

## The Fundamentals of Assessing EFL Writing

Dr. Touria Drid\*

thouriadrid@gmail.com

Developing Psycho-educational Practices Lab(DPEP)

Kasdi Merbah University (Algeria)

Receipt date : 06-02-2018; Acceptance date : 19-02-2018; Publication date: 01-06-2018

**Abstract:** Given the significance of writing ability in academic and professional contexts, assessing this skill has become one of the widely discussed and researched issues in language education. In general terms, it is observed that assessing writing proficiency, which constitutes an integral part of language instruction, is at times said to be randomly performed owing to the instructors' insufficiency of theoretical grounding, inadequate training in this activity or to contextual constraints. The case of assessing writing in English as a foreign language (EFL) is no exception. Hence, EFL writing assessment scholars highlight the need for promoting assessment literacy among its practitioners. The present paper discusses the fundamentals of EFL writing assessment. Its goal is to provide a resource that writing instructors can draw on to enhance their knowledge, skills and practice of this pedagogical task. The paper is also meant to assist in equipping trainees with the requisite knowledge base on the field of assessment.

**Keywords:** assessment, foreign language, test, writing skill.

### Resumé

Compte tenu de l'importance de la capacité d'écrire dans les contextes académiques et professionnels, l'évaluation de cette compétence est devenue l'un des problèmes largement discutés et étudiés dans l'éducation des langues. D'une manière générale, on constate que l'évaluation de la compétence en écriture, qui fait partie intégrante de l'enseignement des langues, est parfois effectuée au hasard en raison de l'insuffisance théorique des instructeurs, d'une formation insuffisante à cette activité ou de contraintes contextuelles. Le cas de l'évaluation de l'écriture en anglais langue étrangère (EFL) ne fait pas exception. Par conséquent, les spécialistes de l'évaluation de l'écriture EFL soulignent la nécessité de promouvoir l'alphabétisation évaluation parmi ses praticiens. Le présent document traite des principes fondamentaux de l'évaluation de l'écriture EFL. Son but est de fournir une ressource sur laquelle les instructeurs en rédaction peuvent s'appuyer pour améliorer leurs connaissances, leurs compétences et leur pratique de cette tâche pédagogique. Le document vise également à aider les stagiaires à acquérir la base de connaissances requise dans le domaine de l'évaluation.

**Mots-clés:** évaluation, langue étrangère, test, compétences en écriture.

\* Corresponding author

## Introduction

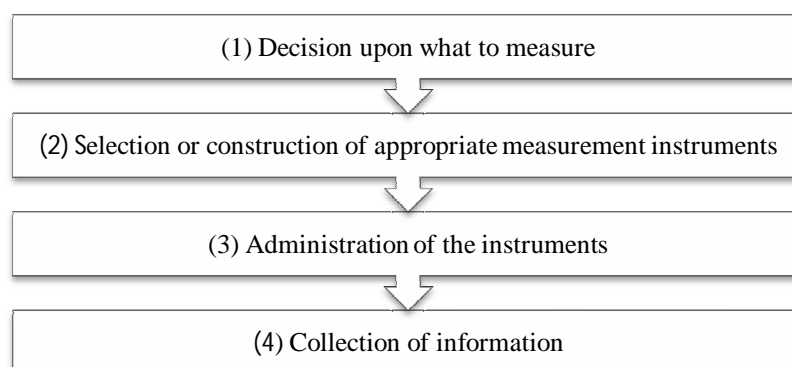
At the heart of educational discussions lies the issue of writing instruction, being a major aspect of literacy. Good writing performance comes to the fore in the academia because it is an indication of students' ability to communicate, critically grasp and display the knowledge transmitted to them in various disciplines (Adler-Kassner & O'Neill, 2010). Even beyond academic circles, writing is cardinal in almost every profession. As the role of writing becomes prominent, the need for an adequate way to assess writing proficiency in language learning contexts appears to be compelling in order to make the right inferences about writing ability and to take subsequent pedagogical and curricular decisions. Against this background, assessment of English as a foreign language (EFL) writing has emerged as one of the heated themes of language teaching and has intrigued substantial research on its various theoretical and practical aspects.

Pertaining to the knotty realm of evaluation, assessing EFL writing is fully-grounded in a multifaceted theory and requires consistency to yield its desired outcomes. However, observation of current routines of EFL writing instructors in various educational contexts indicates that their assessment practices are detached from explicit rationales. In spite of the availability of extensive and technically elaborate literature on writing assessment issues, it appears that EFL writing teachers might be better enlightened about the field if more practical resources are offered. Therefore, the need for succinct and handy material becomes manifest. In this paper we undertake the task of providing a resource roadmap of the theory of EFL writing assessment, which may serve in improving current practices in EFL teaching contexts. The paper first defines the fundamental concepts in assessment, its types and its approaches within the framework of writing instruction. At the heart of the discussion, the standards of assessment are elucidated and the variety of writing assessment methods is explored, including both traditional and inventive techniques. The ultimate objective is to minimize the distance between theory and practice and to address in-service EFL writing instructors' need for a plain and functional resource and even contribute to a more efficient training in writing assessment for pre-service teachers.

## I. Language Assessment: Terminological Distinctions

In the literature, misconception and fuzziness often surround what assessment is. This is attributable to the overlap of the sense of assessment with some related terms. To remove ambiguity, first, a terminological clarification is necessary. Ur (1991) provides a very broad distinction between *assessment*, *evaluation* and *appraisal* in language learning. He explains that the first is connected to the learner, in contrast to the second and third, which relate to courses and teachers respectively. More specifically, for specialists, assessment is regarded as an orderly approach to collecting information about a learner's advancement and accomplishment in learning at different points of the course. This is done on the basis of a range of sources of evidence such as formal tests, essays, interviews, questionnaires, standardized examinations or even informal observations (Hyland, 2003; Richards & Schmidt, 2002; Thornbury, 2006).

