LLT Journal, e-ISSN 2579-9533, p-ISSN 1410-7201, Vol. 26, No. 1, April 2023, pp. 244-254

LLT Journal: A Journal on Language and Language Teaching
http://e-journal.usd.ac.id/index.php/LLT
Sanata Dharma University, Yogyakarta, Indonesia

DEVELOPING ASSESSMENT INSTRUMENTS FOR ENGLISH LITERACY IN BLENDED LEARNING FOR JUNIOR HIGH SCHOOL STUDENTS

Eka Anastasia Wijaya[1*], Ni Made Ratminingsih[2], and I G A Lokita Purnamika Utami[3]
[1,2,3]Universitas Pendidikan Ganesha, Indonesia
ekaanastasia28@gmail.com[1], made.ratminingsih@undiksha.ac.id[2], and lokitapurnamika@undiksha.ac.id[3]
*correspondence: ekaanastasia28@gmail.com
https://doi.org/10.24071/llt.v26i1.5073
received 19 August 2022; accepted 29 March 2023

Abstract
This research intended to develop assessment instruments for English literacy in blended learning. The study employed survey, observation, interview, and document study as the methods with a questionnaire, observation checklist, interview guide, and school document as the instruments to collect the data framed within RnD design with the 4D model by Thiagarajan et al. The developed instruments were categorized as 'very high validity' and 'very good' for their content validity (1.00) and product quality (96% and 98%) by the two judges. It indicates that assessment of learning, assessment for learning, and assessment as learning combined with authentic materials are necessary to improve the learning process as well as the students' English literacy skills. It also implies that English teachers need to highlight the use of assessment for learning and assessment as learning since those assessments lack attention previously.

Keywords: assessment instruments, blended learning, English literacy

Introduction
Nowadays, students are required to have English literacy (Khairuddiniyah, 2017). Literacy means the skills of being literate represented by four basic English skills in the real context (Keefe & Copeland, 2011; Mulia, 2016; Rintaningrum, 2009). Meanwhile, English literacy is a set of English skills that could help students to apply their knowledge in real life.

However, the result of PISA (Argina et al., 2017; Ayunin, Mirizon, & Rosmalina, 2019; Dewantara et al., 2019; Rosana et al., 2020) and preliminary observation and interviews indicated that literacy in Indonesia still becomes a problem. The students were not prepared with the essential skills that could be seen from the lack of basic English skills. Moreover, in the context of Buleleng Regency, the students could not even read English announcements or instructions properly. They also lack tips on listening to English announcements in public places. It is also found that they also lack self-confidence in expressing their opinion.

Related to that facts, Argina et al. (2017) found that the learning evaluation or the assessment is considered one of the causes of Indonesian stagnant position in PISA. It is because the National Examination as the general assessment process in Indonesia only assesses the students' recognition knowledge instead of analysis. Nurdiana (2020) also found that there is a lack of English teachers' knowledge especially about assessing the students' language performance. Therefore, it affects the assessment process.

Moreover, the preliminary observation and interview done as part of the situational analysis indicated that English teachers in Bali have not focused on the use of assessment for learning and assessment as learning in the class. On the other hand, the teachers only emphasized the use of assessment of learning in the form of the multiple-choice test. In line with the COVID-19 pandemic, it required teachers to do the teaching and learning process in the form of blended online learning. Besides, little attention is given to the issue of developing assessment of learning, assessment for learning, and assessment as learning instruments to assess blended English literacy learning.

Based on the above review, the learners could improve their English literacy by using the appropriate assessment. In this case, it seems appropriate to develop those three kinds of assessment instruments for junior high school students to improve their English literacy learning. The focus of this study is limited to developing those three kinds of assessment instruments to assess English literacy in the context of blended learning for seventh-grade students in semester two as part of a larger research.

