



STUDENT'S PERCEPTIONS OF ENGLISH CLASSROOM ASSESSMENT DURING COVID-19 PANDEMIC

¹Yulia Dian Nafisah*, ²Anton Haryadi, ³Junaidi Mistar

1 SMA Islam Almaarif Singosari, Indonesia

2 Transkomunika Research and Training Institute, Indonesia

3 Universitas Islam Malang, Indonesia

ABSTRACT

This research aims at investigating the student perceptions of English classroom assessment at an Islamic Senior High School during Covid-19 pandemic. The research involved 314 students from 20 different classes across three academic years. The instrument used was 30 five-point Likert scale items from Students' Perceptions of Assessment Questionnaire (SPAQ) developed by Waldrip, Fisher, & Dorman (2008). After the data were collected and analyzed for reliability and validity using SPSS20, it was found that the internal consistency/reliability was 0.75. The internal consistency score was high, which means that the average inter-item correlation was high. The discriminant validity was also high, which implies that the instrument was valid. The data were then analyzed descriptively and reported in mean and standard deviation format. It was found that the highest mean belonged to diversity scale and the lowest belonged to student consultation. It means that although the mean of each scale was high, it is expected that the students is consulted and authenticity is improved in this school.

Keywords: *Classroom Assessment Practice; COVID-19 Pandemic; Students' Perception*

E-ISSN: 2621-9158

P-ISSN: 2356-0401

*Correspondence:
bundaaka@gmail.com

Submitted: 2 May 2021

Approved: 15 August 2021

Published: 15 December 2021

Citation:

Nafisah, Y.L., Haryadi, A., & Mistar, J (2021). Student's Perceptions of English Classroom Assessment during Covid-19 Pandemic. *Celtic: A Journal of Culture, English Language Teaching, Literature and Linguistics*, 8(2), 206-218. Doi: 10.22219/celtic.v8i2.16450

ABSTRAK

Tujuan penelitian ini adalah meneliti persepsi siswa terhadap penilaian bahasa Inggris di kelas di suatu Sekolah Menengah Atas Islam. Partisipan penelitian ini adalah 314 siswa dari 20 kelas berbeda lintas tiga tahun akademik. Instrumennya adalah *Students' Perceptions of Assessment Questionnaire* (SPAQ) yang dikembangkan oleh Waldrip, Fisher, & Dorman (2008). Setelah data dikumpulkan, data dianalisis reliabilitas dan validitasnya menggunakan SPSS20. Didapati bahwa konsistensi internal/reliabilitas dan validitas diskriminannya adalah 0.75. Skor konsistensi internalnya tinggi, yang berarti bahwa rata-rata korelasi antar-item juga tinggi. Validitas diskriminan juga tinggi, yang menyiratkan bahwa instrumennya valid. Data survei kemudian dianalisis secara deskriptif dan dilaporkan dalam format rerata dan simpangan baku. Ditemukan bahwa rerata tertinggi adalah skala keberagaman dan rerata terendah adalah skala konsultasi dengan siswa. Artinya, meskipun rerata masing-masing skala termasuk tinggi, siswa perlu diajak berdiskusi terkait aspek penilaian. Selain itu, autentisitas perlu ditingkatkan di sekolah ini.

Kata Kunci: *Pandemi COVID-19; Persepsi Siswa; Praktik Penilaian Kelas*

INTRODUCTION

Far reaching effects of Covid-19 pandemic has affected and interrupted all aspects of life, including education. Almost all schools worldwide are closed to prevent the transmission of Covid-19 (Huber & Helm, 2020). In Indonesia, majority of schools are also closed in an attempt to minimize further spread of Covid-19 (Abidah, Hidaayatullaah, Simamora, Fehabutar, & Mutakinati, 2020). Since health and safety of each citizen is prioritized, policies and regulations are enacted by the central government and Ministry of Education and Culture by shifting from face-to-face interaction to study from home (Wahyono, Husamah, & Budi, 2020). To maintain the safety and well-being of the students at kindergarten to graduate program level, this policy is inevitable and currently the best available option.