As for *evaluation*, its scope is relatively larger than learner achievement. In this connection, Harris and McCann (1994) clarify that it extends to embrace all factors affecting the learning process, including syllabus objectives, course design, materials, methodology and teacher performance. In the same vein, Williams (2003) makes an interesting elaboration. He states that the disparity between assessing and evaluating lies in that assessment covers four interrelated measurement processes. These are represented in Figure 1. Conversely, evaluation entails using the information gathered through such measurements to make judgments about learners and their advancement in terms of learning outcomes.



**Figure 1: Assessment sub-processes**

A further term which is often confused with assessment is *testing* on account of their shared focus on gathering data to estimate learning. Some authors use them interchangeably (Bachman and Palmer, 1996), while others make some distinctions. Brown (2004) differentiates them on the basis of their formality, scope and performer. Tests are planned methods of evaluation prearranged by teachers. By contrast, assessment is more or less an ongoing process which can be incidental or intended and which can be undertaken by the teacher, the peers or the learners themselves. It appears then that tests form an important subset of assessment. Weigle and Malone (2016) hold a slightly dissimilar view. For them, the two terms are positioned on stakes continuum. Assessment is confined to classrooms, has low stakes and is developed by teachers, while testing has higher stakes, is developed by professionals and is standard and applied in larger contexts such as schools, districts, states and so on. In sum, assessment in language learning is seen as evaluation which focuses on measuring learners' performance and which makes use of tests, among other methods. Despite that the aforementioned distinctions are strictly clear-cut, some researchers tend to use the discussed terms interchangeably. It is important then to consider what sense is meant in any discussion of assessment issues.

## II. Types of Assessment

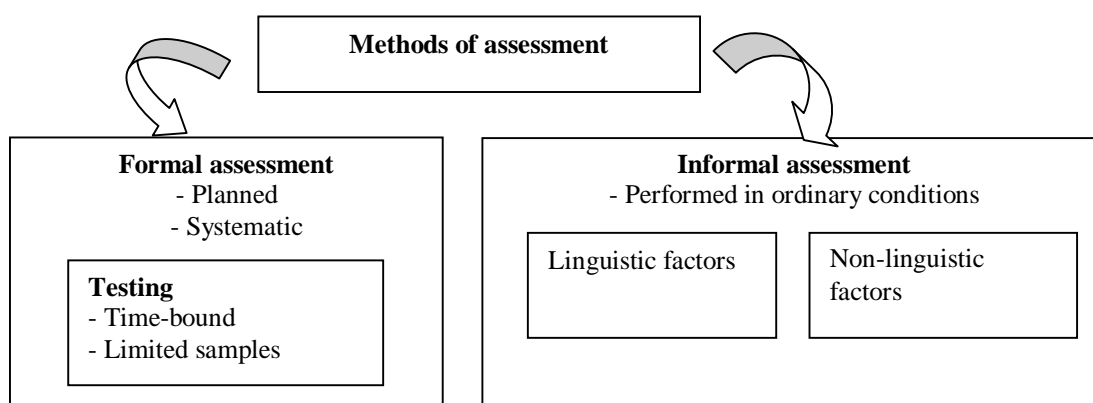
There is no one way in which assessment can be conceived. The diversity of assessment forms and their intersection reflects underlying beliefs about when to measure, how to measure and why to measure. The diversity of assessment types is not restricted to EFL writing but applies to all subjects.

**A. Points to Assess Learning.** Assessment can be undertaken at various stages in a course to accomplish well-defined objectives. According to Shermis and Di Vesta (2011), three essential points can be advantageous for assessing learners: *before learning*, *during learning* and *after learning*. When assessment is undertaken before learning, the teachers seek to decide on the readiness of learners to take a given course based on their current level of knowledge and lacunae. This evaluative account permits teachers to plan content and the level of materials to be used in teaching. When assessment is arranged during learning, the goal is to screen student development along with the employed instructional tools. Insights can be gained on the effectiveness of instruction, necessity of materials' reviewing and pinpointing individual progress of learners. Finally, when assessment takes place after learning, the tendency is to draw on assessment results in making different interpretations and decisions. At this stage, grades can be assigned, instructional and curricular revisions can be effected, and

curriculum efficacy can be determined against school or community objectives or against subject-matter standards.

**B. Methods of Assessment.** Scholars set two broad perspectives for the manner to conduct assessment: *Formal assessment* and *informal assessment*. Obviously, outlining formal assessment essentials is an oversimplification of this massively rich area of research. In some discussions, formal assessment is seen as equivalent to testing (Harris & McCann, 1994). However, some researchers consider them dissimilar. According to Brown (2004), for example, formal assessment includes “exercises or procedures specifically designed to tap into a storehouse of skills and knowledge. They are systematic, planned sampling techniques constructed to give teacher and student an appraisal of student achievement” (p.6). For him, all tests are formal assessment, but there are two features which distinguish them: tests are time-bound and are based on small samples of behaviour, while assessment is not. On these grounds, formal assessment research seems at times to be restricted to discussing testing issues alone.

