

NAVIGATING CHALLENGES IN AUTHENTIC SPEAKING ASSESSMENT: INSIGHTS FROM AN INDONESIAN EFL TEACHER IN AN ELEMENTARY SCHOOL

Ina Safitri Siregar¹, Nur Aisyah Zulkifli²

inasyafitri2002@gmail.com¹, nur'aisyah.zulkifli@uin-suska.ac.id²

Uin Sultan Syarif Kasim Riau

ABSTRACT

Authentic speaking assessment is crucial for assessing students' communication skills; however, its implementation in Indonesian EFL classrooms remains underexplored. This qualitative study examines the challenges faced by Indonesian EFL teachers in conducting authentic speaking assessments and the strategies they use to address them. Data were collected through semi-structured interviews with English teachers in purposefully selected elementary schools. The findings highlight several challenges, such as large class sizes, long learning times, limited student abilities, limited assessment tools, and difficulties in achieving assessment objectives. To address these challenges, teachers use strategies such as peer assessment, structured speaking tasks, simplified rubrics, and informal observation. The study highlights the need for institutionalization, teacher training, and broader training. These findings highlight several challenges, such as large class sizes, long learning times, limited student abilities, limited assessment tools, and difficulties in achieving assessment objectives. To address these challenges, teachers use strategies such as peer assessment, structured speaking tasks, simplified rubrics, and informal observations. This study highlights the need for greater institutionalization, teacher training, and increased assessment resources to improve authentic speaking assessment practices in Indonesian schools.

Keywords: Authentic Speaking Assessment, EFL Teachers, Elementary Schools, Language Assessment, Qualitative Study.

INTRODUCTION

Speaking ability is one of the essential skills students must possess when learning English as a foreign language (EFL). As the primary form of written communication, speaking ability allows one to express thoughts, feelings, and ideas directly to others in an oral context. According to the communicative frameworks of Canale and Swain (1980) and Hymes (1972), effective communication requires not only grammatical structures but also discourse, sociolinguistic, and strategic skills. In the context of EFL teaching, speaking ability presents a particular challenge because students lack an English-speaking environment outside the classroom. Consequently, limited exposure to the target language makes this skill highly beneficial for the ongoing learning process in schools. Therefore, the development of speaking skills in EFL learning needs to receive serious attention, especially from the elementary school level as an initial foundation for language mastery.

Assessment is an essential part of the language learning process and cannot be separated from classroom activities. Through assessment, teachers can assess students' learning progress, identify challenges they face, and provide useful suggestions to support the development of language skills (Brown, 2004; Bachman & Palmer, 2010). In language learning, assessment serves not only as a tool to improve learning outcomes at the end of the course (assessment of learning), but also as a tool to support and facilitate the learning process itself (assessment for learning and assessment as learning). However, speaking skills are considered one of the most important factors in language assessment. This is due to the spontaneous, real-time, and multifaceted nature of speaking, which requires careful consideration of valid and reliable instruments, implementation procedures, and evaluation criteria. However, speaking proficiency assessment is considered one of the most important

factors in language evaluation. This is due to the spontaneous, real-time, and multidimensional nature of speaking, which demands careful consideration of valid and reliable instruments, implementation processes, and evaluation criteria (Luoma, 2004). This complexity is heightened in the context of EFL classrooms, where teachers are concerned with time constraints, large student numbers, and significant differences in student proficiency levels.

Authentic assessment approaches emerge as a more relevant and meaningful alternative to traditional approaches to speaking assessment. According to Wiggins (1998) and Mueller (2006), authentic assessment is defined as a type of assessment that requires students to demonstrate their abilities through tasks that facilitate communication in everyday life. In the context of speaking skills, authentic assessment includes several activities such as role-playing, storytelling, dialogue, presentations, interviews, group discussions, and other simulated communication scenarios that encourage functional and contextual language use. This approach is more capable of improving students' communication skills holistically compared to traditional classroom teaching, which tends to only emphasize linguistic aspects partially. The relevance of speaking assessment is more important in the school context, where language teaching should focus on developing practical communication skills aligned with children's developmental stages (Cameron, 2001). Therefore, it is hoped that implementing assessment at this level can encourage active student participation and improve their ability to communicate more comprehensively and thoughtfully.

