

Opportunities and Questions: A Short Report on Rubric Assessments in Asia and the Middle East

Ronnie Goodwin

Gulf University for Science & Technology (GUST)
Kuwait

Abstract

This qualitative short report considers the viability of the use of rubrics or alternative methods to assess writing in Asia and the Middle East. The background of learning theories, assessment types, and self-assessment literature provides a foundation for further discussion of the appropriate use of rubrics, including the prioritization of criterion, the quality of scoring, the impact of organizational features on scoring, the influence of bias, and the best application of rubric assessment. Relevant points for further study are identified, such as differentiation in research between generalized analytical rating systems and rubric assessment with specific, empirical criterion. The contradictory research regarding the advantages and disadvantages of rubric assessment in comparison with holistic assessment are of particular and crucial interest for global pedagogy. Many of the reviewed Western articles excluded Asian perspectives- except for China- and thus present a limited understanding of social and educational compatibility with new assessments and rubric assessments in particular. The discussion identifies patterns and points of contention and seeks to explore viewpoints rather than limit the scope of inquiry and consideration thus noting that relevant literature suggests that with appropriate teacher training, teachers may appropriately use rubrics as a formative assessment tool for writing in Asia and the Middle East.

Keywords: Asia, education, formative assessment, Middle East, professional development, rater, rubric, scoring, summative assessment

Cite as: Goodwin, R. (2019). Opportunities and Questions: A Short Report on Rubric Assessments in Asia and the Middle East. *Arab World English Journal*, 10 (3) 21-31.

DOI: <https://dx.doi.org/10.24093/awej/vol10no3.2>

Opportunities and Questions: A Short Report on Rubric Assessments in Asia and the Middle East

Since Asia and the Middle East require fluency in two or more languages, better assessment impacts students for many years. Rubrics with specific guidelines and clear expectations facilitate an understanding of the requirements of the school and the teacher and open lines of communication regarding scores between teachers, parents, and students. Although debates continue about the quality of holistic and analytical scoring, the validity of well-designed rubrics often stands above the results for these two continuums of assessment. Despite the resistance expected from many teachers, rubrics may speed teacher efficiency and facilitate opportunities for students to take ownership of their learning outside of the classroom while the teacher remains in control.

Background

In this literature review, the theoretical background regarding learning, assessment, and methods provide a picture of the purpose, types, and applications which might affect teacher scoring in the Middle East and Asia in particular.

Instruction

The method of instruction directly relates to a teacher's evaluation of the mastery of writing skills and correlates to the performance on standardized testing. In many cases, similar teaching approaches foster higher overall scoring results on such large-scale assessments. Other teaching approaches may better suit a given nation's cultural aims of education or a more authentic learning experience. Ismail (2011) explains that many students find writing in their mother tongue painful, and the countries of the Middle East and Asia require mastery of two (or more) languages. He also reports the observations that the writing mastery of the students in the UAE also directly relates to their previous writing experience and acquisition of language(s) and that these students positively accepted constructive feedback more in the classroom setting. They agree that feedback should shape each phase of the writing process (pp. 74-75).

Assessment

The many diverse assessment types continually change and develop with the mastery of desired skills and knowledge. With the advent of easily written translation services and globalized international communications, the Asian emphasis often focuses language mastery upon speaking and listening first, reading second, and writing as a last consideration (Yi, 2009). Hidri and Coombes (2017) explain the many subtler facets of assessment and point out a hidden opportunity to create sustainable development that could be linked to quality assurance.

Assessment Types. Nodoushan (2014) comments that assessment generally falls into the holistic, analytic, and trait-based categories and adds that each category fulfills a specific function which best matches diverse assessment types. A holistic assessment provides a sweeping, 'big picture' view of accomplishment, an analytic assessment includes a scaled approach of scoring overall writing, and trait-based assessment evaluates the mastery of specific goals (p. 122-124). Han (2017) remarks that a holistic writing assessment shows great validity in the context of certification, placement, or research. By contrast, analytic methods assess individual details with

less potential for bias and more impact when applied to formative assessments for use in student education through task completion (p. 124). Firoozi (2019) adds the function-based qualifiers of diagnostic, progressive, and summative types and that selecting the most valid and reliable method remain crucial to the validity of the assessment results.

