

Journal of King Abdulaziz University: Educational and Psychological Sciences

Volume 4 | Issue 4

Article 23

10-1-2025

Language Assessment Literacy: Considerations for Digital and Teacher-led Assessment of Academic Writing

Othman Aljohani

English Language Institute, University of Jeddah Kingdom of Saudi Arabia, oaaljohani@uj.edu.sa

Follow this and additional works at: <https://kauj.researchcommons.org/jeps>

Recommended Citation

Aljohani, Othman (2025) "Language Assessment Literacy: Considerations for Digital and Teacher-led Assessment of Academic Writing," *Journal of King Abdulaziz University: Educational and Psychological Sciences*: Vol. 4: Iss. 4, Article 23.

DOI: <https://doi.org/10.64064/1658-8924.1164>

This Article is brought to you for free and open access by King Abdulaziz University Journals. It has been accepted for inclusion in Journal of King Abdulaziz University: Educational and Psychological Sciences by an authorized editor of King Abdulaziz University Journals.

Language Assessment Literacy: Considerations for Digital and Teacher-led Assessment of Academic Writing

Othman Aljohani

ORCID: 0009-0002-8123-5620

English Language Institute, University of Jeddah

Kingdom of Saudi Arabia

oaaljohani@uj.edu.sa

Abstract

Assessing academic writing in foreign language contexts has been a challenging and complicated task for the teachers. This theoretical study is premised on the notion that language assessment literacy of the teachers involved in the instruction and assessment of academic writing in higher education is crucial not only to the successful course delivery but also to the reliable and valid learning outcomes. Content analysis approach was adopted to determine how language assessment literacy is essential for the teachers of academic writing before they actually attempt to design, assess and interpret the assessment tools, criteria and results. The study offers theoretical overview of the contemporary approaches to both digital and teacher-led or traditional assessment practices to raise awareness of the teachers in regard to the five characteristics of assessment: fairness, accuracy, impact, feedback, and flexibility. The study concluded that robust teacher training in language assessment literacy as well as digital literacy is an imperative to equip teachers with theoretical knowledge and practical skills to engage in performance measurement of academic writing which is aligned with the course learning objectives as well as is accurately interpretive to showcase reliable and valid accounts of students writing development.

Key words: Academic writing; digital assessment; EFL; language assessment literacy; traditional assessment.

Introduction

Academic Writing as a form of discourse is embedded in its subject specificity which is defined by the socio-cultural context, disciplinary preferences and institutional policies. Adherence to these discourse defining elements involve construction of a knowledge system derived from the lexicogrammar, content, rhetoric, genre, academic discipline, and the specific discourse community (Beaufort, 2007). Each of these knowledge-forming descriptors along with the cognitive processes help in the formulation of an accumulated knowledge base which can be employed in the construction of academically appropriate genre that qualifies the conventions and expectations of its specific discourse community. Discourse community conventions and expectations transcend the production of academically appropriate texts to include their assessment which validates students' academic literacy as being on the level of producing community specific discourse.

Premised on this incremental and multidimensional construct of writing, several challenges have emerged in regard to the design, conduct and assessment of reliable and valid summative writing examinations. In most academic settings, writing is subjected to the teach-to-test system and as pointed out by Gibbs (2006), the assessment design encapsulates the entire learning model and the learning processes therein. The instructional materials and the assessment tasks are aligned to successfully achieve the course learning outcomes. This integrated practice of teaching and assessment may help students write academically appropriate texts; however, the development of their academic literacy to produce academic genres other than what they have been taught and assessed on needs to be ascertained. The second challenge refers to the language assessment literacy of the teachers which entails their ability of conducting summative assessment to “showcase the extent to which pedagogic interventions have been successful in achieving course learning objectives (CLO)” (Ahmad, 2021 p.160). Writing teachers who do not have the requisite level of language assessment literacy (LAL) might not produce authentic evaluation of students’ writing. Yet another issue in assessing writing is the test standardization which if not properly agreed upon in terms of the assessment criteria can adversely affect the test results for their reliability and validity.

The study sets to explore the dynamics of conventional and online assessment restricting to the summative assessment (SA) not only because it is the most commonly practiced mode of assessment and tool of gauging students’ writing proficiency in the academia but also because it unfolds the relevance of the teaching methods, instructional materials, and assessment practices (Thomas, Allman, & Beech, 2004 as cited in Ahmad, 2021 p.160).

