A Theoretical Review on the Need to Use Standardized Oral Assessment Rubrics for ESL Learners in Saudi Arabia

Reem Aamer Alaamer[1,2]

[1] Department of English, Faculty of Languages and Translation, King Khalid University, Abha, Saudi Arabia

[2] Department of English Language and Translation, Ar Rass Faculty of Science and Arts, Qassim University, Qassim, Saudi Arabia

Correspondence: Reem Aamer Alaamer, Department of English, Faculty of Languages and Translation, King Khalid University, Abha, Saudi Arabia & Department of English Language and Translation, Ar Rass Faculty of Science and Arts, Qassim University, Qassim, Saudi Arabia.

Abstract

There is a growing need for standardized oral assessment rubrics in learning institutions. This is linked to the growing number of ESL learners not only in Saudi Arabia but other parts of the world. To assess the need to use standardized oral assessment rubrics, this particular study explores various peer reviewed articles that support the use of standardized rubrics while assessing oral skills among ESL learners in Saudi Arabia. Standardized rubrics give students a reference point in regards to what is expected while learning oral skills. As a result, students are able to work towards improving their skills to meet the standards of the rubric. Various scholars have given different definitions for the term rubric. In all the definitions, grading criteria is a common feature. Some experts have stated that, modern rubrics should go beyond grading to guiding students in understanding their expectations in oral tests. When developing standardized rubrics, teachers should ensure that the rubrics meet the required validity and reliability to assist ESL learners in meeting their goals. Literature shows that there is a gap in the current oral assessment rubrics in Saudi Arabia, and it requires a prompt review. Therefore, developing a standardized rubric should take a multidisciplinary approach. Scholars and experts teaching ESL students must be consulted to ensure all important factors are considered and incorporated in the standardized rubric.

Keywords: standardized assessment rubric, ESL learners, speaking skills, oral assessments, teachers, EFL context

1. Introduction

Over the past few decades, the attention of foreign language teaching has shifted from enhancing leaners' linguistic competence to developing communicative competence. Consequently, assessing oral skills has received more attention in second language acquisition (SAL) research. In English as a foreign language (EFL) context, learners are expected to speak fluently and communicate effectively. However, speaking is considered as the most difficult skill to evaluate in comparison to other skills. One reason for this challenge is that, speaking involves cognitive processes that are not equally evaluated in other skills. Additionally, learning speaking skills requires more effort as compared to other skills such as listening, reading, and writing. For most ESL learners, including those in Saudi Arabia, the most stressful exams are those that test speaking skills. Notably, traditional assessment tools lack the ability to evaluate certain aspects of speaking skills (Oosterhof, 2001). This is due to the unclear assessment procedures which make learners develop anxiety. Unlike other skilled-based exams, which clearly state the evaluation methods, the nature of oral exams requires teachers to carefully plan the execution stage of the exam in order to accurately assess a learners' skills. Both teachers and learners believe that assessing oral exams is a challenge that adversely affect the intended learning outcomes. Therefore, assessing oral skills requires special rubrics or criteria to ensure the validity and reliability of oral tests, for ESL learners.

2. Significance of the Study

This study is significant since it informs on the need for Standardized Oral Assessment Rubrics for ESL Learners in Saudi Arabia. Stakeholders in the education center will also understand the benefits of this form of rubric, thus

make policies that will enhance its implementation. Additionally, the study also forms the basis for conducting further studies on the development of standardized oral assessment rubrics that will lead to the improvement of quality of education in Saudi Arabia.

3. Critical Analysis

Assessing speaking skills is a challenging phenomenon for many reasons (Navidinia et al., 2019). First, the nature of speaking requires an examiner to listen carefully to the speaker. However, there are internal and external factors that may hinder the evaluation process. Second, speaking is a complicated skill which often requires assessors to clearly define the assessment benchmark with examples. Hence, the criteria for distinguishing a good speaker from a poor one may differ greatly. Additionally, assessing speaking skills is a difficult task since it involves evaluating individual learners separately (Luoma, 2004). Due to these challenges, universities have not given speaking assessments enough attention in Saudi Arabia and other countries around the globe (Egan, 1999). By developing standardized oral rubrics, more learning institutions will focus on improving their learners' speaking skills.

