

## Assessment of EFL Writing at the Arab University Level under Scrutiny: A Need for a Writing Rubric

Sami A. Al-asadi (M.A, England, Ph.D. Australia)

Al-Qadisiya University \ College of Education

E-mail: san\_san53@ yahoo.com

### Abstract:

At the Arab tertiary level, the impressionistic assessment and marking scheme of students' written products in their standardised examination is the norm. However, such scheme remains problematic. It is criticised for being subjective based on teachers' preferences rather than on demonstrable and predefined criteria to guide and justify their scoring decision. Considering this situation, the purpose of this paper is to develop a criterion-referenced writing rubric as an alternative to the current impressionistic assessment. The rubric is elicited from the best merits of existing relevant literature. To achieve this end, this paper examines the arguments for and against both analytic and holistic evaluation approaches currently in use. This paper is significant in several ways: First, the proposed rubric is hoped to provide writing teachers in the Arab world in general and in Iraq in particular with an instrument to guide them to do the job of assessing and scoring students' written work resulting in more reliable and valid results. Second, to date, there are only few studies conducted in this field at the Arab university-level. This paper is therefore an endeavour to contribute toward expanding the existing literature. Third, it opens up opportunities for further research to investigate how far the application of the writing rubric is useful in comparison to the current impressionistic assessment. This paper highlights that the proposed rubric has some practical implications. It closes with a conclusion.

**Key words:** Analytic scoring scheme; A proposed scoring rubric; Holistic scoring scheme; Writing scoring rubric.

### المستخلص:

التقييم الانطباعي في المستوى الجامعي العربي، واسلوب تصحيح نتائج الطلبة التحريرية في امتحاناتهم الموحدة هو المبدأ المتبع. لكن مثل هذه الخطة تتطوي على مشكلة. فقد جرى انتقادها كونها ذاتية وتستند على ميول المدرس وليس على معايير محددة مسبقا وقابلة للأثبات لكي توجه وتبرر قراراتهم في احتساب درجات الطلبة. اخذين بنظر الاعتبار هذه الحالة، فان الغرض من هذا البحث هو تطوير أداة ذات معايير محددة كبديل للتقييم الانطباعي الحالي استخلصت من أفضل جوانب الدراسات الحالية ذات العلاقة. ولتحقيق هذه الغاية، استقصت الدراسة النقاشات المؤيدة والمعارضة لكلا اسلوبي التقييم، التحليلي والشمولي المستخدمين حاليا. تنبع اهمية هذه الدراسة من عدة جوانب: الاول: تزود الاداة المقترحة لمدرسي الكتابة في العالم العربي عموما والعراق على وجه الخصوص بأداة لتوجيههم للقيام بمهمة تقييم وتصحيح اعمال الطلاب التحريرية مما ينتج عنها نتائج ذات مصداقية ومقبولية. ثانيا: هناك القليل من الدراسات التي اجريت في هذا المجال على المستوى الجامعي العربي. لذلك فان هذه الدراسة هي محاولة لتوسيع الدراسات الحالية في هذا المجال. ثالثا: كما انها كشفت عن العديد من الفرص للقيام بالبحوث والتقصي عن اهمية تطبيق الاداة المقترحة في تقييم وتصحيح الكتابة مقارنة بالتقييم الانطباعي الحالي. وأبرز هذا البحث ان أسلوب التقييم المقترح لها نتائج عملية. واختتم البحث باستنتاج الكلمات المفتاحية: قاعدة تصحيح الكتابة، خطة التصحيح الشاملة، خطة التصحيح التحليلية وقاعدة التصحيح المقترحة.

