Jurnal Penelitian dan Evaluasi Pendidikan
Volume 29, No. 2, December 2025 (217-227)
Online: http://journal.uny.ac.id/index.php/jpep

Developing a socio-cultural literacy test using a deep learning approach for elementary school students

Sovia Lorenza*; Endang Widi Winarni; Badeni
Universitas Bengkulu, Indonesia
*Corresponding author. E-mail: sovialorenza9@gmail.com

ARTICLE INFO

Article History
Submitted:
29 September 2025
Revised:
15 December 2025
Accepted:
18 December 2025

Keywords
instrument development; literacy test; reliability; validity; learning evaluation

Scan Me:

ABSTRACT

This research developed socio-cultural literacy test items integrated with a deep learning approach for fifth-grade elementary school students. The study involved analyzing potential and problems, collecting data and information, developing a design, revising the design, finalizing the draft, testing, analyzing, and reporting. Validation was conducted by six experts, two each in the material, language, and construction aspects. Validity based on Aiken's V-value for the material aspect was 0.625-1, indicating a moderate interpretation and very high validity. The language aspect had a value of 0.65-0.75, indicating a moderate interpretation, and the construction aspect had a value of 0.75-1, indicating a very valid interpretation. Interrater reliability for the material aspect averaged 79%, classified as strong; the language aspect averaged 89%, and the construction aspect averaged 89%, classified as near perfect. The instrument was administered to 25 fifth-grade students and empirically tested in SPSS version 29. Items 1 and 2 were invalid, items 3-15 were highly valid and very highly valid. The reliability was 0.909, a very high value. The difficulty index for items 1-6 was easy, and items 7-15 were moderate. The discrimination power of items 1 and 2 was poor; item 3 was moderate. Items 4, 8, 9, 10, 12, and 13 were classified as very good, and items 5, 6, 7, 14, and 15 were classified as good. Therefore, based on the empirical test results, the developed instrument met the requirements for content validity and was suitable for use.

To cite this article (in APA style):
Lorenza, S., Winarni, E. W., & Badeni, B. (2025). Developing a socio-cultural literacy test using a deep learning approach for elementary school students. *Jurnal Penelitian dan Evaluasi Pendidikan*, *29*(2), 217-227. https://doi.org/10.21831/pep.v29i2.90060

INTRODUCTION

Learning evaluation is a crucial part of the learning process. Evaluation is used to measure the achievement of learning objectives based on the process undertaken. Evaluation is the measurement of student learning outcomes or achievement (Maryadi, 2015). Thus, evaluation is crucial for determining the extent to which students have achieved the learning objectives.

Learning achievement is also assessed through social studies learning outcomes on national cultural diversity. This is crucial because Indonesia is a culturally rich country, and this is a significant concern, especially for elementary school students. It is clear that social studies learning fosters a love of national culture and upholds the concept of diversity as good citizens, in line with the intended learning objectives of social studies. *Ilmu Pengetahuan Alam dan Sosial* (IPAS) combines natural sciences (IPA) and social sciences (IPS) in implementing the Merdeka Curriculum. It is one of the subjects that can support character, competence, and literacy in 21st-century education. The Education Curriculum and Assessment Standards Agency (2024) explains that natural and social sciences are sciences that discuss living and non-living things in the universe and their interactions, and discuss human life as individuals and social beings who interact directly with their social environment. Social studies instruction provides opportunities

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i2.90060

to develop concepts of thought grounded in students' social realities, enabling them to think critically and solve problems effectively (Candra, 2019). The study of national cultural diversity is an important topic in social studies, as regulated in the Decree of the Head of Education Curriculum and Assessment Standards Agency No. 32 of 2024. This instruction aims to enable students to understand culture within their social context and to apply this sociocultural understanding to develop them into good citizens, in line with the objectives of social studies teaching. This aligns with the essence of social studies teaching, as articulated by Sulaiman (2022), which aims to help students develop the knowledge, attitudes, values, and skills necessary to address social problems, participate in society, and become good citizens.

