PJEE
**Premise : Journal of English Education and Applied Linguistics**
https://fkip.ummetro.ac.id/journal/index.php/english
*Rahayu and Rahayu*

# TEACHER'S COGNITIVE AND AFFECTIVE VERSUS TEACHERS' WRITING ASSESSMENT

Endang Mastuti Rahayu
ORCID: https://orcid.org/0000-0002-2374-6716
endangmrahayu63@gmail.com
Endah Yulia Rahayu
ORCID: https://orcid.org/0000-0002-9106-8267
Indahr_99@gmail.com
UNIVERSITAS PGRI ADI BUANA SURABAYA

*Received: January 14, 2019*
*Revised: March 31, 2019*
*Reviewed: April 1, 2019*
*Accepted: April 2, 2019*

*Reviewed: January 15, 2019*
*Reviewed: March 31, 2019*
*Revised: April 1, 2019*
*Published: April 4, 2019*

**Abstract:**

Many English teachers are not confident when they are required to examine their students essay although they have sufficient education and experience in assessing their students' written works. Therefore, they need to be trained over a short period to rate their students' writing. They are also required to improve their knowledge or cognitive in order to literate ESL writing assessment to focus their students on learning to write and to edit their writing. Their affective or attitude to the feedback of writing assessment for writing instruction, the importance of writing assessment, the competence to administer the assessment, time-consuming writing assessment, the confidence of good writing instructor affect their quality in writing assessment. Therefore, in real practice of assessing writing, experienced teachers mostly plan and do their assessment based on what they believe about the assessment. Thus, successful assessing students' writing does not only constitute the major portion of second language writing teachers' workloads but also quantifies teachers affective factors. In addition, in administering assessments, teachers as raters need to care their students' and their own affectiveness, so students will value what they learn and teachers will pay attention more to their students' learning.

***Keyword***s: *teachers' cognitive, teachers' affective, writing assessment*

## INTRODUCTION

Today, writing skill is highly needed in both academic and professional life. However, the measurement of this skill is subjective, and various factors are contributing to variability in ESL's writing scores and rating process. Writing assessment is mostly


*Copyright@2019 by PJEE*

rated by human raters affect scoring variability (Barkaoui, Do ESL essay raters' evaluation criteria change with experience? A mixed-method, cross-sectional study., 2010a; Weigle, 2009). Using Rasch analysis, from 25 trained raters scoring admission and placement tests over seven semesters, five raters show instant bias on both tests (Goodwin, 2016). This analysis suggests any standardized ESL testing programs connect with the writing rubric interpretation and consistency, besides rater training and the use of score. Rasch method can give more detailed result for assessment, rater perception, and small-scale academic testing programs.

Experienced raters quality is better than the novice although both novice and experienced raters can commit bias evaluation (Mostofee, 2016). Another study reveals some raters are more severe and give more importance to linguistic accuracy and refer to evaluation criteria other than listed in the rating scale more frequently than other raters. Since raters are not involved in developing the rating criteria and scale, they tend to shift from content to linguistic accuracy which is often a weak aspect in ESL essay. They also may not receive extensive group training on writing, and there are some problems in writing feature analyses. So for there is little information about how and why raters' evaluation criteria change over time (Barkaoui, Do ESL essay raters' evaluation criteria change with experience? A mixed-method, cross-sectional study., 2010a).

Further studies to investigate what extent, how and why the rater evaluation criteria change over time and across context, or replication with raters from a different linguistic, cultural and professional background, different writing task in different assessment system and context, needs to be conducted. Therefore, exploring raters' evaluation of L2 proficiency can produce variability of raters' judgment (Kuiken & Vedder, 2014a), and only motivated rater participants may start and complete the experiment as they produce less bias measurement (Duijm, Schoonen, & Hulstijn, 2017). Thus, various types of diagnostic assessment by employing rating scales that have been developed for L2 learners can be examined. Potential multiple interpretation and vagueness of some scale descriptors can be unveiled. Discrepancies between the descriptors included in the different scale levels and finding from the SLA literature on the natural acquisition in L2 can be investigated.