## 1. Introduction:

The evaluation of students' writing ability is a topic that has recently received much and varied attention in educational circles because its results are used for a variety of administrative, instructional, and research purposes (Huot, 2002). Researchers and teachers develop various criterion-referenced evaluation tools to evaluate students' performance. One of the most commonly used of these tools is the rubric. Rubric in Merriam Webster's collegiate dictionary (2008) is a scoring guide that describes criteria for student performance and differentiates among different levels of performance within those criteria. Because rubrics set forth specific criteria, define precise requirements for meeting those criteria, and often assign numerical scores to each level of performance The use of rubrics has come into prominence during the past decades as the alternative to traditional measures of writing ability, particularly standardized tests and as a result of a general dissatisfaction with teacher's subjective evaluation and grading procedures on the grounds that they are invalid (Charney, 1984). Relevant literature maintains that there are several benefits that can be achieved through applying scoring rubrics. A rubric with well-structured criteria and sufficiently delineated descriptors that distinguish between students' performance levels has the potential to bring increased consistency and reliability to the task of judging and scoring their work and thus subjectivity becomes more objective (Spandel, 2006; Jonsson & Svingby, 2007). This is based on the assumption the rubric gives the teacher a regulatory device for what to look for when evaluating and scoring students' performance (Morrison & Ross, 1998; Wiggins, 1998). In Spandel's (2006) view, working on agreed and defined criteria likely increases the inter-rater reliability leading to a fair mechanism of scoring in comparison with the case where no rubric is used. Another frequently mentioned advantage of adopting a rubric is that it can be an effective tool in helping students know exactly what is expected of them and redirecting their efforts to satisfy the targeted performance levels "engaging young writers actively in the use of criteria results not only in more effective revisions but in superior first drafts" (Hillocks, 1986, p.160). Along a similar line of argument, Soles (2001, p.3)

states that "research suggests a positive correlation between a students' knowledge of evaluative criteria and the grade the student gets on an essay to which that evaluative criteria is applied". To this end, students should be aware of the evaluation criteria prior to the completion of the writing task to provide them with insights into how they are expected to do well and see why a given piece of writing does or does not meet the evaluation criteria (Montgomery, 2000).

## 2. Statement of the Problem

An important feature of the current evaluation and scoring method used at the Arab tertiary level is that it is done impressionistically. The writing teacher holistically reads each essay and assigns it a single score out of 100-point scale for its overall quality. Such scoring is not guided and justified by explicit and pre-determined criteria. It exclusively draws on personal preferences and standards teachers have developed out of their experience with essay scoring, and/or out of comparing students' test essays to each other (Al-Abed al-Haq and Ahmed, 1994; Barkaoui, 2007).

Such scoring method has two disadvantages. First, each individual teacher has his/her unique standards against which he/she evaluates students' writing. With such human factors, idiosyncratic variations with which different teachers and the same teacher may evaluate their students' written products are highly expected to occur. Such uncontrollable discrepancies has important implications for the reliability of their scoring job (Pearson, 2004; Casanave, 2007). Second, as a result, making a fair and accurate evaluation of students' writings has actually become difficult to reach. This may threat the reliability of the results of the writing test and undermine their validity as a measure to test and measure students' actual writing abilities (Cooper, 1997).

## 3. Significance of the Study

In the light of the above-cited background, this paper is significant for a number of reasons.

First, it emerges out of a pressing need to develop a writing rubric with explicit criteria to improve the reliability and validity of assessing and scoring Arab EFL students' actual writing abilities. Second, to date, there are

only few studies conducted in this field at the Arab university-level. This study is an endeavour to investigate an under-researched area and thus it contributes towards extending and enriching the relevant emerging literature. Third, it opens up opportunities for further research to investigate how far the application of the rubric is useful in comparison to the current impressionistic assessment.

#### 4. Literature Review

In L2 writing assessment, the first analytic scoring rubric to appear is the ESL Composition Profile (Jacob, Zinkgraf, Wormuth, Hartfiel & Hughley, 1981). It has 5 major constructs that the rater looks for when assessing the student's writing ability. They include content, organization, vocabulary, language, and mechanics with each one having four rating levels of very poor, poor to fair, average to good, and very good to excellent. Each level is matched with clear and distinctive descriptors of the writing proficiency as well as a numerical scale (See Appendix1). The Jacobs, et al.'s (1981) 5 constructs are weighted differently: content 13-30, organization 7-20, vocabulary 7-20, language 5-25 and mechanics 2-5.

Weir (1983) developed another rubric that includes seven constructs used to develop an evaluation instrument in ESL writing setting. They are: relevance and adequacy of content, compositional organization, cohesion, adequacy of vocabulary or purpose, grammar, punctuation and spelling. Perkins (1983) proposes a set of criteria students need to meet to produce quality writing including:

- a-The clarity of the topic and how it is stated, developed, and supported.
- b-The effectiveness with which an issue has been raised, treated, and resolved.
- c-Enough support and development of the topic for the reader.
- d-The extent to which the writer has considered the needs of his readers.
- e-The degree of grammatical and lexical cohesion and overall coherence of the piece.
- f-The degree and effectiveness of the use of rhetorical devices.