Theoretical Framework

LAL is premised on the construct that the teachers as assessors have the requisite knowledge about the language they teach and assess and are also familiar with the theoretical and practical perspectives on language testing (Fulcher, 2012; Inbar-Lourie, 2013). Bachman (2004) believes that raising awareness of language teachers about the ethical concerns should be at the center of all LAL programmes. This entails that the language teachers are skilled and experienced to effectively create, conduct, and assess language test items or tasks. LAL includes practical knowledge of various assessment methods as well as cognizance of how their assessment results can impact teaching and learning. LAL is, therefore, crucial in helping teachers not only to identify the context and method for testing but also ascertain contextually effective tools to interpret the results fairly. Assessment of academic writing requires the

assessment design to be in alignment with both the course learning outcomes and learner feedback (Weigle, 2002). This means that as the assessment contexts build up, the teachers need to update their LAL to conform to the standards of traditional and digital assessment systems.

Language Assessment Literacy in the Context of Writing Instruction

Academic writing is distinct from other types of writing in terms of the level of formality, handling of the genre, use of structural and non-structural resources in the creation of texture, creation of the clause relations, and adherence to the notion of academic integrity. More specifically, academic writing is genre based and this feature dictates the pedagogical choices and assessment practices. Academic writing thus involves not only the creation of genre as a final product but also caters for the complex processes that are engaged in the creation of the writing product. Hyland (2018) suggests that teachers of academic writing must have the competence to handle genre specificity which may include assessment of academic essays, term reports, research papers, dissertations, etc. Following Cumming (2001), this diversity in academic genre in the academia requires teachers to be able to not only experts in assessing different genres but also be able to synchronize their assessment outcomes with the course learning objectives.

For reliable and valid test results, LAL can prepare teachers of academic writing to develop effective and comprehensible exam rubrics which are in conformity with the assessment objectives (Fox & Macpherson, 2021). LAL enables teachers to design assessment criteria or benchmarks with measurable descriptors to obtain reliable account of how the students have performed on a variety of the features of academic writing such as the task achievement, grammatical range and accuracy, lexical variation, cohesion and coherence, rhetorical structure etc. (Ahmad, 2019). The role of LAL in raising awareness of teachers for assessment of academic writing has gained further importance in the wake of the inclusion of digital assessment.

Academic writing, especially before the COVID-19 pandemic, has been subjected to traditional or teacher-led assessment methods mostly. Depending on the academic context, level and course design, the writing product whether in the form of timed writing administered in real-time examination setting or in the form of an assignment involved teachers using holistic or analytic criteria to subjectively rate the quality of students' writing (Brown, 2004). The interpretation of the results depends on the skill of the teacher and may vary across different

raters. LAL can minimize variation in rater scoring by enabling the raters to have a skill set whereby they can produce fair, reliable and valid test results by design transparent and consistent assessment criteria.

Recent years have witnessed an enhanced use of integrated technology in both classroom teaching and assessment. Virtual Learning Environment (VLE) such as the Moodle and Blackboard along with score of other digital platforms have introduced a variety of assessment tools. Using these tools to obtain automated grades, feedback, peer reviews etc. involve LAL whereby the teachers could develop a new skill set to cope with the latest trends in assessment. O'Loughlin and McNamara (2018) have found that these digital tools unfold new challenges for the teachers of academic writing not only in terms of administering the assessment but also in interpreting the test scores. Moreover, the similarity check reports generated by plagiarism detection software tools such as iThenticate, Turnitin, SafeAssign etc. need a reviewed insight into how feedback should be interpreted and later adapted in specific pedagogical contexts. More importantly, each writing product is unique insofar it is original in its content, and teachers need to have a very strong cognizance of the author identity and his product while dealing with reliability and validity of digital assessment.

Whether teachers are engaged with traditional or digital assessment, they have an obligation to create a poise between the assessment criteria which is objective in design and their personal approach to using those criteria which is subjective. Lee and Chen (2021) have urged upon teachers to acquire LAL in the use of both traditional and digital assessment methods so that they could accomplish fairness, transparency and quality in assessment.

Developing Language Assessment Literacy

Acquiring LAL is an evolving process which involves institution-led formal training as well as self-regulated professional development initiatives (Hansen & Eriksson, 2019). Studies such as that of Macmillan (2020) have illustrated that individual reflection and professional collaboration may serve as needs analysis interventions to help language teachers identify their needs and wants in regard to LAL. Recent years have witnessed widespread inclusion of digital technologies in assessment, which entails that the language teachers have now need to both equip themselves with the understanding and use of these digital tools and ensure reliable and valid interpretation of the results obtained from them. This training will also help teachers to administer and report assessment which blends both traditional and digital methods.