The umbrella of performance tasks includes the production of writing and essays, and Nodoushan (2014) observes that the development of quality performance assessment evolves from a list of desired objectives to the creation of a motivating task which best facilitates student demonstration of mastery and finally to the development of explicit performance criteria (p. 122). Knoch (2009) quantitatively compares holistic scoring, analytic scales that are ‘intuitively developed’, and analytic scales that are ‘empirically developed’ and found that the empirically-developed analytical scales- the category that criterion-based rubric assessment fits- mitigate many of the weaknesses of other analytical assessment types (pp. 298-299).

De Silva (2014) tells us that a rubric functions as a potential tool of teaching and assessment for both performance and authentic tasks, helps students critically evaluate their work and that of their peers, saves teachers time, accommodates diverse student groups, and allows for easy use (p. 136). Furthermore, the breadth of research regarding analytical scoring often includes less-reliable measures as representative of the continuum of available tools, and more research directed solely at rubric assessment and its potential application to writing classes in Asia and the Middle East poses an area of potential further study.

Self-Assessment Rubric. A wide variety of theories regarding feedback regarding student writing continues to inspire more research, and some researchers go so far as to say that corrective feedback tends to inhibit the flow and creativity of writing and ought to be avoided (Ganji, 2009, p. 118). Rubrics can be used in student self-assessment, as well. De Silva (2014) notes that high school students in their study of in Sri Lanka expressed surprise and disappointment at the scoring of their authentic writing assessments (pp. 137-138). Hale (2015) explains that in East Asia, where the education system is still heavily teacher-centered and controlled, the novelty of these approaches questions if the students benefit more when they have a direct voice in their own grading. Among Saudi students in 2017, they favored rubric assessment with the assurance that they would be involved in the development and clearly understand the expectations of each criterion (Obaid, 2017)

Findings

The background of learning theories, assessment types, and self-assessment literature provides a foundation for further discussion of the appropriate use of rubrics, including the prioritization of criterion, the quality of scoring, the impact of organizational features on scoring, the influence of bias, and the best application of rubric assessment.

Prioritization

A particular assignment might highlight a specific skill or standard for a unit and clearly display the priorities through the lens of class and social culture. Education and social and political influence remain inextricably linked. El Ebyary (2013) observes “how the pressures on the teaching/ learning process is externally managed by some educational bodies with the prime aim

of raising standards” (p. 2170). Firoozi (2019) quotes one definition: “Assessment culture refers to educational evaluation practices that are compatible with current ideologies, social expectations, attitudes, and values” (Inbar-Lourie 2008, p. 285). In America, schools often highly value technological use and innovation. In Korean high schools, communicative listening and speaking for globalized readiness remain high priorities and writing forms an almost-coincidental part of many general assessments (Yi, 2009, pp. 62-64).

In other words, rubrics always show prioritization- even when this occurs unintentionally. One study of 30 Iranian high school students concluded that they ranked grammar and spelling as the most crucial factors in their writing grades and yet this body of students requested a holistic scoring of a composition in place of the use of a rubric. The small sample emerged from within the same Iranian community, so further research has yet to delve deeper into the topic (Tajgozari & Alimorad, 2019). In the preface to their book, Hidri and Coombe (2017) add that the correlation of priorities to grading- and the general process of education itself- create a significant ethical side to the application of all forms of assessment.

Further assessment readiness often occupies one of the highest priorities for in-class writing and feedback. Ganji (2009) explored the effectiveness of corrective feedback in testing preparation, the impact of general guidelines in the IELTS testing, and how feedback from the teacher, from peers, and from self-reflection affect the Iranian upper-intermediate-level writing students’ second attempts. This study concluded that minimal corrective teacher feedback proves more effective than peer feedback or self-assessment. In America, one year later, a similar study conducted with university students found similar results. In 2007, Chinese EFL writing studies upheld this result with the caveat that student revision gives meaning to indirect teacher feedback (pp. 119-124).

Quality

In the evaluation of writing, Gebril and Plakans (2009) write that guidelines for assessing discourse should include diverse lexical, syntactical, rhetorical and pragmatic characteristics, reliable and meaningful application, and clear differences between writing scaled at another level (p. 54). Amini (2018) remarks that the quality translation in EFL classrooms includes accuracy, fluency, and “fitness for the purpose”, and these same characteristics apply to their cognitive translation of ideas into written essays.

Rating Variance. Kimura et al. (2017) stress the relationship between the fluency level of the teacher, instructional efficiency, and the comparative accuracy of scoring. Yi (2009) points out that the teacher’s personal experiences and pedagogical definition of writing excellence influence their holistic grading rationale. Many of the polled high school teachers in Korea described writing acumen using terms of grammar and organization, but one teacher provided an apt offhand description of “accuracy, commitment, good content, creativity, and good paragraphing”, an informal sketch of viable rubric criterion (pp. 53-56, 63).