For Reid (1993), content, purpose and audience, rhetorical features (organization, cohesion, unity), and mechanics (sentence structure, grammar, vocabulary) are the constructs of good writing. Cohen (1994) regards content,

organization, register (appropriateness of level of formality), style (sense of control and grace), economy, accuracy (correct selection and use of vocabulary), appropriateness of language conventions (correct grammar, spelling and punctuation), reader's acceptance (soliciting reader's agreement), and finally, reader's understanding (intelligibility of the text) as the major constructs of successful writing.

In the Arab context, Al-Abed al-Haq and Ahmed (1994) developed a rubric to evaluate argumentative essays composed by Saudi EFL university-level students. It comprises 9 constructs as indispensable components of quality writing including thesis statement and topic sentence, relevance, cohesion, coherence, argumentativeness, quantity, unity, wording, and grammaticality. Their rubric also has 24 sub-criteria including clarity of thesis statement, qualification of thesis statement, indication of further development in the thesis statement, support of thesis statement, clarity of topic sentences, qualification of topic sentences, indication of further development in the topic sentence, support of topic sentences, relevance of treatment, coherence, cohesion, development of argument, support of argument, organization of argument, degree of argumentativeness, logicity of argument, degree of persuasion, completeness of discussion, balance of discussion, unity, wording, grammaticality, spelling, and punctuation.

#### 4.1 Writing Assessment Approaches

Holistic and analytic approaches or a combination of them are the most commonly used to guide evaluate student's writing proficiency. Determining which approach to be used depends on the purpose of assessment, whether for diagnosis, development or promotion (Bacha, 2001) and on a number of practical considerations including the available time, budget, training the raters and the number of the scripts that need to be evaluated (Nimehchisalem, 2010).

##### 4.1.1 Holistic Assessment Approach

As its name suggests, a holistic assessment approach is a qualitative judgement of students' writing ability (Wiseman, 2012). It focuses on assigning a written composition a single global score in its entirety based upon an overall impression it makes on the rater. It does not require raters to

consider the characteristics of quality writing individually when evaluating and scoring a writing sample. They are rather seen as traits that merge into a single entity working together towards its overall quality. Such method assumes that the final-end product or performance being assessed is a whole entity that should be judged as such since “the whole is not equal to the sum of the parts.” Rather, “the whole is equal to the parts and their relationships” (Goulden, 1992, cited in Barkaoui, 2007, p.87). Another important feature of this approach is that it focuses more on what the writer does well than on diagnosing the specific areas of his/ her weakness (Charney, 1984; Cohen, 1994).

#### 4.1.2 Analytic Assessment Approach

Analytic approach is based upon a list of constructs as being the aspects of successful writing, the most common being content, organization, grammar, mechanics and vocabulary (Jacobs, et al, 1981, p. 30). The exclusion of any of them is especially problematic for L2 students since different aspects of the writing skill may develop at different rates for different L2 students. Some of them may have the ability to achieve high level in expressing content and organization but have limited one in the area of grammatical accuracy, while others may have excellent control at the sentence level but are unable to organize their writing at the discourse level (Kroll 1990).

Each construct has definable criteria with a gradation of performance levels on which students' writing abilities are evaluated and scored (Wiggins, 1998). Analytic scoring requires raters to read and examine the written product several times. It involves assigning a score at a time to each level and then summing up the multiple sub-scores to arrive at a single total one that reflects the overall performance (Hamp-Lyons, 1991; Bacha, 2001; Weigle, 2002). Such scoring scheme assumes that “the sum of the sub-scores for the parts is exactly equal to a valid score for the whole and, by evaluating the parts, the rater has evaluated the whole” (Goulden, 1992, cited in Barkaoui, 2007, p.87). Analytic scoring is often used in tests that are formative in nature and it is chosen for placement purposes. It creates a profile on students' performance in the different characteristics of quality writing. Teachers and curriculum designers use the test results and writing assignments for diagnosis of gaps between actual and desired levels of

students' performance. Such diagnostic information can be useful for both students and teachers. It gives students feedback to be aware of their strengths and weaknesses. It also allows teachers to tailor the level and content of writing instruction more closely to the needs of their students to maximise their future learning and eventually improve the quality of their writing assignments (Hamp-Lyons & Kroll, 1997; Becker, 2011).

