JEASP
Journal of English for Academic and Specific Purposes
Volume 3 Number 1, June, 2020

THE ANONYMOUS TEACHERS' FACTORS OF ASSESSING PARAGRAPH WRITING

Endah Yulia Rahayu
(Indahr_99@yahoo.com)

Universitas PGRI Adi Buana Surabaya

ARTICLE

***Keywords*:**

Basic knowledge of writing assessment, efficacy in selecting assessment method, efficacy of scoring accuracy, perception in practicing writing assessment, assessing writing

ABSTRACT

Scoring writing is very subjective and mainly relies on a lot on teachers as raters. They play a significant role to meticulously carry out writing evaluations to adjudicate the linguistic and rhetorical features of their students' written responses. Based on the previous studies, the teachers' factors of knowledge of basic writing assessment, efficacy in selecting assessment method, efficacy in scoring accuracy, and perception in practicing writing assessment can contribute a lot to the quality in teachers' writing assessment. The 56 junior high school English teachers having at least five years of teaching experience, was invited to fill out the questionnaire and scoring paragraph writing. The results were examined with Multiple Linear Regression analysis. Amongst these factors, only the efficacy in scoring writing accuracy predicts the teachers' scoring paragraph writing.

INTRODUCTION

Since teachers play their vital role to judge the linguistic and rhetorical features of their students responds, they have to be literate in administering assessment for learning (AfL), assessment as learning (AsL) and assessment of learning (AoL) integrated well to assign a value to represent the quality of the students' learning outcomes (Mao & Jiang, 2018). However, due to the lack of their English language assessment discernment and training, they still have a problem to collect and interpret the students' learning result in the form of summative and formative assessment. Accordingly, they frequently use any assessments without evaluating or revising them and hardly use statistic procedures to see how the assessment is performing (Bandura, 2006).

The raters' factors in writing assessment are the idiosyncrasy that exists in the rater's cognitive, efficacy, and perception, which is influenced by their robust and appropriate education, experience, and training. Some elementary school teachers and undergraduate students produce variability scores and rating procedures and results, although they followed thorough assessment training before the research (Pas & Bradshaw, 2014). Besides their knowledge in writing

assessment, their efficacy in scoring accuracy, their efficacy in selecting assessment methods, and their perception in practicing writing assessment can determine how good and useful they rate writing assessment (Meissel, Meyer, Yao, & Rubie-Davies, 2017). Thus, individual raters may more reliable than others, even though they receive the same qualified training. By applying the same analytical scoring rubric to assess writing samples, scoring variability happens due to raters' different levels of accuracy, judgment, and interpretation of students' work and the rubric in use (Gonzalez, Trejo, & Roux, 2017). The same scoring rubric is not enough to improve rater reliability. Therefore, specific professional training is needed to make the same point of view towards specific students' work and rubric usage (Bandura, 2006).

Conversely, conducting specialized assessment training, and applying the same rubric and assessment method still varies the raters' scoring across different examinee groups in the placement test and admission test. Therefore, the borderline scoring decision has to be stated by the ESL instructors and the testing director. Teacher-training programs still make teachers confuse to apply the training material and create the writing rubric for their classroom context. Many teacher training does not equip the teachers with the necessary writing assessment components. Meanwhile, by exploring some raters discussion to decide the score of writing performance, these raters have a different degree of personality dynamics to assign an operational score, to appreciate the student effort, and to comprehend the students' intended meaning (Borowski et al., 2011). Rater's dominance can happen during the discussion, so it needs to employ a third rater to handle score discrepant ratings, but more raters will extend a more extended discussion.

Indeed, the each teachers' factor of the basic knowledge of writing assessment, the efficacy of selecting writing assessment method and scoring accuracy, and perception in conducting writing assessment (Pas & Bradshaw, 2014; Soltero-González, Escamilla, & Hopewell, 2012; Borowski et al., 2011; Ghanbari & Barati, 2014; Friedman et al., 2011; Wiseman, 2012; Goodwin, 2016) influence the teachers' quality in scoring writing. However, each of these factors has never been executed and tested simultaneously to measure some teachers' aspects when they assess their students' writing. The researches mentioned above also hardly invited junior high school teachers as their respondents. Thus, the outlook of these teachers' factors impact in junior high school teachers when scoring their students' writing has not been scrutinized. Therefore, in this study, I use all the four factors simultaneously in the form of a questionnaire that was filled out by the 56 junior high school teachers. They also scored two writing pieces. The result of their questionnaire

filling and scoring two paragraph writing pieces was analyzed using Multiple Linear Regression. It found out that only the efficacy of scoring writing positively predicted the qualtiy of the junior high school teachers' scoring quality.

