Assessing the Authenticity and Validity of Tests in CLT-Based English Classrooms in West Java's Formal Education

Fella Rosy[1], Nida Husna[2], Ummi Kulstum[3], Muhammad Farkhan[4]
[1,2,3,4] Master of English Education, UIN Syarif Hidayatullah, Jakarta
Corresponding E-Mail: fellarosy5@gmail.com

Abstract

This study investigates the effectiveness of classroom tests in formal West Java, Indonesia, secondary learning centers that claim to offer Communicative Language Teaching (CLT). Because CLT places significance on fostering communicative competence that contains grammatical, sociolinguistic, discourse, and strategic dimensions, there are concerns that the testing protocols employed may not adequately reflect the communicative goals. The data were collected from four English as a Foreign Language (EFL) instructors through the use of a qualitative multiple case study design by way of document analysis and semi-structured web survey. It was found that, while respondents demonstrated satisfactory theoretical knowledge of communicative testing and the use of performance tasks such as interviews, role-plays, and spoken tasks incorporated into videos, full compliance with the tenets of CLT is restricted. Principal impediments to successful implementation are restricted professional development opportunities, large classes, restricted instructional hours, and institutional policies favoring grammar-based tests. It is the finding of this research that, despite professed adherence to CLT, test validity is inconsistent. It is its recommendation that teacher assessment literacy must be upgraded, wide-ranging training needs to be implemented, and institutional policies need to be redesigned to bring communicative test practices into closer correspondence to CLT ideals.

Keywords: Validity of assessment; CLT; Communicative competence; EFL

Introduction

Communicative Language Teaching (CLT) has been known for decades as one of the most powerful methodologies of English language teaching. It values not only knowledge of grammar but also proficiency in communicating effectively in everyday situations through interaction, focus on the learner, and

communicative tasks. As conceptualized by Canale and Swain (1980) and by Bachman (1990), communicative competence consists of four components grammatical, sociolinguistic, discourse, and strategic that collectively take language teaching out of the realm of memorization of drills toward communicative competence.

Extensive research from Indonesia confirms the positive impact of CLT on English language proficiency for learners. For example, Hasibuan et al. (2021) found that communicative activities in the classroom significantly increased speaking fluency and speaking confidence for learners. Lumy (2018) showed that real-world activities and task-based activities increased the motivation of learners, while Trimadona et al. (2024) showed that pair work and group work increased interactional proficiency for learners. Collectively, these studies indicate the potential of CLT in transforming English language classes in EFL context when fully executed.

Yet, despite these encouraging results, various studies revealed continued problems in CLT implementation. Saragih (2019) commented that CLT was superficially utilized, with restricted student involvement and teacher-oriented pedagogy prevalent in lessons. Yustina and Hikmah (2020) identified another gap between "CLT"-labeled lesson plans and actual classroom procedures, which continued to be translation and grammar-oriented. Wulandari (2021) also identified that open-ended communicative tasks overpowered students due to the lack of sufficient scaffolding and teacher demonstration, minimizing active involvement. These results reveal that CLT effectiveness largely relies on contextual adaptation, teacher preparation, and ongoing assessment reform rather than adopting CLT tag.

Within Indonesia's formal secondary schooling public junior and senior high schools that run under the national curriculum and common examinations the application of CLT has been encouraged by policy reforms towards student-centered learning. The research is centered upon Bandung and Depok (West Java) because they are capital cities in which CLT-based programs have been rolled out and from which access to CLT rooms was practical. Implementation is nevertheless sporadic. Systemic barriers faced by many teachers encompass limited professional development, an examination-centric school culture, and limited autonomy over test design. English tests in these schools tend therefore to become again tests of grammar-based knowledge that do not assess communicative competence, losing instructional coherence and test validity.

This paradox of pedagogy grounded in CLT and regular test-taking protocols reveals a need for research. While numerous works examined the pedagogical value of CLT in the language classroom, few took seriously the validity of classroom tests that claim to comply with CLT principles. In fact, the overwhelming number of users of communicative teaching methodology in the classroom persist in administering tests that focus upon discrete points of

grammar rather than communicative reality. Those procedures directly contravene CLT maxims and the construct validity and coverage of classroom tests.

To fill this gap, the current research tries to explore to what extent West Java's formal CLT-based English classes' classroom tests reflect the communicative competence model. Specifically, it raises the following questions:

1. What kinds of tests are used by English teachers in West Java's formal CLT classrooms?
2. To what extent are these tests calibrated to the communicative competence model?
3. To what extent do these tests of classroom communication reflect communicative competence?