While authentic speaking skills offer several pedagogical advantages, their implementation at the elementary school level presents a number of complex and interrelated challenges. The characteristics of elementary school students as young learners are an important factor to consider. According to Harmer (2007) and Scott & Ytreberg (1990), children in this age group typically have short attention spans, are easily bored, and have few personal learning strategies. In other words, a lack of confidence in speaking a foreign language is a psychological barrier that often prevents students from actively participating in self-assessment of their abilities. Elementary school students' limited vocabulary also prevents them from expressing their ideas clearly and concisely. From a teacher's perspective, classroom management during speaking assessments presents unique challenges, particularly when dealing with individual students in large, dynamic classes. All of these factors collectively create significant challenges for EFL teachers in developing and implementing effective, clear, and consistent teaching methods that are appropriate to the classroom environment.

Several previous studies have examined the application of critical appraisal in EFL teaching, although most have focused on secondary and higher education contexts. These studies have, among others, identified challenges in speaking assessment in secondary schools (Fulcher, 2003; Luoma, 2004), investigated students' perceptions of critical appraisal in higher education (Boud & Falchikov, 2006), and investigated teachers' strategies for implementing speaking assessment in EFL classrooms in secondary schools (Davison & Leung, 2009). However, research specifically examining critical appraisal in school contexts remains quite limited. Only a few studies have addressed children's language learning in general without delving deeper into the perspectives and experiences of EFL teachers in the field. Furthermore, research aimed at methodically investigating EFL teachers in elementary schools particularly the unique challenges they face and the adaptive strategies they employ remains scarce, particularly in the Indonesian educational context. This highlights the need for more focused and contextualized research to understand the dynamics of assessment skills at the elementary school level.

Based on the existing literature, there are significant research findings regarding the difficulties faced by EFL teachers in implementing flipped assessment in elementary schools, particularly in the Indonesian context. Numerous studies have been conducted, but none have fully examined elementary school teachers' perspectives and methods in implementing assessment in the classroom. Understanding their methods and strategies is crucial for developing more effective assessment policies and practices. The unique characteristics of the Indonesian context, such as large class sizes, limited resources, and diverse sociocultural backgrounds, make it difficult to generalize research findings from other countries. Therefore, the purpose of this study is to address this by investigating the difficulties faced by Indonesian EFL teachers in implementing authentic speaking assessments in elementary schools, as well as the strategies they use to overcome these difficulties. Thus, it is hoped that this study can contribute both theoretically and practically to the development of more authentic, relevant, and accessible speaking assessments in the Indonesian educational context.

METODE PENELITIAN

1. Design

This research uses a qualitative approach with a case study design as the primary methodology. According to Creswell and Poth (2018), a qualitative approach was chosen because it aims to understand by considering the perspectives, experiences, and practices of EFL teachers in implementing authentic teaching methods in elementary schools, which cannot be fully understood through quantitative methods. According to Merriam and Tisdell (2016), a qualitative paradigm allows researchers to investigate phenomena contextually and interpretively, thus enabling the development of a more comprehensive practical and pedagogical approach. Within this qualitative framework, the research specifically adopts a single case study design. According to Yin (2018), a case study is a research methodology that focuses on investigating a single unit of analysis in this case, EFL teachers in elementary schools in the context of everyday life. The choice of a single case study design is based on the idea that the cases studied are representative and informative, thus allowing researchers to gain an in-depth understanding of the phenomenon being studied (Stake, 1995). With this design, the goal of the research is not statistical generalization, but rather to produce a generalizing analysis that can provide useful information for educational contexts (Yin, 2018).