Furthermore, the Decree of the Head of Education Curriculum and Assessment Standards Agency No. 32 of 2024 emphasizes the importance of mastering social skills, understanding cultural diversity, and participating in social activities. This integration creates contextual and practical learning that emphasizes not only cognitive aspects but also fosters sociocultural participation. The Education Curriculum and Assessment Standards Agency (2024) also lists the learning outcomes (*capaian pembelajaran* or CP) for phase C of grade V, namely that students can recognize national cultural diversity linked to the context of diversity. From this, it is expected that students will develop an understanding of the importance of maintaining national cultural diversity and recognize various ways of preserving national culture, both through concrete actions in daily life and through conceptual understanding. The learning process must also teach students a deep understanding. Komariyah (2025) found that teachers' understanding of how to implement deep learning-based social studies instruction remains low. In fact, Mu'ti (2025) stated that deep learning is an approach that engages students through fun, conscious, and meaningful learning, leading to an understanding that is not merely abstract conceptual knowledge. Therefore, deep learning is important for teachers to implement to provide a more comprehensive education in the classroom, including in social studies.

To achieve these CPs and, in line with the characteristics of science, instill a love of national culture, teachers play an important role in the social studies learning process by using test questions to assess students' achievement of learning outcomes. The development of social studies learning evaluation questions has been frequently carried out, including at the elementary school level. Fadhilaturrahmi and Ananda (2017) developed a social studies learning evaluation test instrument that yielded test results using a two-dimensional Bloom's taxonomy and was validated and reliable, used to measure learning outcomes. Putri et al. (2023) developed a social studies learning outcome test instrument in the higher-order thinking skills (HOTS) category. Moreover, Putra (2022) developed social studies learning evaluation questions with the Science, Technology, Engineering, Arts, and Mathematics (STEAM) approach. Darmayanti et al. (2021) developed a social studies learning outcome instrument. Nardjosoeripto et al. (2017) developed an online social studies learning evaluation tool based on Classmaker. Not only is social studies learning evaluation used to measure cognitive knowledge, but value-based social studies learning evaluation has also been done, among others, by Naila (2023). Furthermore, Basri (2017) developed evaluations grounded in character education and multiculturalism. However, no specific test instrument has been developed to assess the socio-cultural literacy of fifth-grade elementary school students. Besides, teachers continue to struggle to develop assessment instruments using a deep learning approach. This is supported by Rosmiati et al. (2025), who suggest that training is needed for teachers in developing deep learning assessment instruments and state that teachers need to master test development, test usage, measurement principles and techniques, assessment of learning outcomes, and assigning values to the information obtained to make objective decisions. These assessment activities should be conducted based on the principle that assessments of student learning outcomes are intended as corrective feedback for both students and teachers, with the results obtained increasing student interest and motivation to achieve even better learning outcomes. However, there is no specific test instrument that measures socio-cultural literacy integrated with deep learning for fifth-grade elementary school

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i2.90060

students. This research uses a deep learning approach to support in-depth learning, facilitating students' attainment of deep understanding through the principles of joyful, meaningful, and mindful learning. By developing a socio-cultural literacy test integrated with the deep learning approach, students can gain a deeper understanding of the learning material they are studying.

RESEARCH METHOD

Research Method

This research is a research and development (R&D) according to Borg and Gall (1989), with steps including analysis of potential and problems, data and information collection, product design, design validation, design revision, final draft, trial, analysis, and reporting. This research uses quantitative descriptive analysis. Data were obtained through expert judgment by six experts (two on material aspects, two on language aspects, and two on construction aspects), as well as from the results of the trial of the socio-cultural literacy test. The test developed contains 15 multiple-choice questions containing socio-cultural literacy indicators. The logical feasibility of the evaluation test was established through expert validation, and its empirical feasibility was established through a trial conducted by 25 students. Then, data analysis was carried out using SPSS Version 29 with the following stages:

SPSS Version 29 ⟶ Validity ⟶ Reliability ⟶ Difficulty Index ⟶ Discrimination Power

Validity or logical feasibility encompasses material, construction, and language aspects. To determine the level of expert agreement, the researchers used the Aiken V formula as shown in Formula (1), in which s = r – lo, r = Score given by an expert, lo = Lowest validity assessment score, c = Highest validity assessment score, n = Number of experts (Wasidi, 2020, p. 65).

$$\mathbf{V} = \frac{\sum \boldsymbol{s}}{\boldsymbol{n}\,(\boldsymbol{c}-\mathbf{1})} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

$$Inter-rater\ reliability = \frac{\text{Number of both raters' agreement}}{\text{Number of statements}} x100\% \quad \ldots\ldots\ldots\ldots\ldots \quad (2)$$