Almost no country in the world is prepared enough to plan and organize Covid-19 friendly educational process. Nevertheless, the educational process during this Covid-19 crisis time in Indonesia is considered to be running relatively well (Amalia & Sa'adah 2020). This is the conclusion after a thorough literature review from research articles, news, and books about the educational process during the study from home period. However, there are some challenges in online home learning period in Indonesia, such as unpreparedness of teachers, students, parents, online learning facilities, information and technology, etc. in anticipating such a sudden change.

Ariyanti (2020) and Amalia & Sa'adah (2020) found that internet issue (connection, accessibility, and internet cost) in Indonesia became the major obstacles in providing quality online teaching and learning process. The parents could not afford high internet unlimited access due to current economic decline and relatively expensive data subscription cost. The government has attempted to solve this issue by providing mobile internet quota subsidy. However, the resulting internet usage surge yielded in sluggish internet speeds, which caused long delays and buffers in delivering the learning audio or video contents.

To solve these problems, the Indonesian teachers are then offered to use WhatsApp, WhatsApp Web, Google Classroom, Google Group, TeamLink, Microsoft Teams, Kaizala Microsoft, Zoom Meeting & Webinar, Youtube, Google Hangouts, and others (Anugrahana, 2020), depending on their unique circumstances. The audio and video explanation is only provided on request when the students require so. After the tasks are completed and the exercises are submitted via the mutually agreed platforms, the teachers then assess the student works.

This fact indicates that assessment role is growing more important in teaching and learning process during this period. Assessment is now used to not only score the student learning, but also drive learning and even become learning in itself. Thus, it can be concluded that assessment is now more toward formative assessment than summative assessment. It is in line with what Birenbaum et al., (2015) state that assessment trend now tends toward formative assessment. The trend of assessment for learning and assessment as learning is even accelerated due to the Covid-19. Thus, the classroom assessment now uses a combination of summative and formative assessment.

These combined types of assessment is central process in effective instruction (William, 2013), essential as a part of teaching and learning (Arrafii & Sumarni, 2018), and significantly improve the student English achievement (Umar, 2018). To understand student academic achievement, it is very important to understand characteristics of the assessment tasks as perceived by the students (Alkharusi, 2011). Therefore, it is important to know the student perception on assessment tasks. In an attempt to develop and validate instrument to measure the student perception on assessment task, Dorman & Knightley (2006) identified five scales, i.e. Congruence with planned learning, Authenticity, Student Consultation, Transparency and Diversity. Further development and validation study is then continued by Waldrip, Fisher, & Dorman (2008), which resulted in five similar scales, such as Congruence with Planned Learning, Authenticity, Student Consultation, Transparency, and Diversity.

In Indonesia, there is only one study investigating the student perception on English related classroom assessment task (Rahman, 2020) using SPAQ instrument developed and validated by Waldrip, Fisher, & Dorman(2008). It aimed at exploring how students perceive grammar assessment in the EFL classroom at the English Department of UIN Ar-Raniry. It was found that the students perceived a slight congruence between grammar assessment and the planned learning. In addition, there was inadequate transparency regarding the purpose, authenticity, and assessment forms. In short, the student perceptions of classroom assessment were not good. However, this study was intended for English grammar class at higher education level in Indonesia.

Considering limited number of research about student perception on English language teaching assessment for middle school level in Indonesia, this research is aimed at filling the gap on the research and investigating student perceptions of classroom assessment in one of the Islamic private senior high school in Malang, East Java, Indonesia. In addition, it is also to discuss if the assessment in this school is in accordance with the assessment principles as set out in the applicable Curriculum 2013.

