International Conference on Islam, Law, and Society (INCOILS)
Conference Proceedings 2024

Developing an Engaging English Summative Test with Educaplay: A Game-Based Solution for the Fourth Grade of Islamic Primary School

Muhammad Firnas Hibatulloh[1], Sri Wahyuni[2]
[1,2]Department of English Language Education, School of Graduate Program, State Islamic Institute (IAIN) of Kediri
[1]firnasmuhammad99612@gmail.com, [2]swy090984@gmail.com

ABSTRACT:
The purpose of this study is to develop an interactive English summative test for fourth-grade Islamic primary school students using Educaplay to enhance student engagement and motivation. This study followed a modified seven-step R&D cycle, including information collection, expert validation, product revision, and field testing. The test consisted of 25 multiple-choice questions and was administered to 28 for field testing. Statistical analysis using SPSS 25 showed that 12 questions were valid, the test had a reliability coefficient of 0.765, and the difficulty level of questions was well-balanced. However, 40% of the questions required significant review due to poor discrimination scores. The study highlighted Educaplay's potential to create engaging and effective assessments. The findings support the ongoing use and refinement of digital platforms like Educaplay in educational assessment.

Key words: *English summative test, Educaplay, Islamic primary school*

ABSTRAK:
Tujuan dari penelitian ini adalah untuk mengembangkan tes sumatif bahasa Inggris interaktif untuk siswa sekolah dasar Islam kelas empat dengan menggunakan Educaplay untuk meningkatkan keterlibatan dan motivasi siswa. Penelitian ini mengikuti siklus R&D tujuh langkah yang dimodifikasi, termasuk pengumpulan informasi, validasi ahli, revisi produk, dan pengujian lapangan. Tes terdiri dari 25 pertanyaan pilihan ganda dan diberikan kepada 28 siswa untuk pengujian lapangan. Analisis statistik menggunakan SPSS 25 menunjukkan bahwa 12 soal valid, tes memiliki koefisien reliabilitas sebesar 0,765, dan tingkat kesulitan soal seimbang. Namun, 40% dari soal-soal tersebut memerlukan peninjauan ulang yang signifikan karena nilai diskriminasi yang buruk. Studi ini menyoroti potensi Educaplay untuk menciptakan penilaian yang menarik dan efektif. Temuan ini mendukung penggunaan dan penyempurnaan platform digital yang sedang berlangsung seperti Educaplay dalam penilaian pendidikan.

Kata kunci: *Tes sumatif bahasa Inggris, Educaplay, Sekolah dasar Islam*

INTRODUCTION

In the era of digitalization, the use of technology in education has become increasingly popular, particularly in making learning more interactive and engaging for students. This is especially important for primary school students, who often lack motivation regarding traditional testing methods. Over the years, scholars have emphasized the significance of fostering an engaging learning environment, especially in subjects such as grammar, which are foundational to

language acquisition and comprehension[1]. However, conventional assessment methods, such as paper-based tests, frequently encounter challenges sustaining students' attention and motivation, particularly in settings characterised by diverse cultural and linguistic backgrounds[2]. To address this issue, this article proposes using Educaplay as an alternative platform for creating interactive English summative test.

Educaplay is an online tool or platform that permits teachers to design free educational games or aids with creative and professional results. The site is free to manage and removes the need to apply different software[3]. Educaplay shows positive results in improving students' academic performance with the use of gamification resources that motivate their active participation in collaborative work, and the teacher is the one who generates his resources based on the needs of the students[4]. Teachers register once to create free educational games of different types on Educaplay. Then, they share the games with their students and receive the results. It has 18 different game formats that teachers can apply, from multiple choice quiz, dictation game, word search puzzle, and matching game to Froggy Jump[5].

The platform has been used for teaching various subjects, such as Science in Malaysian primary school[6] and Arabic in a private university in Indonesia[7]. It is also used in teaching English as a Foreign Language and can be used to test students' knowledge of vocabulary and grammar[8]. Similarly, they use Educaplay as a tool for assessing the students.

