Basic Considerations in Writing Instruction & Assessment

[1]Ali Asghar Yousefi Azarfam (Corresponding Author), [2] Seyed Ali Rezvani Kalajahi

[1] Department of Language and Humanities Education, Faculty of Educational Studies, University Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[2]Department of Language and Humanities Education, Faculty of Educational Studies, University Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

E-mail: ali.youssefi8@gmail.com, ali.rezvani85@gmail.com

Abstract –This paper having an overview of the history of ESL/EFL writing assessment, tries to explore in to the different waves of its development. Then, taking into account the direct and indirect methods of writing assessment, it elucidates the emergence period of 'Portfolio Assessment' following the requirements of the educational environment in adjusting the assessment methods with the real needs of ESL/EFL writers. Furthermore, Portfolio Assessment as the innovative process oriented approach to writing assessment is contrasted with the other product & process oriented tools in order to come up with the advantages of it over the other tools. Finally, there is a brief account of the merits of analytic assessment of ESL/EFL writing in the form of portfolios in comparison with the holistic approach to assessment that totally focuses on the requirements of the learners in ESL/EFL contexts to learn better and to be evaluated fairly and authentically.

Keywords – writing instruction; direct method; portfolio assessment; analytic assessment; EFL/ESL writers.

1. An overview to the writing issue

Instruction in English as a Second Language (ESL) has often focused on improving students' skills and abilities in speaking, listening, and reading in the target language (L2) while, as Edelsky and Smith (1989) believe, ignoring the development of the students' writing skills. Harris (1985) concluded that only 2% of ESL instruction was concerned with writing activities, and of this two percent, 72% was related to the mechanical aspects of writing such as syntax, punctuation, and spelling.

As Graves (1984) has pointed out, this lack of attention to writing instruction has been a neglect of research in writing compared to other skill areas. Concerning this issue, Hillocks (1984:12) says, "Only recently has writing for second or foreign language learners received research attention." Educators and administrators are showing increased interest in becoming part of a 'new wave' of assessment in the learning situation, assessment that includes authentic and performance-based measures. S. Murphy (1994:12), pointing out the influential effects of such measures, says, "These methods of assessment allow students to demonstrate desired performance through real-life situations. Such methods of assessment are not limited to multiple-choice and standardized tests, but include projects which require students to demonstrate their 'problem-solving' skills as well as their skills in analyzing and synthesizing information."

The portfolio which is one of these new assessment measures has become increasingly popular, and technology is helping with its creation and management. A language portfolio is a collection of an individual learner's work showing his/her abilities in one or more areas of language skills, his effort and language development over time. The decision to introduce portfolios came about as it was felt that tests which depended upon a "one-shot" attempt failed to show students' true ability. Furthermore, as Hamp-Lyons (1996) observes, the process of generating ideas, drafting, redrafting and editing are vital elements of writing and it became apparent that these important dimensions were not being sufficiently valued and assessed as they were not being recorded in an adequately systematic manner. He maintains:

> Educators hoped that students monitor their own learning and engage in self-reflection which, it was expected, would lead to a more positive attitude towards themselves and their learning. It was felt that in the current academic climate portfolios offered the most likely potential for meeting these criteria (p.153).

2. Research & Practice in EFL Writing

Given the broad general agreement about the importance of learning to write in EFL, it is disturbing to discover that, as Amiran & Mann (1982) believe, most researchers and educators agree that, with rare exceptions, students do not and cannot write well. In a research conducted by them on one hundred EFL learners they found that 90 percent of the respondents considered student writing to be a problem - either a serious problem (40 percent) or a minor problem (50 percent).

Why is it so? Exploring different books and articles about this issue, I may conclude that, as Smith (1991) believes, writing is an area characterized by considerable **divergence** between '*research*' and '*practice*'. He in his article claims that much is known about which practices in teaching writing process are effective; several of these findings are in **conflict** with widespread practices in learning environments. For example, staff of the ERIC Clearinghouse on Reading and Communication Skills reported in 1984 while most authorities of writing agree that learners learn to write by writing, there is a distressing lack of classroom time devoted to extended periods of writing.

Still, writing both in native language or second or foreign language remains a critical area of the curriculum in education centers. Graves (1984:15) identified several ways that writing whether in native or in foreign language is **important** in our lives: these include a variety of purposes that are served by writing and developing writing capability.

- As a contribution to the development of a person, no matter what that person's background and talents... Writing is a highly complex act that demands the analysis and synthesis of many levels of thinking.
- Writing develops initiative. In reading, everything is provided. In writing, the learner must supply everything: the right relationship between sounds and letters, the order of the letters and their form on the page, the topic, information, questions, answers, order.
- Writing develops courage. At no point is the learner more vulnerable than in writing.
- Writing, more than any other subject, can lead to personal breakthroughs in learning.

2.1 An Overview of History of Writing Assessment

Even if by another name, writing assessment has always been at the center of work in writing, "it wasn't called **assessment** then, of course; that language came later." says Hamp Lyon (1996:153). He adds, "During the first of what is identified as three waves in writing assessment, it was called **testing**."

Yancey (1999) in a Conference on College Composition and Communication – CCCC – refers to the history of writing assessment and remarks that writing assessment has changed during the past half century. She says:

> *One way to historicize those changes is to think of them as occurring in **overlapping waves**, with one wave feeding into another but without completely displacing waves that came before. The trends marked by these waves, then, are just that: trends that constitute a general forward movement, at least chronologically, but a movement that is composed of both kinds of waves, those that move forward, those that don't (p.483).*

She puts emphasis on the significance of the metaphor of waves in her definition and claims that it is a useful concept. She suggests that it allows us to mark past non-discrete patterns whose outlines and effects become clearer over time and upon reflection-and whose observation allows us in turn to think in an informed way about issues that might contribute to future waves.
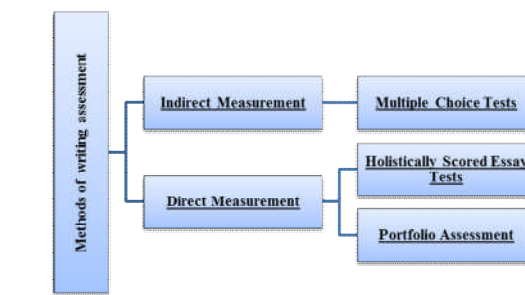

Figure 1: An Overview of History of Writing Assessment

As she suggests during the first wave (1950-1970), writing assessment took the form of 'objective tests'; during the second (1970-1986), it took the form of the 'holistically scored essay'; and during the current wave, the third (1986- present), it has taken the form of 'portfolio assessment' and of programmatic assessment. This is the general history of writing assessment: the one located in method. However, as De Fina (1992) suggests, *other lenses* permit *other views*, particularly when brought together, they allow us to understand differently and more fully. We could also historicize writing assessment, for instance, by thinking of it in terms of the twin defining concepts: **validity** and **reliability**. Seen through this conceptual lens, as Huot (1996) mentions, writing assessment's recent history is the story of back-and-forth shifts between these concepts, with first one dominating the field, then another, now both.