### 5. The Constructs and Features of the ER

EFL/ESL writing teachers, and researchers have had difficulty in finding the right writing rubric that yields reliable results of students' writing ability and evaluates what it claims to measure as often those chosen are inappropriate for to achieve these purposes (Bacha, 2001). In an endeavour to develop a 'right' rubric as an alternative to the impressionistic assessment procedure in the Arab context, the researcher found that borrowing the Jacob, et.al's (1981) 5-point constructs and combining them with the practicality of holistic scoring is the best way to achieve this end for a number of reasons. The 'right' rubric is termed eclectic rubric (ER).

Following the Jacob, et.al's (1981) Profile, the ER consists of the following constructs.

- Content construct has such criteria as excellent content indicating essay that is knowledgeable, substantive with a thorough development of thesis. The essay also exhibits thoroughly explained ideas with well- integrated and sufficient support relevant to the assigned thesis.
- Organization construct has such criteria as compositional organization (logical sequencing of argumentation) including introductory paragraph, body paragraphs and concluding paragraph, cohesive and coherent development and exposition of ideas within and among paragraphs.
- Vocabulary construct has such criteria as effective lexical choices, and voice and tone appropriate for the persuasive purpose of the writing task.
- Language use conventions construct has such criteria as varied sentence constructions, and grammatical correctness (agreement, tense, number, word order/function, articles, pronouns, prepositions).



- Mechanics construct has such criteria as spelling, punctuation, capitalization, paragraphing, and handwriting.

The decision to choose Jacob, et.al's (1981) Profile lies in three reasons: First, it has both content validity and construct validity. It has content validity since it assesses the quality students' argumentative writing; a mode of writing that underlies almost all students' writing assignments and tasks at university level, against the criteria the Profile identifies (Bacha, 2001). It is claimed that a writing rubric has content validity when "it evaluates writers' performance on the kind of writing tasks they are normally required to do in the classroom" (Jacob, et.al, 1981, p. 74). Equally important, the Profile has highest construct validity since it is intended to channel raters to consider those aspects of quality writing nominated in the Profile and control how much weight they give to each aspect when judging students' performance. The Profile also allows raters to distinguish significant differences among students when the scores of their test essays are compared (Bacha, 2001).

Second, the Jacob, et.al's (1981) Profile is considered "one of the best known and most widely used analytic scales in ESL adopted by numerous college-level writing programs" (Weigle, 2002, p. 115). This is due to the assumption that it is an empirically grounded tool whose criteria are developed from a large number of writing samples collected from ESL students (Odell, 1981).

Third, the wording of Jacob, et.al's (1981) constructs can significantly offers teachers a potentially valuable tool for consistent and reliable scoring. The criteria matched with each construct are displayed into a number of carefully worded sentences and phrases to eliminate ambiguity or misinterpretation. They are put in a more descriptive and neutral statements and less based on comparative or evaluative language. Each criterion includes sufficiently rich description of the salient and distinctive qualities of each writing construct to account for what makes discrimination between less and more competent students (Volk, 2002).

The distinguishing features of the ER are: First, its scoring mechanism is substantially different from the one applied in Jacob, et.al's (1981) Profile. The 5 constructs are equally rather than differently weighed. The ER has a

numerical scale of 1-100 points with 100 being the highest score. A scale that conforms to the local one currently in use in Arab universities. An equal-weight scheme shows several advantages over assigning varying weights to different writing aspects for two reasons. First, adhering to a varying weight scheme needs a numerical scale that clearly guides raters to award sub-score for each performance level and how to sum up the multiple scores into an overall score. Such mechanism turns out to be of a complicated nature leading to fatigue and boredom and may confuse raters consequently (Hamp-Lyons, 1991; Attali & Powers, 2008). Critics of analytic scoring also point out that measuring the quality of a text by adding up accumulated sub-scores diminishes its interconnectedness, and gives the false impression that writing can be understood and fairly assessed by analysing it into autonomous features (Hillocks, 1995; White, 1994). Second, the ER requires the teacher to holistically rather than analytically, read each essay focusing on the criteria that match each of the Jacob, et.al's 5 writing constructs to find out how far the entire written script meets them and accordingly assign one impressionistic score for its overall quality rather than awarding sub-scores to each of the constructs. In other words, as Weigle (2002, p.12) puts it, an eclectic writing rubric requires the writing teacher to "read holistically and adjust analytic scores to match holistic impressions".