2. Participant of the Study

Participants in this study were EFL teachers teaching at an elementary school in Indonesia. Participants were selected using purposive sampling, a sampling technique based on specific criteria related to the research objectives, rather than on probability or statistical representation (Creswell & Poth, 2018; Patton, 2015). The main criteria used in the selection process were as follows: (1) active English teachers at the school; (2) have at least two years of teaching experience to understand classroom teaching practices; and (3) have completed a comprehensive assessment of their EFL teaching skills at the school. Purposive sampling is based on the idea that in qualitative research, the quality of information obtained is more important than the quantity of participants (Patton, 2015). The researchers were able to maximize the quality and depth of the data collected by selecting individuals with direct experience related to the research topic. Participants were considered to be able to provide rich, comprehensive, and significant information (containing in-depth information) about the difficulties and approaches associated with implementing speaking proficiency assessment in elementary schools in Indonesia. The primary instrument used in this study was a semi-structured interview. According to Brinkmann and Kvale (2015), semi-

structured interviews were chosen because they allow researchers to explore information flexibly and in-depth using predetermined questions, while allowing participants to develop their knowledge in a clear and understandable manner. This approach was deemed most appropriate for investigating EFL teachers' subjective and personal perspectives, which are complex and contextual. The interview guide was developed based on relevant literature and aligned with the research objectives. The questions in the interview guide covered several key topics, such as teacher guidance on using assessment reviews, challenges encountered during the assessment process, and adaptive strategies developed to address these challenges. To ensure the validity of the instrument, the interview guide was first determined through an expert review process. In addition to the interview guide, the researchers also used human subjects as the primary instrument, capable of collecting, analyzing, and interpreting data critically and reflectively (Lincoln & Guba, 1985).

3. Data collection

In this study, data collection was carried out systematically and in a planned manner in several stages. First, the researcher used the term "participant" to explain the research objectives, data collection procedures, and voluntary consent. At this stage, the researcher also explained the participants' rights, such as the need to assess their own identity and the data provided, and the need to draw conclusions from the research as quickly as possible without special consideration (Creswell & Poth, 2018). Second, after the participants indicated their availability and selected a convenient time, the researcher conducted semi-structured individual interviews. Interviews were conducted in a comfortable and conducive environment, either face-to-face or in person, according to individual preferences and needs. Each interview session lasted 45 to 60 minutes. According to Merriam and Tisdell (2016), all interviews were audio-recorded with the participants' permission to ensure the accuracy of the data obtained. For more comprehensive data analysis, the audio recordings were followed verbatim in the text. The interview process was meticulously conducted to ensure that every statement made by the subjects was documented accurately and completely.

4. Data analysis

The collected data were analyzed using a thematic analysis approach similar to that proposed by Braun and Clarke (2006). Thematic analysis was chosen because it is a flexible and systematic analytical method useful for identifying, analyzing, and describing themes or patterns in qualitative data. The data analysis process was carried out using six relevant stages. The first stage is data recognition, where the researcher read and understood each interview transcript to obtain a comprehensive picture of the data. The first stage is initial discovery, which is the process of systematically identifying and labeling data segments. The third stage is theme discovery, the identified codes were grouped into more comprehensive potential topics. The fourth stage involved examining these topics to ensure that each topic accurately captured the facts and had strong internal coherence. The goal of the fifth stage was to define and analyze these themes clearly and representatively. Finally, the sixth stage involved creating a report in which the themes were presented in a narrative style using facts to support each identified theme (Braun & Clarke, 2006). To ensure data integrity, this study employed participant-checking and peer-review strategies similar to those proposed by Lincoln and Guba (1985).

5. Truthworthiness

To ensure the validity and reliability of this study, the researchers referred to the trustworthiness criteria established by Lincoln and Guba (1985), which include four main aspects: credibility, transferability, dependability, and confirmability.

First, credibility refers to the extent to which research findings accurately reflect participants' perspectives and experiences. To address this aspect, researchers conduct

member checking, which involves returning transcribed and interpreted data to participants for confirmation. Furthermore, researchers implement ongoing engagement by building rapport with participants prior to the interviews, ensuring they feel comfortable sharing their experiences openly and honestly (Lincoln & Guba, 1985).