Yancey (1999) refers to another related approach that constructs the history of writing assessment as the struggle between and among scholars and testing practitioners and faculty, those who consider the terms validity and reliability quite differently: the old expert; the new non-expert. She then concludes from this perspective, the last 50 years of writing assessment can be narrativized as the *teacher-layperson* (often successfully) challenging the (*psychometric*) *expert*, developing and then applying both expertise and theory located not in psychometrics, but in rhetoric, in reading and, increasingly, in writing practice (p.484).

Still another way to trace the history of writing assessment is through its **movement into the classroom**, multiple-choice tests, standing outside and apart from the classroom, have become the portfolios composed within (p.484).

Finally, writing assessment can be historicized through the lens of the **self**. *Which self does any writing assessment permit?* As important, given that tests create that which they purport to measure (Hanson, 1993), *which self does an assessment construct?* **Portfolio assessment**, with its multiple discourses and its reflective text, has highlighted this second question.

Yancey (1999) claims that these lenses don't just frame the past; they point to the future, specifically to three issues which have been put forward by her:

1. *What is the **role** of the person in any writing assessment*?

   The role that the self should play in any assessment is a central concern for educators. It is the self that we want to teach, that we

hope will learn, but that we are often loath to evaluate.

2. *How can we **use** this kind of assessment- which is quite different from the teacher assessment that has focused most of our attention for 50 years- to help students?*

3. *What assessment activities can teach us*?

It is only recently that assessment is seen as a **knowledge making endeavor** which raises a good question: what (else) might we learn from writing assessment? And how would we learn? (ibid: p.485)

It is known from the view of Yancey and other supporters of these overlapping waves that underlying these concerns is a particular construct of writing assessment itself: a **rhetorical act** that is both *humane* and *ethical*. In itself, that understanding of writing assessment is perhaps the most significant change in the last 50 years.

2.2 A Context for a History of Writing Assessment

During the first wave of writing assessment, dating from around 1950 to 1970, as Yancey (1999) remarks, writing assessment was young, complex and conflicted. She says:

> It was a critical time in that most of the issues that currently define the field were identified. It seems that consequently, in our practices today, we can trace the outlines of yesterday's concerns (p.485).

Much of the knowledge about the early days in writing assessment is accurate. Yancey (1999), distinguishing among the different viewpoints of the educators and the testing specialists, states that it is true to consider "**objective**" tests particularly multiple-choice tests of usage, vocabulary, and grammar, as dominating practice. She maintains:

> It's true that most testing concerns focused on sites ancillary to the classroom: typically, on the placement exercise used to place students into appropriate writing courses. And, in general, it is true that at that time, classrooms were defined, at least in part, by what we could call a technology of testing – not only by means of the tests that moved students in and out of classrooms, but also by way of simultaneous efforts to bring "our work" – i.e., the reading and grading of student work – into line with testing theory (p.488).

In practice, the tasks initially divided themselves in two clearly isolated spheres of influence that characterize the first wave of writing assessment: the process of deciding what to teach the students belonged to 'educators', and the process of moving students about, to 'testing specialists'.

2.2.1 Indirect & Direct Measures

From the *perspective of method*, changes in writing assessment appear straightforward and familiar: from first-wave "*objective*" measures like multiple-choice tests, mainly of grammar and usage, to second-wave holistically scored essay tests to third-wave portfolios. As Yancey (1999) suggests, first wave evaluation relied on an "**indirect**" measure - a test of something assumed to be *related* to the behavior, but *not* the behavior *itself* (e. g., items like comma, questions, and pronoun reference corrections).

Within twenty years, during the second wave, scholars began employing a "**direct**" measure - a sample of the behavior that we seek to examine, in this case a text that the student composes. Once the direct measure becomes accepted and even customized as the measure of choice, the "**one essay**" model is soon replaced by a set of texts, so that: a single draft becomes two drafts; two drafts become two drafts accompanied by some authorial commentary; two drafts plus commentary become an undetermined number of multiple final drafts accompanied by "**reflection**" and the set of texts becomes the new: "**portfolio assessment**".

Following the steps in adjusting the testing methods with the requirements of the educational environment, scholars in composition studies conferences like CCCC answered the question *How shall we evaluate writing?* with another question *Which behavior should we examine?* They considered sampling as being critical, in part because sampling was (and is) the stuff of everyday classroom life. Yancey (1999:491) says, "Days in and day out, faculty assign, read, and evaluate student texts. In this sense, *teaching writing is itself an exercise in direct measure*". She declares the measures taken by educators as follows:

- Teachers saw the differences between what they taught in their classes –*writing* and what was evaluated–*selection of homonyms and sentence completion exercises.*
- They thought that differences mattered; and
- They continued to address this disjunction rhetorically, as though the testing enterprise could be altered –first on their own campuses; also at composition studies conferences and testing-focused conferences like the National Testing Network in Writing and the NCTE conferences on portfolio assessment; and concurrently in articles and books (p.494).

Still, it took over 20 years for this analysis to make an impact, over 20 years for the second wave to occur. Murphy (1996:293) puts a reasonable question forward:

> If compositionists saw this disjunction between '*classroom practice*' and '*testing practice*' early, why did it take over two decades to shift from one sampling technique to another, from one methodology to another? And the waves are overlapping, not discreet: why is it that even today, 50 years later, multiple choice tests continue to be routinely used in many assessment exercises.

She poses the responses to these questions by four other questions. They are inter-related, each of them located in or deriving from the methods and sampling issues: Murphy (1996:293) refers to these questions:

- What roles have *validity* and *reliability* played in writing assessment?

- Who is authorized and who has the appropriate expertise to make the best *judgment* about writing assessment issues?
- Who is best suited to *orchestrate* these questions, design an assessment based on the answers, and implement that design? In other words, who will have this power?
- What, after all, is the overall *purpose* of writing assessment in an educational situation?

Each one of these questions points to one understanding of writing assessment; each one identifies a dimension of writing assessment still in dispute.

2.2.2   Validity and Reliability

Writing assessment is commonly understood as an exercise in balancing the twin concepts validity and reliability. Validity means that you are measuring what you intend to measure, reliability that you can measure it consistently.