METHOD

This research employed a survey research to dig into the student perceptions of classroom assessment. Creswell (2009) states that survey research quantitatively or numerically describes trends, attitudes, or opinions of a population by studying a sample of that population. The research site was situated at private Islamic Senior High School in Malang, East Java, Indonesia. The school accreditation is A, which shows that the school is good in terms of curriculum implementation, teaching-learning process, facility, assessment, management, academic staff, and graduate competency standards. Total sampling was employed, where the participants were selected based on whether they have been taught by and acquainted with the English subject. In total, 578 students were from three academic years, and studied in 20 different classes of three majors (language study, social study, and natural science study). They were given an instruction to

complete the Student Perception of Assessment Questionnaire (SPAQ) through Google Form. Then, 314 students completed the self-completed questionnaire by the expected deadline.

In this research, the Student Perceptions of Assessment Questionnaire (SPAQ), which was developed by Waldrip, Fisher, & Dorman(2008), was adapted as the instrument to inquire about student perceptions. There were 30 items, which were divided into five scales: the congruence with planned learning (items 1-5); the authenticity (items 6-12); student consultation (13-18); transparency (items 19-24); and diversity (25-30). Each item was then presented with five Likert scale from strongly disagree to strongly agree.

This instrument was chosen due to two main considerations. First is the theoretical grounding and psychometric quality. The theoretical grounding is similar to those of the current applicable curriculum in Indonesia and psychometric is relatively simple. Second is that the instrument has been tested for validity ($M = .50$) and reliability (Cronbach alpha internal consistency ranging from .68 to .86) during the development and validation phase.

The instrument was then translated into Bahasa Indonesia by an experienced translator. The translated instrument was further checked by both researchers. In addition, a clear explanation on how to rate the statement in the instrument was also provided to make sure any misunderstanding and error were avoided.

In collecting the data, the online questionnaire was distributed to the 20 homeroom teachers, who were consulted beforehand. The preset deadline was also communicated to the students through their homeroom teachers. When it was due, the collected data from Google form were then exported to Microsoft Excel format. The completed forms were checked once again to make sure that the data were intact.

The collected data were arranged from the highest grand mean to the lowest grand mean. After a thorough checking to find missing values, the data were analyzed using SPSS 20 to calculate the internal consistency/reliability and discriminant validity with regard to each scale of the student perceptions of classroom assessment. It was found that the internal consistency/reliability using Cronbach alpha was .942 and discriminatory validity was .75. It can be concluded that the data were valid and reliable. The next step was to analyze the data descriptively to find out the mean and standard deviation. To facilitate easier checking, the data were calculated using SPSS20. To conclude the minimum and the maximum length of this SPAQ 5-point Likert type scale, the following formula, i.e. $(5-1=4)$ was used. It was then divided by 5 as the greatest value of the scale, i.e. $4 \div 5 = 0.80$. The length of the cells is determined below:

- Range from 1 to 1.80 represents very low.
- Range from 1.81 to 2.60 represents low.
- Range from 2.61 to 3.40 represents medium.
- Range from 3.41 to 4.20 represents high.
- Range from 4.21 to 5.00 represents very high.

FINDINGS

There are two parts in findings section, i.e. overall perception and finding for each scale. In the overall perception, general information about student perception on the classroom assessment is presented. The presentation of each scale follows.

Overall Perception

The student perception of the 30 items toward classroom assessment was analyzed descriptively. In Table 1, the highest mean came from Diversity scale with grand mean of 4.09. It means that most participants agreed that they were given equal chances to complete assessment task, various assessments to choose from, and different ways to complete them. The lowest mean was Student Consultation scale with grand mean 3.50. It suggests that some participants perceived that the teachers had explained each type of assessment and its scoring method, but they did not help the class to develop rules for assessment in English language learning activities.

Table 1. The Result of Student Perceptions of Classroom Assessment

Parts of Questionnaire	Grand Mean	Standard Deviation
Diversity	4.09	0.98
Congruence with Planned Learning	3.86	0.98
Transparency	3.81	0.98
Authenticity	3.60	1.04
Student Consultation	3.50	1.05

The standard deviation (SD) of all the items ranges from .98 and 1.05. The SD is standard, indicating that data on the student perceptions of classroom assessment were normally distributed.