LITERATURE REVIEW

Since teachers' knowledge of basic ESL/ELT writing assessment constitutes a significant portion of SLA/EFL teachers' workloads, many English teachers who have been teaching for more than ten years still complain about this (Ghahari & Sedaghat, 2018). The basic knowledge of writing assessment influences their teaching practice and scoring quality by varying their focus on different aspects of language components and paying more attention to lexical accuracy in order to be pronounced professional teachers (Ratnawati, Faridah, Anam, & Retnaningdyah, 2018; Friedman, Farber, & Taylor, 2011). Thus, raters' personality dynamics, appreciation of student effort, comprehension of students' intended meaning are also the prominent factors that influence the process of scoring decisions (Baker, 2016). Finally, teachers as raters have to make sound judgment to their students' learning result by measuring their learning mastery in a valid, objective, fair, integrated, open, systematic, criterion-based, and reliable grading procedure and assessment method. They also have to be able to disseminate the students' learning results to them, schools, parents, and stakeholders (KEMENDIKBUD, 2016).

A high level of raters' self-efficacy is associated with raters' commitment and optimism to scoring accuracy and selecting a writing assessment method that can motivate the students' learning (Chesnut & Burley, 2015). However, overconfidence in scoring accuracy will lead to overestimate the students' writing scores and cause raters to create bias in judging the students' written works (Rahayu & Rahayu, 2019). Therefore, their consistency and judgment influence raters' efficacy in L2 writing evaluation (Mao & Jiang, 2018). Their efficacy determines the variation of their accuracy in rating specific texts based on the assessment method they select. As a result, when evaluating students' responses, they also view the errors in the responses differently, and they may not be fair in judging their students' works. Teachers with different levels of accuracy show different patterns of cognitive and meta-cognitive behaviors (Roscoe, Allen, Johnson, & McNamara, 2018). Therefore, how much teachers' efficacy in scoring accuracy predicting raters' rating writing and how their basic knowledge in writing assessment are explored.

Teachers, as raters, socially practice their assessment methods in their writing classroom to learn and understand the complexity of their classroom teaching and assessment. This effort should

progress teachers' assessment literacy, cognitive, efficacy, and perception, which are intertwined to create an opportunity to reflect what they have done during teaching and assessing activities in achieving the required standards of 21st-century education (Jiang, Sporte, & Luppescu, 2015). They can raise the existing standard only if they have confidence that what they do positively affect their classrooms. For example, for formative assessment, they give a multiple choice for indirect writing assessment, give paragraph for direct writing assessment, integrate writing assessment with other skills for more practical learning, conduct effectiveness teacher-made assessment, and administer alternative assessment like a portfolio, writing self-assessment, writing peer-assessment (Stojiljković, Todorović, Đigić, & Dosković, 2014). The teachers' efficacy in selecting a writing assessment method can influence their quality of writing assessment and improve their student writing competence.

The study of raters' ego engagement in rater judgments in the scoring decision, suggests that raters are influenced by their L2 writing ability since their an understanding with the text and writers mitigate their severity or leniency, and result in positive washback in L2 writing classroom (Wiseman, 2012). The raters' ego is the raters' selves center that manifests in some ways in their personality, which determines their perception or their point of view in judging score. Thus, their engagement with the text has to support the development of L2 writing skills in the classroom. For example, either experienced or novice raters can mark writing responses using a holistic and analytical rubric to qualify their student's writing (Goodwin, 2016). Teachers can comment on their students' responses to show which features mostly influencing the scoring decision (White & Hall, 2014). Rater's comment can be useful when there is disagreement among raters. The differences among raters to the same response can revise the scoring rubric because it reveals areas outside the scoring rubric that raters attend.

In writing assessment practice, there are several ways that teachers or raters can do and undoubtedly affected by their perception. A mixed-method approach provides a way to examine the relationship between the quantitative ratings by raters and their perception and judgemental process in rating students' essays (Wang, Engelhard, Raczynski, Song, & Wolfe, 2017). The quantitative analyses result in a variation of raters scoring essay, and qualitative analysis suggests that raters have an inconsistent perception in textual borrowing, development of the idea, consistency of focus. Periodic retraining can solve this misperception to form a consistent framework for all raters in the scoring decision. Besides that, the flexibility of writing assessment

practices that can support learning opportunities depends on the teachers' perception of understanding the nature of assessment literacy. Their perception of writing assessment still focuses on learning outcomes, not the learning process (Djoub, 2017). Through appropriate training, they can be aware of assessment functions such as a checklist, questionnaire, etc. because they gain more insights into assessment practices by linking assessment theoretical concepts with experience.