Based on the explanation above, the formulated research questions are: 1) What are the needs in developing assessment instruments to assess English literacy in blended learning for grade 7 semester 2 in Buleleng Regency? 2) How are the assessment instruments developed? 3) How is the quality of the developed assessment instruments?

Method

Research design

This research applied research and development (R&D) with the 4D model (Thiagarajan et al., 1974). The 4D model itself contains the Define, Design, Develop, and Disseminate stages. This model was used because the study was product oriented with qualitative and quantitative analysis. Besides, this model has systematic and simple stages that could help to solve the research problems in line with the student's needs and characteristics. However, the Disseminate stage is excluded in the context of this research and it will be continued in the following year due to time limitations as part of a larger research. In the Define stage, the needs analysis of assessment instruments to assess English literacy in blended learning was done. It was done to know the problems, especially the literacy program and the assessment. Further, the blueprint and assessment instruments were developed in the Design stage, followed by the next stage, namely the Develop stage. This stage developed the product in the form of assessment instruments.

Population and sample

The population was all of the junior high school students in Bali. Further, the sample was chosen using the multistage cluster random sampling technique. In this

sampling technique, the researcher chose the sample from the population using smaller groups at each stage, for example, based on geography (cities or states) (Singh & Mangat, 1996). This sampling technique was used to collect data from a large group to avoid time-consuming. The researcher divided and selected the population into clusters and selected some clusters in the first stage. The first stage, which was area or cluster sampling, resulted in twelve schools from the southern and northern parts of Bali. Therefore, six schools both from the southern part of Bali and the northern part of Bali were chosen through a multistage cluster random sampling technique. It was followed by using a simple random sampling through a lottery to choose two schools. In this case, SMP Negeri 1 Singaraja and SMP Negeri 6 Singaraja were chosen as the sample. Those schools were selected based on the sample criteria, which are public junior high schools in Buleleng Regency. Furthermore, as part of a larger research, seventh-grade students from those public junior high schools were chosen as the sample.

Data collection technique

The data collected in the Define stage were obtained through a survey method with a questionnaire, observation method with a checklist and rubric as the instruments, interview method with an interview guide as the instrument, and document study. Moreover, it is followed by the design stage, in which the researcher designed the blueprint and the product draft. Furthermore, the Develop stage was done, and it used expert judgment to find out the product quality.

Data analysis technique

The data were qualitative, especially for the first and the second research questions, and quantitative data, especially for the first and third research questions. In this case, the qualitative data were observation results, interview results, document analysis results, and questionnaire results. The data were analyzed qualitatively by interpreting the result of observation, interview, document analysis, and a questionnaire about the needs analysis of assessment instruments and the development process. Further, the data were interpreted based on the related theory to answer the first and the second research questions. Besides, the result of the questionnaire also perceived quantitative data, especially about the needs analysis. Further, the data about the quality of the developed assessment instruments were analyzed quantitatively by classifying the instruments' quality with certain criteria.

Findings and Discussion

Define stage

In this research, the first stage was done in developing the product called the Define stage, in which the needs analysis was done toward the assessments used by the teachers to assess English literacy in blended learning. It was done to know the problems, especially the literacy program and the assessment. Hariyadi and Yanti (2019) stated that the main purpose of needs analysis is to match the perceptions, beliefs, and expectations of the teachers or other stakeholders with the learners' needs. Moreover, the needs analysis process in this research and development used Thiagarajan's theory about the aspects of the defined stage and Brown's theory about the domain of analysis.

The needs were analysed through present-situation analysis. The result of the needs analysis reflected these two things 1) the necessities of developing an assessment of learning, assessment as learning, and assessment for learning; 2) the necessity of developing assessment instruments that included authentic materials. Therefore, assessment instruments that cover real-life tasks and procedures were developed in this research by using the needs analyzed before as the basis for designing the products.