*Copyright@2019 by PJEE*

Next, scoring rubric furthermore affects raters' rating writing process (Barkaoui, Do ESL essay raters' evaluation criteria change with experience? A mixed-method, cross-sectional study., 2010a), and the complexity of the scoring rubric can be influenced by raters pre-exiting cognition which induce rating consistency (Joe, Harmes, & Hickerson, 2011). Therefore, test developers and rater training should identify the critical features of the rubric that more clearly define the construct and modify the training to address the cognitive validity by simplifying the scoring rubric. It can be exercised in rater training to increase raters; perception agreement towards all descriptors at the rubric, but this effort cannot promise their scoring consistency. Therefore another scoring aids such as the usage of exemplars and scoring rubrics should be maximized. However, before training, raters' experience and expertise in using the rubric should be investigated. After the training, the improvement of scoring practice and a better understanding of how the judgment of language proficiency can be explored. Also, textual features external to scoring rubric also influence raters' scoring decision and perception of the construct of "good writing." (Hall & Sheyholislami, 2013).Their philosophy of teaching, learning and assessment typicals (Classical Humanism, Progressivism, Reconstructionism, Post-Modernism) in their language classroom may show how they bring their interpretation to the rating task (Cheng, 2017). There might be unidentified shared raters' value, as the evidence when all raters award the same score but disagree on the quality of specific features of a text. Therefore, the efforts to improve consistency and to make a rating rubric more explicit and detailed to eliminate raters' misconception from time to time have to be conducted. The efforts include rater monitor, rater training and scoring rubric revision, difficult-to-score essay training and periodic retraining.

Rubric training can lead to raters' reliability and it should be developed and implemented locally (Lovorn & Rezaei, 2011). Therefore rater preparation and certification for high or low-stake writing test need to be initiated because newly-trained raters can exhibit similar measurement to experienced raters since they can learn to rate appropriately and quickly (Attali, 2016; Lim, 2011). In addition to that, Non-expert raters can measure L2 writing test using a scale named functional adequacy adapted by Kuiken and Vedder (Kuiken & Vedder, 2014a), and rating the four components of content, task requirements, comprehensibility, and coherence and


*Copyright@2019 by PJEE*

cohesion. As a result Royal-dawson et al. (2009) consider that teacher experience is not compulsory for rater training because more detailed scoring criteria can be easily assigned by non-teachers. However, since they have hardly any experience, they may give no impact on their students' learning. Thus, the limit of necessary- teaching-experience-qualification should be investigated. In addition to L1 and L2 writing with different features, the relationship between rating quality and the essay is unique to individual writing assessment. Therefore further investigation should focus on individual raters' judgment across subgroups as well as a revision in the assessment procedure and professional development through rater training courses in order to enhance local rating instrument in the context (Ghanbari, Barati, & Moinzadeh, 2012).

Indeed, assessing performance test presents numerous challenges due to the variability of task and rater judgment. In'nami et al. (2015) suggest task and task-related interaction influencing raters' scoring more than rater and rater-related interaction. Task also effects more on children scores' variability, not raters (Kim, Schatschneider, Wanzek, Gatlin, & S.L., 2017). Giving different tasks to different test takers do not affect the scoring by different raters if the score validity exists here. Thus, (Lim, Prompt and rater effect in second language writing performance assessment, 2009) states the scores infer the ability to be measured, and there is no influenced by irrelevant factors being measured. As a result, the scoring discrepancy can be anticipated with rater discussion or negotiation among raters. They can share their inference and rubric construction and also create affordances for them to train those inferences in sharing construction of meaning (Trace, Meier, & Janssen, 2016). However, negotiation does not influence rater severity but reduce measures of rater bias (Trace, Janssen, & Meier, Measuring the impact of rater negotiation in writing performance assessment, 2017). Raters' knowledge, personality dynamics, appreciation of student effort, comprehension of students' are revealed in raters discussion or negotiation (Kim & Lee, 2015).

Therefore, based on the related literature review that raters' knowledge of writing assessment (cognitive) and affective which include personality, attitude and appreciation towards the students, prompt, assessment and rubric, this article has two objectives, namely describing briefly the concepts and principles pertaining to (1)


*Copyright@2019 by PJEE*

Teachers' Problem in Assessing Writing, (2) Teachers' Cognitive in Assessing Writing, and (3) Teachers' Affective in Writing Assessment.