In this context, the emergence of digital platforms tailored for educational purposes presents a promising way to address the challenges of sustaining students' attention and

---

[1] Zoltán Dörnyei and Ema Ushioda, *Teaching and Researching Motivation*, 2nd edn (Harlow: Pearson Education Limited, 2011); Rod Ellis, 'The Study of Second Language Acquisition' (Oxford: Oxford University Press, 2008).

[2] Penny McKay, *Assessing Young Language Learners* (Cambridge: Cambridge University Press, 2006) <https://doi.org/10.1093/elt/ccn063>; Brian K. Lynch, *Language Assessment and Programme Evaluation* (Edinburgh: Edinburgh University Press, 2003) <https://doi.org/10.1016/j.esp.2003.11.003>.

[3] By Vânia Graça, Paula Quadro-flores, and Altina Ramos, 'The Integration of the Digital Platform Educaplay in Interdisciplinary Paths in the 1 St and 2 Nd Basic Education Cycles', *Athens Journal of Education*, 9.3 (2022), 377–91.

[4] Cristina Páez-quinde and Ruth Infante-paredes, 'Educaplay : A Gamification Tool for Academic Performance in Virtual Education during the Pandemic Covid-19', *Revista Cátedra*, 5.1 (2022), 31–44.

[5] Sri Surachmi and Karl Jahniel S Sison, 'Educaplay as Teaching Media Inn Virtual Classes', *The 3rd Bogor English Student and Teacher (BEST) CONFERENCE*, 3 (2021), 1–6.

[6] Mufidah Firzanah Mihat, 'Using Educaplay Apps to Increase Year 4 Pupils ' Interest and Understanding towards Topic Properties of Light', *Journal of Science and Mathematics Education in Southeast Asia*, 46 (2024), 159–72.

[7] Miatin Rachmawati and Fitri Liza, 'Arabic Speaking Skill Using the Educaplay "Froggy Jumps" Application', *Alsuniyat: Jurnal Penelitian Bahasa Sastra Dan Budaya Arab*, 7.1 (2024), 223–34.

[8] L. Castillo-Cuesta, 'Using Digital Games for Enhancing EFL Grammar and Vocabulary in Higher Education', *International Journal of Emerging Technologies in Learning*, 15.20 (2020), 116–29 <https://doi.org/10.3991/ijet.v15i20.16159>.

motivation. Educaplay has gained attention among these platforms for its interactive and gamified approach to learning[9]. While alternatives like Quizizz have gained popularity in various educational settings, Educaplay offers unique features that may be particularly beneficial for engaging Islamic primary school students in Indonesia, who may exhibit limited interest in traditional assessment methods.

Despite offering some unique features compared to other platforms, the use of Educaplay for developing interactive tests has not yet been very massive[10]. Very few studies have addressed the effectiveness of Educaplay in language learning. A mixed-method study of 68 students at a university in Ecuador, found a positive result during his 5 months of use of Educaplay[11]. According to the findings, Educaplay helps improve learners´ grammar in some aspects concerning the use of modals, gerunds, and infinitives. Moreover, students improve their vocabulary knowledge, especially in topics related to jobs and education. In Thailand, Muntrikaeo and Poonpon integrate Educaplay in task-based language teaching for secondary school students[12]. They consider it as the most suitable web-enhanced tool where teachers can create fun and interactive activities and share their learning activities with their students. This study uses matching, fill-in-the-blanks, and dialogue games as preview activities to present the lesson's topic. These activities are designed to assist the students in identifying the subject area and enabling awareness of the context. Helpful vocabulary, words, phrases, and structures of each unit are covered and designed in these three games.

Considering Educaplay's promising features to enhance learning engagement and a few research studies on that platform, this study aims to develop an interactive English summative test for the fourth grade of Islamic primary school by using Educaplay as the tool. The rationale behind using Educaplay is that it is more engaging and effective in capturing students' attention compared to conventional testing methods.

