Readability of English Teacher Assessment Literacy Instrument in Indonesian Context

A.A. Gede Yudha Paramartha[1,*] Ni Luh Putu Eka Sulistia Dewi[1] Luh Gede Eka Wahyuni[1]

[1]*English Language Education Study Program, Universitas Pendidikan Ganesha, Singaraja, Indonesia*
[*]*Corresponding author. Email: yudha.paramartha@undiksha.ac.id*

ABSTRACT

Many studies have pointed out that the assessment practices by teachers in Indonesia were considered low. This arises a question why it happened. That is why the Teacher Assessment Literacy (TAL) need to be studied more deeply. Measurement of TAL is necessary to provide a clearer picture of assessment practices especially by English teachers in Indonesia. With the existence of accurate measurement standard, it is expected that stakeholders can make a form of intervention for teachers to improve the quality of teacher assessment practices. To answer this question, we developed an instrument for TAL based on Indonesian context. The focus of this study is to assess the readability of the developed instrument by consulting the draft to two assessors. Both were asked to read each item in the instrument then gave an assessment of the readability into four criteria: highly readable, readable, somewhat difficult, and very difficult. The results show that most of the items (70%) were categorized 'highly readable', while no item was categorized 'very difficult'. Some items were also discussed as examples to give clearer pictures on the readability of the developed instrument.

Keywords: English, Readability, Teacher assessment literacy.

1. INTRODUCTION

In formal education, the learning process and assessment are two things that are interrelated and cannot be separated. teachers are expected to be able to integrate assessment in learning, so that teachers can apply learning that is targeted and more meaningful [1]. Assessment, especially in English education, is very important because of the washback effect of the assessment on the learning process [2]. The washback effect is described as the impact of the assessment on the learning process, in this case English education. If the assessment is designed in accordance with the learning objectives, the learning process will also run in accordance with the learning objectives, so that the English learning process can run effectively and meaningfully.

In the Indonesian context, assessment is still considered difficult by teachers. Research by Marhaeni, Dantes, & Paramartha [3] shows that there was a wide gap between the expectation of how teachers need to implement assessment according to K-13 and the real implementation of assessment by teachers. This gap occurred at three stages of the assessment, namely planning, implementing, and analysis and reporting the results of the assessment [3]. In another study, it was found that teachers were not familiar with the application of authentic assessment and always used traditional assessment, although teachers agree that authentic assessment was more suitable for measuring student competence [4]. In addition, teachers also experienced difficulties in developing assessment instruments in accordance with learning indicators, as well as designing authentic assessment rubrics [5].

Examining the results of those studies, the question arises why the practice of teacher assessment in Indonesia is still relatively low. Teachers' insights about assessment practices (known as teacher assessment literacy) need to be studied more deeply. Teacher assessment literacy (TAL) is a teacher's fundamental

understanding of concepts and assessment procedures that influence decision-making of an educational outcome [6]. Measurement of TAL is necessary to provide a clearer picture of assessment practices by English teachers in Bali. With the existence of accurate measurement standard, it is expected that stakeholders can make a form of intervention for teachers to improve the quality of teacher assessment practices.

Thus, it is essential to provide an instrument for measuring TAL that is in accordance with the context of English education in Indonesia. This is in line with research by [7] which concludes that the teacher assessment literacy measurement instrument must be made according to the educational context in which this instrument is used. TAL measurement instruments have been developed in various contexts in many countries. For example, Plake, et al [8] as the originator of TAL have developed it in the United States and followed by Mertler & Campbell [9]. Meanwhile, in the Asian Region, the instrument by Plake has also been adapted by Tao [10] in the context of education in Cambodia, Soh & Zhang [11] and Xu & Brown [12] in the context of education in China, and Ölmezer- ztürk & Aydin [13] in the context of education in Turkey. However, reports on the development of TAL measurement instrument in the context of education in Indonesia have not been found to date.