Informal assessment, on the other hand, is performed by teachers in ordinary conditions and normal classroom environments. In this connection, Harris and McCann (1994) explain that every intuitive evaluation of students’ performance of various abilities is part of this kind of assessment. Teachers can straightforwardly see whether students encounter difficulties, what attitudes they have towards learning, how much effort they exercise and how much they are involved in class work. For Brown (2004), informal assessment involves all incidental, spontaneous remarks or impromptu feedback to the student without necessarily compiling results or taking them as a basis of subsequent decisions about competence. Even the very insignificant comments on papers, feedback on draft work or a suggestion on specific strategies are entailed. It is important to note that informal assessment can consider both linguistic (in case language is being assessed) and non-linguistic factors. Harris and McCann (1994) see that non-linguistic factors, though difficult to evaluate, constitute an important section of what goes on in the classroom. It is the task of teachers then to elaborate consistent methods for the appraisal of such factors as learner attitude, cooperativeness, independence, creativity and presentation. The intricacy of the assessment methods is summarized in Figure 3



**Figure 2: Methods of assessment**

**C. Test Purposes.** Broadly speaking, assessment of language by means of systematic testing is meant to be the first step towards decision making at classroom or institutional level. Test designers and teachers can undertake testing for multiple purposes. Two cardinaly distinct paradigms of testing are available: *norm-referenced testing* and *criterion referenced testing*.

The first entails the comparison of individuals' performance to each other. Based on the principle of fairness, the test point in this case is to rank the learners for the purpose of distributing limited resources or positions (e.g. grants, job vacancies). The second paradigm compares individuals' performance to given standard(s) in order to determine whether they have met some pre-established criteria or instructional objectives (Fulcher, 2010).

Another axis on which testing goals can be placed is the *formative-summative* continuum, that is, as a way to probe into strengths and weaknesses during a course or as a means to sum up learner attainment at the end of a course. Cohen (2001) puts this within a larger taxonomy, which encapsulates three broad categories of test functions: administrative, instructional or research functions. Within these categories, the following specific test types are very common:

- *Placement tests*: Aim at supplying information that will help allocate students to proper classes.
- *Diagnostic tests*: Used to spot students' strengths and weaknesses.
- *Achievement tests*: Permit learners to exhibit the growth they have made in their course.
- *Performance tests*: Give information about students' ability to perform particular tasks, usually associated with known academic or workplace requirements.
- *Proficiency tests*: Used to assess a student's broad level of competence, usually to offer certification for employment, university study, and so on.

These extensive categories represent the chief rationales for assessment work, but teachers might utilize tests for other supplementary purposes, such as enhancing learners' motivation, providing exercise opportunities for national or international exams, collecting information about subsequent teaching, or appraising the success of their methods, tasks, or materials (Hyland, 2003).

#### **IV. Assessing Writing**

Related literature and everyday practices of language teachers in a range of educational levels and settings indicate that writing instruction is a very exigent task (Hyland, 2003; Kroll, 1990; Tribble, 1996; Raimes, 1983; Rivers, 1981). The instructors' challenges multiply when it comes to assessing writing. Indeed, specialists argue that the evaluation of a learner's written product, however simple it might appear to be, is undoubtedly a source of enormous frustration for language teachers (Brown, 2001; Neff-Lippman, 2012). It is held that in comparison with teachers of other subjects, where collecting information on learners' grasp and proficiency of given subjects and evaluating is clear-cut because the standards of correctness are themselves straightforward, assessors of writing face manifold complexities.

Weigle (2002) attributes the complicatedness of writing assessment to two principal factors. On the one hand, the variety of writing purposes, styles and genres makes writing ability so extensive. On the other, the bias of the judgments made on samples of writing is most of the time inevitable. In fact, Williams (2003) considers the issue of standard inconsistency in assessing writing the major source of strenuousness because it engenders uneven evaluation. He also sees that the object of writing assessment, i.e. "what to measure", makes teachers confused between assessing the content of writing (topical knowledge), the general ability to write or learners' performance on specific writing tasks.

#### **V. Principles of a Systematic Assessment of Writing**

Assessment of writing is at times performed in a haphazard, unsystematic way. White (1986) takes a very extreme position, stating that a lot of writing instructors are acquainted with roughly nothing about writing assessment and that great numbers of them are arrogant about their lack of knowledge. To lessen the laboriousness of assessing the writing skill, it is

vital for assessors to gain sufficient literacy on the workings of systematic assessment. For the performance of writing assessment to be efficient, the fundamentals of general assessment ought to be attended to, obviously with a careful attention to the nature and intricacies of the writing skill itself. The task of instructors is to balance a number of principles, which are in fact applicable to all other subjects, for optimum outcomes. These include validity, reliability, practicality, authenticity and accountability.

**A. Validity.** The concept of validity is positioned at the centre of assessment enterprise. “Validity refers to whether the test actually measures what it purports to measure” (Cohen, 2001, p. 525). This essentially means that a test, or any assessment method, ought first to be clear about what ability it is measuring and ought to really measure it. In the context of assessing writing, the issue that comes to the fore is determining what to teach and what to test. It is believed by writing assessment scholars that writing instructors have in the first place to teach writing and in the second place to consider what they are *actually* teaching to make their assessment work valid (Williams, 2003). Based on Brown’s (2004) delineation of the aspects of writing, it seems crucial when determining the ability to be taught and assessed to consider which genre the student is supposed to produce, which micro- or macro-skill is addressed as well as which form of performance is required. It is therefore stressed that instructors have to be unambiguous about the domain and form of writing to which generalization is made in assessment; otherwise validity would be threatened.