In line with the necessity of developing those three assessments, it could be assumed that those three assessments are important in the learning process. The necessities themselves come from the nature of those three assessments that could not be separated from the principle of conducting blended learning assessment. It is supported by Yuliyana, Rochmiyati, and Maulina (2021), in which one of the principles for conducting blended learning assessment is including self-reflection as the assessment to reflect learner-centered. Besides, it is important to ask the students' opinions as input on how to assess blended learning. This principle could be seen in assessment as learning.

Moreover, the other principle is how the assessment could encourage students to develop skills and provide feedback from the teachers (Yuliyana, Rochmiyati, & Maulina, 2021). This principle is following the characteristic of assessment for learning. Concerning the assessment of learning, the assessment has to be designed in line with the learning objectives. It is supported by Yuliyana, Rochmiyati, and Maulina (2021), in which the other principle for conducting blended learning assessment is the assessment used must be aligned with the learning objectives.

Besides, it is also important to consider the necessity of developing assessment instruments that include authentic materials. Fithriyah (2015) stated that authentic materials are needed in learning English to help the students to have direct experience with the target language in real situations as much as possible. Therefore, authentic materials could facilitate the learners in learning the target language. Sianipar, Ansari, and Eviyanti (2018) have the same opinion, in which authentic assessment obliges the students to show real-world performance to show their knowledge and skill. It aligns with the developed assessment instruments in this research, which emphasize literacy or real-world task.

In addition, authentic materials could be applied to beginner to advanced learners (Fithriyah, 2015). Therefore, developing assessments with authentic materials for junior high school students could help the students in learning English. It is also said that authentic assessment could be the alternative to avoid assessment that only emphasizes comprehension tests (Surya & Aman, 2016). The developed assessment instruments in this research consider using authentic materials to avoid assessments that only assess the student's comprehension. In line with that, Tosuncuoglu (2018) supported that authenticity is one of the principles of assessment. The authenticity could be presented by including contextualized components and real-world assignments. It is in line with the purpose of this study, in which this study developed assessment instruments that included authentic materials.

Design stage

The second stage in developing the product is called as the Design stage. In this stage, the blueprint for developing the assessment instruments was designed.

The blueprint used Bloom's taxonomy as the guideline to choose the action verbs in proposing the indicators. Octoria et al. (2016) stated that assessment is done to determine the student's learning outcome which could be classified into three domains proposed by Bloom called Bloom's taxonomy. Those three domains of learning are cognitive, affective, and psychomotor domains. In line with that, it is said that the assessment in *Kurikulum 2013* should consider authentic assessment that includes affective, cognitive, and psychomotor (Hutapea, Lubis, & Adisaputera, 2020). The cognitive domain is related to how students acquire, proceed, and use knowledge. Meanwhile, the affective domain refers to students' behavior and attitude during the learning process. The psychomotor domain refers to student's skills during the learning process.

Bloom's taxonomy includes six levels ordered from the simplest level which is a requirement for the most complex level (Bloom et al., 1956). In addition, the six levels in Bloom's taxonomy are considered the levels of skills required in classroom situations (Prasad, 2021). Orey (2010) added that the domain in Bloom's taxonomy is clustered into six subsequent thinking levels. The first three levels belong to the lower-order thinking skills (LOTS) consisting of remembering, understanding, and applying. Meanwhile, the next three levels belong to the higher-order thinking skills (HOTS), consisting of analyzing, evaluating, and creating. Koksal and Ulum (2018) have the same opinion, in which Bloom's taxonomy is hierarchical. Therefore, it is possible to see the lower level at the higher level.