The general objective of this study is developing an instrument for measuring TAL in the context of English education in Indonesia based on the blueprint outlined by Plake [8]. In its development, the instrument developed by Plake [8] was translated into *Bahasa Indonesia* (Indonesian language) and adapted to English education in Indonesian context. In the process, this instrument must meet the requirements of a good instrument, one of which is the readability of the translated instrument [14] [15]. Therefore, the main focus in this article is to examine the readability of the TAL measurement instrument in the context of English education in Indonesia. Readability is important because it connects the reader with the text, and it also affects the fluency while reading the items in the instrument. Moreover, by ensuring the readability of the developed instruments, information misinterpretation can hopefully be avoided to achieve quality instrument to measure TAL.

2. LITERATURE REVIEWS

2.1. Teacher Assessment Literacy Standard

Teacher assessment literacy (TAL) is a teacher's fundamental understanding of concepts and assessment procedures that influence decision-making of an educational outcome [6]. TAL is considered important because it is the key linking the quality of assessment with student achievement. Supporting the above opinion,

Newfields [16] stated that there are three main reasons why TAL is important. First, assessment is a widespread issue in the education system, where teachers spend up to 50% of their time carrying out assessment-related activities. Second, it is very important for teachers to have a broad insight into assessment through a review of related assessment literatures. Third, TAL is needed when teachers need to communicate assessment results to stakeholders.

Related to TAL standard, Plake [8] has made a framework that has been used in many assessment contexts. This standard includes 1) choosing an assessment method that is in accordance with the learning objectives, 2) developing an assessment method that is in accordance with the learning objectives, 3) conducting an assessment, scoring, and interpreting the results of the assessment, whether made by external parties or the assessment method made by the teacher, 4) using assessment results in making decisions about student learning outcomes, teaching planning, curriculum development, and school construction, 5) developing valid ranking procedures, 6) communicating assessment results to stakeholders, and 7) recognizing unethical, illegal, and inappropriate assessment methods and assessment information.

The teacher assessment literacy standard above is in accordance with the standard set by Stiggin in Mertler & Campbell [9]. This standard includes: 1) linking assessment with clear objectives, 2) clarifying achievement expectations, 3) using appropriate assessment methods, 4) developing assessment processes, assessment criteria, and appropriate quality sampling, 5) avoiding bias assessment, 6) communicating student learning outcomes effectively, 7) using assessment for learning improvement.

In this study, the TAL standard by Plake [8] was used. This was done for two reasons. First, the standard set by Plake, et al. has an advantage that Stiggin does not have, namely the measurement of unethical use of assessment information. Teacher assessment literacy is not only about understanding and application, but also about the ethical use of assessment information [16], so assessment ethic can describe literacy deeper. Second, the instrument by Plake [8], called Teacher Assessment Literacy Questionnaire can be accessed online easily, so that by studying the contents of the questionnaire (even in the context of education in the United States), it will reduce the bias in the author's understanding of the standards referred to by Plake [8].

2.2. Translation Readability Scale

The concept of translation readability is a crucial when ones have to translate a text of a source language to a target language. It is needed to check the quality of translation in the target language so that the translation

result can be comprehended by the target reader with ease. A highly readable translated text can be understood by the target readers as if they read a text written by a native writer [17]. The readability of a text can be affected by some factors, such the use of foreign vernacular, the use of ambiguous words, incomplete sentences, and incoherent sentences [18].

The readability test is needed when developing a translated measurement instrument because the target language might have different cultural appropriateness to the source language [14]. The issue of acceptability of the instrument by the audiences in target language environment was also discussed by Durand [15] in which their study sample rate was low because the translated language readability was not appropriate for the audiences. These studies show that the readability test for the current instrument development is required.

Related to this study, a standard by O'Brien [17] was used to test the readability of the developed instrument. O'Brien categorized the readability scale into four qualitative levels: highly readable, readable, somewhat difficult, and very difficult. Those levels were indicated by reading ease and pause.

3. METHOD

In the process of developing the current TAL instrument, the 7 standards of TAL (*choosing, developing, administering, using-decision, using-grading, communicating, recognizing ethics*) were used as the basis. Those standards were then adapted to English education in Indonesia context with the final draft of the instrument consisted of 30 multiple-choice items.

Two linguists were asked to assess readability using the readability instrument developed by O'Brien [17]. Both linguists were asked to read each item along with the answer options, and then gave an assessment of readability based on the indicators as shown in table 1.

After the judges gave their assessments, the developed instrument was improved based on the input from these experts.