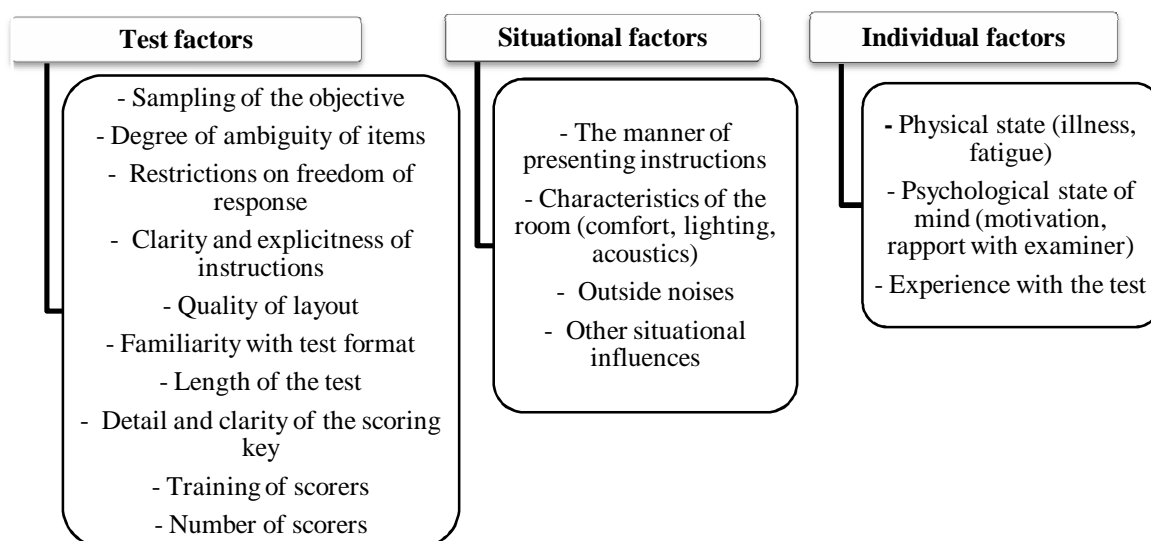
Queries are raised regarding the extent to which validity can be achieved when assessing EFL writing. Many assessment scholars relate this to whether assessment of writing ability is *direct* or *indirect*. Brown (2004) clarifies that indirect (or objective) assessment requires that test-takers do some writing-related activity (e.g. answering questions about grammar, usage and punctuation in multiple choice tests). By contrast, direct assessment means that test-takers are involved in really performing the intended writing task (e.g. producing an essay). By testing writing performance directly, Elliot *et al* (1990) argue that validity will not pose problems. Many writing instructors, however, often violate the validity principle using indirect evaluation methods. They build their evaluation of students’ ability to write on the basis of the scores they get from a test on a micro-skill such as spelling or grammar for instance, while such scores are by no means the sole predictors of writing ability. This raises the issue of whether *writing performance* is being assessed or not. Another common practice in EFL classes in which validity is threatened is the structuring of assessment tasks. The most popular of such ill-structured assignments are those involving optional topics for essay writing (e.g. descriptive, narrative expository or argumentative papers). In this form of assessment, instructors are not clear about what they are measuring as the topics differ in terms of ease and rhetorical features (Williams, 2003). Hyland (2003) also points to some threats to validity when assessing writing such as asking students to write in a genre they have not been exposed to, allowing inadequate time for planning and developing a topic, or requiring learners’ to demonstrate specialist knowledge that they do not possess.

**B. Reliability.** The second principle determining methodical EFL writing assessment is reliability. In general terms, reliability has to do with steadiness and precision of test results. If the result varies when the test is administered a second time, it would not be considered reliable. In simpler terms, if the test result is not the same when the test is scored by a different evaluator or on another occasion, this indicates that the test is not a stable tool of measurement. The necessity of constructing consistent tests is stressed, for this is important in making correct and fair deductions and judgments. A significant issue is the bond between reliability and validity. Bachman and Palmer (1996) demonstrate that reliability is an indispensable but not a sufficient condition for validity. A number of writers suggest that it is

better to design a test that is primarily valid, and then search for ways of making it reliable, rather than creating a reliable test and attempting to make it valid (Johnson, 2001).

In the context of assessing writing, reliability is critically important notably for high-stakes tests, those on which placement or passing are based. However, threats to reliability are numerous and often not observed by practitioners. In general terms, Cohen (2001) distinguishes three factors which might influence reliability of assessment: *test factors* (related to the test itself and rating), *situational factors* (related to the conditions of test administration) and *individual factors* (Related to the state of the test takers). Figure 4 expounds on these factors. Indeed, achieving invariant perception of writing ability is a stumbling block in writing classes. Harris and McCann (1994) mention the case of “impressionistic marking”. This involves marking according to indiscriminate scales which do not explicitly state the evaluation criteria. In that case, examiners might produce inconsistent judgments even if the students’ writing is evaluated by the same teacher on different occasions. Such scales as “Excellent writer”, “marginal writer” or “poor writer” remain highly subjective if not attributed unambiguous criteria with detailed *descriptors* of writing ability. The criteria can cover such features as comprehensibility, grammatical accuracy, spelling, and text organization. According to Nation (2009), another way in which reliability is at risk is restricting assessment to one piece of writing, often evaluated by one examiner. For more consistent results, there should be more than one writing task with at least a second scorer, although this might seem impractical.