### 5.1 Preference for a Holistic Scoring

There are several practical reasons to prefer a holistic scoring to the analytic scoring in the Arab context. Foremost among them are: First, the current writing assessment and scoring procedures at the Arab tertiary level is summative in nature. To put simply, scores students obtain in their standardised tests are used to measure their writing proficiency at a given point in time (at the end of an academic term/ year) to decide their promotion for the subsequent level of education (El-Daly, 1991) and therefore diagnostic information, as in the analytic one, is insignificant (Bacha, 2001). Accordingly, holistic scoring is more valid scoring scheme than analytic one to be used in the Arab context since there is an alignment between holistic scoring and the promotional objectives and goals of the writing instruction.

Holistic scoring has "the highest construct validity when overall attained writing proficiency is the construct assessed" (Perkins, 1983, p.652).

A second reason that lends weight for adopting a holistic scoring scheme is that it is more practical and economical than analytic one. It is usually less time-consuming than analytic one incurring low cost in large scale testing situations as it allows raters to integrate the inherent writing qualities into a single entity contributing equally to its global assessment and assign a single score to the student writing sample (Breland, Camp, Moris, & Rock, 1987; Wright, 1984; Kroll, 1998). Thus, it makes the scoring task less burdensome to raters and lessens the cognitive overload on them (Weigle, 2002; Becker, 2011). In comparison, analytic scoring scheme can cause the scoring process to be substantially slow and expensive "the downside of this sort of procedure is that for a thorough analytic judgment, each writing assignment would need to be scored on a specifically created assessment instrument" (Kroll, 1998, p.228).

It is asserted that the time needed to score essays using analytic scoring is four times the time needed to score them using holistic one. He also concludes that the time needed to train the raters to use the analytic rubric is two times the time needed to train raters to use the holistic one (Bauer, 1981). Another issue of concern with analytic scoring is that scoring one writing aspect can influence how other aspects are scored. If a rater gives an essay a score of 1 for language use, that the same rater may be influenced by this poor score leading him /her to give another score of 1 for content. This might potentially influence the reliability of the assigned (Weigle, 2002).

## 6. Implementation of the ER: Concerns and Recommendations

6.1 Writing syllabus at the Arab university level is exclusively inspired and guided by product approaches (Al-Hazmi, 2006). Under these approaches, Kharma and Hajjaj (1989, p.187) maintain that the aspects of good writing that dominate the teachers' thinking in both teaching writing and assessing students' written work are geared towards "the mechanics of writing, handwriting, spelling, capitalisation, and punctuation; grammatical mistakes and topic development" with little concern given to content and the logical progression of ideas (El-Aswad, 2002)..

Such exclusive pedagogical focus on decontextualized language entities and mechanical skills as being the ultimate objectives of what the test intends to measure is questioned. The mid-sixties brought an increasing awareness of L2 students' needs to produce extended written texts to achieve their communicative ends in academia (Silva, 1990). This line of thought shifts the focus away from the mastery of linguistic competence and brings the textual norms dictated by a particular English-speaking community into play as being crucial aspects for writing a successful argumentative text and thus influence scoring decisions. According to Hyland (2003), such textual conventions have the advantage of showing students the ways in which academic genre is codified in recognisable moves and stages in terms of its purpose and audience and thus ensure that they will be guided to compose a variety of academic texts.

This recommends that EFL Arab students necessarily need to receive explicit and systematic instructions into the organizational patterns of academic genre that the product-oriented writing instruction fails to accommodate (Kamel, 2000). An immediate consequence of this perspective is that writing teachers should incorporate genre pedagogy into their writing instruction and curricula to raise students' conscious awareness of these patterns to become better writers. One possible pedagogical technique to achieve this end is to expose them to authentic model argumentative texts. Teachers should identify their textual moves and stages as well as their typical linguistic features through modelling and explicit instruction (Hyland, 2007). They should also guide students and encourage them to take part in practising the whole process of exploration (Chandrasegaran, 2009). With such training and socialisation, students would gradually gain control over genre norms and eventually transfer and apply them to generate new instances of genre in performing their subsequent writing assignments and tasks in academia (El-Seidi, 2000) 6.2 Applying a criteria-referenced writing rubric is not the norm in EFL writing classroom at the Arab tertiary level. Under these conditions, the application of such rubric may constitute a big challenge to EFL writing teachers since they are well-entrenched in the long-held practice of impressionistic scoring. Consequently, helping them

