

Multidimensional Validation of a HOTS-Based Negotiation Text Assessment for Senior High School Students

Validasi Multidimensi terhadap Penilaian Teks Negosiasi Berbasis HOTS untuk Siswa Sekolah Menengah Atas

Delya Cristiyani Subroto¹ **Ida Nuraeni^{2*}**

Universitas Tadulako^{1,2}

*Corresponding author. Email: idanuraeni.untad@gmail.com

doi: 10.24036/jbs.v14i1.136477

Submitted: Nov 26, 2025

Revised: Mar 26, 2026

Accepted: Mar 30, 2026

Abstract

This study rigorously evaluates the content, construct, and criterion validity of an assessment instrument designed for negotiation text analysis. Employing a quantitative descriptive methodology, data were triangulated through classroom observations, pedagogical document analysis, and performance assessments involving 20 tenth-grade students. Findings demonstrate that the instrument exhibits robust content validity, aligning systematically with national curriculum standards. Regarding construct validity, expert panels yielded high ratings (scores 8–10), although items 2 and 5 required revisions due to identified explicit answer cues. Furthermore, criterion validity confirmed adherence to technical and ethical assessment standards, despite noted redundancies in the response options for items 8 and 9. All participants surpassed the minimum proficiency threshold of 70, with peak scores reaching 100. In conclusion, the instrument is categorized as valid and highly suitable for evaluating analytical competence, following minor refinements to specific items. This validation not only strengthens theoretical perspectives on multidimensional validity in language assessment but also offers practical guidelines for teachers in designing fair and curriculum-aligned negotiation text assessments in Indonesian senior high schools.

Key words: *Multidimensional Validity; Negotiation Text; Indonesian Language Assessment; Analytical Skills*

Abstrak

Penelitian ini bertujuan untuk mengevaluasi secara mendalam validitas isi, konstruk, dan kriteria pada instrumen asesmen analisis teks negosiasi bagi siswa tingkat SMA. Menggunakan metodologi kuantitatif deskriptif, data dikumpulkan melalui observasi kelas, analisis dokumen guru, serta tes performansi dari 20 peserta didik. Temuan penelitian menunjukkan bahwa instrumen tersebut memiliki validitas isi yang sangat baik dan selaras dengan indikator kurikulum nasional. Pada dimensi konstruk, panel ahli memberikan penilaian tinggi (skor 8–10), meskipun terdapat catatan revisi untuk butir soal nomor 2 dan 5 yang teridentifikasi mengandung petunjuk jawaban eksplisit. Sementara itu, validitas kriteria mengonfirmasi bahwa instrumen memenuhi standar teknis dan etika penulisan soal, walaupun ditemukan redundansi pada pilihan jawaban soal nomor 8 dan 9. Seluruh siswa berhasil melampaui ambang batas ketuntasan minimal 70 dengan pencapaian skor tertinggi mencapai 100. Simpulannya, instrumen ini dikategorikan valid dan sangat layak digunakan sebagai alat evaluasi kompetensi analitis setelah dilakukan penyempurnaan minor pada butir soal tertentu. Validasi ini tidak hanya memperkuat perspektif teoretis mengenai validitas multidimensi dalam penilaian bahasa, tetapi juga memberikan pedoman praktis bagi guru dalam merancang penilaian teks negosiasi yang adil dan sesuai kurikulum di sekolah menengah atas Indonesia.

Kata kunci: *Validitas Multidimensi, Teks Negosiasi, Asesmen Bahasa Indonesia, Kemampuan Analitis*

INTRODUCTION

Educational assessment is pivotal in monitoring student learning progress and informing classroom-level instructional decision-making. Within the Indonesian educational context, teachers are required not only to master subject matter but also to design and implement valid assessment instruments that accurately capture students' linguistic abilities (Popham 2014). This proficiency is crucial to ensure that evaluative data authentically reflects student competence, serving as a robust foundation for sound academic decisions (Arikunto 2011).

In Indonesian language pedagogy, negotiation texts represent a genre explicitly mandated by the senior high school curriculum (Kemdikbud 2017). Through this genre, tenth-grade students learn to identify communicative purposes, analyze interactional stages, and evaluate strategies employed to reach consensus. Consequently, assessing students' analytical capabilities requires instruments that

move beyond surface-level comprehension to capture high-order analytical skills involving structure, argumentation, and negotiation strategies.