First and foremost, credibility is determined by the extent to which research findings accurately reflect participants' perspectives and perceptions. To address this aspect, researchers conduct participant analysis, which involves collecting analyzed data and presenting it to participants for confirmation. According to Lincoln and Guba (1985), researchers implement ongoing engagement by building rapport with participants before the interview, ensuring they are aware of the consequences of their actions.

Second, transferability relates to the relevance of the research topic in relation to other contexts with similar characteristics. Although qualitative research does not aim to generalize statistics, researchers provide descriptive and detailed information about the research context so readers can determine whether the research findings are relevant and applicable to their own contexts (Creswell & Poth, 2018).

Reliability also affects the consistency and traceability of the research process. To address this issue, researchers must systematically document each step of the research process, from instrument development and data collection to analysis, to create an audit report that can be reviewed by others to confirm the reliability of the process (Merriam & Tisdell, 2016).

Fourth, confirmability relates to ensuring that research findings are based on the data and not on the researcher's own biases or assumptions. To address this issue, researchers consult with colleagues with a background in qualitative research to critically analyze and interpret the data. Furthermore, researchers practice reflexivity by carefully considering their own positions, assumptions, and biases that may influence the research process (Lincoln & Guba, 1985; Creswell & Poth, 2018).

RESULT AND DISCUSSION

This section presents and discusses the findings derived from thematic analysis of the semi-structured interview conducted with an English teacher at the elementary school level (Sekolah Dasar, SD) in Indonesia. The analysis followed Braun and Clarke's (2006) six-phase thematic analysis framework, generating four major themes that map directly onto the two research questions guiding this study: (RQ1) the challenges faced by the teacher in conducting authentic speaking assessment, and (RQ2) the strategies employed to navigate those challenges. The four themes are: (1) the time constraints and the diversity of learners as a challenge of assessment; (2) the subjectivity and the difficulty of remaining objective; (3) the adaptation of task design to the developmental needs of young learners; and (4) the adaptive strategies of teachers in the practice of assessment. Direct quotes from the participant are given to exemplify each theme and findings are interpreted in dialogue with relevant theoretical and empirical literature.

A. Time Constraints and Learner Diversity as Core Assessment Challenges

The first and most important data point to challenges arising from a combination of limited teaching time and the considerable diversity of student characteristics in the classroom. He identified time as the primary structural challenge in implementing authentic speaking assessments, even though the participants taught relatively small classes of twenty students a context he himself acknowledged as "quite ideal" compared to the large classes documented in many studies of secondary schools in Indonesia. She explains,

"Meskipun jumlah siswa saya hanya 20 orang, tantangan terbesarnya tetaplah keterbatasan waktu. Dengan jam pelajaran yang terbatas misalnya hanya 2x35 menit tetap

sulit untuk menilai satu per satu secara mendalam dalam satu pertemuan. Saya biasanya mengatur waktu dengan cara menilai sebagian siswa di minggu ini, dan sisanya di minggu depan. [Even though I only have 20 students, the biggest challenge is still limited time. With limited lesson time for example only 2x35 minute it is still difficult to assess each student in depth in one meeting. I usually manage this by assessing some students this week and the rest the following week.]"

These findings are consistent with existing literature on the time-consuming nature of oral assessment in language teaching contexts. Brown and Abeywickrama (2010) noted that speaking assessment is inherently more time-consuming than written assessment because each learner must be observed, instructed, and assessed on their oral performance. Participants' solution of spreading the assessment across multiple sessions reflects what Cohen (1994) calls "distributed assessment" a practical compromise that allows for individual assessment without sacrificing the depth and authenticity of the task, but also risks creating inconsistencies in task conditions and assessor circumstances from session to session.

Another challenge, according to the teacher, is the high level of heterogeneity among her students, in addition to time constraints. Specifically, she noted that having both highly active and extroverted students and very shy and introverted students in the same class makes fair assessment very difficult:

"Karakteristik siswa yang berbeda-beda ada yang sangat aktif dan ada yang sangat pemalu membuat saya harus mencari cara agar penilaian tetap adil bagi semuanya. [The varied characteristics of students some very active and some very shy mean I have to find ways to ensure assessment remains fair for everyone.]"