METHOD

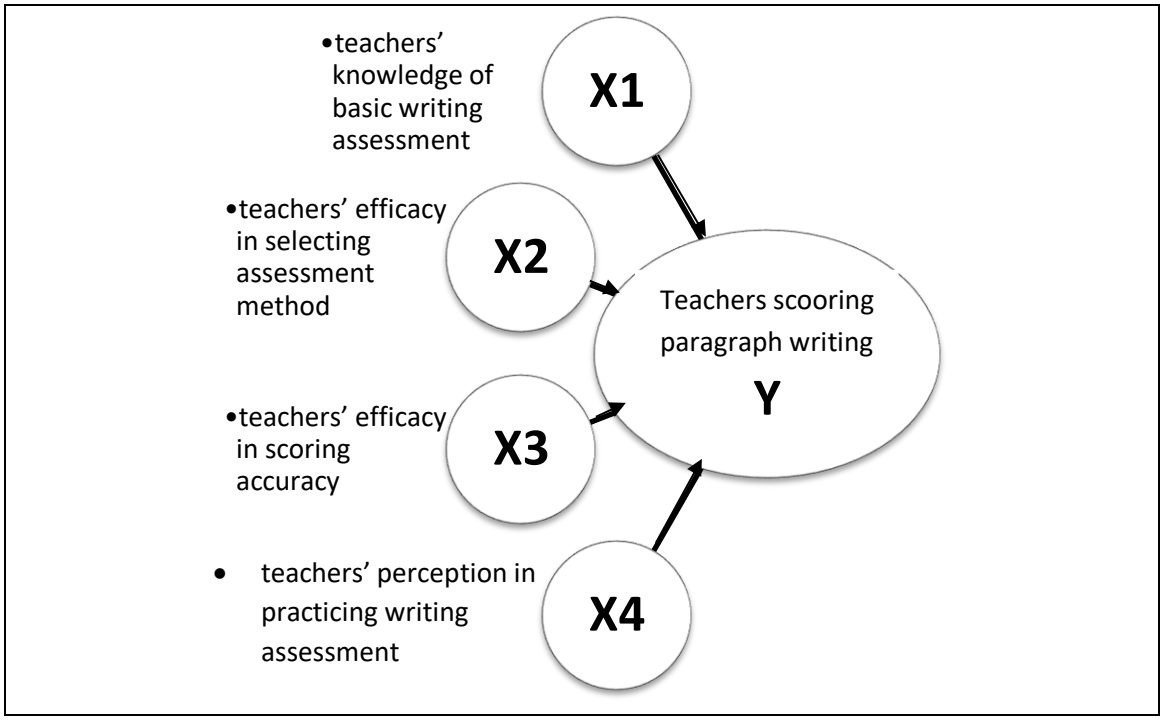
This study was conducted in 2019 by inviting 56 junior high school teachers in Surabaya Indonesia in order to seek how their knowledge of basic writing assessment, their efficacy in scoring accuracy, their efficacy in selecting writing assessment method, and also their perception in practicing writing assessment could have predicted their rating writing. These teachers had at least five years of ELT/ESL teaching experience and also possess teaching license certification. They filled out the questionnaire and score two narrative paragraphs. The questionnaire consisted of four sections – knowledge of writing assessment, the efficacy of scoring accuracy, efficacy in selecting writing assessment, and perception in practicing writing assessment. It uses four Linkert scale – really disagree, disagree, agree, and really agree. They scored the two paragraphs using a narrative paragraph scoring rubric (ReadWriteThink, 2004).

The narrative mode was selected to represent the writing modes because, in Indonesian junior high schools, it is one of the popular writing modes of paragraphs as long functional texts. Teachers in Indonesian junior high school do not teach argumentative writing (KEMENDIKBUD, 2017). The two narrative paragraphs were about the students' impressive experience with good and bad quality. By scoring the narrative paragraphs using the rubric, the teachers or raters' scoring writing quality was measured. These teachers had to produce a score by using an analytical rubric. The analytical rubric for a paragraph is chosen as it has many individual traits or components of written expression with fixed value or score for each component. This fixed score can improve reliability among the raters in measuring the paragraph. Thus, analytic scoring allows the raters to focus on various aspects of personal writing and score some traits higher than others (Brown, 2004).

Next, the average result of the teachers rating narrative paragraphs was regressed with the result of their filling questionnaire containing the four raters' factors - knowledge of writing assessment, the efficacy of scoring accuracy, efficacy in selecting writing assessment, and

perception in practicing writing assessment by applying Multiple Linear Regression (MLR). MLT describes how the factor predictors (teachers' knowledge of basic writing assessment, teachers' efficacy in selecting assessment method, teachers' efficacy in scoring accuracy, and teachers' perception in practicing writing assessment) are numerically related or unrelated to the quality of teachers' scoring writing respectively like the below model of teachers' factors in assessing paragraph writing.

Picture 1. Model of Teachers' Factors in Assessing Paragraph Writing



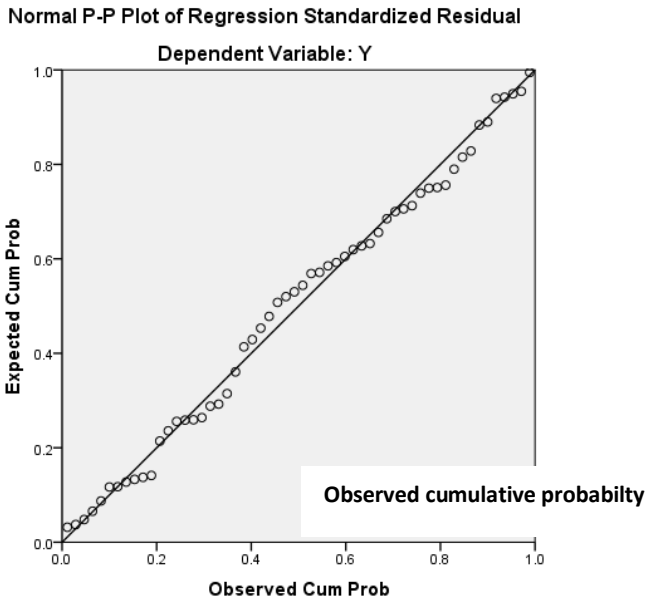
In this study, the observed value (X) is the four raters' factors consisting of teachers' knowledge of basic writing assessment (X1), teachers' efficacy in selecting assessment method (X2), teachers' efficacy in scoring accuracy (X3), and teachers' perception in practicing writing assessment (X4). Meanwhile, the predicted value is teachers scoring paragraph writing (Y). X1, X2, X3, and X4 are regressed with Y. Before the MLR, the classical assumption test was conducted in the form of normality test, Heteroscedasticity test, Multicolinierity Test, and Autocorrelation Test in order to make sure the availability of the lost data during the research (Uyanık & Güler, 2013).