There are numerous benefits of presenting well-designed rubrics. First, they stimulate feedback from students, when applied in classroom as a formative assessment method. Additionally, they can be used in a summative way to help students compare between their self-assessment and a grader's judgment. Oral assessments aim at: measuring language proficiency, identifying strengths and weaknesses of learners, evaluating the course objectives and placing learners in a teaching program. Therefore, in assessing the ability to communicate, a test must be designed in a way that shows the learners' ability to demonstrate their interactional competence in using the target language. Traditionally, a test should have three elements: practicality, validity, and reliability (Nodoushan, 2009a). Lately, learners' task performance is preferred since it involves assessment of learners' performance on an assigned task (Nodoushan, 2008a). To assess student's speaking abilities, teachers should test their interactive and communicative ability by encouraging them to participate in authentic speaking activities. In such tasks, learners' level of mastering speech aspects such as pronunciation, intonation and pragmatic characteristics are measured.

Table 1. Benefits of a well-designed rubrics

Benefits of a well-designed rubric
1. It acts as a summative way to help students compare between their self-assessment and a grader's judgment
2. Stimulates feedback from students
3. Measures language proficiency
4. Identifies strengths and weaknesses of learners
5. Helps in evaluating the course objectives
6. Assists in placing learners in a teaching program

In Saudi Arabia, oral assessments in EFL classroom take different forms including; interviews, pair and group tasks and traditional scoring rubrics. Interviews are used to test learners' speaking skills, in which a face-to-face interaction takes place between the learner and the interviewer. However, such interviews are criticized since they do not represent real-life situations. Similarly, pair and group tasks are used to assess learners' speaking skills (Egyud & Glover, 2001; Nodoushan, 2002). They are advantageous since they save time by assessing more than one learner at a time. According to Ussama's (2013) study, traditional interview scoring rubric based on O'Loughlin's (2001) guidelines is also used to assess the learners' language speaking proficiency.

Teachers may not have the required knowledge to assess students' speaking ability. Therefore, there is a need to develop an authentic oral assessment rubric that measures students' oral performance. This is essential to ensure the validity of the assessment process. For a speaking assessment tool to be valid, it must have the following characteristics. Firstly, it should reflect communicative competence. Second, it should contain a valid representation of the measured constructs in the scoring rubric, and it must have a reliable scoring procedure. Teachers must take into consideration the purposes and contexts of assessment when they develop their scoring rubric. Examples of scoring rubrics include ETS and TOEFL. In ETS TOEFL (2004) scoring rubric, the relationship between pronunciation and fluency is emphasized as well as grammar and vocabulary that are linked to a learners' communicative competence.

There is an urgent need to design a standard speaking assessment rubric that helps in identifying strengths and weaknesses of EFL learners in Saudi Arabia (Nodoushan, 2008a). In assessing the speaking skills of Saudi undergraduates, it is common for teachers to develop or use traditional rubrics. Thus, most teachers often feel frustrated and lost while assessing students' speaking performance in oral tests. In other words, at Saudi

universities, ESL courses do not have standardized rubrics developed by experts, and if any exist, they lack an accurate scoring system. Even though these rubrics provide teachers with the needed scoring criteria, they may not give reliable feedback to improve learners' performance.

In Saudi Arabia, most studies focus on addressing assessment rubrics of other skills other than speaking skills. For example, a study by Alahmadi et al. (2019) investigated the impact of a formative speaking assessment on Saudi undergraduates' performance in a summative test. Similarly, Aldukhayel (2017) explored the clarity and familiarity of three scoring rubrics that were used to assess students' writing achievement in the preparatory year program (PYP) in a Saudi university. Another study by Alshakhi (2019) aimed at exploring the weaknesses of the Saudi writing assessment methods employed at an English Language Institute. The results showed that replacing the existing holistic rubric with an analytic one can enhance contextual-based learning and eliminate cross-grading, to improve student-teacher relationship. In the same vein, Al-Serhani's (2007) study investigated the effect of portfolio assessment method on students' writing performance in EFL classrooms. It is clear from the findings of these studies that, most of the research conducted in the Saudi EFL context is related to the writing assessment hence, more research is required to develop speaking assessment rubrics.