METHODS

Since this study aims to develop an appropriate and engaging grammar test for the fourth grade of Islamic primary school, Research and Development (R&D) design is applied. This design

---

[9] Sujatha Menon, 'First Steps in Online Gamification: Tips for the Language Classroom', *Gedrag & Organisatie Review*, 34.4 (2021), 227–41 <https://doi.org/10.37896/GOR34.04/026>.

[10] Ida Dian Sukmawati and Tri Pujiani, 'Enhancing Classroom Engagement Using Online Games', *Lingua: Jurnal Pendidikan Bahasa*, 19.1 (2023), 75–92 <https://doi.org/10.34005/lingua.v19i1.2616>.

[11] Castillo-Cuesta.

[12] Kuntida Muntrikaeo and Kornwipa Poonpon, 'The Effects of Task-Based Instruction Using Online Language Games in a Flipped Learning Environment (TGF) on English Oral Communication Ability of Thai Secondary Students', *English Language Teaching (ELT)*, 15.3 (2022), 9–21 <https://doi.org/10.5539/elt.v15n3p9>.

is the most appropriate because it naturally aims to fabricate products, not for formulating or testing theory in basic research. According to Borg and Gall, there are ten steps in an R&D cycle; those are (1) research and information collection, (2) planning, (3) preliminary form of product, (4) preliminary field testing, (5) primary product revision, (6) main field-testing, (7) operational product revision, (8) operational field testing, (9) final product revision, and (10) dissemination and implementation[13].

The ten steps above are adapted to match the research objectives, problems found, and time constraints in the study. Accordingly, the steps implemented are research and information collection, preliminary product development, expert validation, product revision, field testing or try-out, revision, and the final product.

The first step of the cycle is research and information collection. In this step, the researcher collects the information related to classroom habits by conducting observation and the needs of teacher and students by conducting interviews. Furthermore, additional information from theories and previous studies is also gained by conducting a literature review. After that, the researcher develops the test based on the information gained. The test is a multiple choice, consisting of 25 items. The topic of the test is "Day and Month". The basic competence of the material gained from the teacher is used here to develop each test item.

After the preliminary test, an expert validates it, that is, the teacher itself. This validation is the third step of the cycle and is carried out to check whether the test completes the criteria of a good test. The instrument used is a validation sheet consisting of two items: content validity and question language and writing.

The fourth step after the expert validation is product revision. The researcher revises the test based on the teacher's validation result. The test that has been revised is then entered into Educaplay. After that, the researcher conducts field testing by trying the test out to the students. The goal is to determine whether or not students benefit from the revised test and whether or not it truly aids in their comprehension of the topic.

After registering for the test, there is a second revision. This is carried out when the information gathered from the revision test attempts continues to demonstrate the challenges or roadblocks students encounter while attempting to meet the comprehension objectives for every unit test item. This also includes the analysis of the test item difficulty, item discrimination, validity, and reliability. The item which does not meet the criteria is replaced or revised again. Finally, the

[13] Meredith D. Gall, Joyce P. Gall, and Walter R. Borg, *Educational Research: An Introduction* (Boston: Pearson Education Inc., 2003).

final product of the test is done and ready to be published for conducting a summative test on the topic "Day and Month" for the fourth grade of Islamic primary school.

RESULT

This section presents the findings from the preliminary study, expert validation, and the testing result. First, related to the findings of the preliminary study, which was done through classroom observation and interviewing the teacher and some students, it was found that the teacher mostly gave the students paper-based tests. Students say this is boring and makes them less enthusiastic about the test. According to the teacher, the tests are taken from the student worksheet (LKS) or the Internet without appropriate adaptation in terms of language and its suitability to the existing basic competence. This happens because of the time constraint had by the teacher. As a result, students have difficulty comprehending and doing the test appropriately.