The concept of validity in language assessment has evolved from the traditional view that separates content, construct, and criterion validity to a “unified validity” framework that places all evidence as part of construct validity. Messick (1989) asserts that validity is not only about the suitability of test content to learning objectives, but also how scores can be interpreted legitimately and used responsibly in educational decision-making. In the context of language assessment in the classroom, this implies the need for evidence that the items truly represent the targeted language ability domain, measure the appropriate cognitive processes, and produce scores that are consistent with established performance standards. Therefore, testing the content, construct, and criterion validity of negotiation text instruments is a prerequisite for ensuring that evaluation data truly reflects students' analytical competence and not factors of chance or technical bias.

In the field of language testing, Bachman and Palmer (2010) emphasize that “useful” tests must demonstrate a balance between a number of qualities, including validity, reliability, authenticity, interactivity, impact, and practicality. Content validity requires a strong link between the test grid and curriculum indicators, while construct validity requires that the items truly operationalize the abilities being measured, such as text structure analysis and negotiation strategies. On the other hand, criterion validity relates to the extent to which test scores correlate with relevant external measures, such as school minimum passing standards or other indicators of learning success. The classroom assessment literature emphasizes that when teachers develop tests themselves, the risk of inconsistency between content, objectives, and question format often increases, making a systematic validation process crucial to ensuring fairness and accuracy of results.

In line with the 2013 Curriculum agenda and national assessment policies, the development of High-Order Thinking Skills (HOTS)-based assessments in Indonesian language learning requires a shift from memorization questions to tasks that demand analysis, evaluation, and creativity. Recent studies on HOTS-based Indonesian language questions show that many test instruments are still dominated by low-level cognitive questions, or appear to be HOTS in format but in substance only require textual information matching. In the context of negotiation texts, HOTS assessments should test students' ability to identify communicative goals, analyze stages of interaction, and evaluate the argumentation strategies used to reach an agreement, rather than simply recognizing definitions or surface characteristics. Thus, multidimensional validity testing of negotiation text instruments which claim to be HOTS is an important contribution to bridging the gap between curriculum requirements and assessment practices in Indonesian language classrooms.

A number of previous studies have emphasized the importance of validity as the main foundation in language test development. Messick (1996) proposed an integrated validity framework that includes content, construct, and criterion evidence, and highlights the consequences of using tests as an integral part of validity itself. This framework was then operationalized in the context of language testing by Bachman and Palmer (2010), who formulated the concept of usefulness with the main components of validity, reliability, authenticity, and impact, while also providing practical guidance on item development and test quality analysis. In Indonesia, Arikunto (2011) emphasizes that the suitability of test materials, learning objectives, and curriculum indicators is a prerequisite for declaring an educational evaluation instrument valid and suitable for use in academic decision-making. These findings confirm that multidimensional validity studies of teacher instruments, as conducted in this paper, have a strong theoretical basis.

In the field of Indonesian language learning, several studies on the development of teaching materials and modules have included validity testing as part of the product development process. Research on a scaffolded writing-based negotiation text module in high school shows that the results of validity testing by experts reached the “valid” category, and the module was declared suitable for improving students' understanding of negotiation text material (Gusvenisa, Yanti, & Gusriani 2023). Other studies (Masnawati et.al 2024) on the development of teaching materials and e-modules for writing negotiation texts using various development models (e.g., ADDIE or discovery learning) generally report similar results: the products are considered valid in terms of content and appearance, and effective in improving students' negotiation text writing skills. However, the focus of these studies was more on the validity of teaching materials and the effectiveness of improving writing skills, rather than on the validity of test instruments that specifically measure analytical skills in negotiation texts as explored in this study.

Another study from Alfiani, Rani, & Wahyuni (2024) critically highlights the gap between material and assessment in negotiation text learning. A descriptive qualitative study by Aisyahara and Sukenti (2025) found a mismatch between the characteristics of negotiation texts, which require dynamic speaking and interaction skills, and the forms of assessment used by teachers, which are still

dominated by non-interactive written questions. The implication of these findings is the need for assessments that are more authentic, interactive, and in line with the demands of real-world negotiation competencies. In line with this, a study analyzing Indonesian language exam questions based on High Order Thinking Skills (HOTS) shows that many instruments claimed to be HOTS are still oriented towards memorization and low-level comprehension, thus not fully measuring students' analytical and evaluative abilities. In this context, this study occupies an important position because it not only develops a negotiation text assessment instrument aimed at measuring analytical abilities, but also systematically tests its content, construct, and criterion validity, while linking it to student learning outcomes at the high school level.