FINDINGS

Normality Test

Before the MLR, the questionnaire data of the 56 junior high school teachers (observed values) and data of 56 junior high school teachers scoring two writing pieces (predicted value) were tested using classical assumption test to check the normality distribution of the data scattered plot. The dependent variable or observe value consists of teachers' knowledge of basic writing assessment (X1), teachers' efficacy in selecting assessment method (X2), teachers' efficacy in scoring accuracy (X3), and teachers' perception in practicing writing assessment (X4) while the predicted value was 56 teachers scoring two writing pieces (Y). The regression data should follow a normal distribution to validate the inferences of the MLR regression. There should be no error data or residuals in terms of no data differences between the observed value of the dependent variables and the predicted value. Picture 1 below is the normal Predicted Probability (P-P) plot, which conforms to the diagonal normality line indicated in the plot.

Picture 2. Normal P-P Plot Diagram



This normal predicted probability (P-P) plot has a diagonal line and a bunch of little circles near the normality line, which is ideal. In the studied data, there was no difference between the dependent variables or observed cumulative probability in terms of the four teachers' factors and

the predicted value or the expected cumulative probability or teachers scoring two writing pieces. The little circles scattered around the normality line, which means normal residual data.

The goodness or fitness of observed values and predicted value is also examined with the Kolmogorov-Smirnov normality test. It tests the normality of the distribution sample values that are standardized and compared with a standard normal distribution. It is equivalent to setting the mean and variance of the reference distribution equal to the sample estimates. In this research, the result of goodness or fit observed and predicted value is the following table of the normality test of Kolmogorov-Smirnov.

Tabel 1. Kolmogorov-Smirnov Normality Test

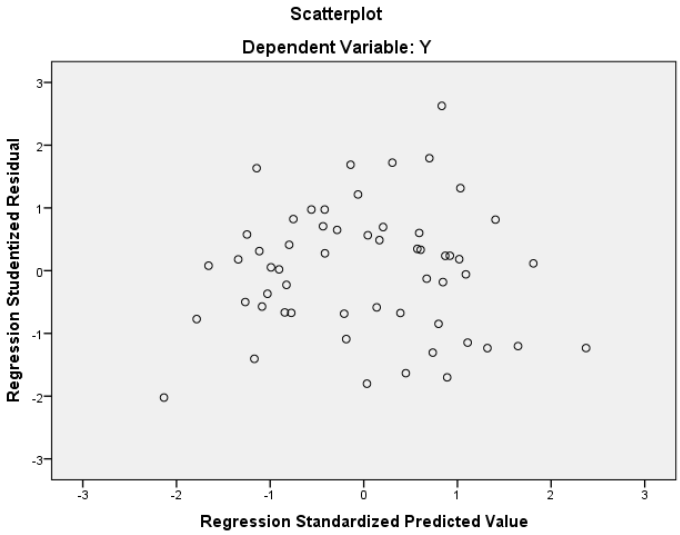
		Unstandardized Residual
N		56
Normal Parameters	Mean	.0000000
	Std. Deviation	2.29229967
Most Extreme Differences	Absolute	.064
	Positive	.064
	Negative	-.062
Kolmogorov-Smirnov Z		.479
Asymp. Sig. (2-tailed)		.976

The goodness of the data of observed and predicted values examined by the Kolmogorov-Smirnov test is 0,976 ($p > 0,05$). It means that the variables or values follow the normal distribution, before Multiple Linear Regression.

Heteroscedasticity Test

In this study, the variability of random variables of predicted and observed is tested with the Heteroscedasticity test. The existence of heteroscedasticity is a significant concern in regression analysis, in the context of the residual or error term. Mainly, heteroscedasticity is a systematic change in the spread of the residuals over the range of measured values. To satisfy the regression assumptions and be able to trust the results, the residuals should have a constant variance, just like below scatterplot of heteroscedasticity test. The vertical range of the residuals increases as the fitted values increases. Below is the heteroscedasticity test of the observed value of teachers' factors and predicted values of teachers scoring writing.

Picture 3. Heteroscedasticity Scattered Plot Diagram



The plots of the observed and predicted value in the Heteroscedasticity test are scattered everywhere, which do not form any shape. Thus, the data of this study fit the Heteroscedasticity assumption.