Concerning the findings in the preliminary study, the researcher develops an engaging English summative test in the form of multiple-choice questions using Educaplay. It consists of 25 question items with four answer choices (A, B, C, D) for each question. After the first draft of the test is finished, the researcher asks the teacher to validate it. Related to the content validity, 20 question items are valid, while the rest (5 question items) are valid enough. Meanwhile, about the question language and writing, 15 question items are very clear, 5 are clear enough, and 5 are less clear. Accordingly, the teacher recommends that the test is appropriate to use with some revisions regarding the 5 question items, which have less clear language and writing.

After the test was revised based on the teacher's recommendation, the test was tried on the fourth-grade students of an Islamic primary school in Kediri, East Java. Twenty-eight students were involved as the test takers. The tryout was conducted to find out about the student's understanding and knowledge based on the topics of the developed test. It was conducted to customize tests that are suitable for students. There was a problem at the beginning of the tryout since some students had problems logging in themselves to the test. However, that problem could be solved well by giving them private assistance.

The tryout result was then calculated using SPSS 25 to check its validity and reliability. To check the validity of the test, the researcher compares the r-table and r-value. If the r-value $\geq$ r-table, the item is declared valid; if the r-value $\leq$ r-table, the item is declared invalid. The r-table is 0,374. From the comparison, it is found that there are 12 valid questions and 13 invalid questions. The valid questions are numbers 5, 6, 8, 9, 11, 12, 15, 16, 17, 18, 20, and 25. Meanwhile, the invalid questions are number 1, 2, 3, 4, 7, 10, 13, 14, 19, 21, 22, and 23. The detail of the r-value can be seen in the table 1 below.

Table 1. The Score of Validity Test

Items	r-value	r-table	Validity
1	0.201	0.374	Invalid
2	0.209	0.374	Invalid
3	0.346	0.374	Invalid
4	0.249	0.374	Invalid
5	0.578	0.374	Valid
6	0.656	0.374	Valid
7	0.012	0.374	Invalid
8	0.423	0.374	Valid
9	0.547	0.374	Valid
10	0.358	0.374	Invalid
11	0.541	0.374	Valid
12	0.601	0.374	Valid
13	0.364	0.374	Invalid
14	0.155	0.374	Invalid
15	0.552	0.374	Valid
16	0.388	0.374	Valid
17	0.541	0.374	Valid
18	0.433	0.374	Valid
19	0.036	0.374	Invalid
20	0.449	0.374	Valid
21	0.248	0.374	Invalid
22	0.222	0.374	Invalid
23	0.020	0.374	Invalid
24	0.283	0.374	Invalid
25	0.477	0.374	Valid

Furthermore, to check the reliability of the test, there are some criteria that must be followed. According to Sudjiono, if the reliability coefficient ($r11$) is ≥ 0.70, then the test is reliable[14]. Meanwhile, if $r11 < 0.70$, then the test is categorized as unreliable. Additionally, if $0.60 > r11 \leq 0.80$, then the test is categorized as the test is categorized as having high reliability[15].

Table 2. The Result of the Reliability Test

Reliability Statistics	
Cronbach's Alpha	N of Items
.765	25

Based on the table above, the alpha coefficient is 0.765. The correlation result means the test is reliable because the alpha is ≥ 0.70. Moreover, the test is categorized as having high reliability because the alpha coefficient is fulfilling the criteria stated by Arikunto, that is $0.60 > r11 \leq 0.80$[16]. Not only checking the validity and reliability of the test, the researcher also finds out the score of

[14] Sudjiono, *Pengantar Evaluasi Pendidikan* (Jakarta: Raja Grafindo Persada, 2011).
[15] Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan*, 2nd edn (Jakarta: Bumi Aksara).
[16] Arikunto.

item difficulty and item discrimination using Microsoft Excel. The item difficulty follows the criteria below:

Table 3. The Criteria of Item Difficulty

Category	
0.00-0.30	Difficult
0.31-0.70	Medium
0.71-1.00	Easy

(Adopted from Gronlund, 1993)

The calculation of the item difficulty score shows that there are 7 questions categorized as Easy (numbers 1, 2, 7, 8, 9, 12, and 14), 11 questions as Medium (number 3, 5, 6, 10, 11, 15, 16, 17, 20, 21, and 25), and 7 questions as Difficult (number 4, 13, 18, 19, 22, 23, and 24). The detailed score of item difficulty can be seen in Table 4 below.