4. RESULT AND DISCUSSION

the main focus in this article is to examine the readability of the TAL measurement instrument in the context of English education in Indonesia. Two linguists were asked to assess readability using the readability instrument developed by O'Brien [17]. The results of readability assessment were then recapped and shown in table 2.

Table 1. The Translation Readability Scale by O'Brien (2010)

Scale	Indicator
Highly Readable	The segment reads as if it were written by a native writer. It is easy to read and you had no cause to pause during reading.
Readable	The segment is relatively easy to read, but you may have had to pause slightly for processing or to jump backwards once in the sentence to re-read something.
Somewhat Difficult	The segment does not read as if it was written by a native writer and you have had to pause once or twice during reading and/or jump backwards to re-read one or two phrases or words.
Very Difficult	The segment is difficult to read because its structure does not conform to what is normally expected of a grammatical sentence in the TL. You would have to re-read to make any sense of it.

Table 2. Readability Assessment from two Assessors

Assessor	Readability Scale							
	Highly readable		readable		Somewhat Difficult		Very Difficult	
	Items	Proportion	Items	Proportion	Items	Proportion	Items	Proportion
I	18	60%	10	33%	2	7%	0	0%
II	21	70%	7	23%	2	7%	0	0%

Based on the data shown in Table 2, the two assessors agreed that there were no items which were very difficult to read. While, two items (Item #8 and #25) were

somewhat difficult to read. Moreover, the they have similar assessment on highly readable items (60% from Assessor I and 70% from Assessor II) as well as readable

items (33% from Assessor II and 23% from Assessor II) with the only differences were from items number #5, #11, and #17. The more detailed description for every readability level is discussed below.

4.1. Highly Readable Items

An item can be categorized as highly readable when it can be comprehended with ease by the reader as if it is originally written by a native writer. Moreover, it can be read without a slight pause when processing the information in the text. Based on the data gathered from the two assessors, it was found that Assessor I considered 18 items (60%) as highly readable, while Assessor II judged 21 items to be so. The followings are three examples of items with highly readable level.

> *"Apakah pertimbangan yang paling penting dalam memilih metode untuk menilai prestasi siswa?"* (Question #1)
>
> *"Seorang guru ingin mendokumentasikan validitas hasil penilaian kelas. Informasi apa yang dapat memberikan bukti atas tujuan tersebut?"* (Question #6)
>
> *"Manakah praktik-praktik penilaian berikut yang hasilnya paling tidak mencerminkan prestasi siswa?"* (Question #21)

These three items are the example of items which are highly readable. Further analysis showed that they used effective and straight forwards sentences, so the two assessors could easily comprehend the intended meaning from the items like what was shared by Assessor II.

> *"This item used a very effective sentence since the it only took two lines, as well as the message can be easily understood."* (Assessor II on Question #1)

Comprehensibility and readability of a text are related to each other. They are two things that can determine the quality of translated text [19]. In the current study, the comprehensibility was achieved by effective sentences that were used by the items. Comprehensibility can also help the translated text to avoid translation error, which can be achieved by using effective sentences [20].

4.2. Readable Items

A readable text can be shown by two characteristics: the readers can easily understand the meaning, but they need a slight pause or jump backward to process the information. Based on the analysis taken from the assessors, it was found that the readable items were mainly those which had long sentences with several information in the items. The followings are the examples.

> *"Seorang guru Bahasa Inggris memulai semester baru dengan materi Recount Text.*

> *Sebelum memulai unit baru, guru meminta siswa untuk menceritakan pengalamannya di masa lampau secara tertulis. Mana dari hal berikut yang menjadi alasan guru melakukan hal di atas?"* (Question #15)
>
> *"Seorang guru Bahasa Inggris ingin siswanya mengatahui hasil tes mereka secepatnya. Beliau mengatakan kepada para siswa bahwa lembar jawaban yang sudah diperiksa akan diletakkan di atas kursi di luar ruangannya pada jam pulang sekolah dan siswa dapat mengambilnya. Apa yang keliru mengenai tindakan guru tersebut?"* (Question #28).