Previous research in Indonesian language education has largely focused on developing teaching materials, modules, and e-modules for negotiation texts, often accompanied by expert validation limited to content and restricted classroom trials. These studies generally conclude that the produced materials are 'valid' and effective for improving students' writing or comprehension skills, yet they tend to treat validity in a narrow, one-dimensional sense and seldom provide a systematic, multidimensional validation framework. Furthermore, much of the work on HOTS-based Indonesian language assessment reveals that items labeled as HOTS frequently remain at lower cognitive levels, indicating a persistent misalignment between curriculum demands for higher-order thinking and the actual constructs measured by classroom assessments. In contrast to these trends, the present study explicitly targets the validation of teacher-developed assessment instruments specifically designed to measure students' analytical competence in negotiation texts at the senior high school level. The novelty of this research lies in its systematic integration of content, construct, and criterion validity within a single study, while simultaneously and empirically linking expert validation results to student performance data and school proficiency benchmarks. This comprehensive approach addresses a significant conceptual and empirical gap concerning how negotiation text assessments can be credibly constructed, scrutinized, and refined to genuinely capture higher-order analytical skills and function as trustworthy tools for instructional decision-making in Indonesian language pedagogy.

However, there is still a notable lack of studies that explicitly target the validation of teacher-developed assessment instruments specifically designed to measure students' analytical competence in negotiation texts at the senior high school level. Few investigations integrate content, construct, and criterion validity within a single study while simultaneously linking expert validation results to empirical student performance data and school proficiency benchmarks. This leaves a conceptual and empirical gap concerning how negotiation text assessments can be systematically constructed, scrutinized, and refined so that they genuinely capture higher-order analytical skills and function as trustworthy tools for instructional decision-making in Indonesian language pedagogy.

Building upon this background, the present study aims to describe the degree of validity of a teacher-developed evaluation tool for analyzing negotiation texts in tenth grade students at a high school in Central Sulawesi. The primary focus is to evaluate the instrument through three dimensions: content, construct, and criterion validity (Fulcher & Davidson 2007). This approach is intended to address the current empirical gap and ensure the availability of reliable measurement tools to support the achievement of learning indicators.

Broadly, this study is expected to provide both theoretical and practical contributions to the field of Indonesian language learning evaluation. Theoretically, it aims to broaden perspectives on evaluation theories focused on negotiation text skills. Practically, the validation results are intended to serve as a concrete guide for educators in developing test instruments that meet technical standards while remaining aligned with pedagogical needs.

METHODS

This study employed a quantitative descriptive design to examine the validity of a teacher-developed assessment instrument for negotiation text analysis. The design was chosen to allow systematic description of empirical score patterns generated from expert validators and student test results, without manipulating any variables. Data consisted of numerical ratings from expert panels on each test item, as well as students' performance scores on the negotiation text assessment, which were then analyzed to portray the instrument's validity profile across several dimensions.

Within this quantitative descriptive framework, the analysis focused on summarizing central tendencies and ranges of validator scores, categorizing items into "accepted," "requiring revision," and "rejected" based on predetermined cut-off values, and comparing student achievement with the school's Minimum Proficiency Threshold (KKM). Patterns of correct and incorrect responses were also inspected to identify potential issues in item difficulty and distractor quality, thereby providing empirical indications of content, construct, and criterion validity. This design enabled an objective,

data-driven evaluation of how well the assessment instrument measures students' analytical competence in negotiation texts.

In this study, the researchers did not select all students as respondents, but instead used a subset of them as a sample. This was done to simplify data collection and processing, given the limitations of time, effort, and cost. Furthermore, the sample, selected using specific techniques, was expected to be representative of the population's characteristics, ensuring the validity of the research results and reflecting the actual situation. The study participants consisted of 20 tenth-grade students. (11 females and 9 males) at a public senior high school in Central Sulawesi. Additionally, an Indonesian language teacher served as the key informant and the developer of the instrument. The research site was selected via purposive sampling, addressing the institution's requirement for standardized assessment tools specifically for the negotiation text genre.

Data were derived from primary and secondary sources. Primary data comprised students' analytical competency test results and teacher insights regarding the assessment design. Secondary data were obtained through classroom observation documentation and instructional practices recorded during the 2023/2024 academic year. All analyzed items were in a multiple-choice format, specifically designed to measure High-Order Thinking Skills (HOTS).