## METHODOLOGY

This article is a conceptual paper which was written based on library study. To carry out the undertaking, various journal articles which are related to the topic of discussion were synthesized. The reviewed articles included both theoretical papers on Writing Assesment research-based papers of a number of recent studies on Teachers' Problem in Assessing Writing, Teachers' Cognitive in Assessing Writing, and Teachers' Affective in Writing Assessment. The former provided solid ground for revisiting for Teachers' Cognitive aspect of Assesing Writing and Teachers' Affective aspect of Assesing Writing with its pertaining principles. Meanwhile, the latter served as the bases to demonstrate current research trends in the area and to orient toward possible future explorations.

## RESULT AND DISCUSSION

### *Teachers' Problem in Assessing Writing*

Many teachers of English, particularly lower secondary school English teachers, are not confident when they are required to examine their students essay although they have a bachelor degree in teaching English. Therefore, they need to be trained over a short period to rate their students' essays to demonstrate knowledge *of the subject, the thesis, most relevant to the topic* consistently. This expectation is not easy because of the lack of transparency in rating scale descriptors can be a factor influencing teachers' performance.

Some teachers factors that could influence raters' interpretation and application of rating scales include raters' experience for novice vs. Experienced raters. It seems to be the most frequently researched factor (Joe et al., 2011; Barkaoui, 2011). Regarding severity in rating, inexperience raters and experienced raters behaved more similarly when using analytic scales compared with when they did by


*Copyright@2019 by PJEE*

applying holistic scales. In other words, raters' experience can affect their ratings differently depending on the type of rating scales used (Joe et al., 2011; Joe et al., 2011). Thus, raters should not be nominated according to their teaching experience as it is not a significant factor (Royal-Dawson & Baird, 2009). Although research findings of rater experience can be varied, many researchers should emphasize the rater training in enhancing the quality of raters' performance because it is the most essential (Lovorn & Rezaei, 2011).

### *Teachers' Cognitive in Assessing Writing*

English language teachers today are required to improve their knowledge or cognitive in order to literate ESL writing assessment to focus their students on learning to write and to edit their writing (Hirvela, 2007). It is imperative that assessing students written works constitutes the major portion of SLA/EFL writing teachers and teachers' workloads and determine their knowledge. Some teachers who have been teaching more than ten years may complain about this (Hirvela, 2007; Ghanbari et al., 2012; Duijm et al., 2017) assert that their knowledge also influences their teaching practice and scoring quality by varying their focus on different aspects of language components and paying more attention to lexical accuracy when rating essay (Fritz & Ruegg, 2013).

AFT, NCME, and NEA or the American teachers' association and council (1990) mention seven standards for teachers' professional development in assessment, comprizing choosing appropriate assessment method for instructional design; developing appropriate assessment method for instructional design; administering, scoring and interpreting the result of teacher-made and externally-made assessment; using assessment result to make decision about individual students, planning teaching, developing curriculum, and improving school; developing valid grading procedure for students' assessment; communicating assessment result to students, parents, other stakeholders; recognizing and using ethical and legal assessment (American Federation of Teachers, National Council on Measurement in Education , National Education Association, 1990). Therefore, teachers' cognition, in term of teachers knowledge of the assessed language, discourse and sociolinguistics have to be comprehended in order to assess their students' writing thoroughly.


*Copyright@2019 by PJEE*

However, what and how to assess students' language performance, indeed also depends on lecturers' or teachers' cognitive and affective to appraise their students' writing competence well. Their value and believe in selecting writing assessment and determine the scoring accuracy of their students' scoring accuracy starts from here (Cheng, 2017; Kuiken et al., 2014a). Therefore, in this study, we find out that teachers could not score writing well although they have sufficient language knowledge (cognitive) and they have a positive attitude toward their students' writing assessment (affective). Unless they have ample training, practice and also knowledge of writing assessment design, they still have difficulty to score their students' work.

***Teachers' Affective in Writing Assessment***

Sinprajakpok (2004) reveals that EFL teachers affective or attitude to feedback of writing assessment for writing instruction, importance of writing assessment, the competence to administer the assessment, time-consuming writing assessment, confidence of good writing instructor, poor student's competence on writing exams and many more really affect their quality in writing assessment (Sinprajakpol, 2004; Borg, 2003). Since these matters deal with teachers' affective, the real practice of assessing writing, experienced teachers mostly plan and do their assessment based on what they believe about the assessment. Pajares (1992) suggests the attitude of experienced language teachers may relate to their practice than less experienced teachers (Pajares, 1992) . The experienced teachers become more embedded with their experience, and thus they might apply the principles more consistently than new teachers.

