text.**
- I decided whether the text is useful to me or to other readers.**
- I responded to open-ended questions to consolidate information learned from the text.**
- I made judgments about the author’s cultural, racial/ethnic, linguistic, socioeconomic, and gender biases based on evidence from the text.**

I thought of the possible consequences of what I read.

2. I think the post-reading strategies I employed in this session helped me in complimenting and deepening my understanding of what I read -----.

To a very little extent

To a little extent

To a moderate extent

To a great extent

To a very great extent