overcome their idiosyncratic biases and adjust themselves to the new scoring job becomes a concern of paramount importance to ensure the reliability of scoring. Consensus among different raters and the tendency of individual rater to give consistent scores across a cohort of students' writing scripts are crucial in establishing the extent to which a scoring rubric can be used with an acceptable degree of reliability (Stemler, 2004). To achieve consensus, relevant literature offers some fundamental recommendations. For example, researchers (Odell, 1981; White, 1985; Hamp-Lyons, 1990) emphasize the significance of training workshops to help raters discuss various issues in the rating process that are best captured by negotiation rather than by an abstract piece of writing. Such workshops are therefore a good platform for raters to come to an agreement on predefined criteria and guide them to control the weight they award to each of them against which students' writing samples are judged and, as a result, scores given by raters may be treated as reliable (Vaughan, 1991).

### Conclusion:

There are several scoring tools currently in use for to assess writing for academic purposes in English as a Foreign/Second language. This article has focused on the two main types: holistic and analytic. It shows that developing a writing rubric can help students be aware of the criteria against which their written work is assessed and scored. It also provides them with feedback to reflect on their own learning progress and for the teacher to measure real progress and identify areas where further learning is required. Inter-rater reliability cannot be guaranteed, unless rating criteria are clarified and discussed extensively. A rubric is useful in enhancing the objectivity and consistency of the teacher's scoring through focusing on well-defined criteria.

An eclectic writing assessment instrument with clearly defined criteria is proposed as an alternative to the impressionistic one that is currently in use at the EFL Arab university-level. The proposed instrument has its roots in the relevant literature. The researcher believes that it can better assess and score Arab students' writing performance for a number of practical reasons that lend weight for adopting it as being efficient and practical to do such job

### 4w3 List of References

Al-Abed al-Haq, F. A., and Ahmed, A. (1994). Discourse Problems in Argumentative Writing. *World Englishes*, 13(3), 307-323. doi: 10.1111/j.1467-971X.1994.tb00318.x.

Attali, Y., and Powers, D. (2008). Construct Validity of e-Rater in Scoring TOEFL Essays (ETS Research Rep. RR-07-21). Princeton, NJ: ETS.

Bacha, N. (2001). Writing Evaluation: What can Analytic versus Holistic Essay Scoring Tell us? *System*, 29 (4), 371–383.

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Barkaoui, K. (2007). Rating Scale Impact on EFL Essay Marking: A Mixed-method Study *Assessing Writing*, 12, 86–107.

Bauer, B.A. (1981). A study of the reliabilities and cost-efficiencies of three methods of assessment for writing ability (ERIC Documentation Reproduction Service No. ED 216357).

Becker, A. (2011). Examining Rubrics Used to Measure Writing Performance in U.S. Intensive English Programs *The CATESOL Journal* (22)1, 113-130. Retrieved from <http://www.catesoljournal.org/wp-content/uploads/2014/06/CJ22>

Breland, H. M., Camp, R., Jones, R. J., Moris, M. M., and Rock, D. A. (1987). *Assessing Writing Skill* (Research Monograph No. 11). New York: College Entrance Examination Board.

Casanave, C. P. (2007). *Controversies in Second Language Writing: Dilemmas and Decisions in Research and Instruction*. Ann Arbor, MI: University of Michigan Press.

Chandrasegaran, A. (2009). What does teaching writing as a process really mean? In L.J. Zhang, R. Rubdy, & L. Alsagoff (Eds.). *Englishes and Literatures-in-English in a globalised world: Proceedings of the 13th International Conference on English in Southeast Asia* (pp. 338-351).

Charney, D. (1984). The Validity of Using Holistic Scoring to Evaluate Writing. *Research in the Teaching of English*, 18, 65-81.

Cohen, A. (1994). Assessing Written Expression. In *Assessing Language Ability in the Classroom* (303-357). Boston: Heinle & Heinle.

Cooper, Charles R. (1977.). Holistic Evaluation of Writing. In Charles, R. Cooper and Lee Odell (Eds.), Evaluating Writing: Describing, Measuring, Judging, (3-31). Urbana, Illinois: National Council of Teachers of English.