Krashen (1981) mentions affective factors comprising motivation, attitude, anxiety, and self-confidence that can influence ESL assessment by varying individual variation scoring (Krashen, 1981). Therefore, successful assessing student writing not only constitutes the major portion of second language writing teachers' workloads but also quantifies teachers' knowledge, beliefs, practices, and affective factors (Ghanbari, Barati, & Moinzadeh, 2012). As a result, in administering assessments, teachers as raters need to care their students' and their affectiveness, so students will value what they learn and teachers will pay attention more to their students' learning (Du, 2009; Crusan et al., 2016).


*Copyright@2019 by PJEE*

Crusan et al. (2016) show the relatively positive impact of affective factors to teachers' writing assessment. However, the survey items of assessment feeling by Crusan et al., only ask about the teacher's anxiety and motivation in assessment. They do not cover attitude and self-confident. In writing assessment practice, there are several ways, that teachers or raters can do. Either experienced or novice raters can mark their students' essay using a holistic and analytical rubric to qualify their student's essay and also see the effects of inter-rater agreement, and raters' severity and self-consistency across marking method (holistic vs. analytic) as cited in Barkaoui (2011). when some teachers or raters rating the same responses and facing scoring discrepancies, they also make some efforts to resolve the score disagreement, like Monte Carlo method of score resolution (Penny, 2011), rater discussion (Kim & Lee, 2015) and rater negotiation (Trace et al., 2016; 2017). Kim et al. (2015) reveal that the agreed scoring decision can be resolved the scoring discrepancy in raters' discussion and negotiation (Kim & Lee, 2015).

Teachers can comment on their students' response to show which features mostly influencing the scoring decision. How strong the comments also illustrate textual features external to the scoring rubric which have to be addressed by teachers during the scoring time (Hall & Sheyholislami, 2013). Rater's comment can be useful when there is disagreement among raters. The differences among raters to the same response can revise scoring rubric because it reveals areas outside the scoring rubric that raters attend to. Besides, raters' evaluation criteria tend to shift from a focus on content to form (linguistic accuracy) which is often a weak aspect of ESL essays, or vice versa. The experienced raters are more likely to comment on the features on student's response which are not listed on the rating scale (Barkaoui, Do ESL essay raters' evaluation criteria change with experience? A mixed-method, cross-sectional study., 2010a). Since there are several practices of writing assessment, the existing surveys still do now cover the major practices (Jianlin, 2017; Crusan et al., 2016).

Raters background and expertise contribute to rater expectations and influence scoring criteria used in rating writing assessment. The think-aloud protocol might explain individual differences in the application of the performance criteria of the essays rubrics. The results further suggest that raters engagement with the text and self-


*Copyright@2019 by PJEE*

monitoring behavior can mitigate rater severity. Wiseman (2012) and Duijm et al. (2017) assert that raters' knowledge and experience influence their rating leniency and vary the focus on different aspects of linguistic features (Wiseman, 2012; Duijm et al., 2017). For example, raters are also sensitive to lexical accuracy when rating essay and they do not always follow the Lexis scale described in the rating scale (Fritz & Ruegg, 2013). However, Lim (2011) states that experience and expertise are raters' temporal dimension because novice raters can learn to rate appropriately and quickly (Lim, 2011). Raters can maintain their rating quality over time depending on rating volume. However Lim's (2011), Wiseman's (2012), Fritz's (2013) and Duijm's (2017) study only use a small number of novice raters in one testing context, and they do not compare NEST and NNEST, so it is difficult to ascertain why might novice raters improvement their rating quality

In a standardized writing test, the score from the newly-trained raters can exhibit similar measurement to experienced raters, due to initial raters' training and screening (Attali, 2016) .In more detailed scoring criteria, rating or teaching experience may not be necessary for raters' selection criteria because it can be relatively easy to be assigned by non-teachers (Royal-Dawson & Baird, 2009; Kuiken & Vedder, Functional adequacy in L2 writing: towards a new rating scale, 2016).