Gomez (1996) refers to the dispute among the advocators of either feature, pointing that they tend to play them off against each other. Accordingly, which one should dominate, assuming only one could be favored, has generated considerable discussion and change.

During the first wave, reliability prevailed; we see this, first, in the kinds of assessments that were commonly employed, and second, by the rationale for using them. That such tests were common is confirmed by various survey data.

The first wave of writing assessment is dominated by a single question as posed by Williamson (1994), not the question we might expect–*What is the best or most valid measure of writing*?–but a question tied to testing theory, to institutional need, to cost, and ultimately to efficiency:

> *Which measure can do the best and fairest job of prediction with the least amount of work and the lowest cost? The answer is: the reliable test (ibid, p.153).*

2.3   The 'Discourse' of a Writing Assessment

But what about validity? This question, raised often enough by the teaching staff of the faculty, dominated the second *wave of writing assessment*. White (1985), as a faculty member, who became an administrator in the Freshman English Equivalency Examination Program in California State University, along with some other instructors elsewhere, saw an obvious 'discrepancy' between what they did with their students in class and what students were then asked to do on tests.

The new concern with validity, then, was also motivated by the fact that by the 1970's, faculty members had begun to identify themselves as '*compositionists*'. They *knew more* about writing: about writing process, about teaching writing process, about writing courses and what they might look like, about what composition studies might be. Yancey (1999:489) concludes by saying, "Given what we were learning, it made increasingly less sense to use tests whose main qualities were reliability and efficiency."

Yancey (1999:492) says:

*The shift to what did seem obvious–**the essay test**–had to be orchestrated, however, and it was, by two rhetorical moves, both of which worked inside psychometric concepts to alter assessment practice: first, to make validity (and not reliability) the testing feature of choice; and second, to undermine the concept of correlation as a criterion for evaluating tests.*

White (1985) took the first approach. He recognized the three variables that had to be accounted for in order to make essay testing feasible:

> *While some...chancellors and the like are resistant to argument, most are not; many of those who employ multiple-choice tests as the only measure of writing ability are properly defensive of their stance but will include actual writing samples if they can be shown that writing tests can be properly constructed, reliably scored, and economically handled (p.84).*

This is exactly what White and some other faculty members set out to do: devise a writing test that could meet the standard stipulated by the testing experts. To do that, they had to solve the *reliability dilemma*: they had to ensure that essay tests would perform the same task as the objective tests.

Administrators like White thus borrowed from the Advanced Placement Program their now-familiar "testing technology", called **'holistic writing assessment'**; the AP assessment was a *classroom-implemented curriculum* culminating in a final essay test that met adequate *psychometric reliability* standards through several quite explicit procedures. White (1985:92) mentions these procedures as follows:

(1) Using writing "prompts" that directed students (see Appendix II).
(2) Selecting "scoring guides" that directed teacher-readers who rated; and
(3) Devising methods of calculating "acceptable" agreement.

The AP testing technology, then, marks the **second wave** of writing assessment by making a more valid, classroom-like writing assessment possible. By applying these procedures, test-makers like White could determine both what acceptable reliability for an essay test should be and, perhaps more important, how to get it.

Pursuing the history of writing assessment, we find out that at the time that administrators and faculty were showing how a more valid measure could also meet an acceptable standard of reliability – and therefore how testing could be more congruent with classroom practice – other administrators and faculty were demonstrating in the language of testing why the '*reliable-only*' test was particularly '*incongruent*'. In 1978, for instance, Rexford Brown made this case not only by appealing to the context of assessment, but also by connecting that test to the context of the larger world:

> *Of course, these [objective] tests correlate with writing ability and predict academic success; but the number of cars or television sets or bathrooms in one's family also correlate with this writing ability, and parental education is*

one of the best predictors there is. All existing objective tests of "writing" are very similar to I.Q. tests; even the very best of them test only reading, proofreading, editing, logic, and guessing skills. They cannot distinguish between proofreading errors and process errors, reading problems and scribal stutter, failure to consider audience or lack of interest in materials manufactured by someone else (p.3).

He maintains that the *correlation* here correlates with more than predictive ability considering them as a measure of wealth, of the *number of cars or television sets or bathrooms in one's family*, and of another variable, parental education.

Brown (1978) implicitly asks if these are items we seek to test. Moreover, given the **discrepancy** between the items on the test and what we in our classrooms teach, he poses this question that what could such scores based on such items really mean, anyway? He says:

> *Meaning is, after all, located in more and other than correlation: it is intellectual and rhetorical substance (ibid, p.4).*

By working both within and against the psychometric paradigm, then, faculty members and administrators moved the educators during the second wave of writing assessment *closer* to classroom practice.

2.4   The Third Wave: New Assessment as Portfolios

The idea of portfolios comes forth from discontent amongst academics and educators about the limitations inherent in structured testing, i.e. the traditional form of evaluation whereby students are asked to perform within a set time limit, to answer specific questions, produce end results according to the expectations of the examiners and have only one opportunity to do so. S. Murphy (1994:47) says:

> *Limitations which have been cited include the fact that structured testing can only evaluate knowledge of facts and keep away all important meta-cognitive knowledge, the processes by which students produce work are ignored, and that such testing fails to show students' true ability. Teaching methodology is constantly being reviewed and modified to meet the demands of a changing society which values meta-cognitive skills and individual thought and expression more than just the ability to perform well in a "pencil and paper" test.*

However, testing methodology as Hamp-Lyons (1999) reminds, has remained surprisingly consistent and seems to rely more on a sharp memory than intellect and sound judgment. There does, indeed, seem to be a *discrepancy* between the innovative teaching methods and traditional testing techniques and it is this discrepancy that portfolios attempt to address. They have been used in such diverse fields as Art, Music, Teaching and Writing and, as will become apparent from the following, if carefully orchestrated, have a considerable potential.

The following are some typical definitions of portfolios:

- A purposeful, chronological collection of student work, designed to reflect student development in one or more areas over time and student outcomes at one or more designated points in time. French (1992:43)
- A chronologically sequenced collection of works that records the evolution of artistic thinking. Adams & Hamm (1992:103)

Although the definitions given above are different, certain underlying principles are evident, namely the emphasis on a collection of work, chronological organization and a purposeful construction.

Reviewing several definitions, we may conclude that a portfolio is essentially a collection of a student's work which can be used to demonstrate his or her skills and accomplishments. As you look into the various uses of the writing portfolios in the teaching of writing, you may feel that an educational portfolio is more than just a group of projects and papers stored in a file folder. It includes other features such as teachers' evaluations and student self-reflections. According to the scholars of the Northwest Evaluation Association (2001), a portfolio is a purposeful collection of student works that exhibits the student's efforts, progress, and achievements. The collection must include student participation in selecting contents, the criteria for selection, the criteria for judging merit, and evidence of student self-reflection.