Achieving the effective levels of LAL is not without its challenges, though. Popham (2009) refers to the rapid inclusion of digital technologies in education which demand constant updating of knowledge and skills to implement the latest tools effectively. LAL deficit in this context may inhibit teachers from successfully ensuring consistency, transparency and objectivity while carrying out formative and summative assessment either traditionally or digitally or blending both. Nevertheless, using digital technologies in assessment is potentially rich in stimulating innovative practices, fostering self-directed learning, and collaborative feedback (Penny & Watson, 2019).

Considerations for the Assessment of Academic Writing

The study assumes that LAL programmes should prioritize knowledge and skill development of the academic writing teachers whereby they are able to implement the notions of fairness, accuracy, impact, feedback and flexibility in their assessment practices.

Fairness in assessment

Fairness in language assessment refers to an equitable assessment context where success of students from diverse cultural and linguistic backgrounds is not hindered by any of the variables of the assessment design. Fairness as a construct in language assessment carries serious implications for test makers, test takers and other stakeholders such as the teachers (Fan & Knoch, 2019 p.118). Test fairness has been conceptualized both subjectively and objectively (Wallace & Qin, 2021). Following Wallace (2018), subjective fairness can be procedural and refers to the uniformity of test administration procedures, or it can be interactional highlighting the quality of interaction between the teacher and the students or can be distributive measuring the extent to which the test scores reflect students' performance. The objective evaluation of test fairness involves use of the statistical procedures such as the psychometric analysis to identify any bias in favour of or against the test takers (Kunnan, 2018).

However, distributional fairness is most commonly employed in academic settings. It ascertains the correlation between students' test scores and their writing performance as measured by the raters both in conventional (rater-mediated) and online assessment. Raters have the most significant role in rater-mediated conventional and online assessment and thereby possess a huge impact on test fairness. Following McNamara (1996), the rater-mediated language assessment is impacted by rater severity, intra- and inter-rater reliability, task complexity, and assessment specifications on the measurement scales. Most research (e.g. Wallace, 2018, Wallace & Qin, 2021) has shown student dissatisfaction with the test scores

they obtained, and this challenges the scoring rubrics, test standardization, and raters' assessment of the writing scripts. Distributional performance is also impacted by the construct of writing and the assessment descriptors that have been operationalized for a certain genre. The assessment criteria for an expository essay, a case study and a term-paper are likely to be different. IELTS assessment specifications for Task 2 contain five main descriptors (Task completion, organization, cohesion and coherence, grammatical accuracy, and lexical range) which are individually defined on a five-point measurement scale. A case study might include content, sources, structure, grammar, and conventions in assessment criteria with further specifications of each descriptor. Some assessment contexts have different percentage for each descriptor which also complicates the rater performance. Even after rigorous test standardization practice, rater performance is likely to be affected while measuring any of the criteria. For instance, the element of cohesion could be problematic to deal with especially in view of the taxonomy of cohesive devices. The raters may have variance in their understanding of the notions of "task completion" or "lexical range" or "balance between the content and authorial stance". Similarly, the use of resources in academic writing cannot be gauged and evaluated fairly unless a plagiarism report from a reliable programme such as Turnitin or iThenticate is obtained.

Fairness has been measured on various frameworks such as that of Kunnan (2004) which analyzes the concept from five dimensions: validity, absence of bias, access, administration, and social consequences. Xi (2010) conceptualized fairness as being aligned with validity for identifiable and relevant groups across all stages of assessment, from assessment conceptualization to the use of assessment results' (p. 154). She expanded her notion of validity argument to include domain description, evaluation, generalization, explanation, extrapolation, and utilization as well as the counter-arguments which would rebut the fairness argument for each of these parameters. For example, in the case of TOEFL iBT, she argued that the counter-arguments for the inference of domain description might include: 1) assessment tasks are not equally representative of the academic domain for different groups; 2) critical English language skills, knowledge, and processes required for some sub-domains are not accessed; and 3) varieties of English included in the test are not representative of the domain (p. 159).

Fairness is achieved through creating standardized prompts and rubrics which are comprehensible and doable for all students undergoing a particular assessment. However, digital assessment has certain additional aspects such as the access to digital tools, ability to use these tools for test taking, and any other disparities that technology may offer. To minimize

fairness deficit, collaborative test development can be introduced to remove the element of subjectivity, thereby creating test items that are reliable and transparent. Similarly, more than one rater can be employed to ensure that not only the results are accurate but also the interpretation of the results is reliable. Apart from being compatible with multiple levels of digital literacy, digital assessment should also ensure smooth working of the entire digital paraphernalia to avoid any disruption in the assessment process.