Finding of Each Scale

In this part, further explanation is given to each scale, consisting of the diversity, congruence with planned learning, transparency, the authenticity, and student consultation.

Diversity

Diversity refers to the extent to which all students have an equal chance at completing assessment tasks. There were six statements in this scale. The table below shows the results in detail.

Table 2. Diversity

Items on Questionnaire	Mean
1. When there are different ways I can complete the assessment.	4.35
2. I have as much chance as any other student at completing assessment tasks.	4.18
3. I am given a choice of assessment tasks.	4.04
4. I am given assessment tasks that suit my ability.	4.03
5. I complete assessment tasks at my own speed.	4.01
6. When I am confused about an assessment task, I am given another way to answer it.	3.96
Grand Mean	4.09

As the table shows, the difference between each highest and lowest mean (4.35 and 3.96) is thin. It implies that the students were given more than one way to complete the assessment. Therefore, the students felt that they had equal chance to complete them due to the various assessment types. In addition, the students were able to complete assessment at their own speed. They were even accommodated while they were having some difficulties to complete a certain assessment task. In this covid-19 pandemic period, the students in this school were assigned assessment tasks via Google Classroom and expected to complete them by the deadline. When they had difficulty, they were offered another way to complete them. It is the reason why the diversity scale is the highest among other scales.

The Congruence with Planned Learning

Congruence with planned learning refers to the extent to which assessment tasks align with the goals, objectives, and activities of the learning. This scale consisted of six statements with the grand mean 3.86. The table below shows the results in detail.

Table 3. Congruence with planned learning

Item on Questionnaire	Mean
1. I am assessed on what the teacher has taught me.	4.10
2. My assignments/tests are about what I have done in class.	4.01
3. How I am assessed is like what I do in class.	3.99
4. How I am assessed is similar to what I do in class.	3.86
5. My English assignments/tests examine what I do in class.	3.66
6. Questions in English subject tests what I know.	3.59
Grand Mean	3.86

As shown above, the highest score is 4.10 and the lowest is 3.59. This scale was high based on the rating criteria. From the top four statements, it was clear that the assessment were congruent with the teaching and learning activities in the class. It indicates that the learning activities and assessment has been well planned. The basic competence, learning contents, and its respective assessment were then communicated to the students. Based on the researcher personal observation, the learning activities were planned and written in lesson plans by the English teachers in this school. The lesson plan was then executed in the learning activities and the classroom assessment practices. The students perceived these activities positively. However, some students felt that there was a slight difference between what they did in the class and what they knew.

Transparency

Transparency refers to the extent to which the purposes and forms of assessment tasks are well defined and clear to the learner. This scale with the grand mean 3.81 consisted of six statements. The table below shows the results in detail.

Table 4. Transparency

Items on Questionnaire		Mean
1.	I am told in advance when I am being assessed.	4.09
2.	I am told in advance on what I am being assessed.	4.03
3.	I am clear about what my teacher wants in my assessment tasks.	3.79
4.	I know what is needed to successfully complete an English lesson assessment tasks.	3.70
5.	I know how a particular assessment task will be marked.	3.68
6.	I understand what is needed in all English assessment tasks.	3.54
Grand Mean		3.81

As shown in Table 4, the highest score is 4.09 and the lowest score is 3.54, which belongs to high category based on the rating criteria. From the two top statements, it indicates that the teachers have done all their tasks such as informing the students when and what would be assessed. As explained in the previous scale, the teachers made a lesson plan, which also include the content and time of assessment. However, as indicated above, some students had some difficulties in understanding what they needed to do to prepare for the assessment task.

The Authenticity

Authenticity refers to the extent to which assessment tasks feature real life situations that were relevant to the learner. This scale consisted of six statements with grand mean 3.60. The table below shows the results in detail.