The approach to testing logical reliability in this study uses the inter-rater reliability calculation using Formula (2) (McHugh, 2012, p. 279). Empirical validity or feasibility encompasses the discrimination power, difficulty, reliability, and validity of the empirical test items in the developed questions. The technique used to measure the validity of the socio-cultural literacy test instrument is the product-moment correlation technique (Winarni, 2018, p. 116) with the following Formula (3):

$$r_{xy} = \frac{N\sum XY-(\sum X)(\sum Y)}{\sqrt{\{\sum X^2-(\sum X)^2\}\{N\sum Y^2-(\sum Y)^2\}}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (3)$$

Reliability is the extent to which an instrument can be trusted as a data-collection tool (Winarni, 2018, p. 137). Reliability is measured using the Cronbach's Alpha formula as presented in Formula (4), in which r_{11} = item's reliability coefficient, K = number of test items, $\sum \sigma_b^2$ = number of test item variants, and σ_t^2 = total of variants. The criterion is that if $r_{11} > r_{table}$, it is reliable, and if $r_{11} > r_{table}$, it is unreliable.

$$r_{11} = \left[\frac{k}{k-1}\right]\left[1 - \frac{\sum \sigma_b^2}{\sigma_t^2}\right] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (4)$$

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i2.90060

The test difficulty level should be balanced or moderate. This balance means having questions that are relatively easy, moderate, and difficult. The difficulty level is assessed based on the student's ability to answer, not from the test maker's perspective. A crucial issue in evaluating difficulty is determining the proportions and criteria for classifying questions as easy, moderate, or complex. The difficulty index is expressed as P (Winarni, 2018, p. 137) using the following Formula (5), in which P = difficulty index.

$$P = \frac{Mean}{Maximum\ Score} \quad \text{..... (5)}$$

The discrimination power of a test is its ability to distinguish intelligent from less intelligent subjects. Formula (6) is used to determine the discrimination power (DP) of test items (Winarni, 2018, p. 137), where A = number of test participants, JA = number of participants in the upper group, JB = number of participants in the lower group, JBA= number of participants in the upper group who answered correctly, and JBB = number of participants in the lower group who answered correctly.

$$DP = \frac{\mathrm{JB}_A}{\mathrm{J}_A} - \frac{\mathrm{JB}_B}{\mathrm{J}_B} \quad \text{..... (6)}$$

The initial stage involved analyzing potential and problems. This stage identified the initial problem, namely the lack of test instruments capable of measuring students' socio-cultural literacy that teachers could use to demonstrate representative student abilities.

Next, data and information were collected. This stage involved collecting data on socio-cultural literacy indicators to develop a test and aligning them with the learning materials, learning outcomes, and classroom learning objectives. The integrated learning material on national cultural diversity aligns with the Grade V learning outcomes set by the Education Curriculum and Assessment Standards Agency (2024), namely that students should be able to recognize national cultural diversity in its diverse contexts. According to Juwita et al. (2021), the sociocultural literacy indicators developed in the sociocultural literacy test instrument are: the ability to recognize, respond, reflect, evaluate, and create knowledge/action plans related to tolerance. Furthermore, the SOLO taxonomy proposed by Biggs and Collis (1982) is applied in accordance with the deep learning approach, namely, to the pre-structural, uni-structural, multi-structural, relational, and extended abstract levels. Therefore, learning outcomes are measured by students' abilities to understand, apply, and evaluate.

Table 1. Grid of Socio-Cultural Literacy Test Items

Socio-Cultural Literacy Indicators	Learning Achievement	Test Indicator	Test Items
Ability to know, respond, reflect, evaluate, and create knowledge/action plans related to tolerance	Students can recognize the diversity of national culture, which is linked to the context of diversity.	Students learn about the types of national culture presented through images as stimuli.	1, 2, 3
		Students can respond to cultural diversity with tolerance when presented with statements as stimuli.	4, 5, 6
		Students can reflect on solutions to social problems in addressing national cultural diversity.	7, 8, 9
		Students can evaluate appropriate attitudes toward national cultural diversity.	10, 11, 12
		Students can develop arguments about national culture using a given situation.	13, 14
		Students can create an action plan to address the national culture in a given situation.	15

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i2.90060

Product design entails developing a sociocultural literacy test question outline tailored to Learning Outcomes, learning materials, learning objectives, and sociocultural literacy indicators. This prepared outline was then validated by two subject-matter experts, who served as validators (expert judges), to provide suggestions in developing the sociocultural literacy test instrument. The sociocultural literacy test question outline for fifth-grade students is presented in Table 1.