On the whole, Weigle (2002) argues that reliability must be integrated into the assessment process by establishing standardized procedures with well-defined *modus operandi* for test construction, supervision and scoring, which would lessen bias in case practical problems arise. Some researchers opt for indirect assessment as a tool to diminish disparity in test results on account of their power to reveal test takers’ knowledge of writing sub-skills. However, others dismiss such tests because they are based on correctness at the expense of communicativeness (Hyland, 2003). On the whole, to enhance reliability in EFL writing assessment, East (2008) argues that examiners have to devise cautiously worded and satisfactorily thorough rubrics. Training raters in scoring and determining the extent to which raters concur about rated scores are also crucial.



**Figure 3: Threats to reliability**

**C. Practicality.** Practicality does not relate to assessment itself but to implementation. According to Crusan (2010), practicality refers to the realistic considerations regarding test administration and execution. The available resources are key determinants of practicality in assessment. Eliot *et al* (1990) see that effective assessment has to judiciously balance between the requirements of validity and reliability on the one hand and the restraints of circumstances on the other. Various resources should be considered when attempting to develop an assessment tool and use it. Bachman and Palmer (1996) differentiate between three types of resources: *human resources* (test writers, scorers or raters, test administrators, and clerical support), *material resources* (space, equipment and materials) and *time* (development time and time for specific tasks). More importantly, these resources should be calculated in terms of their *cost*. Nation (2009) argues that practicality is a chief concern in the assessment of writing, especially in contexts where teachers are non-native speakers of English because marking requires elevated levels of language proficiency. Further, if feedback is needed, writing assessment would become exceedingly time-consuming and laborious. To attain practicality in assessing EFL writing, specialists support the use of other supplementary methods like peer assessment and self assessment.

**D. Authenticity.** Authenticity differs somehow from the previous principles in the sense that it is specific to the assessment of language. This principle is a way to evaluate tests in terms of their representation of the target language use (TLU). Bachman and Palmer (1996) define it as the extent of correlation between the traits of test tasks and those of the target language use. When tests are authentic, their results may be generalized further than performance in the test to true language use in non-test domains. For example, an authentic task that measures writing ability is one that requires learner to write meaningfully as is the case in real situations. As a result of authenticity, they add, test takers would have positive perceptions of the relevance of the assessment tool and would therefore react positively to the tasks. Brown (2004) sees that by insuring the authenticity of language tests, one would present natural, non-contrived language which matches real world language. If writing tests are authentic, the type of writing students will produce would simulate the one they encounter in the factual world, at least in terms of genres and communicative purposes (e.g. writing a letter of complaint, producing a touristic brochure, etc). Weigle (2002) observes that in FL learning contexts, where the target language is scarcely used outside the classroom, it might be thorny to hit upon a writing task that presents an authentic writing situation. Thus, test developers in such contexts sometimes allot authenticity less weight than other principles.

**E. Accountability.** In general terms, the principle of accountability concerns the shared responsibility for improving educational practices through insuring quality. In the context of EFL writing assessment, instructors have to be capable of explaining the justification behind the measuring techniques they use and explain the benefits they will bring to learners (Richards & Schmidt, 2002). According to Harris and McCann (1994), teachers are held accountable for the consequences of tests they design in the sense that they have to be able to present stake-holders (learners, parents, institutions and society in general) with comprehensible indication of what advancement has been achieved through assessment and to explain the reasons of malfunction in case no gains are obtained.

#### **IV. Techniques of Assessing EFL Writing**

The assessment of writing ability can be fulfilled using a wide variety of techniques. The continuum of available techniques rests basically on the objective of assessment and the type of performance required.



**A. Indirect Assessment.** In general, indirect (or objective) assessment uses tasks which are not reflective of real target language use situations but are used to make implications on the ability lying behind performance in the test (Richards & Schmidt, 2002). Indirect assessment is a traditional method of assessing writing which was popular in the 1950s and 1960s. Attempting to measure the sub-skills involved in writing, this type of assessment usually employs multiple choice questions, error spotting or other selected response measures (Weigle, 2012). Indirect assessment reflects the accepted ideas about composition of the time, where focus was placed on such features as grammar, usage and punctuation. Although it is recognized to be consistent and easy to administer and score, writing specialists have noticed important limitations of objective assessment: It seems to decontextualize knowledge and meaning making as it does not require real writing. Narrowing the conception of competence, depriving students from revision opportunities as well as excluding rhetorical and contextual considerations in writing are the most noticed drawbacks of this form of writing assessment (Neff-Lippman, 2012). It should be stressed, however, that indirect assessment is highly reliable and practical although it is deficient in terms of validity and authenticity.

**B. Direct Assessment.** Direct assessment, as its name suggests, requires that learners' writing ability is *directly* evaluated. In broad terms, a direct test refers to "a test that measures ability directly by requiring test takers to perform tasks designed to approximate an authentic target language use situation as closely as possible" (Richards & Schmidt, 2002). Indirect tests appraise the key abilities which are thought to be indicators of the target behaviour, but they do not model the behaviour itself, while direct tests seek to reproduce the real eventual behaviour in the test itself (Johnson, 2001). In assessing writing directly, the test tasks involve production of a sample of writing. Through these tests, students show writing competence rather than spot the right answers without production. Reflecting changes in composition theory, this form of assessment supplanted the indirect paradigm and has become widely used since the 1970s. In fact, it is still used in standardized examinations nowadays and is highlighted as a typical form of large-scale assessment. Weigle (2002) asserts that direct tests are the most widespread and the best researched methods in all contexts of language learning.