Digital assessment of academic writing with the help of software is much more complicated than other types of language assessment which measure the psychometric or statistical significance of the language items or the whole test (Fan & Knoch, 2019). Online assessment minimizes or eliminates the role of the raters but is not without its limitations. In case of itemized assessment tasks such as the true/false or multiple-choice questions (MCQs), online assessment yields reliable results and has, therefore, high distributive fairness as compared with the rater-mediated assessment. But assessment of academic writing involves evaluation of textual features such as synthesis, criticality, thematic progression etc. which an online assessment programme cannot adequately perform. Test fairness from this perspective is likely to be compromised.

Accuracy in assessment

The measure to which assessment design correctly and consistently measures and reports students' linguistic competence may be referred to as accuracy (Bachman & Palmer, 2010). Test reliability, validity and the impact of test results on students' performance are the key variables in gauging the levels of accuracy. Another key feature is the absence of rater bias (Brown & Abeywickrama, 2019) which facilitates in achieving higher levels of accuracy thereby enabling teachers to make informed decisions about students' performance, learning needs and qualification benchmarks (Fulcher & Davidson, 2007).

Traditional or on-campus assessment such as the quizzes or essay writing examinations is administered in highly controlled settings as opposed to the digital assessment where the test takers may indulge in producing plagiarized text or seek help from other sources. To ensure accuracy in assessing academic writing, the notion of academic integrity must form part of the grading system, and teachers must have access to plagiarism detection software such as iThenticate, Turnitin, SafeAssign etc. Equally important is the rater training in implementing these tools as well as the creation of a consistent scoring criteria.

Assessment design determines the extent to which it may produce accurate results. Following Alderson (2005), objective types testing such as that based on multiple choice question (MCQ) or True/False format is not only easy to assess but also produces higher levels of accuracy as compared with, for instance, subjective tests such as the timed essay or speaking task where rater subjectivity may affect the levels of accuracy. In case of academic writing, a test design based on MCQs relating to text analysis of a specific genre not only produces reliable results but also provides insights into the conceptual understanding of the students on the subject. On the other hand, grading an exam essay may involve different raters which can lead to varied levels of accuracy depending on how the raters interpret and implement the assessment criteria. Weigle (2002) suggests that to increase rater reliability, it is imperative to provide training to the raters and create robust assessment criteria based on specific descriptors that raters can consistently adhere to. Students themselves are a crucial variable in the scheme of accuracy in assessment. For example, students' linguistic background, previous test taking experience, test taking strategies, levels of test anxiety etc. have been reported to impact assessment results (Cheng & Fox, 2017; De Jong, 2016).

The concept of accuracy in language assessment, especially in the context of academic writing is marred with complexities, however. Writing tasks designed for standardized tests such as the IELTS or TOEFL may have higher reliability but may showcase deficit in the use of authentic language. On the other hand, timed essays administered in real-time settings may have lower reliability but may reflect authentic language use. Handling washback affect may also pose another challenge as have been witnessed in the case of standardized tests where test takers with better test-taking skills may score higher as compared with those with little or no practice in taking the target tests.

Researchers have proposed a number of measures to ensure higher levels of accuracy in language assessment. These include using a variety of assessment methods blending objective and subjective components judiciously (Brown, 2012); training raters in the art of test marking and grade interpretation (Knoch, 2011); involving automated assessment tools (Chapelle & Chung, 2010); and designing culturally and linguistically contextualized assessment tasks to reduce bias (McNamara, 2000).

Impact in assessment

The notion of impact of assessment transcends the traditional approach to grading. The impact refers to the implications of assessment design on pedagogic practices, learner motivation and

achievement of course learning objectives. While immediate feedback is received in traditional assessment and which can lead to improvements in assessment design through constructive consultation, digital technologies produce written or automated report which needs to be first aligned with the assessment design and then interpreted to ascertain the impact. The impact can be adversely affected due to any disruption in the assessment technology or faulty interpretation of the automated feedback. Assessment designs which incorporate evaluation of higher-order language skills and authentic language use, for instance, are likely to produce positive impact. Digital assessment tools can exploit multimedia reports, collaborative consultations and interactive media to ensure students' active participation. Traditional and digital assessment should focus on reducing test anxiety among the students through the development of transparent rubrics and considerations for holistic learning apart from the examination results.

Impact in assessment is generally referred to as the washback effect and according to Bachman & Palmer (2010) it has significant impact on second language learning, syllabus design, pedagogical methods, students' motivation and institutional policies. Research such as that of Cheng (2014) has established that standardized language tests such as the IELTS and TOEFL and summative assessment as is administered to grade students' performance have generated negative washback since these cause high levels of test anxiety among the students. On the other hand, formative assessment tasks coupled with constructive feedback have been found conducive to effective learning experiences, self-regulated learning and higher levels of motivation (Black & Wiliam, 1998). Dörnyei (2005) suggests that making assessment tasks compatible with the students' language learning needs and authentic use can produce positive washback thereby enhancing students' motivation and second language acquisition.