These sample questions are evaluated as readable since the assessors agreed that in processing the information, they needed to have a slight pause. However, the meaning of the text could still be easily understood. It magnifies the idea shared in the last part which argues that effective sentence is needed for better readability. Apart from the use of effective sentence, Richard and Schmidt [21] also argues that the readability is influenced by the average sentence length and the number of words in the passage. This is in line with a study by Kadayat & Eika [22] who stated that it was not necessary to have short sentences, but too long passage could give too much workload to process the information. Memory work load was also one thing that was needed to be considered for readability when making a long passage [23]. Based on those studies, it is evident that the readability of these items was in in 'readable' level considering the number of words in the items influenced more information work load to the readers.

4.3. Somewhat Difficult Items

An item with somewhat difficult readability level is when the reader feels that the translated text is not written by a native writer. As a consequence, the reader needs to reread several times to process the information. Related to the current study, the assessors agreed that there were two items with this level of readability (item #8 and #25). The reason why these items were considered somewhat difficult was because the choice of words did not meet the general consensus of common Bahasa Indonesia (Indonesian Language). The following was the example of this matter.

> *"Yang mana dari yang berikut ini adalah interpretasi yang valid dari nilai UAS tersebut?"* (Question #25).

The assessors said that because the translated text did not use common Bahasa Indonesia, they need to reread several times to understand the meaning. The phrase *'yang mana dari yang berikut ini...'* ('which of the followings') was not appropriate in Bahasa. So, the assessors suggested to change this phrase which was more appropriate to the common Bahasa that was '*Mana*

dari pernyataan berikut ini…' ('which of the following statements…'). These items were then revised based on the assessors' suggestion.

Cultural appropriateness to a translated text is very important to improve the readability. Language localization of an instrument is needed in large-scale assessments [24]. The data taken from the two assessors signified this idea since they could not easily understand about the idea in the items with 'somewhat difficult' level of readability. Comprehensibility of a translated text was also a factor to achieve culturally and linguistically appropriate text [25]. Thus, revision was needed for these items to meet better readability level.

4.4. Very Difficult Items

A very difficult level of readability is when a text does not follow the expected grammatical structure of the target language. As the consequence, the reader needs to reread many times to make sense of it. Based on the assessors' evaluation, there was no item in the instrument that fell into this level of readability.

5. CONCLUSION

Based on the results of this research, from 30 items assessed for their readability, no items (0%) were categorized 'very difficult'. two item (7%) were 'somewhat difficult'. 70% of the items were categorized 'readable', and 23% were categorized 'highly readable. The items in 'highly readable' level were those which had short and effective sentences, while the items in 'readable' level were those in which the assessors needed slight pause to process the information because of having longer passages. Furthermore, the 'somewhat difficult' items were the items which did not linguistically meet the general consensus of common Bahasa Indonesia (Indonesian Language). Thus, those items need to be rephrased for better readability. In conclusion, the instrument can be easily understood with minor revision, and based on the readability assessment, it has been confirmed that the instrument is ready to be distributed to small sample for validation process.

REFERENCES

[1] J. H. McMillan, Fundamental assessment principles for teachers and school administrators, in:).. Practical Assessment, Research and Evaluation, 2001, vol. 7(8), pp. 2000–2001.

[2] J. McKinley, G. Thompson, Washback Effect in Teaching English as an International Language, The TESOL Encyclopedia of English Language Teaching, 2018, pp. 1–12. DOI: https://doi.org/10.1002/9781118784235.eelt0656

[3] A.A.I.N. Marhaeni, N. Dantes, AA.G.Y Paramartha, Teacher assessment literacy : Discrepancies in authentic assessment practice in EFL context, Proceeding Book of 1st International Conference on Educational Assessment and Policy, 2018, vol. 2, pp.18–21.

[4] Atmazaki, Authentic Assessment in Indonesian Language Learning, In Advances in Social Science, Education and Humanities Research (ASSEHR), 2018, Atlantis Press, pp. 302–309. DOI: https://doi.org/10.2991/icla-17.2018.52

[5] H. Retnawati, S. Hadi, A.C. Nugraha, Vocational high school teachers' difficulties in implementing the assessment in curriculum 2013 in Yogyakarta Province of Indonesia, International Journal of Instruction, 2016, vol. 9(1), pp. 33–48. DOI: https://doi.org/10.12973/iji.2016.914a

[6] W.J. Popham, Assessment literacy overlooked: A teacher educator's confession, Teacher Educator, 2011, vol. 46(4), pp. 265–273, DOI: https://doi.org/10.1080/08878730.2011.605048

[7] Y. Xu, G. Brown, University English Teacher Assessment Literacy: A survey-test report from China, Papers in Language Testing and Assessment, 2017, vol. 6(1), pp. 133–158.