As part of a larger research, the researcher was instructed to design the blueprint for creating assessment instruments to assess English literacy for seventh-grade junior high school students, especially in the second semester. There were four chapters for grade seven in the second semester. The blueprint for developing the assessment instruments itself consists of basic competency, indicator, topic, type of assessment, and mode of learning. In this case, the researcher used the basic competency given in the syllabus to design the indicators. The researcher proposed a minimum of two indicators for each basic competency. Moreover, the researcher adjusted the type of assessment used with the indicators to maximize the learning process. Furthermore, the researcher also added the mode of learning for each assessment instrument developed. The assessment instruments could be applied either offline or in online learning depending on the situation. The researcher added a barcode to ease the user in using the developed assessment instruments for online learning. Besides, the user could also use the developed assessment instruments in the form of a printed book for offline learning. The researcher could use any suitable online platform such as Quizizz, Schoology, Edmodo, Google Classroom, Google Meet, Zoom, etc., for online learning depending on the needs and situation.

The researcher only focused on creating assessment instruments for the cognitive and psychomotor domains based on the basic indicator given in the syllabus. Therefore, the indicators created in the blueprint follow the hierarchy of cognitive and psychomotor domains. The cognitive domain is classified into six hierarchies, called remember (C1), understand (C2), apply (C3), analyze (C4), evaluate (C5), and create (C6) (Ahmad & Hussin, 2017; Nurdiana, 2018; Octoria et al., 2016; Oktaviana & Susiaty, 2020). As the same with the cognitive domain, the psychomotor domain is classified into five hierarchies, called imitation (P1), manipulation (P2), precision (P3), articulation (P4), and naturalization (P5). Based on the syllabus, the basic competency in Chapter 5 until Chapter 8 is varied from

C2 (interpret), C3 (identify), and C4 (compare) for cognitive domains. Meanwhile, the basic competency in Chapter 5 until Chapter 8 belongs to P3 (capture meaning and compose). Therefore, the researcher designed the indicators at the same level or above the level with the basic competency. It is supported by Krashen's theory (1982), in which the transfer knowledge needs to be above the student's current knowledge (i + 1).

After designing the blueprint, the researcher also designed the draft of the assessment instruments. It is important to design the draft of the assessment instruments to align with the required competency. Octoria et al. (2016) have the same opinion, in which the assessment should use techniques that are suitable to the competencies. The techniques could be a test, observation, homework, exercise, paper, project, performance, portfolio, oral questions, etc.

Develop stage

The last stage in developing the product is called the Develop stage. In this stage, the draft of the product was established into the final product by considering the blueprint designed before. In this research, the researcher developed the assessment instruments for grade seven in semester two that consist of four chapters. The product was created using Microsoft Word software by the researcher. The researcher also made the assessment instruments as creative as possible by including colorful backgrounds and topic-related images. Furthermore, the assessment instruments were designed by following the topics and the characteristics of a good assessment. Therefore, the developed assessment instruments could assess the student's English literacy skills. After developing the product, the product was judged by two expert judges. The preview of the final product could be seen in the following figures and the final product could be scanned in the barcode given.



Figure 1. Cover of book chapter



Figure 2. Example of AoL



Figure 3. Barcode of the final product

The researcher chose self-assessment as well as peer assessment as a type of assessment of learning used, in which assessment as learning includes the students dynamically in the assessment process (Lam, 2018). Besides, self-assessment allows the students to observe their learning, achievement, and progress (Mutch, 2012). Further, peer assessment could assist the students to monitor their learning from the perspective of their peers in the form of feedback. Dann (2014) added that the feedback given in the assessment as learning could support the students' learning. The other students, as a peer, could give feedback on assessment as learning. The feedback from a peer could ease the students to monitor their learning since it is given by the same age.

Further, the researcher chose performance assessment in the form of reading, writing, speaking, and listening tests as the types of assessment for learning used. The designed assessment for learning is expected to provide feedback for the students (Sardareh & Saad, 2013). Oyinloye and Imenda (2019) supported that assessment for learning is considered a strategy to improve instruction and learning since the nature of assessment for learning is continuous. Padmadewi and Artini (2019) said that assessment for learning could help teachers deal with assessing the student's progress. The developed assessment for learning instruments was designed to help the student's progress. Unconsciously, it also helps the teachers to monitor and assess the student's progress.