Multicolinierity Test

In this research, to see whether there are any multiple ties, simple correlations, a Variance Inflation Factor (VIF) was examined to see the data of observed and predicted value satisfying the assumption of linear regression analysis. The VIFs of the linear regression indicates that the variance degree in the regression estimates increasing due to multicollinearity. VIF values higher

than 10 indicate that multicollinearity is a problem. Below is the result of the Multicolenierity test of the observed value of teachers factors (teachers' knowledge of basic writing assessment (X1), teachers' efficacy in selecting assessment method (X2), teachers' efficacy in scoring accuracy (X3), and teachers' perception in practicing writing assessment (X4) and the predicted value is 56 teachers scoring two writing pieces (Y)

Table 2. Multicollinearity Test

Model		Collinearity Statistics	
		Tolerance	VIF
1	X1	.996	1.004
	X2	.347	2.885
	X3	.413	2.420
	X4	.584	1.712

The Multicollenierity test by applying VIF to the observed value of this study was 1.004, 2.885, 2.420, and 1.712, respectively, which all were below ten. It means there is no problem in the data of this study towards multiple ties or multicollinearity.

Autocorrelation Test

In this research, the Durbin Watson test was conducted to test the autocorrelation in the residual from a statistical regression analysis. This test has a value ranging from 0 to 4, which means a value of 2.0 - there is no autocorrelation detected in the respondent data. The values from 0 to less than 2 indicate positive autocorrelation, and values from 2 to 4 indicate negative autocorrelation. Below is the result of the Durbin Watson Autocorrelation Test of this study's data.

Table 3. Autocorrelation Test

Durbin-Watson	dU	4-dU
1,903	1,841	2,159

The result of autocorrelation assumption with the Durbin Watson test of this studied data ranged from dU 1,841 tp 4-dU 2,159, meaning that the data of this study was positive autocorrelation (common in time series data) or fulfilling the assumption of autocorrelation.

Multiple Linear Regression

After fulfilling all the classical assumption tests, the data of observed and predicted value was examined with the Multiple Linear Regression (MLR) analysis. The observed value was the four teachers' factors (teachers' knowledge of basic writing assessment (X1), teachers' efficacy in selecting assessment method (X2), teachers' efficacy in scoring accuracy (X3), and teachers' perception in practicing writing assessment (X4), and the predicted value was 56 teachers scoring two writing pieces (Y). Below is the result of the MLR analysis of X1, X2, X3, X4 towards Y.

Tabel 4. Multiple Linear Regression

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	20.887	3.361		6.215	.000
	X1	-.700	.287	-.321	-2.435	.018
	X2	-.039	.087	-.100	-.449	.655
	X3	.061	.061	.206	1.009	.318
	X4	-.026	.061	-.072	-.418	.677

Based on the above regression analysis results, the regression equation of the observed value - X1, X2, X3, X4, and predicted value Y is as followed.

$$y = a + b1\ x1 + b2\ x2 + b3\ x3 + b4\ x4 + e$$

$$y = 20{,}887 - 0{,}700\ x1 - 0{,}039\ x2 + 0{,}061\ x3 - 0{,}026\ x4 + e$$

This regression equation explains as followed:

a. The constant value (a) of 20,887 indicated that X1, X2, X3, and X4 do not influence Y. The teachers' factors of teachers' knowledge of basic writing assessment (X1), teachers' efficacy in selecting assessment method (X2), teachers' efficacy in scoring accuracy (X3), and teachers' perception in practicing writing assessment (X4), indeed do not predict how well the teachers are scoring writing.

b. The coefficient value of X1 reached -0,700. It means the increasing observed value of teachers' knowledge of basic writing assessment would negatively influence the predicted value of teachers scoring writing paragraphs. The higher the observed X1 value, the lower the predicted Y value.

c. The coefficient value of X2 amounted to -0,039. It shows the increasing observed valued of teachers' efficacy in selecting the assessment method would negatively forecast the predicted value of teachers scoring writing paragraphs. The higher the observed X2 value, the lower the predicted value.

d. The coefficient value of X3 is 0,061, which indicates that the increased observed value of teachers' efficacy in scoring accuracy would positively forecast the predicted value of teachers' efficacy in scoring writing paragraphs. The higher the observed X3 value, the bigger the predicted value.

e. The coefficient value of X4 reached -0,024, which implies that the increasing value of the teachers' perception in practicing writing assessment would negatively forecast the predicted value of the teachers scoring writing paragraph. Therefore the higher the observed X4 value, the lower the predicted value.