Assessment in second or language contexts is dictated by institutional and curricular priorities which determine the levels of achievement in terms of grades. Therefore, teach-to-test perspective takes preference over actual language performance (Shohamy, 2001). In order to enhance linguistic competence among the students, language programmes resort to alternative assessment practice which include maintaining portfolio tasks, panel discussions, in-class assignments etc. which have been reported to promote not only language proficiency but also academic literacy (Rea-Dickins, 2004). Following Fulcher (2012), the need to develop the language assessment literacy of teachers becomes more evident than before so that they better understand not only the role of assessment in language learning but also design assessment tasks which genuinely produce positive washback effects.

Feedback in assessment

Academic writing development cannot be fully realized without the feedback which is at the center of language assessment be it traditional or digital. Typically, feedback on writing is based on teachers' comments either individually or in groups, while digital modes can include annotated remarks, audio-video commentary, and digitally produced feedback. Feedback must be focused, workable, and timely. Digital platforms have the added advantage of enabling the raters to track the revisions or identify frequently occurring patterns in writing. Blending teacher-led and digital feedback is likely to generate viable choices for the teachers to modify their teaching practices and develop course design to help students attain academic writing proficiency appropriate to their discourse community standards.

A crucial variable in gauging students' development of writing proficiency as well as the effectiveness of teaching and the instructional material is the feedback. It is primarily aimed at identifying strengths and weaknesses in regard to the teaching methodology, course design and materials and students' progress. Moreover, owing to its exploratory and expository nature, feedback has also been found to enhance self-regulated learning (Hyland & Hyland, 2006). Since students receive an explicit evaluation of their performance and can modify their learning strategies accordingly.

A variety of feedback patterns are used to help students with the development of their academic writing skills. Following Black and William (1998), formative feedback caters for the immediate needs of the students as it is embedded in the in-class teaching and learning process. Based on the circumstantial evidence, the teacher can prioritize areas of focus that may include comments on syntax, lexical ranges, discourse features such as cohesion and coherence, rhetorical organization etc. Summative feedback, on the other hand, involves reports on those test results which are formally graded and count towards the success or failure of the course. Brown and Peterson (2019) opine that the summative feedback provides a holistic overview of the students' competence and is less focused than the formative. Topping (2009) has referred to peer feedback as a very effective tool for enhancing students' critical thinking and writing proficiency. It also promotes self-regulated learning as it empowers the students to take control of their own learning and with low levels of anxiety which is otherwise a feature of other types of feedback. Teacher-led feedback is, however, the most predominant form of feedback. The teacher observes students' performance and based on his skill and experience, he can comment

on the outcomes either on a personalized, group or whole class level. Additionally, he may choose to use written or oral comments, annotations, and audio-visual recordings.

The way the feedback is shared is crucial to better learning experiences for the students. Feedback should incorporate a judicious appraisal of the strengths and weaknesses with strengths being duly appreciated so that the students feel motivated. The weaknesses should be referred to as “areas for improvement” or any other pacifying expression so that the teacher or peer comments are not demotivating. The comments should be clear, comprehensible, logical and workable, and as Sadler (1989) puts it, ambiguous comments can lead to further confusion and may impede the improvements in learning. Similarly, feedback should be timely as students can retain it and can respond to the required amendments or changes appropriately. Zimmerman (2002) suggests that effective feedback always encourages self-reflection which helps the students to fully understand the areas of improvement and thereby adopt relevant strategies for improved learning experiences.

Caution, however, be taken while sharing feedback for excessive and irrelevant comments may cause negative anxiety among the recipients. Bitchener and Knoch (2009) call for selective focus in feedback whereby the students receive feedback which is specific and doable. Moreover, feedback on academic language can be more challenging in view of the varying levels of students’ writing competence, and therefore, the teachers must ensure that they use a variety of feedback strategies to help student diversity. According to Lea and Street (1998), feedback on academic writing involves more complicated issues than a simple commentary on syntactic and lexical features. Academic writing is predominantly genre based, and a range of academic literacy skills are embedded in the process of producing a certain piece of academic discourse. Teachers of academic writing are expected to have a clear understanding of the academic literacies and are appropriately trained to transmit these to their students using a range of feedback patterns.