Example of Holistic Rubric on writing assessment

Articulating thoughts through written communication— final paper/project.

4. **Above Average:** The audience is able to easily identify the focus of the work and is engaged by its clear focus and relevant details. Information is presented logically and naturally. There are no more than two mechanical errors or misspelled words to distract the reader.

3. **Sufficient**: The audience is easily able to identify the focus of the student work which is supported by relevant ideas and supporting details. Information is presented in a logical manner that is easily followed. There is minimal interruption to the work due to misspellings and/or mechanical errors.

2. **Developing**: The audience can identify the central purpose of the student work without little difficulty and supporting ideas are present and clear. The information is presented in an orderly fashion that can be followed with little difficulty. There are some misspellings and/or mechanical errors, but they do not seriously distract from the work.

1. **Needs Improvement**: The audience cannot clearly or easily identify the central ideas or purpose of the student work. Information is presented in a disorganized fashion causing the audience to have difficulty following the author's ideas. There are many misspellings and/or mechanical errors that negatively affect the audience's ability to read the work.

Figure 1. An example of holistic assessment rubric

In higher education contexts around the world, methods of assessing oral communication have received special attention (Panadero & Jonsson, 2013). Specifically, most learning institutions have narrowed their focus on techniques that can achieve the best outcomes in oral assessments (Campbell et al., 2001; Schwartz & Arena, 2013; Stoynoff, 2013). In Saudi Arabia, English language education is highly test-oriented. Therefore, learners use memorization strategies to cope with the requirements of the tests. Since more emphasis is placed on reading and test-oriented writing skills, most ESL leaners in the Kingdom of Saudi Arabia (KSA) are reluctant to speak during the tests. As a result, such students face challenges in exams that assess oral skills, due to their lack of familiarity with oral assessment methods. In my personal experience as a non-native speaker of English and an EFL teacher, I have not only experienced difficulties in assessing students' speaking skills, but I've also noted how current assessment procedures, adversely affect students' oral performance.

In the process of learning and teaching, assessments aim at collecting data about a process, in order to make decisions about improving learners' outcomes and the goals of a program. As literature suggests, the focus of learners' assessment has been on traditional written exams while oral assessment methods have received little attention due to a lack of standardized oral grading rubric. Notably, rubrics are designed to identify the characteristics of communicative competence and the essential features of performance such as strategic competence and pragmatic competence. When evaluating students' oral performance, language teachers tend to have the same evaluation criteria, but they may assess the same student differently. Therefore, standardized rubrics can be utilized to maintain consistency among teachers. This will enhance effective assessment processes and results (Dunbar et al., 2006)

Scholars have identified various features of a standardized rubric that can be employed for precise and consistent learning outcomes. For example, Saeed et al. (2019) considers a rubric as "a grading tool which consists of a set of explicit criteria that are used to assess a specific task performance" (p. 1060). Additionally, rubrics provide "descriptive statements of behaviors that candidates may exhibit in a particular sub-component" (Angelelli, 2009, p. 39, as cited in Samir & Tabatabaee-Yazdi, 2020, p. 102). According to (Sadler, 2009a), rubrics consist of a detailed grading system with either numbers or quality levels. While some rubrics contain standard quality words, such as good or below average, others may present an explanation of the quality in details. Examiners use rubrics to reduce the grading time, decrease the element of subjectivity and give learners appropriate feedback to overcome deficiencies in their performance (Stevens & Levi, 2005).