The form of direct writing assessment is well-defined. Essentially, such measuring devices are administered in a limited time frame (hence the term "timed impromptu writing test"), and the topic is not supplied to writers before the examination. Hamp-Lyons (1991) specifies five additional key features:

- (1) Writers produce one piece of continuous (at least 100 words),
- (2) Writers receive a set of instructions (or *prompt*) but with flexibility given for dissimilar responses,
- (3) Produced samples are read by at least one but normally two or more qualified raters,
- (4) Judgment is tied to common standard (model essays or rating scales),
- (5) Judgment is expressed in numbers.

According to Weigle (2012), both the proper construction of tasks and the appropriate implementation of scoring are important in the use of direct tests, especially to ensure reliability and eliminate variation in the results of assessment. As for the construction of tasks, here are three factors that do influence performance: subject matter (personal Vs non-personal topics, general Vs specialized topics), discourse mode (genre, rhetorical task, cognitive demands) and stimulus material (textual, visual). It is necessary that EFL writing instructors balance these factors in order to make their assessment more systematic and reflective of genuine competence.

Regarding scoring procedures in direct assessment, three approaches are can be utilized: *holistic scoring*, *analytical scoring* and *primary trait scoring*, all of which use a *rating scale* (or a *scoring rubric*). Holistic scoring is developed in such a way that it assesses

writing performance, and it complies with the validity and reliability principles. It starts from the belief that evaluating writing skill does not involve measuring an array of sub-skills, but rather measuring a whole piece of discourse (Williams, 2003). In Holistic scoring, raters give a single score (or point) for the whole script based on trained rater's impression (e.g. 1, 2, 3 or 4). For each point, general overall descriptions of performance are given (descriptors show clear criteria but are usually integrated in a patterned way). The use of such scales requires training raters so that consistent scoring can be achieved, and it is preferred when assessing a large number of tests (Weigle, 2002). Analytical scoring divides writing ability into fundamental elements or criteria (e.g. such as content, word choice, mechanics, organization, grammar, etc) and assesses them independently. Focus is put on traits which are held to be common to all writing. The criteria of assessment are separated and the descriptors for each are supplied independently. Discrete scores are attributed to separate aspects of performance, permitting learners to pinpoint their strengths and weaknesses in precise areas (Brown, 2004). This scale is more appropriate in formative assessment. Primary trait scoring focuses on selected aspects of writing, usually a specific range of discourse (e.g. persuasion or explanation) (Weigle, 2002). The writer's performance on the very particular task at hand is assessed in terms of how much it achieves a given rhetorical goal.

While impromptu timed tests have brought the assessment of writing much rigour, especially in large-scale testing situations, doubts are often raised regarding the faithfulness of this method to reflect learners' real competence. Weigle (2002) argues that direct testing judges a single piece of writing administered under non-ordinary conditions. This seems to present only a partial picture of students' abilities. Further, Neff-Lippman (2012) sees that direct testing discards process and contextual issues and represents a restricted conception of competence. But because direct tests are still widely used, she suggests a number of qualities (e.g. clarity, engagement and audience specification, etc) to be incorporated in their construction for more efficiency.

**C. Alternative Assessment.** In response to the limitations of direct assessment and in line with the changes occurring in composition theory and education at large in the late 1980 and 1990s, alternative methods of assessment have been developed. Writing specialists have become interested in informal classroom assessment methods which cater for learners in a more productive and authentic manner. The alternatives encompasses a wide range of techniques, the chief ones being writing portfolios, protocol analyses, conferences and interviews, journals, peer-assessment, self-assessment and observations.

**1. Portfolios.** A *portfolio* involves a whole record of accomplishments. It is generally defined as "a purposeful collection of work that provides information about someone's efforts, progress or achievement in a given area" (Richards & Schmidt, 2002). In writing assessment contexts, portfolios involve the student's entire writing products which capture his or her full performance, involving even drafts and works selected by learners' themselves (Peñaflorida, 2002). The portfolio approach to writing assessment rests on the premise that writing performance is not uniform. Learners of writing might demonstrate high proficiency in some assignments but not in others. Then, it would be more reasonable to assess a collection of student papers over time and in a variety of genres than to restrict evaluation to one sample (Williams, 2003). In spite of the apparent difficulty of making a consistent assessment of writing through portfolios, Neff-Lippman (2012) argues that their advantage lies in permitting instructors to function as coaches and allowing students to exercise revision of their work in clear rhetorical settings. Further, portfolios involve students in task choice and more essentially give them opportunities to write in authentic contexts. This seems to be a significant step towards learner autonomy. The use of portfolios in assessing writing is subject

to open variation in the modes of accumulating and appraising learner written products (Weigle, 2002).

**2. Protocol analysis.** A *protocol* is “a sample containing observation(s) of a phenomenon which is being described, observed, or measured” (Richards & Schmidt, 2002). Embedded in the cognitively-oriented process approach to writing, the procedure of protocol analysis is counted among the innovative techniques that can be exploited to assess writing in the classroom. In essence, this form of evaluation is “a think aloud activity”. As explained by Peñaflorida (2002), the assessment starts by asking students to record their emerging thoughts all through the writing process. The transcripts are subject to analysis, which can reveal the internal mechanisms of generating ideas, revising and editing texts.

**3. Journals.** A *journal* is a form of writing characterized by extensive freedom and self-reflection on the part of the learner. Brown (2004) defines a journal as “a log (or “account”) of one’s thoughts, feelings, reactions, assessments, ideas, or progress toward goals, usually written with little attention to structure, form, or correctness” (p. 260). Journals are used to elicit regular extensive writing, giving students ultimate autonomy in topic choice and chance to experiment with their abilities in total privacy. According to Stix (2003), journal writing assignments are given to students to see who does or does not understand what is taught. Through writing, misconception comes into clear focus. Writing allows the student to make particular problems explicit. In the context of writing instruction, journals give students opportunities of practicing all aspects of writing, they trigger thinking, and they promote individualization and communication. The assessment of journal entries involves commenting on them in written notes or through conferencing. When assessed, journals go beyond mere scoring to providing constructive feedback to learners.