On the other hand, both experienced and novice raters may not use a rating rubric consistently, but experienced raters' quality is better than the novice (Mostofee, 2016). Experienced raters still have idiosyncratic practice on the explicit rating scale (Ghanbari & Barati, 2014), due to time constraints, mandated curriculum pacing, language learning, and classroom management issues. To mitigate these barriers and to maximize the impact of professional teacher development, teachers' professional development reform should be prioritized (Buczynski & Hansen, 2010).

Specifically, the reform deals with teacher performance licensing and certification, which can reflect and predict teachers' success with their students. Thus, they can show their best teaching and assessing their students' language performance and also improve their preparation, mentoring, and professional development (Darling-Hammond, 2010). Although in certain standardized writing test, teaching experience is not necessary (Attali, 2016; Royal-Dawson & Baird, 2009; Kuiken & Vedder,


*Copyright@2019 by PJEE*

Functional adequacy in L2 writing: towards a new rating scale, 2016), the students cannot value that they learn because their teachers cannot supply with appropriate feedback (Du, 2009).

Relating to the issue of teacher's or lecturer's first language should be linked with their teaching experience and competence, Cruzan et al. (2016) native English speaking teachers (NESTs) are not usually categorized as more competent raters than non-native English speaking teachers (NNESTs) (Crusan et al., 2016). In the previous study conducted by Ling (2001), using the holistic rating rubric, NESTs respond more positively in their criteria to the content and language, whereas the Chinese teachers attended more negatively to the organization and the length of the essays (Ling, 2001). Meanwhile using the analytical rubric, both NESTs and NNESTs rate the written responses at relatively the same quality. The same idea is also suggested by Johnson (2009) that both NESTs and NNESTs can rate equally the same quality in an SLA writing assessment (Johnson, 2009).

## CONCLUSION

Based on the present progress studies in raters or teachers' factor contributing to rating writing assessment, cognitive and affective factors affect how they appraised their students' writing. Cognitive factors include their knowledge and perception about the writing assessment, conduct writing test, scoring accuracy, linguistic and sociolinguistics of how the English language is used. Meanwhile, the teachers' affective factors, in term of their attitude, efficacy, and motivation about the practice of English writing assessment contribute a lot to how they rate their students writing. How good the raters' cognitive and affective factor influence their scoring to their students' assessment also depends on the rater's background, experience, the tested language level, rubric, and prompts. Therefore raters have to equip themselves with writing assessment knowledge, linguistics of English language and positive attitude towards English writing assessment.

## BIO-PROFILE


*Copyright@2019 by PJEE*

Endang Mastuti Rahayu is a senior lecturer in the English Language Education Department of Universitas PGRI Adi Buana Surabaya, Indonesia. She also gives lectures in the Post Graduate Program of Educational Technology in the same university. Her major interests include Instructional Design, Curriculum and Materials Development, Research Methodology

**BIO-PROFILE**

Endah Yulia Rahayu is an ELT specialist who is keen on English language testing and writing assessment. She is on a study leave at Universitas Negeri Malang and is projected to finish in 2019. She works for English Department of Teacher Training Faculty Universitas PGRI Adi buana Surabaya, East Java, Indonesia. Recently she has been conducting her study intensively in rater behavior, perspective and scoring in writing performance. She is also interested in ELT management and leadership.


*Copyright@2019 by PJEE*

## REFERENCES

American Federation of Teachers, National Council on Measurement in Education , National Education Association. (1990). *Standards for teacher competence in educational assessment of students.* Dipetik 12 6, 2017, dari Buros Center for Testing: http://buros.org/standards-teacher-competence-educational-assessment-students

Attali, Y. (2016). A comparison a newly-trained and experienced raters on a standardized writing assessment. *Language Testing, 33*(1), 99-115.

Bacha, N. (2001). Writing evaluation: what can analytic versus holistic essay scoring tell us? *System, 29*(3), 371-383.

Barkaoui, K. (2010a). Do ESL essay raters' evaluation criteria change with experience? A mixed-method, cross-sectional study. *Tesol Quarterly, 44*, 310-357.

Barkaoui, K. (2011). Do ESL essay raters' evaluation criteria change with experience? A mixed-method, cross-sectional study. *Tesol Quarterly, 44*, 310-357.

Barkaoui, K. (2011). Effects of marking method and rater experience on ESL essay scores and rater performance. *Assessment in Education Principles Policy and Practice, 18*(3), 279-293.