Flexibility in assessment

Flexibility caters for assessment design to be adaptive to learner diversity and contextual preferences. Traditional assessment designs which include tests such as the timed essays or in-class quizzes have proved to be less flexible as compared with digital methods. In traditional methods, the assessment design has to be consistent for all target examinees. Creating flexibility in traditional testing may involve creating multiple versions for the same test but this may question issues of reliability and validity. Digital assessment, however, has been

considered more flexible for being asynchronous as it can accommodate a variety of test types. Flexibility can be improved through a number of measures which include but are not limited to allowing students to resubmit, take a makeup test, making submission date flexible, and by providing students with alternative assessment choices such as project writing, video presentation etc.

Yet another perspective on flexibility in language assessment in the higher education sector is to involving the test takers in the assessment processes (Boud & Falchikov, 2006; Nicol & MacFarlane-Dick, 2006). It is premised on the notion of alignment between the learning outcomes and assessment practices which if adequately realized can be achieved through a variety of assessment tasks. For instance, an academic essay tests students' ability to process, organize, analyze and synthesize information in conformity with the target genre using academically appropriate lexicogrammar resources and discourse features. The same ability, although arguable, can be assessed from a blog or a wiki or a video presentation, and therefore, allows students to choose from alternative formats. However, being a fluid concept to-date, flexibility has not received the desired inclusion into the mainstream assessment practices across the academia (Knight, 2002). First, it can be approached from a different set of priorities such as permitting the students' choices in assessment matters relating to the test format and assessment criteria. A few studies report assessment contexts where the students have been individually allowed to choose weightings allocated to the tasks (Jamieson, 2005), and limited control in regard to the test specifications, assessment procedures, and interpretation of the criteria and result (Ellis and Folley, 2009b).

Flexibility in academic writing assessment especially the rater-mediated could be more challenging to implement in view of the complexities involved in the task design, assessment format, and marking criteria. In one study, Kehoe, Tennent, and Windeknecht (2004) found that flexibility was misinterpreted by the tutors who gave a variety of online assessment tasks instead of allowing students control over one assessment type. The result was an overload for the teachers, non-compliance with the assessment criteria by the students. Hence, the conformity of the learning outcomes with the assessment criteria was not achieved. Flexibility in assessment is also vulnerable to generating a limited set of learning outcomes as students are likely to prefer only those assessment components which they feel easy for them. For instance, a choice between an academic essay and an online test will fail to showcase test takers' ability to engage with a writing genre in academic contexts.

Academic writing involves an awareness of and ability to produce a variety of genre which students encounter in their present and future studies as well as in their professions. Universities are responsible for developing both the awareness and the ability which is measured through the assessment components. Giving students unbridled choice in assessment matters can adversely affect the institutional policies. Following Irwin and Hepplestone (2012), flexible assessment is more closely aligned with criterion referenced assessment rather than with norm referenced assessment seeking comparisons between student grades. However, Lindsay (2007) observes the recruiters might wish to compare between students but flexibility does not allow this comparison. Flexibility as an assessment concept can be practiced with the formative tasks where students are allowed to choose between only those test types which have a similar set of learning objectives and share a similar assessment criterion. Summative assessment needs to be assessed according to the disciplinary preferences and curricular choices.

Conclusion

Assuming writing being the most complex of the language skills (Nunan, 2000), and academic writing being one of the most formal manifestations of the skill, the role LAL in preparing teachers for the high standards of pedagogy and language assessment becomes more crucial than ever before. Keeping abreast with the latest trends in the domain of academic writing assessment influenced by the evolving digital platforms, it is imperative for the teachers to garner LAL that is contemporaneous and comprehensive. LAL, therefore, is expected to encourage professional development, empower teachers to create assessment tasks successfully, contribute to the achievement of assessment benchmarks, synchronize teaching and assessment – all to the benefit of the teaching and learning processes involved in the development of academic writing skill. Language assessment for academic writing involves diverse approaches to assessment either through teacher-led or digital interventions. Despite having disparate merits, both approaches should incorporate the notions of fairness, accuracy, impact, feedback, and flexibility in order to produce reliable and valid test results, thereby facilitating the novice writers acquire academic writing proficiency specific to their subject specialism.

References

Ahmad, Z. (2019). Analyzing argumentative essay as an academic genre on assessment frameworks of IELTS and TOEFL. In Hidri, S. (Ed). English language teaching research in the Middle East and North Africa: Multiple perspectives. pp. 279-299, Palgrave Macmillan. <https://doi.org/10.1007/978-3-319-98533-6>

Ahmad, Z. (2021). Teachers' assessment of academic writing: Implications for language assessment literacy. In S. Hidri (Ed). Perspectives on language assessment literacy: Challenges for improved student learning (1st ed). Routledge.