DISCUSSION

Since teachers are considered an essential factor in writing assessment due to its subjectivity nature, there have been many types of research to explore the distinctive features of teachers. The teachers or raters factors deal with extensive coverage such as teacher training quality (Pas & Bradshaw, 2014), teacher' perception and subjectivity (Meissel et al., 2017), raters' judgment (Gonzalez et al., 2017), teachers' perception in conducting writing assessment (Soltero-González et al., 2012), teachers' efficacy and autonomy in conducting writing assessment (Statistik, 2014)(Skaalvik & Skaalvik, 2014) and many more. Amongst the above factors, I sum up that there are three aspects of teachers' factors ranging from teachers' cognitive, teachers' efficacy, and teachers' perceptions. Since the focus of this study is assessing the writing paragraph, the teachers' cognitive relates to the teachers' knowledge of basic writing assessment, and the teachers' efficacy deals with selecting assessment methods and scoring accuracy. Finally, the teachers' perception relates to their perception of administering writing assessment. These four factors were then accumulated in a four-section questionnaire having 99 question items with four Linkert scales. Since the purpose of this research is aimed at examing and predicting the allegedly most influential teachers factors comprizing of teachers' knowledge of basic writing assessment, efficacy in selecting assessment method, efficacy in scoring accuracy, and perception in practicing writing assessment which positively contribute to the quality in teachers' writing assessment. To seek now well the teachers assess writing, the 56 junior high school teachers were ask to scoore two narrative paragraphs using narrative scoring rubric. Next the data of these teachers filling questionnaire and scooring the paragraphs were analyze using MLR.

The classical assumption test is to make sure that the data of this study are normal and fit to the MLR by applying a normality test, heteroscedasticity test, multicollinearity test, and autocorrelation test. The data normality of the observed value of the four teachers' factors and the predicted value or the teachers scoring writing paragraph fulfills the normal distribution, see picture 2 Normal P-P Plot Diagram. Both data on teachers' factors and teachers' scoring writing paragraphs have no difference distribution. Another normality test applied in this study use is Kolmogorov-Smirnov normality. The data fit of this test is 0,976 ($p > 0,05$), meaning the data values follow the normal distribution. The variability of random scattered plot variables of predicted and observed value is tested with the Heteroscedasticity test. The plots of the observed and predicted value in the Heteroscedasticity test are scattered everywhere, which do not form any shape. Thus, the data of this study fit the Heteroscedasticity assumption. Next, to check whether

there are any multiple ties, simple correlations, a Variance Inflation Factor (VIF) was applied to see the data of observed and predicted value satisfying the assumption of linear regression analysis. The Multicollenierity test by applying VIF to the observed value of this study was 1.004, 2.885, 2.420, and 1.712, respectively, which all were below ten. It means there is no problem in the data of this study towards multiple ties or multicollinearity. Finally, to check the autocorrelation in the residual of the data, the Durbin Watson test was conducted. The result of autocorrelation assumption with the Durbin Watson test of this studied data ranged from dU 1,841 tp 4-dU 2,159, meaning that the data was positive autocorrelation (common in time series data) or fulfilling the assumption of autocorrelation.

Next, the MLR analysis found out only teachers' efficacy in scoring accuracy positively the teachers' quality in assessing paragraph writing. The teachers' efficacy is the belief of their professional competence in appraising the quality of their student's writing, and effort is prominent factors that influence the process of scoring decisions. Particularly in ELT context in Indonesia, teachers have to possess a deep comprehension of their students' intended meaning is frequently and far different from the standard English. Otherwise, teachers as raters cannot make sound judgment to their students' learning results. In this study, the 56 respondents scored the paragraph writing pieces using an analytical rubric, which enable them to see the strength and weaknesses of the written works. This rubric can be used easily by both experienced and novice teaches to mark their students' writing and also see the effects of inter-rater agreement, and raters' severity and self-consistency across marking method (Barkaoui, 2011).

However, when teachers face scoring discrepancies, they can make some efforts to resolve the score disagreement by conducting rater discussion (Yuan & Kim, 2018) and rater negotiation (Trace, Meier, & Janssen, 2016). The discrepancy is solved in raters' discussion and negotiation. The differences among raters to the same response can revise the scoring rubric because it reveals areas outside the scoring rubric that raters attend. Also, raters' evaluation criteria tend to shift from a focus on content to form (linguistic accuracy), which is often a weak aspect of ESL paragraphs, 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 (White & Hall, 2014). Teachers should give appropriate comment on their students' response to show which features mostly influencing the scoring decision. The suitable comments need to illustrate textual features of the scoring rubric during the scoring time.

Raters can maintain its rating quality over time, depending on rating volume. Teachers' professional development reform should be prioritized to mitigate these barriers (Anugerahwati & Saukah, 2010). Accordingly, the score from the newly-trained raters can exhibit a similar measurement to experienced raters, due to initial raters' training and screening. In more detailed scoring criteria, 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). Both experienced, and novice raters may not use a rating rubric consistently, but experienced raters' quality is better than the novice when they are required to score their student's writing accurately (Osborne & Walker, 2014).

The other factors of teachers' knowledge of basic writing assessment, teachers' efficacy of selecting assessment method, and teachers' perception in writing assessment did not positively predict the teachers' quality in assessing their student's writing. It may their knowledge of basic writing assessment of the respondents is varied, so in selecting the writing assessment method and administering the writing assessment is hardly conducted. How the 56 respondents conduct and select their writing assessment for their classroom is not recorded in this study. Their efficacy of scoring accuracy determines the variation of their accuracy in rating the narrative paragraph. As a result, when evaluating students' responses, they also view the errors in the responses differently, and they may not be fair in judging their students' works. Teachers with different levels of accuracy show different patterns of cognitive and meta-cognitive behaviors (Wilson & Roscoe, 2019). Therefore, how much teachers' efficacy in scoring accuracy predicting raters' assessing paragraph writing.