Cho (2008) writes that large-scale assessment simply cannot fairly meet the level of rigor of a classroom setting. In a class, the teacher tailors the writing to areas of weakness in prior learning and specifically accesses the students’ previous classroom knowledge, allowing them to

write more confidently, fluently, and specifically. Other than generic personal writing topics, large-scale testing includes limits of personal connection to the student. However, the classroom setting presents conditioned responses in a comfortable environment with known teacher and class expectations, limiting the variety of voice and selection of writing material itself (pp. 49-52).

Zhang, Xiao, and Luo (2015) state that higher-fluency writers perform much better with holistic scoring measures, while lower-fluency writers benefit from the feedback and goal-oriented scoring of analytic scales and rubrics in particular. Especially when the writing occurs in a second or third language, the student writer balances a multitude of cognitive tasks and undergoes assessment at the level of a native speaker. They, too, grow tired as their brain evaluates, organizes, and translates under extreme pressure. Students with greater long-term exposure to writing in their native language and additional languages experience the benefit of multilayer linguistic processing. In many countries, students undergo writing assessment despite the stipulation that instruction begins “wherever conditions permitted” (Ruecker & Crusan, 2018). This paradox means that the students’ former exposure to the language will often widely vary.

As teachers grow tired of reading assessments, especially when the student writers expect expedited results, mental fatigue begins to affect the scoring process, and teachers may also grow more or less lenient in their grading and skim over their reading. This can inspire an unfair holistic scoring, which relies on the impressions of the writings as a whole. Most of the authors in this literature review agree that rubric assessment helps focus raters’ attentions and speeds the grading process, allowing teachers to grade more papers consistently and quickly. However, Zhang, Xiao, and Luo (2015) argue the opposite: more than three rating areas slow the process and negatively impact the quality and reliability of scoring and speed mental fatigue during scoring, slowing the scoring process. Knoch’s quantitative study compared holistic and analytical, trait-based scales with particular emphasis upon rubrics and concluded that a higher number of details and descriptors provided a more reliable baseline for rubric assessment and that the raters themselves ultimately preferred the rubrics to holistic assessment (pp. 298-302). This divergence of evidence and interpretation within the recent literature poses questions for further future study.

Organizational Skills. Foreign language writing assessments require more than linguistic skills; they require prolonged retention of linguistic concepts and the ability to organize these concepts into one cohesive argument. Thus, it requires very high-level cognitive multitasking and quick processing to complete such a task. A well-organized paper may support an argument better than the writing of a student with a broader understanding of the subject and applicable vocabulary. Interestingly, one Korean high school teacher answered that he assessed writing ability “looking at... grammar, flow...coherence, content, and so on” (Yi, 2009, p. 63).

Ruegg and Sugiyama (2013) explore the vast differences in scoring with rating scales and holistic scoring through the lens of the desired objectives being assessed. The authors compared scoring of timed essays and found that the number of paragraphs and cohesive devices as organizational features correlated strongly to higher overall scores even in those marked for deeper textual comprehension and expression. Although the assessments sought to rate content and application, one study found that expert TESOL teachers mentioned handwriting as a factor in thirty percent of their evaluations (Cho, 2008, pp. 53-54).

Meta-Linguistics. Confidence greatly affects the written presentation of their acquired knowledge, and the students' self-reported feelings of accomplishment (Bialik, Martin, Mayo, & Trilling, 2016, pp. 46-47). This hesitance often displays in reliance upon a smaller vocabulary or upon the same simple grammatical features. Cho (2008) explains that the rating of graduate-level TESOL teachers showed a greater emphasis upon meta-linguistic factors than linguistic factors; the raters subconsciously zeroed in on the most common student writing mistakes before considering the content and skill of the writing as a whole (pp. 52-54).

Jeong's 2015 study reached a similar conclusion, noting the link between grammar and mechanics error and lower overall scores as compared to rubric-guided scoring. For these reasons, the holistic, 'red-pen' approach of many traditional teachers provides very specific- albeit discouraging and often negative-feedback and often favors the literal aspects of writing over substance, meaning, and comprehension (Nodoushan, 2014, pp. 123-124). Raters often equate the length of the writing itself with a depth of thought. Beyond the mandatory word range limits commonly given on most assessments, the number of words also strongly predicts the rating given, and holistic rating amplifies this effect (Amini, 2018).