By focusing on assessment validity through the lens of communicative competence, this study contributes an underexplored dimension to the Indonesian EFL literature. It offers evidence on how far classroom assessments have moved beyond traditional grammar testing and suggests implications for teacher assessment literacy, professional development, and institutional policy reforms aimed at aligning classroom tests with CLT principles.

Method

This study employed a qualitative multiple case study method to examine the English tests administered within Communicative Language Teaching (CLT)–oriented classrooms of West Java's formal secondary schools, that is, in Bandung and Depok. The study utilized a case study design as it allows for in-depth examination of multifaceted real-life teaching practices, namely contextualized language testing and assessment phenomena (Merriam & Tisdell, 2016; Yin, 2018). Having multiple cases also facilitated cross-case comparison, and thus the study was able to identify patterns and contrasts more effectively between schools.

Participants

Two junior/senior high schools were selected purposively because they explicitly identified themselves as CLT-based schools and had incorporated communicative assessment practices (i.e., performance tasks and project assessments) into their English courses. These strategic samplings ensured that the schools were settings where CLT measurement was no longer hypothetical but was being used. There were four English teachers participating in the study, two of whom came from a junior high school in Padalarang (Bandung Barat) and two from senior high schools in Kelapa Dua and central Bandung. They were all engaged in the development and administration of English tests for the four skills

in listening, speaking, reading, and writing, and were trained or experienced with communicative language assessment. They volunteered, and recruitment was carried out through a Google Form that was circulated among professional educator networks.

Data Collection Procedures

Data were gathered during March to May 2025 using two main instruments:

Document Analysis: Assessment papers from class, rubrics for tests, and related documents (scoring guides, lesson plans) were obtained from participants themselves and anonymized.

Online Semi-Structured Questionnaire: This was employed to gather teachers' beliefs, understanding, and tacit knowledge regarding communicative assessment. The questionnaire was piloted with two non-participant teachers to test for clarity and face validity before formal use.

Both sources were applied following Creswell and Poth's (2018) methodological triangulation guideline, which ensured the findings' validity by aggregating from different data streams.

Data Analysis

Questionnaire answers and documents were transferred to NVivo for qualitative coding. A three-stage approach: (1) open coding to create initial categories of communicative assessment practice, (2) axial coding to link categories to themes (e.g., task types, coherence with CLT principles), and (3) cross-case analysis to compare themes across schools. Coding was conducted in solitude by the lead researcher and verified by a peer researcher to ensure maximum dependability.

Validity and Trustworthiness

Apart from triangulation, several steps were taken to enhance trustworthiness: member checking of preliminary interpretation with participants, peer debriefing, and record of audit trails of decisions on analysis. These precautions enhanced credibility, dependability, and confirmability of results.

Ethical Considerations

This study received ethical clearance from university prior to data collection. Informed consent and confidentiality were promised by all participants. Pseudonyms were used in place of teacher and school names.

Participant Profile and School Context

The four instructors had varied demographics: three female and one male; two between 20–30, one between 31–40, and one over 40; with teaching experience of less than three years, to more than a decade. This variation yielded a rich data set that exhibited how communicative assessment is conceptualized and implemented in varying institutional environments. Both schools taught a national curriculum but possessed varying levels of autonomy in the design of the tests one prioritizing performance-based activities, the other still combining communicative and traditional testing both presenting contrasting case studies for examination.

Results

Survey Instrument and Conceptual Structure

Data was collected from four English teachers in formal secondary schools in West Java using a semi-structured online survey questionnaire. Its purpose was to investigate, in the context of Communicative Language Teaching (CLT), the validity and authenticity of communicative evaluation procedures. The questionnaire was divided into three sections, as indicated in Table 1: (1) participant demographics; (2) teachers' assessments of communicative competence based on Canale and Swain's (1980) model; and (3) Likert-scale items modified from Bachman and Palmer's (1996) test usefulness framework, emphasizing construct validity and authenticity.

By mapping teachers' theoretical knowledge, task design techniques, and perceived problems, this structure enabled the study to address all three research topics.

Table 1. Research Instrument Design and Source Framework

Section	Item Type	Description	Source Framework
1	Demographic	Age, Gender, Teaching Experience, School Location	Self-developed
2	Open-ended	Definitions of Communicative Assessment, Task Types	Canale & Swain (1980)
3	Likert-scale	Frequency of Real-Life Tasks, Fluency vs Accuracy Focus	Bachman & Palmer (1996)

Knowledge of Communicative Assessment

The results showed that each of the four participants had a solid theoretical grasp of communicative assessment. Teachers consistently stressed the contextual and interactive character of assessment over a limited grammar focus across all open-ended responses. This fits the paradigm proposed by Canale and Swain (1980).