CONCLUSION AND SUGGESTION

Teachers' factors in assessing writing have been widely discussed because of its distinctive subjective nature, such as teachers' professional development(Pas & Bradshaw, 2014), perception, judgment, (Soltero-González et al., 2012)autonomy, and many more in conducting writing assessment. These factors were synthesized into three aspects of teachers' factors ranging from teachers' cognitive, teachers' efficacy, and teachers' perceptions, in terms of teachers' knowledge of basic writing assessment, and the teachers' efficacy in selecting assessment methods, and scoring accuracy, finally, teachers' perception of administering writing assessment. After undergoing the classical assumption test, these factors were regressed with the teachers scoring writing using Multiple Linear Regression analysis, in order to seek which factor that might

contribute the quality of teachers' writing assessment, only teachers' efficacy in scoring accuracy predict the teachers' quality in writing assessment. The teachers' knowledge of basic writing assessment, efficacy of selecting assessment method, and perception in writing assessment contra predict the teachers' quality in assessing their student's writing. This finding contradicts the previous research because all of the factors positively support the quality of teachers assessment.

This study only invited junior high school teachers from some private school in a big city in Indonesia and only asking teachers to score one writing mode, so the result cannot be generalized. Further research needs to invite most teachers of a specific area with different school and writing types in order to obtain a more explicit model of teachers' factors that predict their writing assessment quality. It is undeniable because by having a clear picture of teachers' factors modeling in supporting writing assessment, the ELT writing assessment context, such as in Indonesia, teachers can have a comprehensive understanding of their students' intended meaning and know well how to fix, improve and assess it. Without acknowledging the teachers' factor, teachers as raters cannot make sound judgment to their students' learning results.

By knowing the weakness and strength of their factors, which include their cognitive, affective, perception and their attitude in writing assessment, teachers or raters can resolve their score disagreement by conducting rater negotiation and agreement. The experienced raters can share their writing assessment experience to the novice teachers. Experience teachers are usually like to comment on the features on student's responses which are not listed on the rating scale. Teachers can maintain its rating quality over time, depending on rating volume and demand. Teachers' professional development reform should be prioritized the improvement of teachers' factors because newly-trained raters can exhibit a similar measurement to experienced raters, due to initial raters' training and screening. However, both experienced and novice raters may not use a rating rubric consistently, but experienced raters' quality is better than the novice when they are required to score their student's writing accurately.

REFERENCES

Anugerahwati, M., & Saukah, A. (2010). Professional competence of English teachers in Indonesia : a profile of exemplary teachers. *Indonesian Journal of English Language Teaching*, *6*(2), 107–119.

Baker, K. M. (2016). Peer review as a strategy for improving students' writing process. *Active Learning in Higher Education*, *17*(3), 179–192. https://doi.org/10.1177/1469787416654794

Bandura, A. (2006). Guide for Constructing Self-Efficacy Scales. In *Self-Efficacy Beliefs of Adolescents*, (pp. 307–337). Information AgePublishing. https://doi.org/10. 1080/09243453. 2012. 680892

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. https://doi.org/10.1080/0969594X.2010.526585

Borowski, A., Carlson, J., Fischer, H. E., Henze, I., Gess-Newsome, J., Kirschner, S., & Van-Dal, J. (2011). Different Models and Methods To Measure Teachers'Pedagogical Content Knowledge. In Bruguière, Catherine, Tiberghien, Andrée, Clément, & Pierre (Eds.), *ESERA 2011 Conference: Science learning and Citizenship* (pp. 1–12). European Science Education Research Association. Retrieved from http://vbn.aau.dk/files/78578296/ebook_esera2011_Strand13.pdf#page=29

Brown, D. (2004). *Teaching by principles.*

Chesnut, S. R., & Burley, H. (2015). Self-efficacy as a predictor of commitment to the teaching profession: A meta-analysis. *Educational Research Review*, *15*, 1–16. https://doi.org/10.1016/j.edurev.2015.02.001

Djoub, Z. (2017). Revisiting EFL assessment. https://doi.org/10.1007/978-3-319-32601-6

Friedman, I. A., Farber, B. A., & Taylor, P. (2011). Professional predictor of teacher burnout as a. *most*, *86*(1), 28–35.

Ghahari, S., & Sedaghat, M. (2018). Optimal feedback structure and interactional pattern in formative peer practices: Students' beliefs. *System*, *74*(1), 9–20. https://doi.org/10.1016/j.system.2018.02.003

Ghanbari, N., & Barati, H. (2014). Iranian EFL writing assessment : the agency of rater or rating scale ?, *4*(2), 204–228.