Table 4. Item Difficulty Score

Items	**Mean**	**Interpretation**
1	0.75	Easy
2	0.86	Easy
3	0.61	Medium
4	0.07	Difficult
5	0.54	Medium
6	0.32	Medium
7	0.82	Easy
8	0.79	Easy
9	0.89	Easy
10	0.46	Medium
11	0.54	Medium
12	0.86	Easy
13	0.29	Difficult
14	0.82	Easy
15	0.39	Medium
16	0.50	Medium
17	0.46	Medium
18	0.25	Difficult
19	0.07	Difficult
20	0.54	Medium
21	0.64	Medium
22	0.11	Difficult
23	0.29	Difficult
24	0.29	Difficult
25	0.36	Medium

The last aspect to be considered is the score of item discrimination. It follows the criteria as presented in the table below:

Table 5. The Criteria of Item Discrimination

Range	Grade	Recommendations
D ≥ 0.40	Excellent	Preserve
0.30 ≤ D ≤ 0.39	Good	Possibilities for enhancement
0.20 ≤ D ≤ 0.29	Average	Need to verify/review
0.00 ≤ D ≤ 0.19	Poor	Reject or review in-depth
D ≤-0.01	Worst	Remove

(Adopted from Crocker & Algina, 2008)

The calculation of the item discrimination score shows that there is no question categorized as worst. However, 10 questions categorized as Poor (numbers 1, 4, 7, 13, 14, 18, 19, 22, 23, and 24) should be rejected or reviewed deeply. Moreover, 6 questions are Average (numbers 2, 3, 8, 9, 10, and 11), 4 questions are Good (number 6, 12, 16, and 21), and 5 questions as Excellent (number 5, 15, 17, 20, and 25). The detailed score of item difficulty can be seen in Table 6 below.

Table 6. Item Discrimination Score

Items	Score	Interpretation
1	0.04	Poor
2	0.21	Average
3	0.25	Average
4	0.12	Poor
5	0.43	Excellent
6	0.38	Good
7	0.01	Poor
8	0.25	Average
9	0.27	Average
10	0.32	Average
11	0.28	Average
12	0.36	Good
13	0.17	Poor
14	0.01	Poor
15	0.65	Excellent
16	0.37	Good
17	0.47	Excellent
18	0.11	Poor
19	0.12	Poor
20	0.43	Excellent
21	0.31	Good
22	0.03	Poor
23	0.17	Poor
24	0.17	Poor
25	0.44	Excellent

DISCUSSION

The findings show that several vital results need to be addressed. First, the number of invalid questions is more significant than the number of valid questions, where the former is 13 and the latter is 12. This result is interestingly different from most studies on a similar topic, which

show the opposite of the results found by the current researcher. Noviasmy and Risma, who conducted a study in a junior high school, show that 80% of their test items are valid[17]. Furthermore, Liando et al. find even better result in which all of the items in their test are categorized as valid[18]. However, this finding is not surprising since the test conducted by the researcher is still the first draft that can be later revised to get a high validity score. This aligns with Sintawati and Wulandari, who conducted a validity analysis on an English summative test of the first grade of primary school and found that 6 of 10 questions were categorized as valid[19]. In this study, there is a tight gap between valid and invalid questions, although the number of valid questions is still more significant.

Although 13 items of the test are invalid, interestingly, the reliability analysis shows that the test is reliable, although it is not that high. The coefficient is 0.765, higher than the minimum coefficient of 0.70. It is different from the findings of Liando et al., Akçay and Önal, and Riski and Wahyuni, which show a high coefficient reliability[20]. The former study was conducted on junior high school students by giving them 50 items of multiple choice, which resulted in 0.82 coefficient reliability. Meanwhile, conducted in primary schools of 412 students in Turkey, the second study has a 0.888 coefficient reliability. The last study has an even higher reliability coefficient by the core 0.934. All three studies are similar in that most of their tests are valid.