Ahmad, Z. (2020). Summative assessment, test scores and text quality: A study of cohesion as an unspecified descriptor in the assessment scale. European Journal of Educational Research, 9(2), 523 535. <https://doi.org/10.12973/eu-jer.9.2.523>

Bachman, L. (2004). Statistical analyses for language assessment. Cambridge University Press.

Bachman, L. F., & Palmer, A. S. (2010). Language assessment in practice: Developing language assessments and justifying their use in the real world. Oxford University Press.

Beaufort, A. (2007). College writing and beyond: A new framework for university writing instruction. Utah State University Press. <https://doi.org/10.2307/j.ctt4cgnk0>

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7-74.

Bitchener, J., & Knoch, U. (2009). The contribution of written corrective feedback to language learners' development of language knowledge. Journal of Second Language Writing, 18(4), 248-265.

Boud, D. and Falchikov, N. (2006). Aligning assessment with long-term learning. Assessment & Evaluation in Higher Education 31, no. 4: 399-413.

Brown, H. D. (2004). Language assessment: Principles and classroom practices. Pearson Education.

Brown, J. D. (2012). Developing, Using, and Analyzing Rubrics in Language Assessment with Case Studies in Asian and Pacific Languages. Peter Lang.

Brown, H. D., & Peterson, P. (2014). Principles of language learning and teaching. Pearson Education.

Brown, H. D., & Abeywickrama, P. (2019). Language Assessment: Principles and Classroom Practices (3rd ed.). Pearson Education.

Chapelle, C. A., & Chung, Y. R. (2010). The Routledge Handbook of Second Language Acquisition and Language Testing. Routledge.

Cheng, L. (2014). Examining the impact of high-stakes language testing on school pupils in the UK. Assessing Writing, 21, 72-82.

Cheng, L., & Fox, J. (2017). Assessment in the Language Classroom: Teachers' Supporting Assessment. Palgrave Macmillan.

Cumming, A. (2001). Learning to write in a second language: Two decades of research. *The TESOL Quarterly*, 35(2), 91-119.

De Jong, J. H. (2016). A Framework for Language Proficiency Development and Alignment. Springer.

Dörnyei, Z. (2005). The psychology of the language learner: Individual differences in second language acquisition. Lawrence Erlbaum Associates.

Ellis, C. and Folley, S. (2009b). The use of scoring rubrics to assist in the management of increased student assessment choice. Paper presented at ALT-C 2009, September 8-10, in Manchester, UK.

Gibbs, G. (2006). How assessment frames student learning. *Innovative Assessment in Higher Education*. 23-36. 10.4324/9780203969670.

Fan, J & Knoch, U. (2019). Fairness in language assessment: What can the Rasch model offer? *Papers in Language Testing and Assessment*, 8(2), 117-142

Fox, J., & Macpherson, R. (2021). Designing fair writing rubrics: A critical review of assessment literacy. *Journal of Writing Research*, 16(3), 411-430.

Fulcher, G., & Davidson, F. (2007). *Language Testing and Assessment: An Advanced Resource Book*. Routledge.

Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113-132.

Hansen, R., & Eriksson, M. (2019). Enhancing language assessment literacy through professional development. *Journal of Teacher Education*, 70(1), 14-28.

Hyland, K., & Hyland, F. (2006). Feedback on feedback: The rights and responsibilities of students and teachers. In P. K. Matsuda & T. Silva (Eds.), *Second language writing research: Perspectives on the process of writing* (pp. 171-188). Lawrence Erlbaum.

Hyland, K. (2018). Academic writing and identity: The discursive construction of knowledge. Longman.

Inbar-Lourie, O. (2013). Guest Editorial to the special issue on language assessment literacy. *Language Testing*. 30. 301-307. 10.1177/0265532213480126.

Irwin, B. and Hepplestone, S. (2012). Examining increased flexibility in assessment formats. *Assessment & Evaluation in Higher Education*, 37 (7), 773-785.

Jamieson, G. (2005). Weighting assessment tasks according to student preferences: Differences relating to a student's primary language. Paper presented at Making a Difference: 2005 Evaluations and Assessment Conference, December 1, in Sydney, Australia.

Kehoe, J., Tennent, B. and Windeknecht, K. (2004). The challenge of flexible and non-traditional learning and teaching methods: Best practice in every situation? *Studies in Learning, Evaluation, Innovation and Development* 1, 1: 56-63.

Knight, P. T. (2002). Summative assessment in Higher Education: Practice in disarray. *Studies in Higher Education* 27, no. 3: 275-286.