Portfolios are tried out as a new vehicle for teaching, learning and assessment as well as a means for students to negotiate their own syllabuses within the given framework. Hence, student empowerment, learner-centeredness and self-directed learning are the key concepts of this experiment. Arter & Spandel (1992) talking about the writing process says:

> *Portfolio should include a description of its purpose and goals as well as of the criteria for selection and assessment. Preferably, it should also contain the student's own reflection on and evaluation of both the selected work and the process of studying and learning (p.36).*

Accordingly, the portfolio is an individual and learner-centered tool for learning and assessment. Lucas (1992), pointing to the different functions of portfolios, notes that its purpose and function may vary, and so may its form and contents as well as the process of producing and compiling it. She says:

> *At first, however, a portfolio is often a working portfolio, also called a process portfolio, which contains a variety of pieces of work with all their versions. The working portfolio thus documents the whole process of studying and learning (ibid, p.9).*

Normally, students select some of their pieces of work at the end of a course or a term, for example. Students also assess the selected pieces and state their criteria for selection. According to her, the final showcase portfolio thus usually represents the students' best work showing their strengths. Naturally, the criteria for selection depend on the *purpose* of the portfolio.

A central idea of the portfolio experiment is to promote learner-centered and self-directed learning. The students should take greater responsibility for their own work but also to have the freedom and power to make decisions concerning their studying. There is also a need for them to learn to set their own goals as well as to assess their

work and learning and state their criteria. And, most importantly, the students need to be directed in order to feel ownership of their learning and of their abilities to communicate in English. It is called **empowerment** of students as learners of English.

As it is inferred from the literature on portfolio use, in the late 1980s and into the 1990s, other portfolio systems developed; individual faculty began using writing portfolios, sometimes as a means of formal assessment, sometimes as a way of learning. All of these portfolio assessments expressed a **direct connection** to classroom practice.

Portfolio experiences illustrated a *new function* identified for writing assessment during the third wave: *creating knowledge about assessment*, of course, but also about our own practices, as Allen & Yancey (1997:86) say:

> *When writing assessment is located within practice, its validity is enhanced, to be sure. But equally important, it reflects back to us that practice, the assumptions undergirding it, the discrepancy between what it is that we say we value and what we enact. It helps us understand, critique and enhance our own practice; in other words, because of its location–in practice–and because it makes that practice visible and thus accessible to change.*

They as theorist-practitioners argue that the humanistic endeavor requires a student-informed and -informing assessment and the expertise that can create it.

Faculty experience with portfolios as an assessment technology has focused the attention from yet another perspective: that of **practice**. The effect of this practice, as Yancey (1999) argues, has been to suggest new understandings about the kinds of expertise that might inform our assessment practices, with the specific effects of localizing and grounding expertise of some kinds. She indicates them as:

- First, **student expertise**. Through the reflective texts in portfolios, students are asked to demonstrate a kind of expertise about their own work; their 'secondary' reflective texts are used as confirming evidence of student achievement as documented in a primary text. Writing well is thus coming to mean twofold: *writing well* and *being an expert on one's writing*.
- Second, **reader expertise**. Assessment specialists are looking more carefully at what they are calling "expert" readers, based on a second wave holistic model that Bill Smith used at Pittsburgh and was later adapted for portfolio assessment by Haswell in 1994. In this model, readers are experts–authoritative about the relationship between a student and a specific course, one that the teacher-reader has very recently taught. Conceived of this way, reliability is not a function of agreement, directed or otherwise, among raters so much as it is a function of rater experience with particular curricula.
- Third, **theoretical expertise** that grows out of and is integrated with practice. The practical work in assessment undertaken during the

third wave has created a body of rich data permitting theories of writing assessment to emerge. The theories are developing in two ways: as elaboration and new applications of assessment theory generally; and as readings of practice suggest (ibid, p.496).

2.5 Organizing Assessment

Closely related to the issue of expertise is that of **power**. During the first wave of writing assessment, faculty seemed content to allow testing specialists to direct the tests while they directed classroom activities. During the second wave of writing assessment, faculty began to see writing assessment as something that wasn't peripheral to the classroom, but important in its own right, as Daniel Fader (1997:83) suggests:

> *.... Writing assessment is to be taken seriously because its first purpose is to determine quality of thought rather than precision of form. As our students, our readers, and our network of cooperating teachers have told us, it matters because it tries to test something that matters so much.*

Assessment within the classroom thus took on increased emphasis and importance. Two examples–one focused on the '**role of error**' and another on '**response to student texts**'–in the history of writing assessment illustrate how assessment concerns began to move inside the classroom, become transformed in that context, and generate new questions for assessment specialists and compositionists alike. During the **first wave** of writing assessment, '*error*' (by means of test items) outside the classroom determines which classroom a student enters. During the second wave, error comes inside the classroom. Shaughnessy (1977:63) says:

> *Taken together, errors weave a pattern open to teacher observation and intervention. Still understood as mistakes, they become clues allowing a teacher to plan where to start and what to do. During the third wave, pattern of error is its own discourse.*

Hull and Mike (1990) in an article in CCC considering error originally an external marker of discrepancy, suggest that it thus moves into the classroom and becomes its own reasonable text, a means of knowing for both student and teacher.

A similar kind of movement occurs with '**response** to student writing'. During the first wave of writing assessment, considerable comment is provided on how important response is in helping students: assessment as a discipline, located outside the classroom, includes no provision for response.

During the *second wave*, we see the first formal study of *response*, conducted by Sommers in 1981. Located not outside the classroom but inside, Sommers study is based in and oriented toward recommending good classroom practice.

During the *third wave* of writing assessment, as Chandler in one of composition studies conferences in 1997 declared, modes of response and their functions–when to praise, how to help students move toward *reflective* writing, and how students interpret our

comments to them–have become a central concern. Harris (1997) responds the current question of: should preferred response always be non-directive, or should it be situated? In this way:

> *Response is theorized newly, not as an evaluative end, but rather as an inventive moment in composing. It is a text in its own right; another place to continue the opportunity for writers to change not only their phrasings but their minds when given a chance to talk about their work with other people (ibid, p.68).*

Moving inward now–into the classroom and then into and within composing itself–writing assessment becomes a **social** act. As a social act, writing assessment exerts enormous influence, explicitly and implicitly, often in ways, both faculty and students do not fully realize. Certainly, writing assessment has been used *historically* to exclude entire groups of people: White (1985) makes the point that a primary motivation for holistic scoring was explicitly political, to enlarge and diversify the student body.