Table 5. Authenticity

Item on Questionnaire		Mean
1.	My English assessment tasks are useful in everyday things.	3.72
2.	I can show others that my learning has helped me do things.	3.71
3.	Assessment in English examines my ability to answer every day questions.	3.58
4.	I find English assessment tasks are relevant to what I do outside of school.	3.55
5.	Assessment in English tests my ability to apply what I know to real-life problems.	3.52
6.	Assessment in English examines my ability to answer every day questions.	3.49
Grand Mean		3.60

As shown above, the highest score is 3.72 and the lowest score is 3.49, which indicates a thin difference. This scale is also high based on the rating criteria. The two highest statements indicate that the English assessment was useful in everyday things and the students could explain and show them to the other parties. It also tested the student ability to implement the English to the real life problems and examine the student ability to answer everyday question. In this school, the teachers sometimes assigned the students to look for the example of certain topic, such as congratulating, in the internet. The teacher then would

discuss with the respective students if this is correct or not. In addition, the material was sometimes adapted from website, such as the current song lyrics.

Student Consultation

Student Consultation means that students were consulted and informed about the assessment task types being assigned. This scale consisted of six statements and had the grand mean 3.50. The table below shows the results in detail.

Table 6. Student Consultation

Items on Questionnaire	Mean
1. My teacher has explained to me how each type of assessment is to be used.	3.90
2. I am aware how my assessment will be marked.	3.60
3. I have a say in how I will be assessed in English lesson.	3.54
4. I can select how I will be assessed in English lesson.	3.44
5. In English lesson, I am asked about the types of assessment that are used.	3.40
6. I have helped the class develop rules for assessment in English lesson.	3.12
Grand Mean	3.50

As shown in Table 6, the highest score is 3.90 and lowest score is 3.12. This scale in general is high according to the rating criteria. Based on the top two statements, the teachers had explained about the types of assessment to be used and scoring method. As explained earlier, the lesson plan included the type of assessment and its scoring technique. The teachers then communicated them to the students. However, the students were not always consulted in terms of assessment types and assessment rules. Thus, this decision was from the teachers. This fact is due to Covid-19 pandemic where the teacher-student communication was just through Google Classroom and sometimes WhatsApp. Therefore, the students were not consulted by the teachers.

DISCUSSION

As discussed earlier, this research is driven by the lack of empirical research about student perceptions of classroom assessment, especially within the EFL context at middle school in Indonesia. A number of findings have improved our understanding about the nature of classroom assessment as perceived by the students.

First is about diversity. It was found that the mean score was 4.09 (SD: 0.98). This is the highest mean score among five subscales. One possible explanation to this fact is that the variety of assessment types employed by the English teachers. As insiders to this school, it is true that the teachers combined summative and formative assessment. For example, the teachers assigned many alternative assessments, such as portfolio, performance, product based, project based, self-assessment, peer assessment, etc. The implementation of various alternative assessment types is in accordance with the spirit of curriculum 2013 (Azhar, 2018). By administering various assessments, it is more likely that it fulfills student individual characteristics and needs. Therefore, it is the possible reason why the highest numbers of students perceive that they were given an equal

chance and different ways to complete assessment tasks. Variety of assessment types improve students' motivation (Seale, Chapman, & Davey, 2000). The various tasks enable the students to express themselves based on their own preferred types of assessment. In turn, it also increases students achievement (Umar, 2018).

Second is about congruence with planned learning. It was found that the mean score is 3.82 (SD .92), which suggests that the assessment was perceived as congruent with the planned learning. The assessment in this school was considered good and in line with the first principle of assessment, i.e. to assess the learning the students have experienced. It is in accordance with Curriculum 2013, where the teachers are expected to plan the learning activities and the subsequent assessment. In this school, the teachers prepared the lesson plan before teaching the class so the learning activities and the subsequent assessment are well planned and implemented accordingly. Therefore, the students know that what they learn will be assessed during or after the learning activities. The students become more enthusiastic in teaching learning activities when they realize that what they are learning in classroom will be tested on assessment tasks (McMillan & Nash, 2000; Santhanam, 2002; Brookhart & Bronowicz, 2003). Student enthusiasm and motivation determines to what extent the students are willing to invest their time and attention to the lesson. The more enthusiastic and the more motivated the students are in learning, the more successful they can be in their language acquisition endeavor (Purwanti, Puspita, & Mulyadi, 2019; Rosmayanti & Yanuarti, 2018).