Allen and Tanner (2006) define rubric as "a type of matrix that provides scaled levels of achievement or understanding, for a set of criteria or dimensions of quality for a given type of performance..." (p. 197). In other words, a grading rubric consists of scores, performance levels, criteria, and descriptors. It also shows what is expected of learners in an assignment or a test by stating the criteria and deciding on the levels of quality, ranging from excellent to poor. Assessment rubrics can be used by teachers to score learners' performance or work. Due to their effectiveness, rubrics are crucial instruments in assessing EFL learners' oral skills. To effectively assess learners' communicative competence, teachers need to understand the relationship between the content of speech courses and their objectives. Designing and developing rubrics is mostly done by publishers and researchers (Timmerman et al., 2010). In addition, rubric banks have been provided by online sources (Dornisch & McLoughlin, 2006). This makes it easier for teachers to develop rubrics to assess their students (Andrade & Du, 2005; Boud & Soler, 2015).

For a long time, 'assessment rubrics' have been used in books and research papers. According to Dawson (2017, p. 348), there are certain elements that a rubric must have. These include: "evaluative criteria, quality definitions for the criteria at particular level and a scoring strategy". In this regard, various studies have been conducted in an attempt to evaluate rubrics. However, they may not present functional definitions of how such rubrics are used or what they intend to achieve.

Rubrics have different characteristics. Firstly, rubrics can be specific or generic (Tierney & Simon, 2004). Secondly, there is a stark distinction between rubrics created by teachers and those created with students' collaboration (Reddy & Andrade, 2010). Thirdly, a rubric can be used by students in self- and peer-assessment as well as in questioning a task requirement (Andrade & Du, 2005; Panadero & Romero, 2014). In terms of the scoring system, some rubrics use holistic strategies while others use analytic scoring strategies (Sadler, 2009a). Since scores can be multiple, certain procedures must be included to resolve disputes. There are two methods for scoring a test: the holistic (impressionistic) method, which gives an overall score to a learner's performance (Luoma, 2004), and the analytic method which gives one score for different criteria such as vocabulary, communicative ability, fluency, and pronunciation. The analytic rating score increases the reliability of the assessment process and the assessment tool (Strikaew et al., 2015). Therefore, it is preferred in highly sensitive assessments. Using holistic rubrics for evaluating students may be disadvantageous since the performance of some students may not suit a certain category. Therefore, combining both holistic and analytic rubrics is preferable for assessing students in an EFL context. Although analytic rubrics are "complicated and time consuming" (Taufiqulloh, 2009, p. 187), they are very effective in assessing communication skills. According to Han (2017), holistic assessment methods are largely valid in the context of research, placement, or certification. On the contrary, analytic scales are less biased and assess specific details when employed for the formative assessment of learners. In some rubrics, software can be employed in making judgments about learners' performance (Dimopoulos et al., 2013). Nevertheless, many rubrics lack clear evaluative criteria related to "organizations, mechanics, word choice and supporting details.

Quality descriptors or definitions are used in rubrics to represent an evaluative criterion at a certain level (Sadler, 2009b). These quality definitions need the examiners or assessors to give judgments in line with the scoring strategy. The levels of quality in a rubric may be based on learning outcomes taxonomies or grade levels (e.g., Timmerman et al., 2010). According to Biggs and Tang's (2007, it is acceptable to combine both levels. For example, their rubric includes A-D grade descriptors and four levels of SOLO taxonomy. However, some researchers recommend the use of a single-point rubric since it demonstrates only the standard performance of learners (Fluckiger, 2010).

Firoozi (2019) states that choosing valid and reliable assessment methods is essential to achieve valid results. To ensure the validity and reliability of rubrics, several processes can be used. For instance, Timmerman et al. (2010) used a quality process in which they performed statistical tests to guaranty the reliability of grading rubric. From the tests, they concluded that validity of rubrics can be achieved by consulting experts in a particular field and

comparing criteria to the ones used in similar tasks. According to a study by Dunbar et al. (2006), a competent speaker rubric should have the ability to identify a suitable time for speaking, to speak in a clear and expressive way, to present organized and understandable ideas, to listen attentively, to use effective medium for communication, to structure an appropriate message, to recognize other people's level of receptivity to the spoken message, and to support given information with examples. In another study conducted by Saeed et al. (2019), an oral rubric was developed based on Canale and Swain's communicative approach to L2 teaching and testing. This approach consists of four competencies, namely: grammatical, strategic, discourse and sociolinguistic. The content of the rubric was validated by experts in the field and the four competencies were employed to develop an effective assessment criterion. For better understanding, rubrics can be presented using a table, grid, or matrix (Sadler, 2009a). However, rubrics can have other forms of presentation such as images or other non-text representations.