The challenge of fairly assessing heterogeneous learners is most pronounced at the elementary school level, where individual differences in personality, emotional regulation, and developmental readiness are most pronounced. In the context of second language assessment, MacIntyre et al. (1998) found that foreign language anxiety was one of the strongest affective barriers to oral performance, and numerous studies have confirmed that young learners are particularly vulnerable to performance anxiety due to their heightened self-consciousness and limited metalinguistic awareness. Cameron (2001) further argues that the limited and moderate cognitive abilities of young learners mean that the demands of speaking tasks must be carefully calibrated to avoid creating conditions in which anxiety, rather than genuine communicative effort, becomes the response. Participants' descriptions of students who "already knew what they wanted to say but didn't know the English" and the immediate results of silence underscore assessment designs that reduce cognitive load while maintaining authenticity.

B. Subjectivity and the Difficulty of Maintaining Objectivity in Assessment

The second theme relates to teachers' honest understanding of the challenges they face in determining the objectivity and consistency of student writing. This theme emerged naturally in response to questions about the impact of personal relationships on assessment and the challenges of grading. As one participant stated:

"Iya, terkadang sulit memisahkan kedekatan pribadi dengan nilai akademis. Terkadang saya merasa tidak tega memberi nilai rendah kepada siswa yang sudah sangat berani mencoba meski tata bahasanya masih banyak yang salah. [Yes, it is sometimes difficult to separate personal closeness from academic grades. Sometimes I feel reluctant to give a low mark to a student who has been very brave in trying, even though their grammar still has many errors.]"

This study illustrates the difference between two similar assessments: the recognition and appreciation of communicative effort and affective engagement on the one hand, and

the maintenance of strict criterion-based assessment standards on the other. Importantly, Thorndike (1920) discussed the halo effect, which suggests that the rater's global impression of the ratee in this case, the teacher's admiration for the student's courage influences ratings on separate dimensions of performance. In the context of assessing young students, this relationship may be further strengthened by the teacher's genuine attitudes about the student's emotional development and its potential dangers, which may be reinforced by low scores on the still-developing and fragile self-concept.

The teachers mentioned above also identified characteristics as a source of assessment subjectivity, suggesting that conflicting communication between introverted and extroverted students may not necessarily result in biased assessments if left unaddressed. This study supports Fulcher's (2003) assertion that task-based speaking performance is highly sensitive to construct-irrelevant variation that is, variation in scores caused by factors such as personal characteristics, anxiety, or familiarity rather than language ability. Stiggins (1991) linked this assessment challenge directly to assessment literacy, explaining that a teacher who does not receive explicit instruction in the principles of performance assessment is unable to establish a clear connection between students' communication skills and their linguistic proficiency. Importantly, the researcher highlights the critical level of self-awareness involved in this, emphasizing the need for careful observation and suggesting that reflective practice can serve as a partial countermeasure to systemic bias even without formal analysis.

C. Adapting Task Design to Suit Young Learners' Developmental Needs

The third theme describes the teacher's approach to creating appropriate speaking tasks for elementary school students. A hallmark of the participant's practice was her effort to support speaking tasks within the everyday context of early childhood education a design principle consistent with the theory of task authenticity in early childhood education. She illustrated her task design philosophy as follows:

"Saya merancang tugas yang dekat dengan keseharian mereka, seperti role play tentang membeli jajan atau storytelling tentang hewan peliharaan. Saya juga menyesuaikan tingkat kesulitan; siswa yang sudah lancar saya minta menggunakan kalimat lengkap, sedangkan yang masih kesulitan boleh menggunakan frasa pendek. [I design tasks that are close to their daily lives, such as role play about buying snacks or storytelling about pets. I also adjust the level of difficulty; for students who are already fluent I ask them to use full sentences, while those still struggling are allowed to use short phrases.]"