After the socio-cultural literacy test outline was compiled and validated by subject-matter experts, the next stage was to validate the instrument's design, which involved six experts: a subject-matter expert, a construction expert, and a language expert. These six validators were experts in elementary education, based on their concentration and expertise, and therefore qualified to assess the developed socio-cultural literacy test. The validation score used a scale of 1-5 with the following criteria: 5 = very good if four descriptors appear, 4 = good if three descriptors appear, 3 = moderate if two descriptors appear, 2 = poor if one descriptor appears, and 1 = very poor if none of the descriptors appear.

The validation instruments for the subject matter, construction, and language aspects are presented in Table 2, Table 3, and Table 4. Each table is followed by the descriptors.

Table 2. Validation Questionnaire for the Socio-Cultural Literacy Test Items for the Material Aspect

No.	Tested Aspects	Number 1 Multiple Choice				
		1	2	3	4	5
1.	The questions use contextual stimuli (images/graphs/text/explanations, etc.) and are engaging.					
2.	The question-and-answer stimuli meet the requirements for the basic competencies of socio-cultural literacy knowledge.					
3.	The questions' requirements align with the level of thinking and the dimensions of knowledge.					
4.	The answer choices are homogeneous and logical with respect to the material.					
5.	The answer key is clear and does not cause ambiguity or bias.					

Descriptors of each aspect:

Aspect 1:
a. The stimulus is relevant to the competency requirements being measured.
b. The stimulus is in the form of images/graphs/text/data.
c. The stimulus is interesting.
d. The stimulus is appropriate to the context of the question.

Aspect 2:
a. The stimulus is relevant to the competency requirements being measured.
b. The answer requirements are appropriate to the competency requirements being measured.
c. The question stimulus serves as a trigger to determine the answer.
d. The question stimulus does not indicate the answer.

Aspect 3:
a. The question stimulus is appropriate to the level of thinking/cognitive level.
b. The question stimulus is appropriate to the factual/conceptual knowledge dimension.
c. The answer choices are appropriate to the cognitive level.
d. The answer choices are appropriate to the knowledge dimension.

Aspect 4:
a. Answer choices are homogeneous in terms of content.
b. Answer choices are relevant to the question stimulus.
c. Answer choices are logical in terms of depth of material.
d. Answer choices are logical in terms of breadth of material.

Aspect 5:
a. The answer key is evident according to the question indicators.
b. The answer key does not create double meanings/bias.
c. The answer key is clear in terms of material content.
d. The answer key is clear in terms of material context.

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i2.90060

Table 3. Validation Questionnaire for the Socio-Cultural Literacy Test Items for the Construction Aspect

No.	Tested Aspect	Item Number Multiple Choice				
		1	2	3	4	5
1.	Questions should be structured in a neat format, allowing participants to read and understand them quickly without being hampered by layout or other technical errors.					
2.	Questions should be easily understood and not confusing.					
3.	The choices provided should consist of one correct answer and several incorrect answers.					

Descriptors of each aspect:
Aspect 1:
a. Questions are structured with sufficient spacing between sections (appropriate spacing).
b. The font type used must be easy to read.
c. Punctuation is used correctly and consistently.
d. The font size used must be easy to read.
Aspect 2:
a. Use easy-to-understand and straightforward language.
b. Questions should not be wordy or use long, convoluted sentences.
c. No question can be interpreted in more than one way.
d. The language used in the questions is appropriate to the participants' level of knowledge.
Aspect 3:
a. Correct answers must be consistent with the objective of the question.
b. Incorrect answers or distractors must be designed in such a way that they appear reasonable or logical.
c. A balanced number of answer choices.
d. Correct answer choices are appropriate for testing understanding.