A relevant cognitive framework for HOTS in this study can be grounded in the upper levels of the revised Bloom's taxonomy, namely *analyzing*, *evaluating*, and *creating*. In the context of negotiation text assessment, *analyzing* involves the ability to identify communicative purposes, decompose the text into its structural stages, and recognize the relationships between arguments, counterarguments, and negotiation strategies. *Evaluating* refers to making reasoned judgments about the effectiveness, coherence, and fairness of negotiation moves, including assessing the soundness of evidence and the appropriateness of language choices. *Creating* may encompass designing alternative negotiation scenarios, reconstructing more effective arguments, or proposing improved strategies to reach consensus.

Operationally, HOTS-oriented items should therefore require students to move beyond locating explicit information or recalling definitions, and instead engage them in interpreting textual patterns, comparing competing positions, and justifying their choices with reference to textual evidence. Such items typically demand multi-step reasoning, integration of textual and contextual cues, and reflection on possible consequences of different negotiation strategies. In this way, the cognitive framework ensures that the assessment genuinely targets higher-order analytical competence, rather than merely rebranded forms of lower-order comprehension.

The data analysis followed a systematic process involving collection, identification, classification, description, and conclusion. Validity assessment was conducted through the following procedures:

- **Content and Construct Validity:** Expert panel scores were analyzed descriptively to determine the mean, minimum, and maximum values for each item. Items were categorized as "accepted" (≥ 8), "requiring revision" (6–7.9), or "rejected" (< 6).
- **Item Analysis:** to identify potential issues regarding item difficulty, which were subsequently cross-referenced with qualitative feedback from the validators.

FINDINGS

This section presents empirical findings on the validity of assessment instruments used to measure the competence of students in class X in analyzing negotiation texts. The evaluation was conducted comprehensively, covering three main dimensions: content validity, construct validity, and criterion validity. In general, the developed instrument has met the eligibility criteria, although there are several items that require technical revision.

1. Content Validity

Based on the analysis results, the evaluation instruments developed by teachers were in line with the targeted competency achievement indicators. Content validity ensured that the measuring instruments were representative of the negotiation text material domain being taught.

Matrix kuantitatif. Results of Content Validation of Negotiation Text Analysis Instruments		
Assessment Aspects	Empirical Findings	Scientific Reasons
Indicator Suitability	The questions are in line with the curriculum guidelines.	Instruments with good content validity must be able to accurately measure specific learning objectives.

Content Composition	The tested content covers the aspects of urgency, relevance, and continuity.	This is crucial to ensure that the evaluation results provide a comprehensive picture of student competency.
----------------------------	--	--

2. Construct Validity

Construct validity was tested to ensure that each item truly measured students' analytical abilities, rather than other external factors. The score data from the validator panel showed a range of values between 7 and 10.

- **Maximum Score:** Four validators gave a perfect score (10), indicating that most items effectively operationalized the skill indicators.
- **Revision Notes:** The lowest score (7) was given by three validators, suggesting that some items were less than optimal.
- **Identification of Weaknesses:** Researchers found that item 2 contained a snippet of the correct answer, while item 5 provided explicit clues that made it easy for students to guess the answer.

The weaknesses in questions 2 and 5 risk distorting the assessment results because students no longer use their critical thinking skills but only rely on the clues in the questions.

3. Criterion Validity

The validity of the criteria in this instrument covers the technical aspects of question writing and the quality of distractors. The validation show several important points:

- **External Criterion (Criterion Variable):** A "gold standard" measure with established validity, such as actual performance, a clinical diagnosis, or the behavioral outcome to be predicted.
- **Correlation Analysis:** The use of statistical methods (such as Pearson correlation) to determine the strength and direction of the relationship between test scores and the criterion. The higher the correlation, the higher the criterion's validity.
- **Objectivity:** The criterion should be a measurable, tangible, and objective outcome (e.g., sales figures, test scores, attendance).

4. Student Performance Analysis

Empirical data from the implementation of the instrument shows that all 20 students (11 female and 9 male) achieved or exceeded the minimum passing grade of 70. The scores ranged from 70 to 100. This achievement not only demonstrates the students' mastery of the material, but also confirms that the instrument functions reliably as a measure of analytical competence in the genre of negotiation texts.