This different approach to task design allows for a more comprehensive understanding of structural language learning, consistent with Vygotsky's (1978) framework of the zone of proximal development, where students are guided by their current abilities with appropriate support structures. 16 The use of culturally and experientially familiar scenarios buying a snack, describing a pet is directly consistent with what O'Malley and Pierce (1996) illustrate as situational judgment tasks that correspond to the real world that learners actually encounter in their everyday lives. Skehan (1998) also argues that task-based language learning should provide students with content that can be easily analyzed cognitively and pragmatically. This is because it reduces irrelevant cognitive barriers and maximizes students' ability to use their native language.

The participant task series which includes two dialogues, a "Demonstrate and Explain" activity, role-playing, and storytelling constitutes a series of easy-to-understand structures that collectively highlight several aspects of young EFL learners' communication skills. Cameron (2001) specifically identified "Demonstrate and Explain" as one of the most effective speaking tasks for young EFL learners because it uses familiar objects from the child's environment, reduces work-related stress through the use of real-life tools, and

provides a communicative purpose for writing. More specifically, the integration of role-playing provides opportunities for the development of pragmatic competence the ability to use language appropriately in social contexts identified by Canale and Swain (1980) as an integral component of communicative competence that is often presented in conventional formats.

D. Teachers' Adaptive Strategies in Managing Assessment Challenges

The fourth and final theme documents the various adaptive strategies individuals used to address the challenges identified in Themes 1 through 3. Rather than abandoning authentic assessment in the face of practical constraints, teachers presented several creative, goal-oriented coping strategies that challenged pedagogical norms and contextual pragmatism.

The most consistently applied strategy was the use of simple rubrics to maintain focus and consistency in the analysis. "I use simple rubrics to keep the assessment focused," the teacher explained. This practice aligns with Mertler's (2001) recommendation that classroom-based rubrics should prioritize the use of practical skills over psychometric skills, especially in situations where teachers must collaboratively assess student learning and assessment without the need for formal evaluation. Most importantly, participant rubrics prioritize communication skills and the ability to move beyond mere linguistic accuracy. This is a pedagogical principle explicitly stated in the text for reforming speaking assessment:

"Saya ingin mengubah fokus penilaian dari akurasi tata bahasa menjadi keberanian berkomunikasi. Saya ingin penilaian berbicara di SD terasa seperti percakapan biasa yang menyenangkan, bukan ujian yang menakutkan, agar anak-anak jatuh cinta dulu dengan bahasa Inggris sejak dini. [I want to change the assessment focus from grammatical accuracy to communicative courage. I want speaking assessment at the elementary level to feel like an enjoyable ordinary conversation, not a frightening examination, so that children fall in love with English from an early age.]"

This vision is grounded in solid pedagogy and theory. According to Arnold (1999), a learning environment that prioritizes accuracy over fluency and evaluates linguistic issues can foster authentic communication. This is especially true for young learners whose linguistic systems are still in the early stages of development. Participants' desire to view speaking assessments as "fun, casual conversations" rather than "scary tests" aligns with Shohamy's (2001) critique of the power dynamics that emerge from the context inherent in conventional testing and her support for assessment approaches that empower rather than intimidate learners. In addition, giving bonus points to students who demonstrate communication skills in addition to linguistic proficiency is an effective type of learning environment used to reduce the effectiveness of learning filters (Krashen, 1982) and create an environment in which even the most gifted students are eager to try writing.

Regarding peer assessment, teachers' experiences vary. Although she has experimented with peer assessment as a way to distribute the assessment workload and develop students' evaluative metacognition, she found that elementary school students tend to give higher grades to their close friends, thereby reducing the data generated by their peers. She stated: "I have done peer assessment, but the results are not very objective because they tend to give high grades to close friends. Therefore, I consider myself a major contributor to informal communication." This document highlights the documented limitations of peer assessment in early childhood learning: Sluijsmans et al. (1999) and Topping (1998) both point out that peer assessment relies on learners' developmental ability to apply evaluative criteria independently of social connections, an ability that elementary school children lack. Participants' reliance on informal, systematic observation is a

pragmatic, context-based response that highlights the formative value of ongoing assessment while acknowledging the developmental limitations of peer evaluation at this age level.