**4. Conferences and Interviews.** Conversational in nature and rooted in the process approach to writing, *conferences* involve discussion of learner’s written work, portfolios or journals with teachers and peers in order to fine-tune ideas, talk about difficulties, highlight strengths and weaknesses or receive feedback (Thornbury, 2006; Richards & Schmidt, 2002). It is claimed that conferences are a formative assessment tool whose the chief function is to offer affirmative washback. An *interview* is a carefully constructed type of conference in which teachers interrogate students about a specific assignment using focused probes. The use of both conferences and interviews in assessing writing informally calls for caution in order conform to the principles of assessment. Both are shown to be of low practicality, while their reliability rests on a clear specification of objectives and procedures (Brown, 2004).

**5. Observations.** Observations constitute an integral part of teaching. When used as a writing assessment tool, they permit teachers to scrutinize performance as it occurs in its authentic settings and evaluate competence accordingly. Gould and Roffey-Barentsen (2014) show that observations depend on detailed recording with sufficient clarity as to the observed facts on performance. Checklists, scaling rates or anecdotal records can be used in order to itemize the competences to be observed. According to Brown (2004), observation as an alternative assessment tool is conducted in a non-systematic way and covers both verbal and non-verbal behaviours of students in such a natural way that might not be noticed by learners. It is stressed, however, that it should be guided by clear objectives and conducted with relative systematicness.

## 6. Self-assessment and Peer Assessment

Self-assessment is the ability to judge one's own progress. Assessment specialists today assert that learners need to be trained in this mode of evaluation (Thornbury, 2006). Self-assessment enhances learners' reflection about their own work. The theoretical justification of this evaluation mode, according to Brown (2004), is derived from two established principles of second language acquisition: autonomy and intrinsic motivation. In the context of writing, Harris and McCann (2012) see that self-assessment is a direct and resourceful way to probe the problematic areas of language for students. It gives them an elevated degree of control over their own accomplishment and makes them motivated to challenge themselves towards more skillful writing. The benefit of self assessment, according to Hathaway (2014) is that students take possession of the evaluation process. This would add a personal dimension to learning and reduce sensitivity to feedback. Peer-assessment, which involves all activities in which learners evaluate each other's performance, is seen as a very important formative evaluation source that writing instructors can draw on. It rests on the principle of cooperative learning (Brown, 2004). Peer assessment concretizes the role of authentic audience in writing and encourages the growth of interpersonal intelligence based on clear expectations (Noel, 2017). But for the use of self-assessment and peer assessment to be efficient and productive, it is necessary that the learners themselves are guided with clear objectives and criteria determined by the teacher.

On the whole, the tendency in writing theory now is to use alternatives in the classroom as a way to triangulate measures of writing competence, a reaction to the long-established one-shot method of essay testing. Alternative assessment has the merit of being more formative, authentic and process-oriented. These are indicators of an elevated extent of communicativeness and fairness (Lanski & Verbruggen, 2010).

## Conclusion

We have attempted in this paper to provide a holistic picture of the essentials of EFL writing assessment, whose understanding might enhance current practices. This is in a way an attempt to reaffirm the issue of promoting a sufficient and urgent writing assessment literacy among EFL writing instruction practitioners and developing their pedagogical repertoire with inventive active methodologies. Against this background, some recommendations are provided to mend the malfunctioning parts of the assessment apparatus and to eliminate the widely observed unproductive, static assessment routines in EFL writing classes. The following pointers are provided:

- EFL writing teachers ought to be fully acquainted with the technical distinctions and the wide variety of assessment purposes.
- For EFL writing assessment to yield its desired outcomes, the principles of general assessment have to be observed.
- While institutional restrictions may not always permit teachers to use the assessment tasks they would favour, practicality issues should be cautiously treated in such a way that validity is ensured through assessing writing performance.
- Reliability issues should not seek consistency of measurement at the expense of preparing learners for a more authentic use of the target language.
- Writing assessors have to receive sufficient training in test construction methods and scoring procedures in both pre-service and in-service contexts in order to ensure fair and effective assessment. This can be achieved via enhancing a collaborative exercise and designing appropriate benchmarking of texts to achieve consistency.
- In order for writing assessment to have a positive influence on teaching and to promote learner progress, alternative tools have to be integrated into EFL writing classes.

In the end, it should be stated that the enterprise of assessing EFL writing follows an intricate network of principles and approaches derived from the vast field of general assessment. These are tailored to fit the nature of the writing skill and the context of language teaching simultaneously. In fact, an appropriate practice of assessing EFL writing must be grounded in a thorough knowledge of assessment fundamentals. Not equipped with adequate assessment literacy, EFL writing instructors may fall in the trap of rendering this activity a mere psychometric, statistical process, which discards important aspects of language learning and which provides no direct feedback to teaching. Written language is in the first place a medium of communication, and if assessing writing does not help in preparing EFL writers for wider communications, the role of writing programmes in developing literacy would be negligible.