Gonzalez, E., Trejo, N., & Roux, R. (2017). Assessing EFL university students' writing: A study of score reliability. *Revista Electronica de Investigacion Educativa*, *19*. https://doi.org/10.24320/redie.2017.19.2.928

Goodwin, S. (2016). A Many-Facet Rasch analysis comparing essay rater behavior on an academic English reading/writing test used for two purposes. *Assessing Writing*, *30*, 21–31. https://doi.org/10.1016/j.asw.2016.07.004

Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher Perspectives on Evaluation Reform. *Educational Researcher*, *44*(2), 105–116. https://doi.org/10.3102/0013189x15575517

KEMENDIKBUD. (2016). *Panduan Literasi SMP*. Jakarta.

KEMENDIKBUD. (2017). *Silabus Bahasa Inggris SMP Revisi 2017*.

Mao, Z., & Jiang, L. (2018). *Classroom Writing Assessment and Feedback in L2 School Contexts by Icy Lee (review)* (Vol. 73). https://doi.org/10.3138/cmlr.599

Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). S*ubjectivity of teacher*

judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education*, *65*, 48–60. https://doi.org/10.1016/j.tate.2017.02.021

Osborne, J., & Walker, P. (2014). Just ask teachers: BUILDING expertise, trusting subjectivity, and valuing difference in writing assessment. *Assessing Writing*, *22*, 33–47. https://doi.org/10.1016/j.asw.2014.06.002

Pas, E. T., & Bradshaw, C. P. (2014). What affects teacher ratings of student behaviors? the potential influence of teachers' perceptions of the school environment and experiences. *Prevention Science*, *15*(6), 940–950. https://doi.org/10.1007/s11121-013-0432-4

Rahayu, E., & Rahayu, E. (2019). Teacher's cognitive and affective. *Premise Journal*, *8*(1), 102–116.

Ratnawati, R., Faridah, D., Anam, S., & Retnaningdyah, P. (2018). Exploring academic writing needs of Indonesian EFL undergraduate students. *Arab World English Journal*, *9*(4), 420–432. https://doi.org/10.24093/awej/vol9no4.31

ReadWriteThink. (2004). Rubric for a narrative writing piece.

Roscoe, R. D., Allen, L. K., Johnson, A. C., & McNamara, D. S. (2018). Automated writing instruction and feedback: Instructional mode, attitudes, and revising. *Proceedings of the Human Factors and Ergonomics Society*, *3*, 2089–2093. https://doi.org/10.1177/1541931218621471

Royal-Dawson, L., & Baird, J. A. (2009). Is teaching experience necessary for reliable scoring of extended english questions? *Educational measurement: issues and practice*, *28*(2), 2–8. https://doi.org/10.1111/j.1745-3992.2009.00142.x

Skaalvik, E. M., & Skaalvik, S. (2014). Teacher self-efficacy and perceived autonomy: relations with teacher engagement, job satisfaction, and emotional exhaustion. *Psychological Reports*, *114*(1), 68–77. https://doi.org/10.2466/14.02.PR0.114k14w0

Soltero-González, L., Escamilla, K., & Hopewell, S. (2012). Changing teachers' perceptions about the writing abilities of emerging bilingual students: Towards a holistic bilingual perspective on writing assessment. *International Journal of Bilingual Education and Bilingualism*, *15*(1), 71–94. https://doi.org/10.1080/13670050.2011.604712

Statistik, B. P. (2014). Teacher self-efficacy: substantive implications and measurement dilemmas. *Katalog BPS*, *XXXIII*(2), 81–87. https://doi.org/10.1007/s13398-014-0173-7.2

Stojiljković, S., Todorović, J., Đigić, G., & Dosković, Z. (2014). Teachers' self-concept and empathy. *Procedia - Social and Behavioral Sciences*, *116*(February), 875–879. https://doi.org/10.1016/j.sbspro.2014.01.313

Trace, J., Meier, V., & Janssen, G. (2016). "I can see that": Developing shared rubric category interpretations through score negotiation. *Assessing Writing*, *30*, 32–43. https://doi.org/10.1016/j.asw.2016.08.001

Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. In *Procedia -*

Social and Behavioral Sciences (Vol. 106, pp. 234–240). https://doi.org/10.1016/j.sbspro.2013.12.027

Wang, J., Engelhard, G., Raczynski, K., Song, T., & Wolfe, E. W. (2017). Evaluating rater accuracy and perception for integrated writing assessments using a mixed-methods approach. *Assessing Writing*, *33*(March), 36–47. https://doi.org/10.1016/j.asw.2017.03.003

White, K. M., & Hall, A. H. (2014). Examining teachers ' perceptions of effective writing strategies and barriers to implementation. *Clemson University TigerPrints*. Retrieved from https://tigerprints.clemson.edu/eugene_pubs/24

Wilson, J., & Roscoe, R. D. (2019). Automated writing evaluation and feedback: multiple metrics of efficacy. *Journal of Educational Computing Research*. https://doi.org/10.1177/0735633119830764

Wiseman, C. S. (2012). Rater effects: Ego engagement in rater decision-making. *Assessing Writing*, *17*(3), 150–173. https://doi.org/10.1016/j.asw.2011.12.001

Yuan, J., & Kim, C. M. (2018). The effects of autonomy support on student engagement in peer assessment. *Educational Technology Research and Development*, *66*(1), 25–52. https://doi.org/10.1007/s11423-017-9538-x