Table 2. Teachers' Definitions of Communicative Assessment

Respondent	Definition
R1	"Assessment that measures contextual use of language, not just grammar."
R2	"Assessment to measure ability to communicate in English."
R3	"An assessment designed to evaluate interaction with peers or teachers."
R4	"Oral evaluation focusing on contextual use and accuracy."

Direct quotes illustrate how teachers framed assessment:

R1: "It's not about right/wrong grammar only; it's about how well students can express ideas in real situations."

R3: "We try to simulate authentic interaction, like interviews, role-plays or presentations, to see actual language use."

Types of Assessment Tasks and Language Aspects

Three of the four teachers reported frequent use of performance-based tests (role-plays, vlogs, project presentations, structured discussions) which align with Bachman and Palmer's (1996) authenticity and interaction criteria. One teacher continued to rely primarily on written grammar-based tests, illustrating a theory–practice gap. 75% of participants (3/4) reported they "always" use interactive or performance tasks at least once per unit. 25% (1/4) admitted using discrete-point grammar tests as their main assessment because of institutional requirements. Language aspects emphasized included fluency, vocabulary, coherence, and context appropriateness features central to communicative competence. Teachers also reported balancing grammatical accuracy with the ability to sustain real-life communication.

Real-Life Interaction and Task Authenticity

All teachers included at least some student–student or student–teacher interaction in their tests. Three out of four teachers rated their use of authentic tasks as 5 out of 5 on the Likert scale, indicating high commitment to real-world language application. Examples included: "Mini-interviews" simulating job or airport situations. "Student vlogs" reporting on personal experiences. "Debate tasks" on current issues in English. This aligns with Richards' (2006) view that meaningful, authentic contexts enhance language learning.

Implementation Challenges

Despite strong theoretical understanding, several constraints limited full implementation of communicative assessment: Large class sizes (over 35 students in some classes). Time restrictions for performance-based tests. Low student motivation in non-graded speaking activities. Direct participant comments highlight the tension: R2: "I want to do more role-plays, but with 40 students and only 90 minutes, it's impossible to assess everyone." R4: "Students still think grammar tests are easier and safer, so they put minimal effort into speaking tasks." These barriers mirror previous findings by Saragih (2019) and Masduqi & Prihananto (2020) on systemic constraints to CLT assessment.

Practicality and Professional Development

Teachers varied in their ability to integrate technology to support assessment. Two teachers reported using Google Classroom, WhatsApp, or video submissions to manage large-class assessments. However, grading time and preparation remained major obstacles. Only one teacher (25%) had received formal training in communicative assessment, although all four expressed strong interest in future workshops, peer learning, and mentoring. This suggests significant potential impact from focused professional development programs.

Reliability Across Data Sources

Comparing questionnaire responses with collected test documents revealed partial alignment between reported practices and actual test formats. While teachers described interactive tasks, some test papers were still dominated by grammar and vocabulary items. This indicates inconsistent implementation and highlights the need for follow-up classroom observations or test-task audits to strengthen reliability of findings.

Discussion

To answer each research question, this discussion incorporates the key findings with the theoretical framework of the study and pertinent literature. It underlines the consequences for policy and practice, compares teaching methods with CLT principles, and uses Table 3 as the anchor.

Table 3. Summary of Teachers' Practices, Perceptions, and Challenges in CLT-

Based Communicative Assessment

Respondent	Types of Communicative Tests Used	Perception of Communicative Competence	Aspects Assessed	Implementation Challenges	Assessment Training
T1	Vlogs, interviews, group discussions, role plays, presentations, projects	Meaning-focused; fluency and interaction prioritized over grammar	Grammar, fluency, vocabulary, content, relevance to topic	Student motivation; ensuring engagement in performance tasks	No
T2	Interviews, pair/group tasks, written projects	Emphasizes message delivery and fluency	Vocabulary, fluency, coherence, interaction	Large class sizes (~40); limited time	No
T3	Role play, speaking tasks, writing tasks, interviews	Communication over correctness; students assessed on performance	Grammar, fluency, idea development	Time-consuming to prepare; limited reliability	No
T4	Class discussions, written tasks, small speaking groups	Based on curriculum but recognizes importance of interaction and content	Fluency, vocabulary, comprehension, creativity	Only feasible with small classes; digital tool limitations	Yes

Types of Communicative Tests Used by Teachers

Each of the four participants reported using performance-based assessments that mimic authentic communication, such as vlogs, interviews, presentations, group debates, and projects. These fit the task authenticity and interaction criteria of Bachman and Palmer (1996), indicating a common goal to represent real-life language use in evaluation. Teacher 4's continued heavy reliance on written assignments and curriculum-bound tests, however, suggests inconsistent implementation.