Third is related to transparency. It was found that the mean score is 3.81 (SD:0.98). Discrepancy of this scale and congruence with planned learning is very thin. It implies that this scale is somehow related to planned learning or lesson plan. When the assessment is planned in advance, the teachers have the chance to explain the students about the aspects of assessment. It indicates that the assessment is well defined and clear to the learner. Transparency has a positive impact on student learning (Settiawan & Hilmawan, 2016). Therefore, there is an increasing need for a greater transparency in assessment processes (Rust, Price, & Berry, 2003). The transparency and democracy should be part of teacher assessment literacy and practice (Giraldo, 2018), by letting students know about what is expected from them, assessment time, aspects of assessment, scoring rubrics, grading technique, etc.

Fourth is authenticity issue. It was found that the mean score was 3.67 (SD:1.04). Authenticity here refers to the extent to which assessment tasks feature real life situations those are relevant to the learner. This authenticity enables the students to be more motivated since the lessons they are learning and the assessment to test their learning progress simulate real life experience. Authentic assessment is a central element in communicative language teaching (Esfandiari & Gawhary, 2019). Therefore, more English language instructions now integrate authentic learning situation, authentic material, and hence authentic assessment (Jaelani & Umam, 2021). In addition, authentic assessment is a critical component in curriculum 2013 (Hamidah, 2013). However, authenticity is lower than the

other three scales. The possible explanation to this fact is that pandemic situation forced the teachers mostly to rely on the student worksheet (LKS). Thus, the teacher did not frequently use authentic assessment.

Fifth is related to student consultation. It was found that the mean score is 3.50 (SD: 1.05). The mean score for student consultation was the lowest among five scales. It means that some students felt that they did not participate in developing their assessment criteria. The similar finding was empirically supported by the previous studies (Cheng, Rogers, & hu, 2004; Wang et al. 2013; Cheng, Wu, & Liu, 2015). In Indonesian context, this phenomenon is similar to a research finding by Rahman (2020) that the students were not consulted before deciding the assessment criteria. It is important to consult the students to make sure the assessment is fair and reliable (Dancer & Kamvounias, 2005; Rust et al., 2003). In addition, student participation is really encouraged and expected in Curriculum 2013 (Pusat Kurikulum dan Perbukuan, 2014). However, during the pandemic, the teacher were not always in a good time and atmosphere to consult the assessment with some students. The teachers were exhausted with all the burdens, either professional or personal ones. This may explain why this scale was the lowest.

In short, the mean scores of five subscales inform that the assessment in this school during home learning is perceived to be good by the students. The grand mean of each scale was more than 3.4 out of 5, which implies the student perception on classroom assessment is high. However, the finding shows that the authenticity and student consultation was the lowest among the other three scales. Therefore, it is expected that these authenticity and student consultation scales be improved.

CONCLUSION

Based on the findings of this research, the students agree that classroom assessment in their English subject were congruent with planned learning, authentic, transparent, had been consulted with students, and diverse. The assessment met the criteria of good classroom assessment practice. As the scale is aligned with the applicable curriculum 2013, it means that the English assessment has implemented the directions set by the national curriculum.

This research informs the stakeholders such as teachers and educational administrators that student perceptions of English classroom assessment were high. However, some improvements need to be focused on the area of student consultation and authenticity. Since the finding shows that student consultation and authenticity scales had the lowest grand mean, it is expected that the students need to be consulted in relation to types of assessment, scoring method, use of assessment, and especially rules of assessment. In addition, it is expected that the authenticity should be improved so the students learn everyday English and they can implement what they learn at school at their daily activities.