El-Aswad, A. (2002). A Study of the L1 and L2 Writing Processes and Strategies of Arab Learners with Special References to Third-year Libyan University Students. Unpublished PhD thesis, University of Newcastle.

El-Daly, M. H. (1991). A Contrastive Analysis of the Writing Proficiency of Arabic and Spanish Speakers: Linguistic, Cognitive, and Cultural Perspectives. Unpublished Doctoral Dissertation, University of Pittsburgh.

El-Seidi, M. (2000). Meta-discourse in English and Arabic Argumentative Writing: A Cross-linguistic Study of Texts Written by American and Egyptian University Students. In I. Zeinab, S. Aydelott and N. Kassabg (Eds.), Diversity in Language: Contrastive Studies in Arabic and English Theoretical and Applied Linguistics, (111–126). Cairo: The American University Press.

Hamp-Lyons, L. (1991). Scoring Procedures for ESL Contexts. In L. Hamp-Lyons (Ed.), Assessing Second Language Writing in Academic Contexts (241-278). Norwood, NJ: Ablex.

Hamp-Lyons, L., and Kroll, B. (1997). TOEFL 2000 Writing: Composition, Community and Assessment (TOEFL Monograph Series No 5 Princeton, NJ: Educational Testing Service.

Hillocks, Jr., G. (1986). Research on Written Composition: New Directions for Teaching." Urbana, IL: National Council of Teachers of English (ERIC Document Reproduction Service No. ED 265 552).

Huot, B. (2002). (Re)Articulating Writing Assessment for Teaching and Learning. Logan, Utah: Utah State University Press.

Hyland, K. (2003). Genre-based pedagogies: A Social Response to process. Journal of Second Language Writing, 12, 17-29.

Jacobs, H.L., Zinkgraf, S.A., Wormuth, D.R., Hartfiel, V.F. and Hughey, J.B. (1981), Testing ESL Composition: A Practical Approach. Newbury House, Rowley, MA.

Jonsson, A., and Svingby, G. (2007). The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences. Educational Research Review, 2, 130–144.

Kamel, S.A. (2000). Categories of Comprehension in Argumentative Discourse: A Cross-Linguistic Study. In I. Zeinab, S. Aydelott & N. Kassabgy (Eds), Diversity in Language:



Contrastive Studies in Arabic and English Theoretical and Applied Linguistics (193–235). Cairo: The American University Press.

Kharma, N., and Hajaj, A. (1989). Errors in English among Arabic Speakers: Analysis and Remedy. Longman Group UK Limited.

Kroll, B. (1990). Second Language Writing: Research Insights for the Classroom. New York: Cambridge University Press.

Montgomery, K. (2000). Classroom Rubrics: Systematizing what Teachers do Naturally. The Clearinghouse 73 (6), 324-328.

Morrison, G. R., and Ross, S. M. (1998). Evaluating Technology-based Processes and Products. New Directions for Teaching and Learning 74, 69–77.

Nimehchisalem, V., and Mukundan, J. (2012). Determining the Evaluative Criteria of an Argumentative Writing Scale. English Language Teaching, 4(1), 58-69

Odell, L. (1981). Defining and Assessing Competence in writing. In C.R., Cooper (Ed.), The Nature and Measurement of Competency in English (95–138). Urbana, IL: National Council of Teachers of English.

Perkins, K. (1983). On the Use of Composition Scoring Techniques, Objective Measures, and Objective Tests to Evaluate ESL Writing Ability. TESOL Quarterly, 17, 651–671.

Reid, J. (1993). Teaching ESL Writing. New Jersey: Prentice Hall Regents.

Silva, T. (1990). Second Language Composition Instruction: Developments, Issues, and Directions in ESL. In B. Kroll (Ed.) Second Language Writing Research: Insights for the Classroom (11-17). New York: Cambridge University Press.

Soles, D. (2001). Sharing Scoring Guides. Paper presented at the annual meeting of the Conference on College Composition and Communication, Denver, CO, March 15 (ERIC Document Reproduction Service No. ED 450 379).

Spandel, V. (2006). In Defence of Rubrics. English Journal, 96 (1), 19–22.