Outside Sources. The inclusion of outside sources and the ability to appropriately cite them indicates less about the student's deeper knowledge of the topic and writing skills than it does the student's grasp of complex grammatical features and higher-level analysis. Thus, a rubric often corrects this natural imbalance in scoring between form and substance. Gebril and Plakans (2009) found that- with holistic grading- the correct citation of outside sources in writing correlated with higher scores than those given to writing which conveyed original ideas and a deeper and more sophisticated understanding of the reading (pp. 63-65). Ezza (2017) suggests that more specialized writing rubrics emphasize the specific expectations, vocabulary, and previous knowledge of writing directed to niche or trade groups. A persuasive paper must cite previous evidence, such as case law, literary examples, and medical case studies, and ought to conform to the relevant conventions. Thus, exposure to native language writing in the target style supplements understanding of complex concepts which might be included in a rubric (pp. 196-197).

Application

Although the majority of research about rubrics studies their development, criterion selection, efficacy, and best use, the rubrics prove ineffective if the teachers cannot effectively and accurately score the writing. Best practices which utilize rubrics also gradually shift to become more goal-aligned, clear, and reflective of the experience of teachers, students, and parents.

Calibration. Jeong (2015) notes that despite the common use of rubrics in writing and performance-based assessments, raters rarely receive input or training regarding appropriate rubric development as a tool for assessing specific standards. Teachers literally have limitless options of foci for a rubric, especially if they create it for a specific goal or assignment. Ezza (2017) included ten of the most common writing traits in their survey of teacher rating behaviors: audience, text structure, ideas, persuasive devices, vocabulary, cohesion, paragraphing, sentence

structure, punctuation, and spelling. Notably, more than half of those traits analyze choices rather than the content itself (pp. 196-197).

Firoozi (2019) recommends teacher training in the use of rubrics as one of the greatest tools of reading and writing literacy and that they receive training in the higher-order, critical thinking cognitive skills which non-native language users require to best apply their learning. These skills apply both to teachers and students. Nonetheless, many educational organizations, such as the Center for Curriculum Redesign, offer suggested lists and even provide the websites where flexible, tested assessment rubrics can be found (Bialik, Martin, Mayo, & Trilling, 2016, pp. 14-16).

Teacher Preference. Teachers and students must fully appreciate and understand these rubrics. In many Asian and Middle Eastern countries, the teacher often may direct every aspect of the classroom with little interference or unwanted explanation. Since these schools typically remain teacher-driven and teacher-controlled, the give-and-take of rubrics may seem alien, incomprehensible, or doubtful. El Ebyary (2013) records the impressions of teachers in the Middle East and records a general trend to mistrust formative assessment as an unfamiliar method which conflicts with the educational perspective of their own education and of their previous teacher training (pp. 2170-2173). Zhang, Xiao, and Luo (2015) also in favor of holistic scoring on the grounds that teacher training for analytical writing assessment requires at least twice as much time and possibly more time in detail-oriented measures, such as rubric use. They also question analytic assessment as representative of the mastery of specific skills and not as a measure of the overall status.

Recalling the case study of Korean high school writing raters, they frequently expressed consternation at explaining their grading process and reasoning (Cho, 2008). For the calibration of better rubrics and grading, the students must know what facets of writing the teacher will assess. De Silva (2014) reminds teachers that providing a rubric may not fully clarify the desired outcomes. For example, the of 'Neatness' could refer to using pen and not pencil, the handwriting, the organization and proofreading of the paper, etc. In a 2017 study of Saudi students, eighty-six percent of students expressed the belief that rubric assessments of writing would further clarify their needs, but ninety-six percent of the same respondents feared that they would not have time to sufficiently review and comprehend each of the rubric assessment areas before writing (Obeid, 2017). In their Sri Lanka case study, high school students scored over twice as high in all areas when the teacher reviewed the rubric with the class beforehand (pp. 135-138). The students' feedback on the strengths and weaknesses of the rubric assessment also inform better teacher calibration to achieve multiple goals.

Student Self-Reflection. Nodoushan (2014) notes: "If performance criteria are well defined, the student will then understand what he or she must do to improve" (p. 122). Thus, providing a rubric clarifies areas of assessment, speeds teacher scoring, and clarifies the rationale behind the scoring to the parents and students, providing opportunities for independent reflection and academic growth. Hale (2015) concluded that self-assessment with rubrics makes students feel trusted, develops independent responsibility, and encourages more objective self-reflection about the quality, strengths, and weaknesses of their written work.