Portfolios, for many, are developed from similar impulses, as Catharine Lucas (1992:4) notes:

> *Portfolios provide for what I call reflective evaluation, a kind of formative feedback the learners give themselves. Through this technology, then, students' own evaluative perception is allowed to develop.*

2.6 Writing Assessment and the Self

As Berlin (1984) has suggested, **education ultimately and always is about identity formation**, and this is no less true for writing assessment than for any other discipline. What we are about, in a phrase, is *formation of the self* and "writing assessment", because it exerts so much power, plays a crucial role in what self, or selves, will be permitted–in our classrooms, in our tests, ultimately, in our culture. Yancey (1999) in her article makes it clear that the self also provides a lens through which we can look backward and forward at once, to inquire as to how it was constructed during the three waves of writing assessment. Here is a summary of her observation:

- During the **first wave** of writing assessment, the *tested self* of course took very narrow forms. In multiple-choice tests, the self is a **passive**, forced-choice response to an external expert's understanding of language conventions. Agency is neither desired nor allowed.
- During the **second wave**, the self becomes a producer–of a holistically scored essay–and thus an agent who creates text. Still, there is *less agency* there than it appears. The text that is created is conventionally and substantively determined by an expert who constrains what is possible, by creating the *prompt* and designing the *scoring guide* used to evaluate the text. Given these constraints, the authorship of such a text is likely to be a static, single-voiced self who can only anticipate and fulfill the expert's

expectations, indeed whose task is to do just that. At best, agency is limited; a self-in-writing is permitted, but it is a very limited self, with very circumscribed agency. The text does not admit alternative discourses conceptually or pragmatically: it is text as correct answer.
- During the **third wave** of writing assessment, *the self emerges*, and it is often multiple, created both through diverse texts and through the reflective text that accompanies those texts. Reflective texts in **portfolios** invite the readers to "fictionalize" authors.

This narrativizing tendency constitutes one of our primary ways of understanding, one of our primary ways of making sense of the world, and is an essential strategy in comprehension. The story of writing assessment is a narrative of uninterrupted progress. Yancey (1999) focuses on by saying that it is rather of narrative of some waves: the early wave, governed by the objective measure; the second wave, which saw the move to the more valid holistically scored essay; the third wave, where portfolios contextualized our students' work and invited us to consider how we read, how we interpret, how we evaluate.

2.7 Writing as a Process

Going through the history of the writing assessment, we may argue that holistic writing assessment procedures put emphasis on the product of the writing while at the same time some researchers were doubtful about the effectiveness of doing so and were searching for a scheme to increase the amount of achievement of the learners during the writing course. Hillocks (1984:32) claims, "the major general finding from the research on teaching writing is that student achievement is higher when the teaching approach emphasizes writing as a **process** rather than writing as a **product**."

B. Kroll (1990) argues that in the traditional product-oriented approach, form and correctness are the major concerns. He says:

> *The teacher provides drill work on specific skills, makes many of the major writing decisions for the students (topic, form, length, etc.), and serves as the sole audience/judge. Learning involves following rules, conforming to formula, and achieving technical mastery of formal conventions and modes. Students work alone on their writing assignments, and while trying to discover what they want to say, are reminded of such technical matters as using topic sentences and avoiding writing sentence fragments and run-ons (p.4).*

This approach has been noted in comprehensive classroom observations, in examination of ESL/EFL textbooks, and in a review of teachers' scoring of ESL/EFL compositions.

The student in the product-oriented writing class tries to get it right the first time, because the paper turned in will be the only version. The teacher carefully marks all the mechanical errors in red ink and writes notes in the

margins about the logic and clarity of the essay. Because the student will be doing nothing further with the piece, he/she often pays little attention to the teacher's comments. As Kroll (1990) notes, under these conditions there is not much of a sense of **ownership** or **investment** in the writing. Virtually all the various subparts of the traditional approach have been shown to be ineffective in producing capable EFL writers. Kroll (1990:9) identifies several reasons for the failure of this approach:

- It emphasizes form and mechanics before, and often at the expense of, ideas and meaning.
- It focuses on the *product* rather than the *process*.
- It seriously neglects the *earliest stages* of the writing process.
- It offers too many *artificial contexts* for writing.
- It isolates mechanical skills from the context of writing.
- Rather than being an outgrowth of research and experimentation, the traditional approach is based on sheer historical momentum of out-of-date theoretical assumption.

From the experience of the educators and from the research conducted during the past 15 years, there has emerged a **process-oriented** approach to teaching writing in ESL & EFL. Recognizing that writing is a complex, recursive, dynamic nonlinear process, experts in the field of composition have developed and tested instructional methods more in keeping with the true nature of the act of writing. Looked at this way, Hillocks (1984) believes that the writing process has a number of distinct stages. He suggests the following six stages on   page 142 of his book:

- **Prewriting:** The writer gathers information and plays with ideas during the prewriting stage. Prewriting activities may include drawing, talking, thinking, reading, listening to tapes and records, discussion, role playing, interviews, problem-solving and decision-making activities, conducting library research, and so on. Research shows that students who are encouraged to engage in an array of prewriting experiences evidence greater writing achievement than those enjoined to "get to work" on their writing without this kind of preparation.
- **Drafting:** The writer develops his/her topic on paper (or a computer screen) during the drafting stage. Beginning may be painful and difficult, producing false starts and frustration in the writer. In the process-oriented approach, the focus is on content, not the mechanics of writing.
- **Revising:** During this stage, the writer makes whatever changes he/she feels are necessary. Revision may involve additions and deletions; changes in syntax, sentence structure, and organization; and in some cases, starting over completely.
- **Editing:** Polishing of the draft takes place in the editing stage. The writer gives attention to mechanics such as spelling, punctuation, grammar, and handwriting, and may also make minor lexical and syntactic changes.

- **Publication:** Publication refers to the delivery of the writing to its intended audience. Classmates, other students, parents, and community members are among the potential audiences for students' written work.

According to him, the revision stage is most productive of superior final products if it includes input from teachers or fellow students. He has also found that student motivation and achievement are enhanced when student work is 'published' for a larger audience than the teacher.

2.8 The Major Components of Writing Ability

Attitudes and priorities in ESL/EFL writing assessment have changed significantly since 1990. Reliability and the traditional validities still have importance, but it is better understood that a good assessment demands more than this.