Table 4. Validation Questionnaire for the Socio-Cultural Literacy Test Items for the Language Aspect

No.	Tested aspect	Item Number Multiple Choice				
		1	2	3	4	5
1.	Language comprehension					
2.	Simplicity and practicality					
3.	Proper use of punctuation					

Descriptors of each aspect:
Aspect 1:
a. The language used must be clear.
b. Sentences must not be ambiguous or long-winded.
c. Word choice must avoid double meanings.
d. The language must be easy for students to understand.
Aspect 2:
a. Questions must be written in simple, practical language.
b. Jargon or technical terms that are irrelevant to the question topic or that students may not understand.
c. Sentences must not be too long or complicated.
d. Questions must be tailored to the questions.
Aspect 3:
a. Correct use of punctuation is crucial for constructing understandable sentences.
b. Moderate use of punctuation.
c. Correct use of punctuation.
d. Unambiguous use of punctuation.

After validating the socio-cultural literacy test instrument, revisions were made in response to the validator's suggestions. This revision was carried out to develop a socio-cultural literacy test instrument suitable for use and capable of measuring students' socio-cultural literacy in a representative manner. After revisions, a final draft was obtained for the pilot testing phase.

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i2.90060

Pilot testing of the test instrument was conducted in a fifth-grade class at SDN 17 Kepahiang, with 25 respondents. This small-scale pilot test aimed to obtain an overview of the use of the socio-cultural literacy test instrument in fifth-grade classes. Following the pilot testing, the next stage was analysis and reporting of the results. The analysis included logical validity and reliability based on expert judgment, as well as validity, reliability, discrimination power, and difficulty index obtained from empirical testing of the questions using SPSS version 29.

FINDINGS AND DISCUSSION

Findings

The research results demonstrate the level of validity achieved through expert assessments across the material, construction, and language aspects. The validity of the test instrument for the material aspect is presented in Table 5.

Table 5. Validity of the Test Items on the Material Aspect

Item Number	Aiken's V Value for Material Aspect	Interpretation
1, 2, 3, 4, 5, 14	0.75	Moderate
6, 7, 8, 9, 10, 11, 12, 15	1	Very Valid

Table 5 presents Aiken's V for the socio-cultural literacy test instrument, which ranges from 0.625 to 1, indicating moderate interpretability and high validity. This indicates that the socio-cultural literacy test instrument meets the content validity criteria for the material aspect.

Table 6. Validity of the Socio-Cultural Literacy Test Items on the Language Aspect

Item Number	Aiken's V Value for Language Aspect	Interpretation
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15	0.75	Moderate

Table 6 presents Aiken's V for the socio-cultural literacy test instrument, which ranges from 0.625 to 0.75, indicating a moderate level of validity. This indicates that the socio-cultural literacy test instrument meets the content validity requirements for language.

Table 7. Validity of the Socio-Cultural Literacy Test Items on the Construction Aspect

Item Number	Aiken's V Value for Construction Aspect	Interpretation
1, 2, 6, 7, 8, 9, 10, 11, 12, 15	0.875	Very Valid
3, 4, 5, 14	0.75	Moderate

Table 7 presents Aiken's V for the socio-cultural literacy test instrument, which ranges from 0.75 to 1, indicating interpretation validity at the moderate and very valid levels. This indicates that the socio-cultural literacy test instrument meets the feasibility criteria for its construction aspect.

In addition to logical validity, logical reliability was assessed to evaluate the feasibility of the socio-cultural literacy test instrument. Logical reliability examines the instrument's consistency and reliability with respect to its material, language, and construction.

Table 8. Reliability of the Socio-Cultural Literacy Test Items on the Material Aspect

Item Number	Percentage of Material Agreement
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15	60% - 100%
Average	**79%**
Interpretation	**Strong**

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i2.90060

Table 8 presents the interrater reliability of the socio-cultural literacy test instrument, with agreement ranging from 60% to 100% across materials. The average score was 79%, with a strong interpretation. This indicates that the socio-cultural literacy test instrument is reliable and suitable for use.

Table 9. Reliability of the Socio-Cultural Literacy Test Items on the Language Aspect

Item Number	Language Agreement Percentage
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15	67% - 100%
Average	**89%**
Interpretation	**Nearly Perfect**

Table 9 presents the interrater reliability of the socio-cultural literacy test instrument, with agreement ranging from 67% to 100% for the language aspect. The average score was 89%, with near-perfect interpretation. This indicates that the socio-cultural literacy test instrument is reliable and suitable for use in terms of language.