Knoch, U. (2011). Rating Scales for Diagnostic Assessment of Writing: A Study on the Impact of the Rating Scale Design on Performance Scores. Peter Lang.

Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. J. Weir (Eds.), European language testing in a global context: Proceedings of the ALTE Barcelona Conference (pp. 27-48). Cambridge: Cambridge University Press.

Kunnan, A. J. (2018). Evaluating language assessments. Routledge.

Lea, M. R., & Street, B. V. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23(2), 157-172.

Lee, C., & Chen, H. (2021). Online writing assessments in the post-pandemic era: Opportunities and challenges. *Computers and Composition*, 58, 101669.

Lindsay, R. (2007). Review of 'Towards inclusive learning in higher education. (ed.) Adams, M. and Brown, S., Routledge, 2006.' *British Journal of Educational Technology* 38, no. 6: 1133-1134.

Macmillan, D. (2020). Reflective practices in teacher education: Building language assessment literacy. *TESOL Journal*, 11(2), e4500.

McNamara, T. (1996). Measuring second language proficiency. London: Longman.

McNamara, T. (2000). Language Testing. Oxford University Press.

Nicol, D. and MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education* 31(2): 199-218.

Nunan, D. (2000). Language teaching methodology: Textbook for teachers. Malaysia: Longman

O'Loughlin, K., & McNamara, T. (2018). Assessing language proficiency in the digital age. *Language Assessment Quarterly*, 15(2), 89-108.

Penny, C., & Watson, R. (2019). Digital assessment in language education: Benefits and challenges. *Language Learning & Technology*, 23(1), 1-18.

Popham, W. J. (2009). Assessment for educational leaders. Pearson Education.

Rea-Dickins, P. (2004). Understanding teachers as agents of assessment. *Language Testing*, 21(3), 249-258.

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

Shohamy, E. (2001). The power of tests: A critical perspective on the uses of language tests. Pearson Education.

Topping, K. (2009). Peer assessment. *Theory into Practice*, 48(1), 20-27.

Wallace, M. P. (2018). Fairness and justice in L2 classroom assessment: Perceptions from test takers. *Journal of Asia TEFL*, 15(4), 900- 1238.
<http://www.doi.org/10.18823/asiatefl.2018.15.4.11.1051>

Wallace, M.P. and Quin, C.Y. (2021). Language classroom assessment fairness: Perceptions from students. *LEARN Journal: Language Education and Acquisition Research Network*, 14(1), 492-521.

Weigle, S. C. (2002). Assessing writing. Cambridge University Press.

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing*, 20(10), 1-24.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, 41(2), 64-70.

مهارة تقييم اللغة: دراسة نظرية للتقييم الرقمي والتقييم الموجه بواسطة المعلم لمهارة الكتابة الأكademية باللغة الإنجليزية

د. عثمان أحمد الجهني

معهد اللغة الإنجليزية - جامعة جدة

oaaljohani@uj.edu.sa

ملخص:

يُعد تقييم الكتابة الأكademية في سياقات اللغات الأجنبية مهمة شاقةً ومعقدةً للمعلمين. ترتكز هذه الدراسة النظرية على فكرة أن إتقان المعلمين المشاركين في تدريس وتقييم الكتابة الأكademية باللغة الإنجليزية في التعليم العالي لمهارات التقييم اللغوي أمرٌ بالغ الأهمية، ليس فقط لنجاح تقديم المقررات الدراسية، بل أيضاً لتحقيق نتائج تعلم موثوقة وصالحة. وقد تم اعتماد نهج تحليل المحتوى لتحديد مدى أهمية إتقان مهارات التقييم اللغوي لعلمي الكتابة الأكademية قبل محاولتهم تصميم أدوات التقييم ومعاييره ونتائجها وتقييمها وتفسيرها. وتقدم الدراسة لمحنةً نظريةً شاملةً عن المناهج المعاصرة لمهارات التقييم الرقمي والتقليدي والتقييم الموجه بواسطة المعلم، وذلك بهدف رفع مستوى وعي المعلمين بخصائص التقييم الخمس: الإنصاف، والدقة، والتأثير، والتغذية الراجعة، والمرنة. وخلصت الدراسة إلى أن التدريب الجيد للمعلمين في مجال إتقان مهارات التقييم اللغوي، بالإضافة إلى إتقان مهارات التقييم الرقمي، أمرٌ ضروري لتزويدهم بالمعرفة النظرية والمهارات العملية اللازمة لإتقان قياس أداء الكتابة الأكademية بما يتماشى مع أهداف تعلم المقرر، ويقدم تفسير دقيق لعرض تقارير موثوقة وصالحة لتطور مهارات الكتابة لدى الطلاب