Related to the item difficulty, there are 7 questions (28%) that are categorized as easy, 11 questions (44%) as medium, and 7 questions (28%) as difficult. According to Arifin (2012), a well-constructed test should not be too easy or too difficult. To obtain good learning achievement, one of the options for proportion between the difficulty levels of the questions is spread evenly, with 25% of the items being difficult, 50% being moderate, and 25% being easy. Considering that aspect, this test is well-constructed since there is a similar proportion of the difficulty items that is 1:2:1. This finding is different from Suek and Semiun et al., who find that their tests are dominated

---

[17] Yessicka Noviasmy and Risma, 'An Analysis of The English Summative Test: EFL Teacher-Made Test', *Jelita: Journal of Education, Language Innovation, and Applied Linguistics*, 5.1 (2024), 41–50 <https://doi.org/10.56185/jelita.v5i1.618>.

[18] Nihta V. F. Liando, Eunike Serhalawan, and Ceisy Wuntu, 'Analysis of Teacher-Made Tests Used in Summative Evaluation at SMP Negeri 1 Tompaso', *Jurnal Ilmiah Wahana Pendidikan*, 7.8 (2021), 480–93 <https://doi.org/10.5281/zenodo.5775342>.

[19] Fairus Sintawati and Ratih Ayu Wulandari, 'Analysis of Summative Test at First Grade of SDIT Raudhatul Jannah Cilegon', *JEDLISH: Journal of Education and English Language Teaching*, 3.1 (2023), 15–20.

[20] Liando, Serhalawan, and Wuntu; Işıl Akçay and Nezih Önal, 'Developing An Achievement Test For Primary School English Course: Validity And Reliability Study', *Journal of Teacher Education and Lifelong Learning*, 5.2 (2023), 778–88 <https://doi.org/10.51535/tell.1341862>; Yunita Ayu Riski and Sri Wahyuni, 'Developing MALL-Based Reading Comprehension Test for Seventh Grade Using Quizizz', *JALL: Journal of Applied Linguistics and Literacy)*, 7.2 (2023), 321–34.

by easy and medium level and have a very small number of difficult items[21]. Therefore, their tests are not considered well-structured. Meanwhile, for the item discrimination of the current test, 10 questions (40%) are categorized as poor, 6 questions (24%) are average, 4 questions (16%) are good, and 5 questions (20%) are excellent. This finding shows that there are still 10 questions that need to be deeply reviewed or then rejected. Fortunately, no question needs to be directly removed.

Regarding the use of Educaplay in the test administration, based on the researcher's observation, it is found that the website helps the teacher keep the students engaged in the test. By providing an attractive interface that is easy to use, students become more interested and motivated to take the test. This aligns with the study of Castillo-Cuesta and Páez-Quinde et al., who found that students perceived using digital games in Educaplay as motivating for learning English[22]. These games promoted students' active and dynamic learning since Educaplay helped create and share engaging educational activities, thus increasing their interest in learning the target language.

Considering the benefits of digital platforms toward language learning, teachers should integrate the use of digital platforms more into their learning, especially for assessing students. Educaplay definitely can be one of the options. This is in line with Graça et al., who state that digital platforms serve as pedagogical tools that bring added value to the learning process and stimulate self-confidence, the achievement of skills, understanding of knowledge, reflection and argument in communication, among other essential skills for the citizen of tomorrow[23].

Even so, there are still shortcomings in Educaplay, just like what occurs in general with the use of digital platforms. Despite its exciting features and its ability to provide various learning opportunities, students feel burdened since they had to prepare a large quota when using the Educaplay program since it was consumed immediately. Furthermore, they found it challenging to respond since, while listening to the Educaplay Arabic Speaking skills review material, they needed to grasp Arabic terminology[24]. This also happens in the current study in which the students initially have difficulty understanding the workflow of Educaplay. Fortunately, this problem was solved by giving clear directions and explanations to the students. Therefore, in applying digital platforms to students, there must be guidance and supervision from the teacher to make it successful.