Table 10. Reliability of the Socio-Cultural Literacy Test Items on the Construction Aspect

Item Number	Construction Agreement Percentage
1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15	67% - 100%
Average	**89%**
Interpretation	**Nearly Perfect**

Table 10 shows the interrater reliability of the socio-cultural literacy test instrument, with a constructional agreement ranging from 67% to 100%. The average score was 89%, with near-perfect interpretation. This indicates that the socio-cultural literacy test instrument is reliable and suitable for use.

In addition to assessing logical validity through interrater agreement, an empirical validity test was conducted in SPSS version 29.00, based on results from a fifth-grade test with 25 respondents. Table 11 presents the empirical validity of the socio-cultural literacy test instrument.

Table 11. Validity of the Socio-Cultural Literacy Test

Item Number	R_{table}	Category
1 - 2	0.3961	*Not Valid
3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15		Valid

Table 11 presents the empirical validity of the socio-cultural literacy test instrument, based on the test results. The output indicates that Items 1 and 2 are invalid, that Item 3 has a moderate level of validity, and that Items 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, and 15 have valid conclusions. Therefore, Items 1 and 2 cannot be used in this study.

Table 12. Reliability of the Socio-Cultural Literacy Test

Reliability Statistics	
Cronbach's Alpha	**N of Items**
.909	15

Based on reliability calculations for the socio-cultural literacy test instrument, a Cronbach's alpha of 0.909 was obtained, indicating very high reliability. This indicates that the developed socio-cultural literacy test instrument is suitable for use and meets the requirements for reliability/consistency.

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i2.90060

Table 13. Difficulty Level of the Socio-Cultural Literacy Test Instrument

Item Number	Difficulty Level	Criteria
1, 2, 3, 4, 5, 6	0.88	Easy
7, 8, 9, 10, 11, 12, 13, 14, 15	0.44 - 0.68	Moderate

Table 13 shows that the test items' difficulty levels range from 0.44 to 0.88. Therefore, based on the discrimination power, Items 1, 2, and 3 are at an easy level (0.88). Items 4, 5, and 6 are at an easy level (0.76), and Items 7, 8, 9, 10, 11, 12, 13, 14, and 15 are at a moderate level.

Table 14. Discrimination Power of the Sociocultural Literacy Test Items

Item Number	Discrimination Power	Criteria
1, 2	-0.174 – 0.044	Poor
3	0.357	Moderate
4, 8, 9, 10, 12, 13	0.720 - 0.755	Excellent
5, 6, 7, 11, 14, 15	0.592 – 0.695	Good

Table 14 shows the highly variable discrimination power of the test items. Items 1 and 2 had poor discrimination power; Item 3 had adequate discrimination power; Items 4, 8, 9, 10, 12, and 13 had excellent discrimination power; and Items 5, 6, 7, 11, 14, and 15 had good discrimination power. This indicates that the socio-cultural literacy test item instrument can effectively differentiate between high- and low-ability students. Based on empirical test results for the 15 items, including validity, reliability, discrimination power, and difficulty index, Items 1 and 2 are ineligible for use in the study because they do not meet the eligibility criteria for test items. Therefore, the socio-cultural literacy test comprises Items 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, and 15.

Discussions

Based on the development of the socio-cultural literacy test instrument, valid, reliable, and suitable scores were obtained. The validity level for the material aspect was assessed as valid and moderate; for the language aspect, as moderate; and for the construction aspect, as valid and moderate. This indicates that the developed instrument meets the elements of content validity based on the scores given by experts (expert judgment). This can serve as a reference for developing valid question instruments for social studies learning in elementary schools. This is consistent with research by Marselina et al. (2021), which yielded valid results for the social studies learning outcome test administered to fifth-grade elementary school students. In line with the findings of research (Utami & Wardani, 2020), which demonstrated the item validity and reliability of the developed instrument for the cognitive assessment of social studies learning. Research by Wahidmurni et al. (2021) on the development of assessment instruments with social studies teachers. This research obtained reliability estimates of 79% for the material aspect and 89% for the language and construction aspects, based on expert assessments. This indicates that the developed instrument is reliable, as evidenced by expert validation. Furthermore, based on empirical testing, 13 items were valid, and two items were invalid. The instrument's reliability was assessed using Cronbach's alpha, which was 0.909. The difficulty index was at an easy-to-moderate level. The discriminative power of the two items was low, and the remaining 13 items were usable. This study shows that the instrument can be used to measure the socio-cultural literacy of fifth-grade elementary school students. This study developed questions that integrated Deep Learning with the SOLO taxonomy to assess levels of student understanding. The SOLO taxonomy by Biggs and Collis (1982) has five levels of understanding: pre-structural, uni-structural, multi-structural, relational, and extended abstract (deep understanding). Pre-structural is the stage at which students have not yet developed an