### References

- AdlerKassner, L. & O'Neill, P. (2010). Reframing writing assessment to improve teaching and learning. Logan, UT: Utah State University Press.
- Bachman, L.F. & Palmer, A.S. (1996). Language testing in practice: Designing and developing useful language tests. Oxford: Oxford University Press.
- Bridgeman, B., & Carlson, S. (1983). Survey of academic writing tasks required of graduate and undergraduate foreign students (TOEFL Research Report No. 15). Princeton, NJ: Educational Testing Service.
- Broughton, G., Brumfit, C., Flavell, R., Hill, P. & Pincas, A. (1980). Teaching English as a foreign language (2nd ed.). London and New York: Routledge.
- Brown, H. D. (2001). Teaching by principles: An Interactive approach to language pedagogy. New York: Longman Inc. NY: Longman.
- Brown, H. D. (2004). Language assessment: Principles and classroom practices. White Plains, NY: Pearson Education.
- Clapham, C. (2000). Assessment and testing. In M. Byram (Ed.) Routledge encyclopedia of language teaching and learning. London: Routledge, pp 48-53.
- Cohen, A.D. (2001). Second language assessment. In M. Celce-Murcia (ed.), Teaching English as a second or foreign language (3rd ed., pp. 515-534). Boston: Heinle.
- Crusan, D. (2010). Assessment in the second language writing classroom. Ann Arbor, MI: The University of Michigan.
- East, M. (2008). Dictionary use in foreign language writing exams: impact and implications. Amsterdam: John Benjamins Publishing Company.
- Elliot, N., Plata, M., & Zelhart, P. (1990). A program development handbook for the holistic assessment of writing. Baltimore: University Press of America.
- Fulcher, G. (2012). Assessment literacy for the language classroom. Language Assessment Quarterly, 9(2), 113–132.
- Fulcher, G. (2010). Practical language testing. London: Hodder Education.
- Gould, J., & Roffey-Barentsen, J. (2014). Achieving your diploma in education and training. London: Sage Publications Ltd.
- Hamp-Lyons, L. (1991). Scoring procedures. In L. Hamp-Lyons (Ed.), Assessing second language writing in academic contexts (pp. 241-276). Norwood, NJ: Ablex
- Harris, M. & McCann, P. (1994). Assessment. Oxford: Heinemann Publishers.
- Hathaway, J. (2014). Writing strategies for fiction. United States: Shell Educational Publishing.
- Hyland, K. (2003). Second language writing. New York: Cambridge University Press.
- Hyland, K. (2013). ESP and writing. In Brian Paltridge, B. & Starfield, S. (Eds.) The handbook of English for specific purposes (pp. 95-114). Oxford: Wiley-Blackwell.

- Johnson, K. (2001). *An introduction to foreign language learning and teaching*. Harlow: Longman Pearson.
- Kroll, B. (Ed.) (1990). *Second language writing: Research insights for the classroom*. New York, NY: Cambridge University Press.
- Lenski, S. D., & Verbruggen, F. (2010). *Writing instruction and assessment for English language learners K-8*. New York: Guilford Press.
- Nation, I. S.P. (2009). *Teaching ESL/EFL reading and writing*. New York : Routledge
- Neff-Lippman, J. (2012). *Assessing Writing*. In I.L. Clark et al (Eds). *Concepts in composition: Theory and practice in the teaching of writing* (2<sup>nd</sup> Ed, pp. 145-167). New York: Routledge.
- Noel, E. (2017). *From a Culture of Testing to a Culture of Assessment: implementing Writing Portfolios in a Micro Context*. In R. Al-Mahrooqi, C. Coombe, F. Al-Maamari & V. Thakur (Eds). *Revisiting EFL assessment: Critical perspectives* (pp. 221-236). Switzerland: Springer.
- Peñaflorida, A. H. (2002). *Non-traditional forms of assessment and response to student writing: A step toward learner autonomy*. In J.C. Richards & W. A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice* (pp. 344-353). Cambridge: Cambridge University Press.
- Raimes, A. (1983). *Techniques in teaching writing*. New York: Oxford University Press.
- Richard, J.C. & Schmidt, R. (Eds.). (2002). *Longman dictionary of language teaching and applied linguistics* (3rd ed.). London: Longman.
- Rivers, W. M. (1981). *Teaching foreign language skills* (2<sup>nd</sup> ed.). Chicago, IL: University of Chicago Press.
- Shermis, M. D., & Di Vesta, F. J. (2011). *Classroom assessment in action*. Lanham Md.: Rowman & Littlefield Publishers
- Stix, A. (2003). *Social Studies Strategies for Active Learning*. Shell Educational Publishing.
- Taylor, L. (2009). *Developing assessment literacy*. *Annual Review of Applied Linguistics*, 29, 21-36
- Thornbury, S. (2006). *An A-Z of ELT: A dictionary of terms and concepts used in English language teaching*. Oxford, UK: Macmillan Education.
- Tribble, C. (1996). *Writing*. Oxford: Oxford University Press.
- Ur, P. (1991). *A Course in Language Teaching*. Cambridge: Cambridge University Press.
- Weigle, S. C. (2012). *Assessment of writing*. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-7). Hoboken, NJ: Blackwell Publishing.
- Weigle, S.C. (2002). *Assessing writing*. Cambridge, UK: Cambridge University Press.
- Weigle, S.C. & Malone, M. (2016). *Assessing English for academic purposes*. In K. Hyland and P. Shaw (Eds.), *The Routledge handbook of English for academic purposes* (pp. 608-620). London, UK: Routledge.
- White, E. (1986). *Teaching and assessing writing*. San Francisco, CA: Jossey-Bass.
- Williams, J. D. (2003). *Preparing to teach writing: Research, theory, and practice* (3rd ed.). Mahwah, NJ: Erlbaum.