Additionally, De Silva's sources argue that the sense of ownership and accountability drives them to put more effort into the first such attempt and actively invest in their grades with fewer rationalizations for a poor grade (2014, p. 137). Multiple studies among different age groups across the world uphold the importance of specific teacher feedback with fewer notes on grammar and spelling for improving student scoring trends, but some recent research indicates that the subsequent improvement in students' writing only occurs when students must revise their work or otherwise take ownership of the identified weaknesses and apply them in a way which meaningfully integrates the feedback into their personal writing processes (Ganji, 2009, pp. 124-125). Thus, teacher utilization of rubric assessment as a means of nurturance of student creativity, growth, and self-reflection ought to apply it as part of a formative growth strategy which develops writing in carefully-planned stages.

Bias

Han (2017) points out that accurate holistic assessment remains the most trusted form of evaluation and that such methods can often produce the most accurate large-scale view of student progress. He goes on to say that holistic assessment shows a greater tendency toward swings of scoring due to personal bias. This bias can consist of many different influences. As already discussed, the commonplace holistic bias toward overemphasis upon grammar, spelling, and other organizational and meta-linguistic features and briefly discussed teacher perspectives as a key factor which more heavily affects holistic scoring than rubric scoring (or that of other analytic measures).

Fernandez and Siddiqui (2017) also point out that rater scoring differences of writing often result from the severity or leniency of their grading preferences. In their 2017 study of fifteen Pakistani writing raters, the respondents rated the same three essays holistically according to their usual scoring practices. Below, Table 1 demonstrates the extensive possible variations of holistic rating.

Table 1. *Pakistani Rater Scoring of Essays*

Rater	Essay 1	Essay 2	Essay 3	Total
M1	6	4	7	17
M2	7	6	8	21
M3	12	9	10	31
M4	11	8	10	29
M5	6	6	9	21
M6	7	6	9	22
M7	5	6	7	18
M8	7	6	9	22
M9	8	7	8	23
M10	10	9	9	28
M11	5	6	6	17
M12	6	7	6	19
M13	10	10	9	29
M14	8	7	6	21
M15	9	7	7	23

This gap in scoring proves particularly devastating in high-stakes testing for course placement, employment, or crucial educational assessments. Referencing Table 1 above, if students become accustomed to the grading of an atypically-lenient rater, such as M3, but a strict

rater, such as M11, grades their writing, students will justifiably feel confused and often question the teacher and/or the test itself. Quintero, Trejo, Guzman, and Gonzalez (2017) studied the use of criterion-based, empirical rubrics and found that rubric use narrowed the gap between the severity or leniency of raters but did not eliminate it. These students and raters (in Mexico) more reliably predicted the variations of scoring and adjusted quickly due to their greater exposure to rubrics and analytical writing assessments as a means of summative assessment. The author did not discuss or study rubric use as a formative assessment tool, indicating that the results regarding rubric validity and reliability applied specifically to the context of widespread summative assessment.

Conclusions

From the pedagogical background to the assessment types and appropriate use of rubrics, this qualitative review seeks to understand the limits of the viability of rubric assessments of writing in Asia and the Middle East. The research indicates that appropriate rubric use includes a careful alignment with standards of the culture, classroom, and relevant testing. The teacher's own writing skill heavily influences their teaching strategies and perceptions of student mastery; holistic grading often results in inconsistent scoring. While rubrics undoubtedly clarify goals, explicitly justify scoring, speed the grading process, and provide opportunities for teacher and student development, the review of the literature indicates that their use in conjunction with specific goals and student involvement often determines their value as tools of writing assessment.

In the Middle East and Asia, teachers often express a misunderstanding of this relationship between the purpose of rubrics and their best use. As many of these schools, especially language or business schools, globalize their educational approach, the importance of students' investment in their own writing shows a need to align assessment to these goals. When used properly as a formative assessment tool or as a clear summative assessment guide, rubrics allow teachers to bridge their direction to student understanding and would be uniquely suited to these schools if raters undergo proper training. Perhaps the future of rubrics in Asia and the Middle East remains best-suited to formative development of individual skills and the identification of areas of strength or weakness or as reinforcement of teacher-directed goals and student writer internalization of their areas of development.