4. Recommendations

Although the scope of this paper is limited to literature on using rubrics in assessing students' oral performance, rubrics can be designed and used to achieve other goals in other contexts or domains, such as in peer review of language teaching (Magno, 2012) and research evaluation (Wong et al., 2013). Thus, conducting more studies on the design, implementation and effectiveness of speaking rubrics can improve educational outcomes.

5. Conclusion

The review of literature has shown that, assessment rubrics have a pivotal role in improving learners' performance, whereas inaccurate speech assessment rubrics may cause poor evaluation of the learners' skills. Since most speaking/oral tests in the Saudi context lack a standardized, validated, and reliable scoring rubric based on which teachers can assess EFL leaners objectively, there is a need to develop standardized rubrics that will be used to evaluate undergraduate students in Saudi universities and other higher learning institutions. In addition, empirical research shows that, there is a need for more applied studies that address the issue of assessment criteria and performance bands. According to Dawson (2017) "Practitioners designing a rubric could also fruitfully reflect on the design elements as a way of revealing assumptions about rubrics" (p. 358). Therefore, designing a standardized scoring rubric is essential to provide learners with feedback with respect to the goals of the course they are learning. Consequently, students' outcomes will be improved and more learners will be encouraged to participate in oral assessment thus improving their speaking skills.

References

Alahmadi, N., Alrahaili, M., & Alshraideh, D. (2019). The impact of the formative assessment in speaking test on Saudi students' performance. *Arab World English Journal (AWEJ), 10*(1), 259-270. https://doi.org/10.24093/awej/vol10no1.22

Aldukhayel, D. M. (2017). Exploring Students' Perspectives toward Clarity and Familiarity of Writing Scoring Rubrics: The Case of Saudi EFL Students. *English Language Teaching, 10*(10), 1-9. https://doi.org/10.5539/elt.v10n10p1

Allen, D., & Tanner, K. (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE—Life Sciences Education, 5*, 197-203. https://doi.org/10.1187/cbe.06-06-0168

Al-Serhani, W. F. (2007). The Effect of Portfolio Assessment on the Writing Performance of EFL Secondary School Students in Saudi Arabia [Online Submission]. *Online Submission*. Retrieved from https://eric.ed.gov/?id=ED596227

Alshakhi, A. (2019). Revisiting the Writing Assessment Process at a Saudi English Language Institute: Problems and Solutions. *English Language Teaching, 12*(1), 176-185. https://doi.org/10.5539/elt.v12n1p176

Andrade, H., & Du, Y. (2005). Student Perspectives on Rubric-referenced Assessment. *Practical Assessment, Research & Evaluation, 10*(3), 1-11. Retrieved from https://scholarsarchive.library.albany.edu/cgi/viewcontent.cgi?article=1001&context=edpsych_fac_scholar

Angelelli, C. V. (2009). Using a rubric to assess translation ability: Defining the construct. In C. V. Angelone & Holly E. Jacobson (Eds.), *Testing and Assessment in Translation and Interpreting Studies: A call for dialogue between research and practice* (pp. 13-47). https://doi.org/10.1075/ata.xiv.03ang

Biggs, J., & Tang, C. (2007). *Teaching for Quality Learning at University*. Maidenhead: Open University Press.