This is consistent with the findings of Elmiana (2018) and Rizqi (2020), who discovered that time constraints, class size, and teacher confidence frequently cause communicative evaluation procedures in Indonesia to become fragmented. In contrast, worldwide research (e.g., Fulcher, 2021; Koh, 2021) demonstrates that instructors are better equipped to create realistic, high-validity exams when schools offer systemic assistance, such as lower student-teacher ratios and regular training. This implies that the full implementation of CLT-based evaluation in Indonesian classrooms is still hampered by structural factors.

Coverage of Communicative Competence

A theoretical concept of communication competence as contextual, meaningful, and interactional was shared by the teachers. The four competences (grammatical, discourse, sociolinguistic, and strategic) identified by Canale and Swain (1980) are reflected in this. In actuality, though, the coverage was selective. While Teachers 3 and 4 focused on grammar or curriculum-linked objectives, Teachers 1 and 2 employed more thorough rubrics. Interestingly, in all four cases, there was a lack of strategic competence (managing communication breakdowns or meaning negotiation).

This gap between theory and practice aligns with Dewi et al. (2025), who reported that Indonesian speaking tests prioritize performance over interactive negotiation. In international settings, strategic competence is often embedded in role-play or spontaneous tasks (e.g., "repair strategies" scoring in Singapore's oral exams), suggesting one actionable recommendation: Indonesian teachers could adapt rubrics to explicitly include strategic competence.

Validity Issues and Implementation Challenges

Teachers' intent to design authentic, purposeful tasks was constrained by situational factors time limits, student motivation, class size, and technological preparedness. These factors undermine the validity principles of Bachman and Palmer's (1996) test usefulness model, particularly construct validity and authenticity.

For example, Teacher 2's 40-student class size makes individualized oral testing impractical, and Teacher 3 cited long preparation times for project-based testing. This highlights the need for institutional measures such as rotating oral assessment schedules, digital task banks, or co-assessment with colleagues.

The absence of professional training only Teacher 4 had received official assessment training was another significant problem. According to Saragih (2019) and Masduqi & Prihananto (2020), quality improvement requires institutional support and teacher assessment literacy. This implies a policy implication: to assist teachers' professional development, ministries or school boards could provide structured training programs and mentorship programs with a CLT focus.

Bridging the Gap Between Theory and Practice

This study highlights a clear disconnect between teachers' conceptual understanding and their classroom implementation. While the theoretical underpinnings are strong, practical realities dilute fidelity to CLT principles. Such a gap mirrors findings in other EFL contexts where CLT adoption has been policy-driven but under-resourced (Koh, 2021). One way forward is to provide concrete, classroom-friendly assessment frameworks e.g., task templates, scoring rubrics including strategic competence, or shared online banks of authentic tasks. Teachers also expressed enthusiasm for workshops, simulations, and peer mentoring. These forms of professional development could be aligned with the Ministry of Education's "Merdeka Belajar" policy to ensure both top-down and bottom-up support.

Implications for EFL Teaching and Policy

For Teachers: Develop rubrics that explicitly assess strategic competence; use digital platforms (Google Classroom, Padlet, Flipgrid) for large-class speaking tasks; and incorporate peer-assessment to ease grading loads.

For Schools: Adjust schedules to allow for performance-based assessment and provide co-assessment opportunities.

For Policy Makers: Integrate communicative assessment training into teacher certification programs, and create incentives for schools to adopt authentic testing practices.

For Future Research: Explore longitudinal effects of teacher training on test validity, and compare rural vs. urban schools to identify systemic disparities.

Study Limitations

This study focused on four teachers in two urban centres in West Java, limiting generalizability to rural or private school contexts. Data collection relied on self-report and document analysis; direct classroom observation or student outcomes would enrich future research. Despite these limits, the study offers valuable insight into the lived realities of implementing CLT-based assessment in Indonesian classrooms.

Conclusion

This study shows that English teachers in West Java's formal secondary schools possess a solid theoretical grasp of communicative assessment and are beginning to implement promising practices such as authentic tasks and interaction-oriented testing. Nonetheless, three persistent challenges insufficient training, limited instructional time, and institutionally entrenched test norms continue to restrict the full realization of CLT-aligned assessment in the classroom. To strengthen the validity and effectiveness of communicative testing, three priority actions are recommended:

1. Professional capacity building through sustained workshops, peer mentoring, and hands-on training in communicative test design and scoring.