REFERENCES

- Abidah, A., Hidaayatullaah, H. N., Simamora, R. M., Fehabutar, D., & Mutakinati, L. (2020). The Impact of Covid-19 to Indonesian Education and Its Relation to the Philosophy of "Merdeka Belajar." *Studies in Philosophy of Science and Education*, 1(1), 38–49. <https://doi.org/10.46627/sipose.v1i1.9>
- Alkharusi, H. (2011). Development and Datametric Properties of a Scale Measuring Students' Perceptions of The Classroom Assessment Environment. *International Journal of Instruction*, 4(1), 105–120.
- Amalia, A., & Sa'adah, N. (2020). Dampak Wabah Covid-19 Terhadap Kegiatan Belajar Mengajar Di Indonesia. *Jurnal Psikologi*, 13(2), 214–225. <https://doi.org/10.35760/psi.2020.v13i2.3572>
- Anugrahana, A. (2020). Hambatan, Solusi dan Harapan: Pembelajaran Daring Selama Masa Pandemi Covid-19 Oleh Guru Sekolah Dasar. *Scholaria: Jurnal Pendidikan Dan Kebudayaan*, 10(3), 282–289. <https://doi.org/10.24246/j.js.2020.v10.i3.p282-289>
- Ariyanti, A. (2020). EFL Students' Challenges towards Home Learning Policy During Covid-19 Outbreak. *IJELTAL (Indonesian Journal of English Language Teaching and Applied Linguistics)*, 5(1), 167. <https://doi.org/10.21093/ijeltal.v5i1.649>
- Arrafii, M. A., & Sumarni, B. (2018). Teachers' Understanding of Formative Assessment. *Lingua Cultura*, 12(1), 45. <https://doi.org/10.21512/lc.v12i1.2113>
- Azhar, F. (2018). Authentic Teaching and Assessment As the Solution To Educational Evaluation in Reference To Asean Economic Community in Indonesia. *International Journal of Educational Best Practices*, 2(1), 26. <https://doi.org/10.31258/ijebp.v2n1.p26-38>
- Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., ... Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education*, 13(1), 117–140. <https://doi.org/10.1177/1478210314566733>
- Brookhart, S. M., & Bronowicz, D. L. (2003). "I don't like writing. It makes my fingers hurt": Students talk about their classroom assessments. *Assessment in Education: Principles, Policy and Practice*, 10(2), 221–242. <https://doi.org/10.1080/0969594032000121298>
- Cheng, L., Rogers, T., & hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing*, 21(3), 360–389. <https://doi.org/10.1191/0265532204lt288oa>
- Cheng, L., Wu, Y., & Liu, X. (2015). Chinese university students' perceptions of assessment tasks and classroom assessment environment. *Language Testing in Asia*, 5(1). <https://doi.org/10.1186/s40468-015-0020-6>
- Creswell, J. W. (2009). Research Design Qualitative, Quantitative, and Mixed Methods Approaches. In V. Knight (Ed.), *Sage Publication. Inc* (third edit, Vol. 3, p. 295). <https://doi.org/10.2307/1523157>
- Dancer, D., & Kamvounias, P. (2005). Student involvement in assessment: A project