Finally, participant services to support the institution include practical, tailored teacher training for each child on assessment, as well as access to existing rubrics and task banks, minimizing the systemic dimensions of the challenges faced. I clearly articulated this need:

"Kami sangat butuh pelatihan guru yang fokus pada teknik asesmen yang praktis dan menyenangkan bagi anak SD. Dukungan berupa bank soal atau contoh rubrik yang siap pakai dari pemerintah atau sekolah juga akan sangat membantu efisiensi kerja kami. [We really need teacher training focused on assessment techniques that are practical and engaging for elementary school children. Support in the form of question banks or ready-to-use example rubrics from the government or school would also greatly improve our working efficiency.]"

This article highlights important issues in professional infrastructure that can impact individual teacher practice. According to Inbar-Lourie (2008), assessment literacy the knowledge base necessary to design, implement, interpret, and use assessments effectively cannot be developed through ad hoc or recognized training, but rather requires systematic, context-based professional development embedded within the professional teacher community. Participants' calls for practical, child-appropriate training resources underscored the need for professional development programs in EFL elementary schools in Indonesia to emphasize general principles of teaching and, in particular, emphasize the developmental characteristics, need for effectiveness, and communication goals in learning English as a second language.

E. Summary of Findings

Overall, an emerging theme from this analysis is the potential for nuanced and multi-layered authentic assessment in Indonesian EFL schools. In this study, teachers' commitment to authentic, student-centered assessment and their pragmatic ability to adapt their practices in response to structural and effective constraints in their contexts were examined. Three key tensions shaping authentic speaking assessment at the elementary school level are: the tension between assessment depth and time availability; the tension between relational warmth and evaluative objectivity; and the tension between communicative and linguistic authenticity. Tensions are not limited to the Indonesian context; they are also influenced by the country's institutions, development, and culture, particularly regarding the teaching of English as a Foreign Language (EFL) at the highest level in Indonesia. The following section highlights the findings of this study and their implications for policy, teacher development, and future research.

CONCLUSION

This study explored the challenges and strategies experienced by an elementary school English teacher in conducting authentic speaking assessment in the Indonesian EFL context. The findings revealed that authentic speaking assessment is a complex and multidimensional process shaped by structural, pedagogical, and developmental factors.

The study identified several major challenges faced by the teacher, particularly limited instructional time, learner diversity, and difficulties in maintaining objectivity during assessment. Even in relatively small classes, assessing students' speaking performance individually within restricted lesson hours remained difficult. In addition, differences in students' personalities, confidence levels, and language abilities complicated the process of ensuring fair and balanced assessment. The findings also showed that teachers often struggle to separate emotional considerations from academic evaluation, especially when assessing

young learners who demonstrate strong communicative effort despite linguistic inaccuracies.

To address these challenges, the teacher implemented various adaptive strategies. These included designing authentic and developmentally appropriate speaking tasks, using familiar real-life contexts, simplifying assessment rubrics, and prioritizing communicative courage over grammatical perfection. The teacher also emphasized creating a supportive and enjoyable assessment environment to reduce students' anxiety and encourage active participation in English communication. Furthermore, the findings highlighted the importance of reflective practice, continuous observation, and practical teacher training in improving assessment quality.

Overall, this study demonstrates that authentic speaking assessment at the elementary school level requires flexibility, sensitivity to learners' developmental needs, and strong assessment literacy. The study also highlights the need for institutional support, including practical training programs, ready-to-use assessment resources, and professional development opportunities for teachers. These findings contribute to a deeper understanding of authentic assessment practices in Indonesian EFL classrooms and provide important implications for policy, teacher education, and future research.