For the assessment of learning, the researcher designed a summative test alongside another chapter to be given at the end of the semester (final test). Mutch (2012) supported that assessment of learning assessment summarizes and outlines the students' achievement at a period. This assessment will be given at the end of the learning process at a certain unit (Lee & Mak, 2012).

In conclusion, the type of assessment of learning used is a summative test consisting of a complete test and a short-answer test conducted at the end of the semester. In this case, assessment of learning could be used to review and report the students' achievement at a period (Lee & Mak, 2012; Mutch, 2012). Moreover, the types of assessment for learning used were reading, writing, speaking, and listening test. Lam (2018), Mutch (2012), Sardareh and Saad (2013) stated that assessment for learning could be used to improve the learning process through feedback given. Further, a study by Oyinloye and Imenda (2019) revealed that the assessment for learning approach significantly affects the learners' performance compared to learners with normal instruction. Meanwhile, the types of assessment as learning used were self-assessment and peer assessment. In this case, assessment as learning deals with the students' thinking, feelings, actions, and motivation related to the learning process (Lam, 2018).

Content validity and product quality

After the product was developed, it is important to check the quality. The data could be seen in the results of the expert judge's form. The researcher used Gregory's formula to find out the criteria for content validity (Candiasa, 2010).

Table 1. Gregory's formula

		Expert 1	
		Irrelevant	Relevant
Expert 2	Irrelevant	A = 0 items	B = 0 items
	Relevant	C = 0 items	D = 10 items

Content validity =

$$\frac{D}{A+B+C+D} = \frac{10}{0+0+0+10} = 1.00\ (very\ high\ validity)$$

The formula above indicated that the content of the developed product was qualified as 'very high validity' with a score of 1,00.

Meanwhile, the researcher used the criteria and the formula below to classify the product quality.

The Formula (Expert 1)

$P = \frac{\sum x}{IMS}$ x 100%

$P = \frac{48}{50}$ x 100% = 96% (very good)

The Formula (Expert 2)

$P = \frac{\sum x}{IMS}$ x 100%

$P = \frac{49}{50}$ x 100% = 98% (very good)

Note:

P = percentage of product quality
∑x = total score from an expert
IMS = Ideal Maximum Score

The percentage and the formula above indicated that the quality of the developed product was qualified as 'very good' by the two judges with percentages of 96% and 98%.

In developing a product, it is important to check its quality. Therefore, the researcher also developed the instrument to check the product's quality. In this case, there were two judges as the expert on the content and the quality of the product. The results of the content validity and product quality indicated that the developed product was in the 'very high validity' and 'very good' criteria. For the result of content validity, Judges 1 and Judges 2 gave scores of 94% and 96% that classified into 'very high validity' within a range of 81% - 100% belongs to 'very high validity'. Meanwhile, Judges 1 and Judges 2 gave scores of 96% and 98% for the product quality that classified as 'very good' within a range of 81% - 100% belongs to 'very good'.

In line with the results and discussion above, there are implications as the logical consequences from the conclusion of the research. This study could give impacts by applying the developed assessment instruments in the classroom, both in the context of offline and online learning. In other words, the developed assessment instruments integrated the use of technology that is suitable for today's conditions. The developed assessment instruments consisted of summative tests as part of the assessment of learning, performance tests as part of the assessment for learning, also self-assessment and peer assessments as part of the assessment of learning. In this case, this study emphasizes assessment for learning and assessment as learning that is neglected by most English teachers. Therefore, English teachers could emphasize the use of assessment for learning to improve their teaching. Besides, the students could also improve their learning through assessment as learning. In conclusion, English teachers could use this study as a suggestion to develop blended assessment instruments and to give emphasis to those three assessments that could improve the teaching and learning process.