Stemler, Steven E. (2004). A Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Inter-rater Reliability. Practical Assessment, Research & Evaluation, 9(4). Retrieved from <http://PAREonline.net/getvn.asp?v=9&n=4>.

Vaughan, C. (1991). Holistic Assessment: What Goes on in the Raters' Minds? In L. Hamp-Lyons (Ed.), Assessing Second Language Writing in Academic Contexts (111–125). Norwood, NJ: Ablex.

Weigle, S. (2002). Assessing Writing. Cambridge, UK: Cambridge University Press.

Weir, C. J. (1990). Communicative Language Testing. Englewood Cliffs, NJ: Prentice Hall.





---

Weir, C.J. (1983). Identifying the Language Needs of Overseas Students in Tertiary Education in the United Kingdom. Unpublished PhD Thesis, University of London, Institute of Education.

White, E. M. (1984). Holisticism. College Composition and Communication, 35, 400-409.

White, E.M. (1985). Teaching and Assessing Writing. San Francisco, CA: Jossey-Bass.

Wiggins, G. P. (1998). Educative Assessment: Designing Assessments to Inform and Improve Student Performance. San Francisco: Jossey-Bass Publisher

Appendix 1: Jacobs et al.'s (1981) Scoring Profile (Source: Hughes, 2002, p.104)

ESL COMPOSITION PROFILE				
STUDENT	DATE	TOPIC		
SCORE	LEVEL	CRITERIA	COMMENTS	
CONTENT	30-27	EXCELLENT TO VERY GOOD: knowledgeable • substantive • thorough development of thesis • relevant to assigned topic		
	26-22	GOOD TO AVERAGE: some knowledge of subject • adequate range • limited development of thesis • mostly relevant to topic, but lacks detail		
	21-17	FAIR TO POOR: limited knowledge of subject • little substance • inadequate development of topic		
	16-13	VERY POOR: does not show knowledge of subject • non-substantive • not pertinent • OR not enough to evaluate		
ORGANIZATION	20-18	EXCELLENT TO VERY GOOD: fluent expression • ideas clearly stated/ supported • succinct • well-organized • logical sequencing • cohesive		
	17-14	GOOD TO AVERAGE: somewhat choppy • loosely organized but main ideas stand out • limited support • logical but incomplete sequencing		
	13-10	FAIR TO POOR: non-fluent • ideas confused or disconnected • lacks logical sequencing and development		
	9-7	VERY POOR: does not communicate • no organization • OR not enough to evaluate		
VOCABULARY	20-18	EXCELLENT TO VERY GOOD: sophisticated range • effective word/ idiom choice and usage • word form mastery • appropriate register		
	17-14	GOOD TO AVERAGE: adequate range • occasional errors of word/idiom form, choice, usage <i>but meaning not obscured</i>		
	13-10	FAIR TO POOR: limited range • frequent errors of word/idiom form, choice, usage • <i>meaning confused or obscured</i>		
	9-7	VERY POOR: essentially translation • little knowledge of English vocabulary, idioms, word form • OR not enough to evaluate		
LANGUAGE USE	25-22	EXCELLENT TO VERY GOOD: effective complex constructions • few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions		
	21-18	GOOD TO AVERAGE: effective but simple constructions • minor problems in complex constructions • several errors of agreement, tense, number, word order/function, articles, pronouns, prepositions <i>but meaning seldom obscured</i>		
	17-11	FAIR TO POOR: major problems in simple/complex constructions • frequent errors of negation, agreement, tense, number, word order/ function, articles, pronouns, prepositions and/or fragments, run-ons, deletions • <i>meaning confused or obscured</i>		
	10-5	VERY POOR: virtually no mastery of sentence construction rules • dominated by errors • does not communicate • OR not enough to evaluate		
MECHANICS	5	EXCELLENT TO VERY GOOD: demonstrates mastery of conventions • few errors of spelling, punctuation, capitalization, paragraphing		
	4	GOOD TO AVERAGE: occasional errors of spelling, punctuation, capitalization, paragraphing <i>but meaning not obscured</i>		
	3	FAIR TO POOR: frequent errors of spelling, punctuation, capitalization, paragraphing • poor handwriting • <i>meaning confused or obscured</i>		
	2	VERY POOR: no mastery of conventions • dominated by errors of spelling, punctuation, capitalization, paragraphing • handwriting illegible • OR not enough to evaluate		
TOTAL SCORE	READER	COMMENTS		