- designed to assess class participation fairly and reliably. *Assessment and Evaluation in Higher Education*, 30(4), 445–454. <https://doi.org/10.1080/02602930500099235>
- Dorman, J. P., & Knightley, W. M. (2006). Development and validation of an instrument to assess secondary school students' perceptions of assessment tasks. *Educational Studies*, 32(1), 47–58. <https://doi.org/10.1080/03055690500415951>
- Esfandiari, M., & Gawhary, M. W. (2019). From Genuineness to Finder Authenticity in Communicative Language Teaching. *International Journal of English and Cultural Studies*, 2(1), 36. <https://doi.org/10.11114/ijecs.v2i1.4149>
- Giraldo, F. (2018). Language assessment literacy: Implications for language teachers. *Profile: Issues in Teachers' Professional Development*, 20(1), 179–195. <https://doi.org/10.15446/profile.v20n1.62089>
- Hamidah, E. (2013). Indonesian Efl Secondary School Teachers " Perception and Preferences on Authentic Speaking Performance Assessment. *English Language and Literature International Conference (ELLiC)*, 90–96.
- Huber, S. G., & Helm, C. (2020). COVID-19 and schooling: evaluation, assessment and accountability in times of crises—reacting quickly to explore key issues for policy, practice and research with the school barometer. *Educational Assessment, Evaluation and Accountability*, 32(2), 237–270. <https://doi.org/10.1007/s11092-020-09322-y>
- Jaelani, A., & Umam, A. (2021). Preparing EFL pre-service teachers for curriculum 2013 through authentic materials and assessment integration. *Journal of English Educators Society*, 6(1), 171–177. <https://doi.org/10.21070/jees.v6i1.829>
- McMillan, J. H., & Nash, S. (2000). Teacher classroom assessment and grading practices decision making. *Paper Presented at the Annual Meeting of the National Council on Measurement in Education*, 39.
- Purwanti, D., Puspita, H., & Mulyadi, M. (2019). the Correlation Between English Learning Motivation and English Proficiency Achievement of English Study Program Students. *Journal of English Education and Teaching*, 3(1), 79–94. <https://doi.org/10.33369/jeet.3.1.79-94>
- Pusat Kurikulum dan Perbukuan, K. (2014). *Pedoman Guru Mata Pelajaran Bahasa Inggris (Guideline for English Teachers)*. Jakarta: Kementerian Pendidikan dan Kebudayaan.
- Rahman, F. (2020a). Undergraduate Students' Perception towards Grammar Assessment in the EFL Classroom. *SAGA: Journal of English Language Teaching and Applied Linguistics*, 1(2), 127–136.
- Rosmayanti, D., & Yanuarti, H. (2018). The Relationship Between Students' Motivation and Their Learning Achievement. *Professional Journal of English Education*, 1, 783–788. <https://doi.org/10.32865/fire202062188>
- Rust, C., Price, M., & Berry, O. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment and Evaluation in Higher Education*, 28(2), 147–164.

- <https://doi.org/10.1080/02602930301671>
- Santhanam, E. (2002). Congruence of teaching, learning, assessment and evaluation. *Focusing on the Student. Proceedings of the 11th Annual Teaching Learning Forum*, 159–166. Perth: Edith Cowan University.
- Seale, J. K., Chapman, J., & Davey, C. (2000). *The influence of assessment in students' motivation to learn in a therapy degree course* *The influence of assessments on students' motivation to learn in a therapy degree course*. (May 2018). <https://doi.org/10.1046/j.1365-2923.2000.00528.x>
- Settiawan, D., & Hilmawan, R. (2016). Increasing transparency in assessment to improve students' learning at Language Development Centre of UIN Suska Riau. *PROSIDING ICTTE FKIP UNS 2015, 1*, 364–368.
- Umar, A. M. A. (2018). *The Impact of Assessment for Learning on Students' Achievement in English for Specific Purposes A Case Study of Pre-Medical Students at Khartoum University: Sudan*. *11(2)*, 15–25. <https://doi.org/10.5539/elt.v11n2p15>
- Wahyono, P., Husamah, H., & Budi, A. S. (2020). Guru profesional di masa pandemi COVID-19: Review implementasi, tantangan, dan solusi pembelajaran daring. *Jurnal Pendidikan Profesi Guru*, *1(1)*, 51–65.
- Waldrip, B. G., Fisher, D. L., & Dorman, J. P. (2008). Students' perceptions of assessment process: Questionnaire development and validation. *5th International Conference on Science, Mathematics and Technology Education*, 561–568.
- William, D. (2013). Assessment: The Bridge between Teaching and Learning. *Voices from the Middle*, *21(2)*, 40.