Hamp-Lyons (1990:69) considers the writing assessment from a perspective that focuses on the '*validity*' of the assessment:

> *It is convenient to think of four components of a* **direct** *assessment of writing for which validity must be established: the task, the writer, the scoring procedure, and the reader(s). As our perceptions of the task of assessing writing have expanded, those four components have remained relevant. Here I shall put the writer first, recognizing the greater attention in the present critical humanist orientation of applied linguistics and language teaching to the people lies at the heart of assessments. I shall argue for an approach to writing assessment that takes into account who the learner is, the social context the learner has come from, and the target context in which the learner will have to function.*

2.8.1 The writer:

In writing assessment research, the writer has too often been forgotten, probably because *researchers* are more '*distant*' from actual writing classrooms than they should be. Hamp-Lyons (1990) believes that classroom teachers, when they prepare essay tests or other kinds of writing tests and assignments for their students, do not forget the human beings they work with, and who will be taking the test. He continues in this way:

> *Their consciousness of the people being tested shapes their responses to choices of topics and of reading material (if any) to be used for content input; it gives them clear views on the amount of time writers will need to carry out a task, and the criteria by which the writing should be judged (p.71).*

It seems that what we do already know is that ESL or EFL learners are greatly varied in language background, socioeconomic status, personality, learning style, as well as all the other factors which apply equally to L1 learners. Hamp-Lyons (1990) believes that if a writing assessment

is to be humanistically, as well as psychometrically, defensible, all of these factors should be accounted for.

As the work of Hamp-Lyons & Zhang (2001) with Asian learners in Australia showed, we do not yet possess sufficient knowledge of culturally determined writing behaviors to be able to teach students what to change in their writing in order to conform to expectations, should they wish to do so. Although it is now fairly well accepted that written text production is in part culturally determined, there is still far too little research for it to be possible to make assertions about the "usual" cultural patterns and/or problems of writers from any particular background.

However, for the classroom teacher/assessor of writing the details of research on differences is unimportant. What matters is the understanding that all writers are influenced by who they are and how they became this person: Race, gender, ethnicity, culture, language background, level of education in L1 and in L2, stage of cognitive development, learning style, motivation, degree of support in the home background—all this contributes to individuality. If we are serious about respecting the rights of writers, and ensuring that our assessments do no harm and allow writers to show the best they are capable of, these standards of informativeness must be met.

An additional writer issue is that in interpreting a task and creating a response to it, each writer must create a "fit" between his or her world and the world of the essay test topic. Hamp-Lyons (1990) suggests that each writer needs to take the *other-initiated* test task and transform it into a *self-initiated* topic - that is, make it his or her own. In order to match her or his response to the tester's expectations, the writer must follow the steps of attending to, understanding, and valuing the task. If this process breaks down, the writer will replace the task with a related or a different one, but will not respond to the topic intended. Hamp-Lyons (1990:82) says:

> The problem here is that it may not be clear to the rater whether the writer has done this deliberately or accidentally due to lack of topic knowledge or linguistic incompetence.

Hamp-Lyons & Zhang (2001) observed and recorded a number of second-language writers reading a prompt on a large-scale essay test; they reported four cases as examples of the ways in which individual writers 'read' a prompt. Of the four writers only one 'read' in the expected way and wrote an essay that wholly "fit" the test maker's and the reader's expectations. Not surprisingly, this writer was rewarded, while the other three suffered to various degrees for their inability to match the test maker's expectations.

Additionally, all the other four elements (to be discussed in the next sections) significantly influence the writer, either through the task the writer has to perform, or through the judgment and the consequences of that judgment. Thus, even when the writer is not discussed directly in discussing other key constructs in writing assessments, the writer should never be perceived as a forgotten element.

2.8.2   The scoring procedure:

Over the last thirty years, there have been a great many developments in scoring procedures for writing assessments. Hamp-Lyons (1990) believes that most scoring procedures can be placed into one of the following three categories:
(1) holistic scoring; (2) multiple trait or analytic scoring; and (3) primary trait scoring.

a) Holistic scoring:

Hamp-Lyons (1990:84) defines it in this way:

> A true holistic reading of an essay involves reading for an individual impression of the quality of the writing, by comparison with all other writings the reader sees on that occasion.

An equally serious problem with holistic assessments is that as Hamp-Lyons & Zhang (2001) believe the context in which they are carried out–where the average time to score a single 2-page handwritten essay of about 500 words may be less than one minute–means that it is not possible to capture performance data from raters as they are doing the rating task, and therefore it is impossible to get far enough into that performance to be able to understand fully what is going on. Without research which enables us to understand the **processes** actually used by individual raters, we shall never be able to find ways to make judgments more reliable and at the same time more valid (i.e., the 'new validities' described above). But we will turn to raters in the final section of this part of the discussion.

b)   Primary trait scoring:

Primary trait scoring involves deciding what **one** aspect of writing is key to success on this task, developing a highly-detailed set of descriptors for performance on that aspect (**trait**), and training teachers/raters in its use. Hamp-Lyons & Zhang (2001:106) claim:

> Primary trait scoring has not been often used because its key element, the development of a single scale on a single feature of writing seen to be most salient for a very specific task, makes it very resource-intensive. Primary trait scoring scales have to be developed afresh for every context.

However, the primary trait approach was the source for Hamp-Lyons' (1990) development of multiple trait scoring.

c)   Multiple trait scoring:

In striking contrast to holistic assessment, multiple trait assessment of writing defines a procedure that is **'context-sensitive'** at all stages and in all dimensions of the test development, implementation, scoring, and score reporting. Hamp-Lyons & Zhang (2001:109) suggest:
As its name implies, multiple trait scoring treats the construct of writing as complex and many-sided, and allows teachers or test developers to identify the qualities or traits of writing that are important in a particular context or task type, and evaluate writing according to the most important traits in a specific context. Multiple-trait scoring also allows raters to pay attention to the relative strengths and weaknesses in an individual writer's text and score some traits higher than others.

While there may be some arguments in favor of holistic scoring of the writing of quite advanced native users of a language, cost and practicality being the main ones, when we turn to the specific problems and needs of assessing *second-language* writing, those arguments are much weaker than the arguments in favor of multiple trait assessment. Furthermore, as Hamp-Lyons & Zhang (2001) point out, multiple trait scoring is a useful tool for **researching** what is going on inside a writing assessment because of the detailed way it opens up the process; it is also a vital tool for teachers of second or foreign language writers because it provides teachers with rich information that will facilitate decisions about remedial courses, selection of course types, etc.

It has become increasingly common for test developers designing *new* ESL/EFL writing assessments to use multiple-trait scoring.

2.8.3   The complexities of scoring:

In the past ten years or so, research has certainly helped to identify ways in which the scoring of writing can be made more reliable and more valid; but it has also uncovered many new complexities. As specialists in writing assessment, such as Zhang (1999), have increasingly turned their attention to scoring, to what qualities of writing are judged, how they are judged, who the judges are, and many other issues scoring is a far more complex process than was ever imagined. In the research of the last decade the strength and complexity of the link between the scoring procedure and the human beings who apply the procedure has become remarkably clear.