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i2.90060

understanding of the learning material, typically evidenced by incorrect answers. Unistructural understanding occurs when students can grasp a single piece of initial information. Multistructural understanding occurs when students already understand several parts of the material being studied but are unable to connect the relationships among those parts. Relational understanding is students' ability to connect one part of the material to another correctly. Extended abstract understanding is students' ability to generate new ideas from information obtained from the material under study. This sociocultural literacy test instrument can measure students' understanding using the SOLO taxonomy.

CONCLUSION

Based on the research findings, the developed test can serve as a reference for elementary school teachers in developing socio-cultural literacy test instruments that support deep learning through the SOLO taxonomy. This socio-cultural literacy test has undergone expert judgment and content validity, and, based on statistical results, is suitable for teachers to measure the socio-cultural literacy of fifth-grade elementary school students. Further research should be conducted on other learning materials using a larger sample to develop a valid and reliable test.

Conflict of Interests

The authors declare that they have no conflict of interest to disclose.

REFERENCES

Basri, I. (2017). Evaluasi pembelajaran sekolah dasar (SD) berbasis pendidikan karakter dan multikultural. *Jurnal Ilmiah Sekolah Dasar*, *1*(4), 247–251. https://doi.org/10.23887/jisd.v1i4.12593

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the observed learning outcome).* Academic Press.

Borg, W. R., & Gall, M. D. (1989). *Educational research: An introduction* (5th ed.). Longman.

Candra, F. (2019). *Pengajaran ilmu pengetahuan sosial di SD*. UNIPMA Press. https://eprint.unipma.ac.id/89/1/29.%20Pengajaran%20IPS%20SD.pdf

Darmayanti, L. E., Suarni, N. K., & Arnyana, I. B. P. (2021). Pengembangan tes hasil belajar IPS dan kuesioner motivasi belajar siswa kelas IV SD. *Jurnal Penelitian Dan Evaluasi Pendidikan Indonesia*, *11*(1), 21–31. https://doi.org/10.23887/jpepi.v11i1.274

Decree of the Head of Education Curriculum and Assessment Standards Agency No. 32 of 2024. (2024). Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia.

Education Curriculum and Assessment Standards Agency. (2024). *Capaian pembelajaran mata pelajaran ilmu pengetahuan alam dan sosial (IPAS) fase A–C untuk SD/MI/program paket A*. Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia.

Fadhilaturrahmi, F., & Ananda, R. (2017). Evaluasi pembelajaran IPS berbasis Taksonomi Bloom dua dimensi di sekolah dasar. *Jurnal Basicedu, 1*(2), 12–21. https://doi.org/10.31004/basicedu.v1i2.154

Juwita, H., Nur, S., Rahmah, A., Syaikhuddin, & Zubaidah, I. (2021). *Modul perbaikan pembelajaran literasi sosial budaya.* Direktorat KSKK Kementerian Agama Republik Indonesia. https://fliphtml5.com/ayhae/ewmp/basic

ISSN (print) 2685-7111 | ISSN (online) 2338-6061

10.21831/pep.v29i2.90060

Komariyah, S. (2025). Deep learning dalam upaya meningkatkan kompetensi sosial siswa melalui pembelajaran IPS. *Jurnal Sosialita*, *20*(1), 43-50. https://doi.org/10.31316/js.v20i1.7742

Marselina, K. T., Lasmawan, I. W., & Dantes, N. (2021). Pengembangan instrumen kemampuan berpikir kritis dan hasil belajar IPS pada siswa kelas V SD. *Jurnal Penelitian Dan Evaluasi Pendidikan Indonesia*, *11*(2), 105–114. https://doi.org/10.23887/jpepi.v11i2.620

Maryadi. (2015). *Evaluasi hasil belajar: Pengetahuan dan teknik*. Magnum Pustaka.