DISCUSSION

The primary focus of this study is to verify that the assessment instrument accurately measures students' negotiation text analysis competence in accordance with curricular mandates. Findings indicate that content validity was adequately established, as each test item was systematically aligned with the targeted learning indicators. This reinforces the theoretical premise that content validity serves as a critical safeguard, ensuring the instrument is both representative and comprehensive of the measured domain. As emphasized by Arikunto (2011), congruence between test material and specific learning objectives is an absolute prerequisite for an evaluation tool to generate accurate data.

Regarding criterion validity, the results confirm that the instrument was developed with a mature consideration of technical, ethical, and objective dimensions. The use of neutral language, free from socio-cultural biases, and the inclusion of logical distractors are primary strengths that ensure fairness for all examinees. Nevertheless, the repetitive phrasing observed in the response options for items 8 and 9 warrants serious attention. Maintaining option homogeneity is essential for optimizing the function of distractors, thereby providing the instrument with the sharp discriminative power required to differentiate nuanced levels of student analytical proficiency.

From a construct validity perspective, the instrument generally succeeds in operationalizing analytical thinking through its systematic item structure. However, an analysis of items 2 and 5 revealed technical deficiencies that potentially diminish this validity. The presence of "partial definitions" and explicit cues within these items allows students to identify correct answers without engaging in rigorous analysis. Revising these segments is crucial to ensure the instrument consistently evaluates pure linguistic competence and remains free from extrinsic biases that could misinform instructional decision-making.

The alignment between expert validation results and empirical student performance underscores the instrument's functional reliability. Indicates that the tool is not only theoretically sound but also pedagogically effective in measuring competency within the negotiation text genre. Collectively, these findings demonstrate that through a systematic validation process, educators can develop credible assessment tools that provide objective insights into the development of student literacy.

This research is a first step toward multidimensional validation, which currently focuses only on curriculum alignment (Content) and expert assessment and basic item analysis (Construct & Item Analysis). It was openly acknowledged in the discussion that Messick's unified validity test would not be fully established until further psychometric testing (such as reliability) was conducted.

The results also speak to Bachman and Palmer's notion of *test usefulness*, which emphasizes that validity must be considered alongside practicality, reliability, authenticity, and impact. The evidence that all students reached or exceeded the minimum passing grade, combined with expert judgments on item quality, suggests that the instrument has at least a baseline level of usefulness for classroom decision-making. However, the identification of explicit answer cues in items 2 and 5 and redundant wording in items 8 and 9 highlights potential threats to construct representation and score meaning. These technical flaws indicate that, in its current form, the test may partially inflate students' apparent analytical performance by allowing them to rely on surface clues rather than genuine reasoning, thus challenging the strength of the construct-related evidence envisioned in international validation frameworks.

Furthermore, the emphasis on HOTS-oriented multiple-choice items situates this work within broader debates about how well selected-response formats can capture higher-order language abilities. International literature has long cautioned that multiple-choice tasks risk privileging and may underrepresent complex, interactive skills. The present findings confirm that careful item writing, particularly the removal of key-word clues and redundant options is essential if such items are to approximate the analytical and evaluative processes described in higher-order taxonomies. By documenting both the strengths and limitations of the current instrument, and by recommending further psychometric analyses such as reliability estimation, this study contributes to ongoing international conversations about how classroom-based language assessments can be made more valid, fair, and instructionally informative, especially in contexts where teachers themselves design the tests they use.

CONCLUSION

This study establishes that the assessment instrument developed to measure negotiation text analysis competence among senior high school students (n=20; 11 females, 9 males) satisfies the established validity criteria. The instrument validity across content, construct, and criterion dimensions. Content validity is evidenced by the systematic alignment of test items with curriculum indicators, while construct validity confirms the instrument's capacity to operationalize students' analytical skills effectively.

The effectiveness of the instrument is further strengthened by student performance data and although all students passed the KKM, these high scores must be interpreted with great caution given the weaknesses in some of the test items, wherein all participants reached or exceeded the Minimum Proficiency Threshold (KKM) of 70, with scores ranging from 70 to 100. These results not only reflect student mastery of the subject matter but also for generating substantive instructional insights within the negotiation text genre.