Boud, D., & Soler, R. (2016). Sustainable Assessment Revisited. *Assessment & Evaluation in Higher Education, 41*(3), 400-413. https://doi.org/10.1080/02602938.2015.1018133

Campbell et al. (2001). Peer versus self-assessment of oral business presentation performance. *Business Communication Quarterly, 64*(3), 23-42. https://doi.org/10.1177/108056990106400303

Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education, 42*(3), 347-360. https://doi.org/10.1080/02602938.2015.1111294

Dimopoulos, I., Petropoulou, O., & Retalis, S. (2013). Assessing Students' Performance Using the Learning Analytics Enriched Rubrics. Paper presented at *the proceedings of the Third International Conference on Learning Analytics and Knowledge*, Leuven. https://doi.org/10.1145/2460296.2460335

Dornisch, M. M., & A. S. McLoughlin. (2006). Limitations of Web-based Rubric Resources: Addressing the Challenges. *Practical Assessment, Research & Evaluation, 11*(3), 1-8.

Dunbar et al. (2006). Oral Communication Skills in Higher Education: Using a Performance-Based Evaluation Rubric to Assess Communication Skills. *Innovative Higher Education, 31*(2), 115-128. Springer. https://doi.org/10.1007/s10755-006-9012-x

Egyud, G., & Glover, P. (2001). Oral testing in pairs: A secondary school perspective. *ELT Journal, 55*(1), 70-76. https://doi.org/10.1093/elt/55.1.70

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). https://doi.org/10.1186/s40468-019-0078-7

Fluckiger, J. (2010). Single Point Rubric: A Tool for Responsible Student Self-assessment. *Delta Kappa Gamma Bulletin, 76*(4), 18-25. Retrieved from https://digitalcommons.unomaha.edu/tedfacpub/5/

Han, T. (2017). Examining the Impact of Scoring Methods on the Institutional EFL Writing Assessment: A Turkish Perspective. *PASAA, 53*, 112-147. Retrieved from https://files.eric.ed.gov/fulltext/EJ1153666.pdf

Jonsson, A., & Svingby, G. (2007). The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences. *Educational Research Review, 2*(2), 130-144. https://doi.org/10.1016/j.edurev.2007.05.002

Lazaraton, A. (1992). The structural organization of a language interview: A conversation analytic perspective. *System, 20*, 373-86. https://doi.org/10.1016/0346-251X(92)90047-7

Luoma, S. (2004). *Assessing speaking*. Ernst Klett Sprachen. https://doi.org/10.1017/CBO9780511733017

Magno, C. (2012). Assessing Higher Education Teachers through Peer Assistance and Review. *The International Journal of Educational and Psychological Assessment, 9*(2), 104-120. Retrieved from https://ssrn.com/abstract=2287173

Morreale, S. P., Moore, M. R., Surges-Tatum, D., & Webster, L. (2007). *'The competent speaker' speech evaluation form* (2nd ed.). Washington, DC: National Communication Association. Retrieved from https://www.natcom.org/uploadedFiles/Teaching_and_Learning/Assessment_Resources/PDF-Competent_Speaker_Speech_Evaluation_Form_2ndEd.pdf

Morreale, S. P., Rubin, R. B., & Jones, E. (1998). *Speaking and listening competencies for college students.* Washington, DC: National Communication Association. Retrieved from https://www.natcom.org/uploadedFiles/Teaching_and_Learning/Assessment_Resources/PDF-Speaking_and_Listening_Competencies_for_College_Students.pdf

Navidinia, H., Mobaraki, M., & Malekzadeh, F. (2019). Investigating the effect of noticing on EFL students' speaking accuracy. *International Journal of Instruction, 12*(1), 83-98. https://doi.org/10.29333/iji.2019.1216a

O'Loughlin, K. J. (2001). *Study in language testing 13.* Cambridge: Cambridge University Press.