---

21 Leni Amelia Suek, 'Item Analysis of an English Summative Test', *PEJLaC: Pattimura Excellence Journal of Language and Culture*, 1.1 (2021), 9–18 <https://doi.org/10.30598/pejlac.v1.i1.pp9-18>; Thresia Trivict Semiun, Maria Wihelmina Wisrance, and Merlin Helentina Napitupulu, 'English Summative Test: The Quality of Its Items', *English Education:Journal of English Teaching and Research*, 7.2 (2022), 119–27 <https://doi.org/10.29407/jetar.v7i2.18347>.

22 Castillo-Cuesta; Páez-quinde and Infante-paredes.

23 Graça, Quadro-flores, and Ramos.

24 Rachmawati and Liza.

CONCLUSION

In conclusion, the statistical analysis using SPSS 25 revealed a mix of valid and invalid questions, with 12 out of 25 questions meeting the validity criteria. The overall reliability of the test is acceptable, with a Cronbach's Alpha of 0.765, indicating that the test was consistent and dependable. The analysis of item difficulty categorized the questions into easy, medium, and difficult, aligning well with established standard for balanced test construction. However, item discrimination analysis indicated that 40% of the questions needed significant review or rejection to improve effectiveness. Furthermore, the use of Educaplay makes the test more engaging for students. The platform's interactive nature and user-friendly interface motivated students and enhanced their interest in doing the test. Despite the initial test having more invalid questions, the study underlines the potential of digital platforms like Educaplay in creating effective and engaging assessments. The findings suggest the need for ongoing refinement and validation of test items to ensure high validity and reliability. Future research could explore further improvements and the long-term impact of such digital assessments on student learning outcomes.

ACKNOWLEDGMENTS

The author would like to thank my lecturer who is also the co-author of this article in Advanced Assessment in English Language Instruction class, Dr. Sri Wahyuni, M.Pd., for her encouragement and guidance that make us enthusiastically write and submit this paper. Any mistakes remain mine alone. Furthermore, I send my best gratitude for my head of study program, Dr. Toyyibah, S.S., M.Pd., for giving me the information of INCOILS 2024 and encouraging me to join and submit my paper to this event.

BIBLIOGRAPHY

Akçay, Işıl, and Nezih Önal, 'Developing An Achievement Test For Primary School English Course: Validity And Reliability Study', *Journal of Teacher Education and Lifelong Learning*, 5.2 (2023), 778–88 <https://doi.org/10.51535/tell.1341862>

Arikunto, Suharsimi, *Dasar-Dasar Evaluasi Pendidikan*, 2nd edn (Jakarta: Bumi Aksara)

Castillo-Cuesta, L., 'Using Digital Games for Enhancing EFL Grammar and Vocabulary in Higher Education', *International Journal of Emerging Technologies in Learning*, 15.20 (2020), 116–29 <https://doi.org/10.3991/ijet.v15i20.16159>

Crocker, L., and Algina, J., *Introduction to Classical and Modern Test Theory* (Ohio: Cengage Learning, 2008)

Dörnyei, Zoltán, and Ema Ushioda, *Teaching and Researching Motivation*, 2nd edn (Harlow: Pearson Education Limited, 2011)

Ellis, Rod, *The Study of Second Language Acquisition*, (Oxford: Oxford University Press, 2008)

Gall, Meredith D., Joyce P. Gall, and Walter R. Borg, *Educational Research: An Introduction* (Boston: Pearson Education Inc., 2003)

Graça, By Vânia, Paula Quadro-flores, and Altina Ramos, 'The Integration of the Digital Platform Educaplay in Interdisciplinary Paths in the 1 St and 2 Nd Basic Education Cycles', *Athens Journal of Education*, 9.3 (2022), 377–91

Gronlund, N., *How to Make Achievement Tests and Assessments*, (Boston: Allyn and Bacon, 1993)