Conclusion

This study developed assessment instruments (assessment of learning, assessment for learning, and assessment as learning) integrated with authentic materials. The process of developing the product used the 4D model by Thiagarajan et al. with a limitation, which is up to the third stage, the Develop stage. The last stage, which is the Disseminate stage, will be continued in the following year as part of a larger research. Therefore, the researcher intends to give suggestions for the further researcher to consider the weakness as the basis for further research. Besides, further research needs to continue this study to the Disseminate stage. Moreover, for English teachers, it is suggested that English teachers should implement those three assessments in the teaching and learning process by considering their importance. The teachers could use this study as a guideline to develop their blended assessment instruments. Besides, it is suggested to emphasize the use of assessment for learning and assessment as learning to help the teachers and the students to improve the teaching and learning process from the assessment given.

References

Ahmad, T. S. A. S., & Hussin, A. A. (2017). Application of the Bloom's taxonomy in online instructional games. *International Journal of Academic Research in Business and Social Sciences, 7*(4), 1009-1020. http://dx.doi.org/10.6007/IJARBSS/v7-i4/2910

Argina, A. W., Mitra, D., Ijabah, N., & Setiawan, R. (2017). Indonesian PISA result: What factors and what should be fixed? *The 1st Education and Language International Conference Proceedings Center for International Language Development of Unissula*, 69-79.

Ayunin, Q., Mirizon, S., & Rosmalina, I. (2019). PISA reading literacy performance and its correlation with engagement in reading activity and reading interest. *International Seminar and Annual Meeting BKS-PTN Wilayah Barat, 1*(1), 573-585.

Bloom, B. S. (1956). *Taxonomy of educational objectives handbook: Cognitive domain*. New York: McKay.

Candiasa, I. M. (2010). *Statistik univariat dan bivariat disertai aplikasi SPSS*. Singaraja: Universitas Pendidikan Ganesha.

Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory, policy and practice. *Assessment in Education: Principles, Policy & Practice, 21*(2), 149-166.

Dewantara, I. P. M., Suandi, I. N., Rasna, I. W., & Putrayasa, I. B. (2019). Cultivating students' interest and positive attitudes towards Indonesian language through phenomenon-text-based information literacy learning. *International Journal of Instruction*, *12*(2), 147-162.

Fithriyah, F. (2015). The importance of authentic materials in developing appropriate and effective listening skills. *PIONIR: Jurnal Pendidikan, 4*(2), 1-15. http://dx.doi.org/10.22373/pjp.v4i2.180

Hariyadi, A., & Yanti, D. R. (2019). The importance of needs analysis in materials development. *Jurnal Ilmiah Profesi Pendidikan, 4*(2), 94-99. https://doi.org/10.29303/jipp.v4i2.88

Hutapea, L. S., Lubis, M., & Adisaputera, A. (2020). The development of authentic assessment instruments on anecdotal text material for class X students of SMK Yapim Indrapura, Batu Bara Regency. *Budapest International Research and Critics in Linguistics and Education Journal, 3*(4), 1826-1814. https://doi.org/10.33258/birle.v3i4.1379

Keefe, E. B., & Copeland, S. R. (2011). What is literacy? The power of a definition. *Research & Practice for Persons with Severe Disabilities, 36*(3-4), 92-99. https://doi.org/10.2511/027494811800824507

Khairuddiniyah, K. (2017). The implementation of literacy strategies in teaching English by the English teacher at MAN Insan Cindekia Padang Pariaman. *English Education: Jurnal Tadris Bahasa Inggris*, *10*(2), 242-257.

Koksal, D., & Ulum, O. G. (2018). Language assessment through Bloom's taxonomy. *Journal of Language and Linguistic Studies, 14*(2), 76-88.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon Press Inc.

Lam, R. (2018). Understanding assessment as learning in writing classrooms: The case of portfolio assessment. *Iranian Journal of Language Teaching Research, 6*(3), 19-36.

Lee, I., & Mak, P. (2014). Assessment as learning in the language classroom. *Assessment and Learning, 3,* 66-78.