Oosterhof, A. (2001). *Classroom applications of educational measurement* (3rd ed.). Upper Saddle River, New Jersey, USA: Merril Prentice-Hall.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review, 9*, 129-144. https://doi.org/10.1016/j.edurev.2013.01.002

Panadero, E., & Romero, M. (2014). To Rubric or not to Rubric? The Effects of Self-assessment on Self-regulation, Performance and Self-efficacy. *Assessment in Education: Principles, Policy & Practice, 21*(2), 133-148. https://doi.org/10.1080/0969594X.2013.877872

Reddy, Y. M., & Andrade, H. (2010). A Review of Rubric Use in Higher Education. *Assessment & Evaluation in Higher Education, 35*(4), 435-448. https://doi.org/10.1080/02602930902862859

Sadler, D. R. (2009a). Indeterminacy in the Use of Preset Criteria for Assessment and Grading. *Assessment & Evaluation in Higher Education, 34*(2), 159-179. https://doi.org/10.1080/02602930801956059

Sadler, D. R. (2009b). Transforming Holistic Assessment and Grading into a Vehicle for Complex Learning. In Joughin, G. (Ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 1-19). Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-8905-3_4

Saeed, K. M., Ismail, S. A. M. M., & Eng, L. S. (2019). Malaysian Speaking Proficiency Assessment Effectiveness for Undergraduates Suffering from Minimal Descriptors. *International Journal of Instruction, 12*(1), 1059-1076. https://doi.org/10.29333/iji.2019.12168a

Salmani Nodoushan, M. A. (2002). *Text-familiarity, reading tasks and ESP test performance: A study on Iranian LEP and Non-LEP University Students*. PhD Dissertation, University of Tehran, Iran.

Salmani Nodoushan, M. A. (2008a). Performance assessment in language testing. *Journal on School Educational Technology, 3*(4), 1-7. https://doi.org/10.26634/jsch.3.4.640

Salmani Nodoushan, M. A. (2009a). Measurement theory in language testing: Past traditions and current trends. *Journal on Educational Psychology, 3*(2), 1-12. https://doi.org/10.26634/jpsy.3.2.1023

Samir, A., & Tabatabaee-Yazdi, M. (2020). Translation Quality Assessment Rubric: A Rasch Model-based Validation. *International Journal of Language Testing, 10*(2), 101-128. Retrieved from https://www.ijlt.ir/article_118019.html

Schwartz, D., & Arena, D. (2013). *Measuring What Matters Most. Choice-Based Assessments for the Digital Age.* Cambridge, Massachusetts. London, The MIT Press. https://doi.org/10.7551/mitpress/9430.001.0001

Srikaew, D., Tangdhanakanond, K., & Kanjanawasee, S. (2015). Development of an English-speaking skill assessment model for grade 6 students by using portfolio. *Procedia-Social and Behavioral Sciences, 191*, 764-768. https://doi.org/10.1016/j.sbspro.2015.04.715

Stevens, D. D., & Levi, A. (2005). Leveling the field: Using Rubrics to achieve greater equity in teaching and grading. *Essays on Teaching Excellence*.

Stoynoff, S. (2013). Looking backward and forward at classroom-based language assessment. *ELT Journal, 66*(4), 523-532. Oxford University Press. https://doi.org/10.1093/elt/ccs041

Tierney, R., & Simon, M. (2004). What's Still Wrong with Rubrics: Focusing on the Consistency of Performance Criteria across Scale Levels. *Practical Assessment, Research & Evaluation, 9*, 1-10. https://doi.org/10.7275/jtvt-wg68

Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2010). Development of a 'Universal' Rubric for Assessing Undergraduates' Scientific Reasoning Skills Using Scientific Writing. *Assessment & Evaluation in Higher Education, 36*(5), 509-547. https://doi.org/10.1080/02602930903540991

Ussama, R. (2013). *Conversation proficiency assessment: A comparative study of two-party peer interaction and interview interaction implemented with Thai EFL learners* (Doctoral dissertation, Prince of Songkla University).

Wong, G., Westhorp, G., Pawson, R., & Greenhalgh. T. (2013). Realist Synthesis: RAMESES Training Materials. Retrieved from https://www.ramesesproject.org/indexphp?pr=Project_outputs

Zaim, M. (2020). Authentic Assessment for Speaking Skills: Problem and Solution for English Secondary School Teachers in Indonesia. *International Journal of Instruction, 13*(3), 587-604. https://doi.org/10.29333/iji.2020.13340a