2.8.4   The reader:

A large body of research investigating the work of readers' practices in writing assessment led to the emphasis on reader training that became so common in writing assessment programs from the early 1980's on. Newcomb (1977) looking at raters' responses to native speakers' essays, showed that where raters deviated from their own typical response patterns, these deviations could be explained by affective interactions between the rater and the text.

Similar findings from a variety of studies resulted in a concern to create training procedures to counteract the possibility that readers would bring different agendas to the reading and assessing of the same text.

In the EFL context, using ethnographic methods, Hamp-Lyons et al (2001) found results that were similar to Newcomb (1977) findings related to individual reader agendas. They found that readers of EFL essays *responded* to cultural differences in them, and did so *differentially* in ways that appeared to be partially attributable to their *experiential* backgrounds and to their response to the *students' linguistic/rhetorical* backgrounds. They looked at the decision-making behaviors of expert and novice raters rating EFL writing, and found that the *expert* raters spent more of their attention on *higher-order* aspects of the writing, and were more *reflective* about their own processes, while *novice*

raters focused more on *lower-order* aspects of the writing.

By the early 1980s, White (1985), writing about L1 settings, felt the need to determine the effectiveness of holistic assessments of writing, and to assert not only the

reliability of the holistically-derived scores but also the validity of holistic reading processes. As J.D. Brown (1991) believes, in L2 context, we have learned a great deal more in the past ten years about the behavior of readers through quantitative scoring comparisons and detailed ethnographic studies of essay raters. The evidence from such studies suggests that, left to their own judgments, cannot agree on the absolute quality or the relative quality of essays, nor can they agree on the specific qualities in essays that make them good, worse, or worst.

However, more interestingly, these rich studies have shown the complexity of the rating process that White (1985) asserted, and have taken it further. It shows that raters are influenced by their own cultural contexts and learning/teaching experiences perhaps as much as by the variation in quality of student essays. Zhang (1999) argues that even the most experienced and skilled raters act as individuals, using their own values, even in situations where there is good and extensive rater training and clearly defined criteria. He says:

> *Clearly, we are a long way yet from being able to characterize what it is that raters do and explain when, how and why they are able to do it consistently (p.31).*

2.9   Developments in Writing Assessment

The alternative assessment movement has been largely driven by, and certainly powerfully guided by, developments in writing assessment. As long ago as 1989, Cooper said:

> *Locally developed, holistically scored writing tests enable participating academic communities to define standards of 'good' writing that can be responsive to the particular strengths of their own student population, rather than crushing such particularities and the creative spirits of the students who express them under the yoke of internationally imposed standards of correctness (p. 579).*

Since then, we have increasingly come to realize that **'local'** development and implementation, when done well, is a powerful force for positive educational change. This is one of the benefits of portfolios that have made them so popular in L1 literacy assessment, and there are strong reasons why the argument would be equally or more relevant in ESL/EFL contexts.

2.9.1   A New Look: Four Types of Indices of Writing Quality

According to Gomez (1996), there are four types of indices of writing quality:

(i)   countable micro-indicators of quality (Percent of Correctly Spelled Words, Percent of Correct Word Sequences),

(ii) analytic ratings (Topic Development, Internal Organization, Conveying Meaning, Sentence Construction, Mechanics),
(iii) overall holistic ratings of communicative effectiveness, and
(iv) writing productivity (total words written).

A total of nine different scores were obtained from these four types of indices, summarized in Table 1.

Type of Index	Indicators
Micro Indicators	% Correctly Spelled Words (% CSWd)
	% Correct Word Sequences (%CWSeq)
Analytic Scoring	Topic Development (Topic);
	Mechanics;
	Organization of Thoughts (Organization);
	Conveying Meaning (Meaning);
	Sentence Construction (Sentence).
Holistic Scoring	Overall Quality and Clarity of Communication to Reader (Holistic).
Productivity	Simple Frequency count of Words Written (Total Words)

Table 1: Four Types of Indices of Writing Quality

Gomez (1996), in view of the changes in the writing pedagogy, talks about the different waves from a perspective related to assessment indices, and specially focuses on the analytic scoring procedures, and considers five analytic traits as mentioned in the table above.

2.9.2 Disadvantages of Holistic Scoring

It is inferred from the works of Gomez (1996) and some other researchers that some shortcomings in the procedures in holistic scoring can be alleviated in analytic scoring through multiple traits. The adherents of analytic scoring talk about the weak points of holistic scoring in this way:

• It Assumes that all relevant aspects of writing ability develop at the same rate and can thus be captured in a single score.
• In this mode of scoring a single score may cover an uneven writing profile and may be misleading for placement.
• It constitutes a sorting or ranking procedure and is not designed to offer correction, diagnosis, or feedback.
• Single scores do not permit raters to differentiate features of writing such as depth and extent of vocabulary, aspects of organization and control of syntax.
• Scores in this mode are not often readily interpretable, as raters do not always use the same criteria to arrive at the same scores.

Considering these shortcomings, the supporters of analytic rating assume some advantages for this method of assessment. The following is a brief review of them.

2.9.3 Advantages of Analytic Assessment & Scoring

• There is more observation, hence improved reliability.
• There is a vast range of writing performances.
• Norm-referencing is discouraged.

• There is a greater discrimination across wider range of assessment bands.
• There is the removal of tendency to assess imprecisely.
• This kind of scoring procedure provides more research data/information.
• *It is especially **more appropriate** for ESL/EFL writers as different features of writing develop at different rates.*

In **multiple trait** scoring that is a kind of analytic scoring, raters are required to judge the selected features or skills of writing. This subjective scoring involves the separation of the various features of a composition into components for scoring purposes. An analytical scale focuses raters' scoring and thus ensures reasonable agreement among raters to permit a reliable score to be obtained from collective multiple ratings.

The use of analytic scales has two very practical advantages: Firstly, it permits a *profile* of the areas of language ability that are rated, and secondly, they tend to *reflect* what raters do when rating samples of language use.

Reviewing the existing literature on the application of analytical scoring procedure, it is known that it can be easily and effectively used as a tool that leads to *greater* reliability as each candidate is awarded a number of scores. Furthermore, it is said that in Cambridge ESOL Research Notes (2003) analytic scoring can allow for more precise diagnostic reporting, particularly in the case where a candidate's skills may be *developing* at differing rates reflecting a marked profile. In this research, analytical scores can be used for correlational research, growth measurement, prediction, placement, and program evaluation.

In addition, analytic scores act as useful guides for providing *feedback* to students on their compositions and to formative evaluation which may be used.