McHugh, M. L. (2012). Lessons in biostatistics interrater reliability: The Kappa statistic. *Biochemica Medica*, *22*(3), 276–282. https://www.biochemia-medica.com/assets/images/upload/xml_tif/McHugh_ML_Interrater_reliability.pdf

Mu'ti, A. (2025). *Pembelajaran mendalam menuju pendidikan bermutu untuk semua*. Pusat Kurikulum dan Pembelajaran Badan Standar, Kurikulum, dan Asesmen Pendidikan Kementerian Pendidikan Dasar dan Menengah Republik Indonesia. https://bpmpkaltara.kemdikbud.go.id/wp-content/uploads/2025/02/nasmik-deep-learning-2025-full_10-feb.pdf

Naila, I. (2023). Evaluasi pembelajaran IPS berbasis nilai untuk SD/MI. In O. A. Suciptaningsih (Eds.), *IPS kependidikan dasar*. Nawa Litera Publishing. https://books.google.co.id/books?hl=en&lr=&id=HGzFEAAAQBAJ&oi=fnd&pg=PA95&dq=info:Z6Gcf_F3NTcJ:scholar.google.com&ots=MK58vUlljN&sig=9oXbUagGwTOIdmfdX0NOARAPlCY&redir_esc=y#v=onepage&q&f=false

Nardjosoeripto, P., Sari, N., & Purbosari, P. M. (2017). Analisis ahli dalam pengembangan evaluasi pembelajaran IPS online berbasis classmaker. *Edudikara: Jurnal Pendidikan Dan Pembelajaran*, *2*(3), 268–278. https://doi.org/10.32585/edudikara.v2i3.58

Putra, I. G. A. D. (2022). *Pengembangan instrumen evaluasi interaktif berbasis pendekatan STEAM pada kompetensi pengetahuan IPS siswa kelas V SD No. 3 Sibanggede*. Undergraduate Thesis, Universitas Pendidikan Ganesha. https://repo.undiksha.ac.id/10563/

Putri, R. S., Sanjaya, W., & Fitria, Y. (2023). Penyusunan instrumen penilaian HOTS dalam pembelajaran IPS sekolah dasar. *Jurnal Ilmiah Universitas Batanghari Jambi, 23*(2), 1318-1322. http://dx.doi.org/10.33087/jiubj.v23i2.3475

Rosmiati, Susiloningsih, W., Rusminati, S. H., & Juniarso, T. (2025). Pelatihan penyusunan instrumen penilaian proses dan hasil pembelajaran deep learning. *Kanigara*, *5*(2), 89–101. https://doi.org/10.36456/rkqedx61

Sulaiman, U. (2022). *Pembelajaran IPS SD/MI*. Alauddin University Press. https://repositori.uin-alauddin.ac.id/23249/1/Pembelajaran%20IPS%20SD-MI.pdf

Utami, D. A. P., & Wardani, N. S. (2020). Pengembangan instrumen penilaian kognitif dalam pembelajaran tematik kelas V sekolah dasar. *Lentera: Jurnal Ilmiah Kependidikan*, *13*(1), 1-18. https://jurnal.stkippgribl.ac.id/index.php/lentera/article/view/483

Wahidmurni, W., Marandy, Y. S., & Hani'ah, Z. (2021). Demografi dan kompetensi pengembangan instrumen penilaian hasil belajar berbasis Higher Order Thinking Skills guru Ilmu Pengetahuan Sosial. *J-PIPS (Jurnal Pendidikan Ilmu Pengetahuan Sosial)*, *8*(1), 90–103. https://ejournal.uin-malang.ac.id/index.php/jpips/article/view/13957/0

Wasidi. (2020). *Tes dan pengukuran pendidikan*. Halaman Moeka Publishing.

Winarni, E. W. (2018). *Teori dan praktik penelitian kuantitatif, kualitatif penelitian tindakan kelas (PTK), research and development (R&D)*. Bumi Aksara.

ISSN (print) 2685-7111 | ISSN (online) 2338-6061