Despite these strengths, this study has several limitations that should be acknowledged. First, the sample was restricted to 20 students from a single class at one senior high school, which constrains the generalizability of the findings to broader populations and school contexts. Second, the validation process focused primarily on descriptive analyses of expert ratings and student scores, without incorporating more advanced psychometric procedures such as reliability estimation, item discrimination indices, or factor analysis that could provide deeper evidence of construct validity. Third, the instrument was limited to multiple-choice items, so the conclusions about students' analytical competence are confined to what can be captured through selected-response formats and do not encompass productive skills such as written or oral negotiation performance.

Based on the findings, the following recommendations are proposed to optimize the instrument's quality:

- **HOTS assessment developers** should strictly minimize explicit answer cues, which are often inadvertently embedded in long text-based questions."
- **Retain the recommendation** to expand the psychometric evaluation), but add a recommendation for future researchers to test these instruments using Item Response Theory

(IRT) or Classical Test Theory (CTT) with a much larger and more representative sample size to ensure the stability of item quality.

- **Editorial Enhancement:** Item formulation should consistently prioritize High-Order Thinking Skills (HOTS) over verbatim textual citations to minimize the influence of rote memorization.
- **Expansion of Psychometric Evaluation:** Future research should expand the scope of validation to include statistical reliability testing, thereby strengthening the instrument's overall psychometric foundation.

Taken together, these findings affirm that a systematically constructed and carefully validated assessment instrument can serve as a robust lever for aligning classroom practice with curriculum demands for higher-order analytical engagement with negotiation texts. By foregrounding multidimensional validity evidence and explicitly linking expert judgment to empirical student performance, this study offers a concrete model for how teacher-developed tests in Indonesian language education can be strengthened to support fair, transparent, and instructionally meaningful decisions about students' literacy development.

REFERENCES

- Aisyahara, Erindi, and Desi Sukenti. 2025. "Analysis of Hots-Based Indonesian Language Examination Questions in Class X Kansai Vocational School." *Mimbar Pendidikan* 10 (1): 32–47. <https://doi.org/10.17509/mimbardik.v10i1.81046>.
- Alfiani, Calvin Reindira, Abdul Rani, and Sri Wahyuni. 2024. "Ketidaksesuaian Materi Dan Asesmen Pada Pembelajaran Bahasa Indonesia Teks Negosiasi." *Imajeri: Jurnal Pendidikan Bahasa Dan Sastra Indonesia* 7 (1): 101–13. <https://doi.org/10.22236/imageri.v7i1.15756>.
- Arikunto, Suharsimi. 2011. *Dasar-Dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Bachman, Lyle F., and Adrian S. Palmer. 2010. *Language Testing in Practice: Designing and Developing Useful Language Tests*. Oxford: Oxford University Press.
- Brown, H. Douglas, and Priyanvada Abeywickrama. 2010. *Language Assessment: Principles and Classroom Practices*. 2nd ed. White Plains, NY: Pearson.
- Fulcher, Glenn, and Fred Davidson. 2007. *Language Testing and Assessment: An Advanced Resource Book*. London: Routledge.
- Gusvenisa, Yulsi, Zherry Putra Yanti, and Atika Gusriani. 2023. "Validitas Pengembangan Modul Menulis Teks Negosiasi Berbasis Scaffolded Writing Pada Siswa Kelas X SMAN 6 Padang." *Jurnal Pendidikan Tambusai* 7 (3): 20672–77. <https://doi.org/10.31004/jptam.v7i3.9544>.
- Kemdikbud. 2017. *Bahasa Indonesia Kelas X: Teks Negosiasi (KD 3.10/4.10)*. Jakarta: Kementerian Pendidikan dan Kebudayaan Republik Indonesia.
- Masnawati, Muhammad Hanafi, Jamaluddin Ahmad, Rustam Efendy Rasyid, Yusmah Yusmah, and Andi Astina Adnan. 2025. "Pengembangan Bahan Ajar Menulis Teks Negosiasi Dengan Model Discovery Learning Berbantuan E-Book Di Kelas XI MAS DDI Al Mujahidin Bendoro." *Qalam : Jurnal Ilmu Kependidikan* 13 (2): 9–17. <https://doi.org/10.33506/jq.v13i2.3785>.
- Messick, Samuel. 1989. "Validity." In *Educational Measurement*, 3rd ed., edited by Robert L. Linn, 13–103. New York: Macmillan.
- Messick, Samuel. 1996. "Validity and Washback in Language Testing." *ETS Research Report Series* 1996 (1): 1–18.

Popham, W. James. 2014. *Classroom Assessment: What Teachers Need to Know*. 7th ed. Boston: Pearson.