Mulia, V. L. C. (2016). The contributions of literacy skills to national development. *Leksema, 1*(2), 111-118. https://doi.org/10.22515/ljbs.v1i2.101

Mutch, C. (2012). Assessment for, of and as learning: Developing a sustainable assessment culture in New Zealand schools. *Policy Futures in Education, 10*(4), 374-385. https://doi.org/10.2304/pfie.2012.10.4.374

Nurdiana, N. (2018). An evaluation of revised Bloom's taxonomy on a general English test for non-English major students. *Journal of English Teaching as a Foreign Language*, *4*(1), 33-46.

Nurdiana, N. (2020). Language teacher assessment literacy: A current review. *Journal of English Language and Culture, 11*(1), 66-74. http://dx.doi.org/10.30813/jelc.v11i1.2291

Octoria, D., Sudiyanto, S., Witurachmi, S., & Wardani, D. K. (2016). The implementation of Bloom's taxonomy in the assessment instrument for economics learning to improve the students' cognitive competencies. *Proceeding The 2nd International Conference on Teacher Training and Education, 2*(1), 655-665.

Oktaviana, D., & Susiaty, U. D. (2020). Development of test instruments based on revision of Bloom's taxonomy to measure the students' higher order thinking skills. *Jurnal Ilmiah Pendidikan Matematika, 9*(1), 21-29. http://doi.org/10.25273/jipm.v9i1.5638

Orey, M. (2010). *Emerging perspectives on learning, teaching, and technology*. Switzerland: Jacobs Foundation.

Oyinloye, O. M., & Imenda, S. N. (2019). The impact of assessment for learning on learner performance in life science. *EURASIA: Journal of Mathematics, Science and Technology Education, 15*(11), 1-8. https://doi.org/10.29333/ejmste/108689

Padmadewi, N. N., & Artini, L. P. (2019). Assessment instruments for improving English teaching skills through microteaching in Indonesia. *Asian EFL Journal Research, 21*(2), 49-77.

Prasad, G. N. R. (2021). Evaluating student performance based on Bloom's taxonomy levels. *Journal of Physics: Conference Series,* 1-10. https://doi.org/10.1088/1742-6596/1797/1/012063

Rintaningrum, R. (2009). Literacy: Its importance and changes concept in the concept and definition. *TEFLIN Journal, 20*(1), 1-7.

Rosana, D., Widodo, E., Setyaningsih, W., & Warno, D. S. (2020). Developing assessment instruments of PISA model to measure students' problem-solving skills and scientific literacy in junior high schools. *Indonesian Journal of Science Education, 8*(2), 292-305. https://doi.org/10.24815/jpsi.v8i2.17468

Sardareh, S. A., & Saad, M. R. M. (2013). Defining assessment for learning: A proposed definition from a sociocultural perspective. *Life Science Journal, 10*(2), 2493-2497.

Sianipar, V. M. B., Ansari, K., & Eviyanti, E. (2018). Development of authentic assessment instrument based on curriculum 2013 on material of poetry by students class X SMA Negeri 2 North Rantau academic year 2017/2018. *International Journal of Education, Learning and Development, 6*(8), 79-90.

Singh, R., & Mangat, N. S. (1996). *Elements of survey sampling*. Dordrecht: Springer.

Surya, A., & Aman, A. (2016). Developing formative authentic assessment instruments based on learning trajectory for elementary school. *Research and Evaluation in Education, 2*(1), 13-24. https://doi.org/10.21831/reid.v2i1.6540

Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. Indiana: Indiana University.

Tosuncuoglu, I. (2018). Importance of assessment in ELT. *Journal of Education and Training Studies, 6*(9), 163-167.

Yuliyana, M., Rochmiyati, R., & Maulina, D. (2021). Blended learning assessment instrument for elementary school. *Edunesia: Jurnal Ilmiah Pendidikan, 2*(3), 668-676. https://doi.org/10.51276/edu.v2i3.189