[8] B.S. Plake, J.C. Impara, J.J. Fager, Assessment Competencies of Teachers: A National Survey, Educational Measurement: Issues and Practice, 1993, vol. 12(4), pp. 10–12. DOI: https://doi.org/10.1111/j.1745-3992.1993.tb00548.x

[9] C. Mertler, C. Campbell, Measuring Teachers' Knowledge & Application of Classroom Assessment Concepts: Development of the "Assessment Literacy Inventory", 2005.

[10] N. Tao, Development and Validation of Classroom Assessment Literacy Scales:English as a Foreign Language (EFL) Instructors in a Cambodian Higher Education Setting, Victoria University, 2014, DOI: https://doi.org/10.1038/132817a0

[11] K.C. Soh, L. Zhang, The Development and Validation of a Teacher Assessment Literacy Scale: A Trial Report, Journal of Linguistics and Language Teaching, 2017, vol. 8(1), pp. 91–116.

[12] Y. Xu, G. Brown, University English Teacher Assessment Literacy: A survey-test report from China, Papers in Language Testing and Assessment, 2017, vol. 6(1), pp. 133–158.

[13] E. Ölmezer-Öztürk, B. Aydin, Toward measuring language teachers' assessment knowledge: development and validation of Language

Assessment Knowledge Scale (LAKS), Language Testing in Asia, 2018, vol. 8(1). DOI: https://doi.org/10.1186/s40468-018-0075-2

[14] J.A. Balogun, C.E. Mbada, A.O. Balogun, U.A. Okafor, Development and evaluation of the readability, stability and internal consistency of a psychometric instrument designed to assess physiotherapists' knowledge and attributes of professionalism, International Journal of Medical and Health Sciences Research, 2017, vol. 4(6), pp. 102-117.

[15] M.A. Durand, J. Song, R.W. Yen, K. Sepucha, M.C. Politi, S. Dhage, G. Elwyn, Adapting the breast cancer surgery decision quality instrument for lower socioeconomic status: improving readability, acceptability, and relevance, MDM policy & practice, 2018, vol. 3(2).

[16] T. Newfields, Teacher development and assessment literacy, Authentic Communication: Proceedings of the 5th Annual JALT Pan-SIG Conference, 2006, pp. 48–73.

[17] S. O'Brien, Controlled language and readability, Translation and cognition, 2010, vol. 15, pp. 143-168.

[18] B.Y. Pelawi, Dampak teknik, metode, dan ideologi penerjemahan terhadap hasil terjemahan teks the gospel according to matthew dalam teks bahasa Indonesia, Unpublished Disertation University of Gadjah Mada, Yogyakarta, 2014.

[19] A. Acar, K.U. İŞİSAĞ, Readability and comprehensibility in translation using reading ease and grade indices, International Journal of Comparative Literature and Translation Studies, 2017, vol. 5(2), pp.47-53.

[20] M. Popović, Relations between comprehensibility and adequacy errors in machine translation output, In Proceedings of the 24th Conference on Computational Natural Language Learning, 2020, pp. 256-264.

[21] J.C. Richards, R. Schmidt. Longman Dictionary of Language Teaching and Applied Linguistics, New York, The Guilford Press, 2002

[22] B.B. Kadayat, E. Eika, Impact of sentence length on the readability of web for screen reader users, In International Conference on Human-Computer Interaction, 2020, pp. 261-271.

[23] J. Mikk, Sentence length for revealing the cognitive load reversal effect in text comprehension, Educ. Stud., 2008, vol. 34(2), pp. 119–127.

[24] S. Dept, A. Ferrari, B. Halleux, Translation and cultural appropriateness of survey material in large-scale assessments, Implementation of large-scale education assessments, 2017, pp. 168-192.

[25] S. Lin, M. Ji, Assessing linguistic comprehensibility of healthcare translation using the POCA model. In Multicultural Health Translation, Interpreting and Communication, Routledge, 2019, pp. 69-84.