2.9.4 Analytic Assessment in IELTS Writing

A valid documentation of the application of analytic scoring is detected in the revision of assessment criteria and rating scale descriptors for the IELTS Writing Modules, which began in January 2002 (Cambridge ESOL, 2003). Exploring the data in the literature of the ESL/EFL writing assessment, we may observe the outstanding attempts and consequences of this research.

The move to *analytical scales* in the process of this revision was for reasons of consistent examiner focus and multiple observations. In a recent internal study to investigate variability in General Training Writing, the performance of the study markers (26 IELTS examiners – all of whom used the analytic marking approach) – was matched against the original markers (who varied in their use of the analytic approach according to the current marking guidelines). According to Cambridge ESOL (2003), in this study inter-correlations varied markedly, with fully analytic-marked scripts achieving the highest values. These findings suggest 'that a move to analytic scoring will bring with it a higher degree of consistency. The **benefits** of analytical assessment in relation to the IELTS examination, according to Shaw (2003), cited in

the Research Notes of Cambridge ESOL (2003:10) are as follows:

1) enhanced reliability through increased observations,
2) wide range of writing performances,
3) greater discrimination across wider range of assessment bands (6-9 Bands),
4) provision of a greater control over what informs the impressions of raters,
5) removal of the tendency to inexplicit assessment, and
6) active discouragement of norm-referencing and the provision of research data/information.

These benefits suggest that analytic assessment outweighs any advantages offered by a holistic approach to assessment.

Shaw (2003), revealing these advantages, goes on declaring that through this system of assessment trainers encourage raters to profile rather than global mark for reasons of thoroughness and consistency and currently some centers routinely profile mark. He concludes: "Consequently, a decision to remove the element of choice by ensuring compulsory analytic marking of all tasks would seem a logical step" (p.11).

3. Conclusion

This paper attempted to provide an overview of the history of ESL/EFL writing assessment and tried to explore in to the different waves of its development. Then, taking into account the direct and indirect methods of writing assessment, it elucidated the emergence period of 'Portfolio Assessment' following the requirements of the educational environment in adjusting the assessment methods with the real needs of ESL/EFL writers. Furthermore, Portfolio Assessment as the innovative process oriented approach to writing assessment was contrasted with the other product & process oriented tools in order to come up with the advantages of it over the other tools. Finally, there was a brief account of the merits of analytic assessment of ESL/EFL writing in the form of portfolios in comparison with the holistic approach to assessment that totally focuses on the requirements of the learners in ESL/EFL contexts to learn better and to be evaluated fairly and authentically.

References

1. Amiran, E. & Mann, J. (1982), *Writing Composition: Literature Synthesis and Report*. Portland, OR: Northwest Regional Educational Laboratory.
2. Adams, D.M. & Hamm, M.E. (1992) "Portfolio Assessment and Social Studies: collecting, selecting and reflecting on what is significant". Social Education 56. WPA.
3. Allen, M. & Yancey K. B. (1997) "Outside Review of Writing Portfolios: An On-Line Evaluation". *WPA, 20*(3), 64-88.
4. Arter, J.A. & Spandel,V. (1992) "Using Portfolios of Student Work in Instruction and Assessment". Instructional Topics in Educational Management, National Council on Measurement in Education, Spring: 36-44.
5. Brown, R. (1978) "What We Know Now and How We Could Know More about Writing Ability in America.". *Journal of Basic Writing, 1*(4), 1-6.
6. De Fina, A. (1992), *Portfolio Assessment: Getting Started*. New York: Scholastic Professional Books.
7. Edelsky, C. & Smith, K. (1989), *Is that writing - or are those marks just a figment of your curriculum*. Washington, DC: National Education Association: 183-193.
8. Fader, D. (1997) "Writing Samples and Virtues". Greenberg et al: 79-92.
9. Gomez, L. (1996) "The Bilingual Research Journal", *20* (2), 209-233.
10. Graves, D. H. (1984), *A researcher learns to write*. In M. Rose. New York: Guilford Press: 1-18.
11. Hamp-Lyons, L. & Zhang, B. (2001), *World Englishes: Issues in and from academicwriting assessment*. Cambridge: Cambridge University Press: 101-116.
12. Hanson, P. (1993) "Writing Assessment and The Self". Journal of Basic Writing 1.5.
13. Harris, J. (1997) "A Teaching Subject: Composition Since 1966". Upper Saddle River: Prentice.
14. Harris, M. (1985), *Diagnosing writing process problems*: A pedagogical application of speaking-aloud protocol analysis. In M. Rose (Ed.). When a writer can `t write. New York: Guilford Press: 166-181.
15. Hillocks, G. (1984) "Research on written composition." Urbana, IL: ERIC Clearinghouse on Reading and Communication.
16. Hull, G. & Mike R.(1990) "This Wooden Shack Place: The Logic of an Unconventional Reading". Conference on College Composition and Communication 41: 287-98.
17. Huot, B. (1996) "Toward a New Theory of Writing Assessment". Conference on College Composition and Communication 47: 549-567.
18. Kroll, B. (1990), *What Does Time Buy? ESL student performance on home versus class compositions*. Cambridge University Press.
19. Lucas, C. (1992) "Introduction: Writing Portfolios. Changes and Challenges" Portfolios in the Writing Classroom: Urbana: NCTE, , 1-12.
20. Murphy, S. (1996), *Portfolio Approaches to Assessment: Break-through or More of the Same?* Assessment of Writing: Politics, Policies, Practices. New York: MLA,. 284-301.
21. Newcomb, J. S. (1977) "The influence of readers on the holistic grading of essays". Unpub. University of Michigan.
22. Shaughnessy, M. (1977) "Errors and Expectations". New York: Oxford UP.
23. Shaw, S. D. (2003), *The benefits of analytical assessment in relation to the IELTS examination*. Cambridge ESOL Research Notes (2003) Cambridge and New York: Cambridge University Press.
24. Smith, M. (1991) "Put to the test: The effects of external testing on teachers". Educational Researcher, 20: 5: 8-89.
25. White, E. (1985) "Pitfalls in the Testing of Writing". San Francisco. Jossey-Bass.
26. Williamson, M. (1994) "The Worship of Efficiency: Untangling Practical and Theoretical Considerations in Writing Assessment". Assessing Writing.
27. Yancey, K. B. (1999), *Looking Back as We Look Forward: Historicizing Writing Assessment as a Rhetorical Act*: Conference on College Composition and Communication 50. New York: MLA.
28. Zhang, W. (1999) "The rhetorical patterns found in Chinese EFL student writers' examination essays in English and the influence of these patterns on rater response". Unpub. Hong Kong Polytechnic University.