About the Author:

Dr. Ronnie Goodwin is an Assistant Professor in the English Department at Gulf University for Science & Technology (GUST). Dr. Ronnie specializes in teaching Business Writing, English Composition, and Linguistics to college-level, high school, and adult learners.
ORCID, <https://orcid.org/0000-0002-0377-5173>

References

- Amini, M. (2018). How to evaluate the TEFL students' translations: through analytic, holistic or combined method? *Language Testing in Asia*, 8(10).
- Bialik, M., Martin, J., Mayo, M., & Trilling, B. (2016). Evolving Assessments for a 21st Century Education. *Assessment Research Consortium*.
- Cho, D. (2008). Investigating EFL Writing Assessment in a Classroom Setting: Features of Composition and Rater Behaviors. *The Journal of Asia TEFL*, 5(4), 49-84.

- De Silva, R. (2014). Rubrics for Assessment: Their Effects on ESL Students' Authentic Task Performance. *CELC Symposium*: Open Sri Lanka University.
- El Ebyary, K. M. (2013). Profiling Formative Assessment Culture in EFL Teacher Education Programs in the Middle East. *Theory and Practice in Language Studies*, 3(12), 2169-2177.
- Ezza, E. (2017). Criteria for Assessing EFL Writing at Majma'ah University. *Evaluation in Foreign Language Education in the Middle East and North Africa*. Springer International Publishing, 185-200.
- Fernandez, M., & Siddiqui, A. M. (2017). Markers' criteria in assessing English essays: an exploratory study of the higher secondary school certificate (HSCC) in the Punjab province of Pakistan. *Language Testing in Asia*, 7(6).
- Firoozi, T. (2019). The language assessment literacy needs of Iranian EFL teachers with a focus on reformed assessment policies. *Language Testing in Asia*, 9(2).
- Ganji, M. (2009). Teacher-correction, Peer-correction and Self-correction: Their Impacts on Iranian Students' IELTS Essay Writing Performance. *Journal of Asia TEFL*, 6, 117-139.
- Gebril, A., & Plakans, L. (2009). Investigating Source Use, Discourse Features, and Process in Integrated Writing Tests. *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 7, 47-84.
- Gustilo, L. E. (2013). An analysis of writer's performance, resources, and idea generation processes: the case of Filipino engineering students. *Language Testing in Asia*, 3(2).
- Hale, C. C. (2015). Self-assessment as academic community building: a study from a Japanese liberal arts university. *Language Testing in Asia*, 5(1).
- Han, T. (2017). Examining the Impact of Scoring Methods on the Institutional EFL Writing Assessment: A Turkish Perspective. *PASAA*, 53, 112-147.
- Hidri, S. B. B., & Coombe, C. (2017). Evaluation in Foreign Language Education in the Middle East and North Africa. *Springer International Publishing*.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3).
- Ismail, S. (2011). Exploring Students' Perceptions of ESL Writing. *English Language Teaching*, 4(2), 73-83.
- Jeong, H. (2015). Rubrics in the classroom: do teachers really follow them? *Language Testing in Asia*, 5(6).
- Kimura, Y., Nakata, Y., Ikeno, O, Naganuma, N., & Andrews, S. (2017). Developing classroom language assessment benchmarks for Japanese teachers of English as a foreign language. *Language Testing in Asia*, 7(3).
- Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275-304.
- Nodoushan, M. A. S. (2014). Assessing Writing: A Review of the Main Trends. *Studies In English Language and Education*, 1(2), 119-129.
- Obaid, R. (2017). Second Language Writing and Assessment: Voices from Within the Saudi EFL Context. *English Language Teaching*, 10(6), 174-181.
- Quintero, E. F. G., Trejo, P., Guzman, R. R., & Gonzalez, E. F. (2017). Assessing EFL University Students' Writing: A Study of Score Reliability. *Revista Electronica de-investigacion Educativa*, 19(2), 91-103.

- Ruecker, T., & Crusan, D. (2018). *The Politics of English Second Language Writing Assessment in Global Contexts*. Taylor & Francis Publishing.
- Ruegg, R., & Sugiyama, Y. (2013). Organization of ideas in writing: What are raters sensitive to? *Language Testing in Asia*, 3(8).
- Tajgozari, M., & Alimorad, Z. (2019). Iranian EFL students' perceptions of criteria for assessing students' written performance. *Global Journal of Foreign Language Teaching*, 9, 2-9.
- Yi, J. (2009). Defining Writing Ability for Classroom Writing Assessment in High School. *Pan-Pacific Association of Applied Linguistics*, 13(1), 53-69.
- Zhang, B., Xiao, Y., & Luo, J. (2015). Rater reliability and score discrepancy under holistic and analytic scoring of second language writing. *Language Testing in Asia*, 5(5).