REFERENCES

- Arnold, J. (Ed.). (1999). *Affect in language learning*. Cambridge University Press.
- Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. <https://doi.org/10.1080/0969595980050102>
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399–413. <https://doi.org/10.1080/02602930600679050>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brinkmann, S., & Kvale, S. (2015). *InterViews: Learning the craft of qualitative research interviewing* (3rd ed.). SAGE Publications.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Longman.
- Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). Pearson Education.
- Cameron, L. (2001). *Teaching languages to young learners*. Cambridge University Press.
- 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/1.1.1>
- Clandinin, D. J., & Connelly, F. M. (1996). Teachers' professional knowledge landscapes. *Educational Researcher*, 25(3), 24–30.
- Cohen, A. D. (1994). *Assessing language ability in the classroom* (2nd ed.). Heinle & Heinle.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393–415. <https://doi.org/10.1002/j.1545-7249.2009.tb00242.x>
- Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Lawrence Erlbaum.
- Farrell, T. S. C. (2015). *Promoting teacher reflection in second language education: A framework for TESOL professionals*. Routledge.
- Fulcher, G. (2003). *Testing second language speaking*. Pearson Longman
- Hamid, M. O., & Baldauf, R. B. (2008). Will CLT bail out the bogged down EFL? *English Today*, 24(4), 16–22.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning — tensions and

- synergies. *The Curriculum Journal*, 16(2), 207–223.
- Harmer, J. (2007). *The practice of English language teaching* (4th ed.). Pearson Longman.
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected readings* (pp. 269–293). Penguin.
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 385–402. <https://doi.org/10.1177/0265532208090158>
- Jannah, M., & Cahyono, B. Y. (2019). Assessment practices of Indonesian EFL teachers for the 2013 curriculum. *TEFLIN Journal*, 30(1), 71–91.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE Publications.
- Luoma, S. (2004). *Assessing speaking*. Cambridge University Press.
- MacIntyre, P. D., Clément, R., Dörnyei, Z., & Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation. *The Modern Language Journal*, 82(4), 545–562.
- McKay, P. (2006). *Assessing young language learners*. Cambridge University Press.
- Mercer, S., & Dörnyei, Z. (2020). *Engaging language learners in contemporary classrooms*. Cambridge University Press.
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). Jossey-Bass.
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research, and Evaluation*, 7(25), 1–10.
- Moon, J. (2000). *Children learning English*. Macmillan Heinemann.
- Mueller, J. (2006). *Authentic assessment toolbox*. North Central College. <http://jfmuller.faculty.noctrl.edu/toolbox>
- Newfields, T. (2006). Teacher development and assessment literacy. In *Authentic Communication: Proceedings of the 5th Annual JALT Pan-SIG Conference* (pp. 48–73). JALT.
- Nunan, D. (1999). *Second language teaching and learning*. Heinle & Heinle Publishers.
- O'Malley, J. M., & Valdez Pierce, L. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. Addison-Wesley Publishing.
- Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). SAGE Publications.
- Pinter, A. (2006). *Teaching young language learners*. Oxford University Press.
- Pinter, A. (2011). *Children learning second languages*. Palgrave Macmillan.
- Rea-Dickins, P. (2001). Mirror, mirror on the wall: Identifying processes of classroom assessment. *Language Testing*, 18(4), 429–462. <https://doi.org/10.1177/026553220101800407>
- Renandya, W. A., & Widodo, H. P. (Eds.). (2016). *English language teaching today: Linking theory and practice*. Springer.
- Scott, W. A., & Ytreberg, L. H. (1990). *Teaching English to children*. Longman.
- Shohamy, E. (2001). *The power of tests: A critical perspective on the uses of language tests*. Pearson Education.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Sluismans, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self-, peer- and co-assessment. *Learning Environments Research*, 1(3), 293–319.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249–276.
- Stake, R. E. (1995). *The art of case study research*. SAGE Publications.
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappan*, 72(7), 534–539.
- Stiggins, R. J. (2005). *Student-involved assessment for learning* (4th ed.). Pearson Prentice Hall.
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4(1), 25–29.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research, and*

- Evaluation, 2(2), 1–3.
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. Jossey-Bass.
- Xu, Y., & Liu, Y. (2009). Teacher assessment knowledge and practice: A narrative inquiry of a Chinese college EFL teacher's experience. *TESOL Quarterly*, 43(3), 493–513. <https://doi.org/10.1002/j.1545-7249.2009.tb00246.x>
- Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.