Basturkmen, H. (2012). Review of research into the correspondence between language teachers' state beliefs and practices. *System, 40*, 282-295.

Black, P., & Wiliam, D. (2010). Inside the backbox: raising standard through classroom assessment. *Phi Delta Kappan, 92*, 81-90.

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching, 36*, 81-109.

Buczynski, S., & Hansen, B. (2010). Impact of professional development on teacher practice: Uncovering connections. *Teaching and Teacher Education, 26*(3), 559-607.

Butler, J., & Britt, M. (2011). Investigating Instruction for improving revision of argumentative essays. *Written Communication, 28*(1), 70-96.

Cheng, L. F. (2017). *Assessment in the language classroom.* London: Palgrave.

Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, belief, and practice. *Assessing Writing, 28*, 43-56.

Darling-Hammond, L. (2010). *Evaluating Teacher effectiveness: how teacher performance assessments can measure and improve teaching.* Washington DC: Center for American Progress.


*Copyright@2019 by PJEE*

Du, X. (2009). The affective filter in second language teaching. *Asian Social Science, 5*(8), 162-165.

Duijm, K., Schoonen, R., & Hulstijn, J. (2017). Professional and non-professional raters' responsiveness to fluency and accuracy in L2 speech: an experimental approach. *Language Testing*, 1-27. doi:https://doi.org/10.1177/0265532217712553

Eckes, T. (2012). Operational Rater Types in Writing Assessment: Linking Rater Cognition to Rater Behavior. *Language Assessment Quarterly, 9*, 270-292.

Fritz, E., & Ruegg, R. (2013). Rater sensitivity to lexical accuracy, sophistication and range when assessing writing. *Assessing Writing, 18*(2), 173-181.

Ghanbari, B., Barati, H., & Moinzadeh, A. (2012). Problematizing rating scales in EFL academic writing assessment: voices from iranian context. *English Language Teaching, 5*(8), 76-90.

Ghanbari, N., & Barati, H. (2014). Iranian EFL Writing Assessment: The agency of rater or rating scale? *Iranian Journal of Language Testing, 4*(2), 204-228.

Gonzales, E., Trejo, N., & Roux, R. (2017). Assessing EFL university students' writing: a study of score reliability. *Redie (Revista Electronica de Investigacion Educativa), 19*(2), 91-103.

Goodwin, S. (2016, October). A many-facet rasch analysis comparing essay rater behavior on an academic English reading/writing test used for tw purpose. *Assessing Writing, 30*(2), 21-31.

Hall, C., & Sheyholislami, J. (2013). Using appraisal theory to understand rater values: an examination of rater comments on ESL Test Essays. *Journal of Writing Assessment, 6*(1), 1-17.

Harsch, C., & Martin, G. (2012). Adapting CEF-descriptors for rating purposes: Validation by a combined rater training and scale revision approach. *Assessing Writing, 17*, 228-250.

Hirvela, A. B. (2007). Writing scholar as teacher educator: exploring writing teacher education. *Journal of Second Language Writing, 16*(3), 125-128.

In'nami, Y., & Koizumi, R. (2015). Task and rater effects in L2 speaking and writing: A synthesis of generalizability studies. *Language testing, 33*(3), 341-366.

Jeong, H. (2017). Narrative and expository genre effects on students, raters, and performance criteria. *Assessing writing, 31*, 113-125.

Jianlin, C. (2017). Factors affecting Chinese raters' rating of high-stakes English exam Essays. *Chinese Journal of Applied Linguistics, 40*(2). doi:https://doi.org/10.1515/cjal-2017-0013


*Copyright@2019 by PJEE*

Joe, J., Harmes, J., & Hickerson, C. (2011). Using verbal report to explore rater perceptual processes in scoring: a mixed method application to oral communication assessent. *Assessment in Education: Principles, Policy & Practice, 18*, 239-259.

Johnson, D., & Brackle, L. (2012). Linguistic discrimination in writing assessment: How rater react to Arican American "errors", ESL errors, and standard English Errors on a state-mandated writing exam. *Assessing Writing, 17*(1), 35-54.

Johnson, K. (1999). *Understanding language teaching: reasoning in action.* Boston: Heinle & Heinle.