Liando, Nihta V. F., Eunike Serhalawan, and Ceisy Wuntu, 'Analysis of Teacher-Made Tests Used in Summative Evaluation at SMP Negeri 1 Tompaso', *Jurnal Ilmiah Wahana Pendidikan*, 7.8 (2021), 480–93 <https://doi.org/10.5281/zenodo.5775342>

Lynch, Brian K., *Language Assessment and Programme Evaluation* (Edinburgh: Edinburgh University Press, 2003) <https://doi.org/10.1016/j.esp.2003.11.003>

McKay, Penny, *Assessing Young Language Learners* (Cambridge: Cambridge University Press, 2006) <https://doi.org/10.1093/elt/ccn063>

Menon, Sujatha, 'First Steps in Online Gamification: Tips for the Language Classroom', *Gedrag & Organisatie Review*, 34.4 (2021), 227–41 <https://doi.org/10.37896/GOR34.04/026>

Mihat, Mufidah Firzanah, 'Using Educaplay Apps to Increase Year 4 Pupils ' Interest and Understanding towards Topic Properties of Light', *Journal of Science and Mathematics Education in Southeast Asia*, 46 (2024), 159–72

Muntrikaeo, Kuntida, and Kornwipa Poonpon, 'The Effects of Task-Based Instruction Using Online Language Games in a Flipped Learning Environment (TGF) on English Oral Communication Ability of Thai Secondary Students', *English Language Teaching (ELT)*, 15.3 (2022), 9–21 <https://doi.org/10.5539/elt.v15n3p9>

Noviasmy, Yessicka, and Risma, 'An Analysis of The English Summative Test: EFL Teacher-Made Test', *Jelita: Journal of Education, Language Innovation, and Applied Linguistics*, 5.1 (2024), 41–50 <https://doi.org/10.56185/jelita.v5i1.618>

Páez-quinde, Cristina, and Ruth Infante-paredes, 'Educaplay : A Gamification Tool for Academic Performance in Virtual Education during the Pandemic Covid-19', *Revista Cátedra*, 5.1 (2022), 31–44

Rachmawati, Miatin, and Fitri Liza, 'Arabic Speaking Skill Using the Educaplay "Froggy Jumps" Application', *Alsuniyat: Jurnal Penelitian Bahasa Sastra Dan Budaya Arab*, 7.1 (2024), 223–34

Riski, Yunita Ayu, and Sri Wahyuni, 'Developing MALL-Based Reading Comprehension Test for Seventh Grade Using Quizizz', *JALL: Journal of Applied Linguistics and Literacy)*, 7.2 (2023), 321–34

Semiun, Thresia Trivict, Maria Wihelmina Wisrance, and Merlin Helentina Napitupulu, 'English Summative Test: The Quality of Its Items', *English Education:Journal of English Teaching and Research*, 7.2 (2022), 119–27 <https://doi.org/10.29407/jetar.v7i2.18347>

Sintawati, Fairus, and Ratih Ayu Wulandari, 'Analysis of Summative Test at First Grade of SDIT Raudhatul Jannah Cilegon', *JEDLISH: Journal of Education and English Language Teaching*, 3.1 (2023), 15–20

Sudjiono, *Pengantar Evaluasi Pendidikan* (Jakarta: Raja Grafindo Persada, 2011)

Suek, Leni Amelia, 'Item Analysis of an English Summative Test', *PEJLaC: Pattimura Excellence Journal of Language and Culture*, 1.1 (2021), 9–18 <https://doi.org/10.30598/pejlac.v1.i1.pp9-18>

Sukmawati, Ida Dian, and Tri Pujiani, 'Enhancing Classroom Engagement Using Online Games', *Lingua: Jurnal Pendidikan Bahasa*, 19.1 (2023), 75–92 <https://doi.org/10.34005/lingua.v19i1.2616>

Surachmi, Sri, and Karl Jahniel S Sison, 'Educaplay as Teaching Media Inn Virtual Classes', *The 3rd Bogor English Student and Teacher (BEST) CONFERENCE*, 3 (2021), 1–6