2. Structural support such as smaller class sizes, dedicated assessment time, and digital tools to make authentic assessment feasible.

3. Assessment literacy development for both teachers and policymakers to ensure stronger alignment between pedagogy, curriculum, and testing systems.

Future research should broaden the sample to include diverse schools and incorporate students' perspectives alongside classroom observations. Applying mixed-method validation techniques such as inter-rater reliability checks, item analysis, and longitudinal tracking will provide richer evidence of the validity and impact of communicative assessment in Indonesian EFL contexts. By addressing these areas, teachers and policymakers can move closer to achieving truly communicative, valid, and context-responsive assessment practices.

References

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1–47. https://doi.org/10.1093/applin/I.1.1

Cohen, L., Manion, L., & Morrison, K. (2017). *Research methods in education* (8th ed.). Routledge. https://doi.org/10.4324/9781315456539

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.

Daar, J., & Ndorang, D. (2020). The impact of role play and pair work on students' speaking skills. *Jurnal Pendidikan Bahasa Inggris Indonesia, 8*(1), 45–56.

Dewi, R. S., Hidayati, N., & Putri, R. (2025). Enhancing students' listening and speaking through CLT. *Language Teaching Journal, 10*(1), 34–47.

Effendi, H., & Novita, R. (2024). Barriers to implementing CLT in Indonesian classrooms. *ELT Journal of Indonesia, 15*(2), 88–99.

Elmiana, N. (2018). Teaching English in an Indonesian vocational high school: A classroom-based study. *Indonesian Journal of Applied Linguistics, 8*(3), 576–585. https://doi.org/10.17509/ijal.v8i3.15272

Fulcher, G. (2003). *Testing second language speaking*. Pearson Longman.

Hasibuan, H., Lubis, A. H., & Rangkuti, R. (2021). Enhancing students' communicative competence through CLT approach. *Jurnal Ilmiah Mandala Education, 7*(4), 412–421. https://doi.org/10.36312/jime.v7i4.2323

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research, 15*(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Lestari, H., & Margana, M. (2024). CLT implementation challenges in urban high schools. *Journal of English Language Teaching and Linguistics, 9*(1), 102–117.

Lumy, M. (2018). The effectiveness of CLT in increasing students' motivation and speaking fluency. *Journal of English Language and Education, 4*(1), 1–10. https://doi.org/10.31004/jele.v4i1.100

Masduqi, H., & Prihananto, H. (2020). Challenges of applying CLT in Indonesian high schools. *Indonesian Journal of English Language Teaching and Applied Linguistics, 4*(2), 201–216. https://doi.org/10.21093/ijeltal.v4i2.588

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education/Macmillan.

Rafiud Ilmudinulloh, M., Azzahro, M., & Wibowo, F. (2022). Kampung Inggris: An informal CLT-based learning model. *Journal of English Language Studies, 7*(2), 77–85. https://doi.org/10.30870/jels.v7i2.13045

Richards, J. C. (2006). *Communicative language teaching today*. Cambridge University Press.

Rizqi, M. A. (2020). The implementation of communicative language teaching in senior high school. *Indonesian Journal of English Language Teaching and Linguistics, 5*(2), 245–264. https://doi.org/10.21093/ijeltal.v5i2.627

Saragih, A. (2019). Teachers' perception of CLT implementation in senior high school. *International Journal of English Teaching and Learning, 7*(1), 55–70. https://doi.org/10.32939/ijetl.v7i1.273

Savignon, S. J. (2002). *Interpreting communicative language teaching: Contexts and concerns in teacher education*. Yale University Press.

Salam, F., & Luksfinanto, R. (2024). National exam policy and CLT conflict. *Journal of English Pedagogy, 12*(1), 44–59.

Trimadona, R., Wati, S., & Azzahra, N. (2024). Enhancing speaking competence through pair work in CLT classrooms. *Indonesian Journal of Educational Studies, 11*(1), 22–35.

Wulandari, S. (2021). Teacher challenges in implementing CLT in junior high schools. *Journal of Language and Literature Education, 7*(1), 101–112. https://doi.org/10.23887/jlle.v7i1.29877

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.

Yustina, N., & Hikmah, R. (2020). Investigating classroom implementation of CLT in public schools. *TEFLIN Journal, 31*(1), 13–31. https://doi.org/10.15639/teflinjournal.v31i1/13-31