Johnson, K. (2009). *Second language teacher education: a sociocultural perspective.* New York: Routledge.

Kim, S., & Lee, H. (2015). Exploring rater behaviors during a writing assessment discussion. *English Teaching, 70*(1), 97-121.

Kim, Y., Schatschneider, C., Wanzek, J., Gatlin, B., & S.L., O. (2017). Writing evaluation: rater and task effects on the reliability of writing scores for children in Grades 3 and 4. *Reading and Writing, 30*(6), 1287-1310.

Krashen, S. D. (1981). *Second Language Acquistion and Second Language Learning.* Pergamon Press Inc.

Kuiken, F., & Vedder, I. (2014a). Raters' decisions, rating procedures and rating scales. *Language Testing, 31*(3), 279-284.

Kuiken, F., & Vedder, I. (2016). Functional adequacy in L2 writing: towards a new rating scale. *Language Testing, 34*(3), 321-336.

Lim, G. (2009). Prompt and rater effect in second language writing performance assessment. *Dissertation of Phd*. Michigan: University of Michigan.

Lim, G. (2011). The development and maintenance of rating quality in performance writing assessment: A longitudinal study of new and experienced raters. *Language Testing, 28*(4), 543-560.

Ling, S. (2001). Native- and nonnative-speaking EFL teachers' evaluation of Chinese students' English writing. *Language Testing, 18*(3), 303-325.

Lovorn, M. G., & Rezaei, A. R. (2011). Assessing the assessment: Rubrics training for pre-service and new in-service teachers. *Practical Assessment, Reseach & Evaluation*, 1-18.

Mitchell, E. (2005). *The influace of belief on the teaching practice of high school foreign language teachers.* Amherst: University of Massachusetts.

Moss, P. (1994). Can there be validity without reliability? *Educational Researcher, 23*, 5-12.


*Copyright@2019 by PJEE*

Mostofee, S. G. (2016). Examining five behavior conducted by two groups of novice and experienced raters in two rating processes. *International Journal of Applied Linguistics & English Literature, 5*(4), 199-211.

NG, T., & Felman, D. (2009). How broadly does education contribute to job performance? *Personnel psychology, 62*, 89-134.

Pajares, M. (1992). Teachers' belief and educational research:cleaning up a messy construct. *Review of Education Research, 62*(4), 307-331.

Penny, J. J. (2011). The accuracy of performance task scores after resolution of rater disagreement: A Monte Cartlo study. *Assessing Writing, 16*(4), 221-236.

Royal-Dawson, L., & Baird, J. (2009). Is teaching experience necessary for reliable scoring of extended English question? *Educational Measurement: Issue and Practice, 28*, 2-8.

Sinprajakpol, S. (2004). *Teachers' believe about language learning and teaching: the relationship between believes and practice.* Buffalo: University of New York.

Taylor, L. (2010). Developing assessment literacy. *Annual Review of Applied Linguistics, 29*, 21-36.

Trace, J., Janssen, G., & Meier, V. (2017). Measuring the impact of rater negotiation in writing performance assessment. *Language Testing, 34*, 3-22.

Trace, J., Meier, V., & Janssen, G. (2016, October). "I can see that": developing shared rubric category interpretation through score negotiation. *Assessing Writing, 30*(3), 32-43.

Ucelli, P., Dobbs, C., & Scotts, J. (2013). Mastering Academic Language: Organization and stance in the persuasive writing of high school students. *Written Communication, 30*(1), 36-62.

Weigle, S. C. (2009). *Assesing Writing.* (C. J. Alderson, & B. Lyle F, Penyunt.) UK: Cambridge University Press.

White, E. (2009). Are you assessment literate? Some fundamental questions regarding effective classroom-based assessment. *OnCUE Journal, 3*(1), 3-25.

Winke, P., & Lim, H. (2015). ESL essay raters' cognitive processes in applying the Jacobs et al. rubric: An eye-movement study. *Writing Assessment, 25*, 38-54.

Wiseman, C. (2012). Rater effects: Ego engagement in rater decision-making. *Assessing Writing, 17*(3), 150-173.

Zhang, J. (2016). Same test different processing? Exploring how raters' cognitive and meta-cognitive strategies influence rating accuracy in essay scoring. *Assessing Writing, 27*, 37-53.


*Copyright@2019 by PJEE*