JIGE 6 (2) (2025) 1101-1107
JURNAL ILMIAH GLOBAL EDUCATION
ejournal.nusantaraglobal.ac.id/index.php/jige
DOI: https://doi.org/10.55681/jige.v6i2.4005

Development of a Speaking Competence Evaluation Instrument for A2-Level Students through a Corpus-Based German–Indonesian Slang Dictionary in the German Language Education Program

Tri Edliani Lestari[1*], Ninuk Lustyantie[1], Fathiaty Murtadho[1]
[1] Pascasarjana Universitas Negeri Jakarta, Jakarta, Indonesia
Corresponding author email: tri.edliani@mhs.unj.ac.id

Article Info

Article history:

Received June 9, 2025
Approved June 28, 2025

Keywords:

speaking competence; German slang; corpus linguistics; lexicography; foreign language evaluation.

ABSTRACT

Speaking fluently and appropriately in German remains a persistent challenge for Indonesian learners at the A2 level, especially in contexts that demand spontaneous, pragmatic communication. Many existing evaluation tools focus narrowly on grammatical accuracy, often neglecting fluency, lexical diversity, and sociopragmatic competence. This study aims to develop a corpus-informed speaking evaluation instrument specifically designed for assessing A2-level German learners in Indonesian higher education. Employing a Research and Development (R&D) framework using the ADDIE model (Analysis, Design, Development, Implementation, Evaluation), the instrument was systematically constructed through needs analysis, corpus-based task design, expert validation, and iterative refinement. Data collection involved interviews, classroom observations, and document analysis, supported by expert judgment and inter-rater reliability testing. The final product includes task-based prompts grounded in authentic communicative scenarios, a Likert-based rubric covering fluency, accuracy, vocabulary range, and pragmatic appropriateness, and lexical input informed by spoken German corpora. Results demonstrate the instrument's validity and reliability, while qualitative feedback supports its pedagogical relevance. This study offers a novel contribution to corpus-informed language assessment and serves as a model for localized speaking evaluation tools aligned with CEFR standards.

How to cite: Lestari, T. E., Lustyantie, N., & Murtadho, F. (2025). Development of a Speaking Competence Evaluation Instrument for A2-Level Students through a Corpus-Based German–Indonesian Slang Dictionary in the German Language Education Program. *Jurnal Ilmiah Global Education*, *6*(2), 1101–1107. https://doi.org/10.55681/jige.v6i2.4005

INTRODUCTION

Speaking competence remains a fundamental challenge in foreign language learning, particularly for Indonesian students studying German as a foreign language. While students are generally able to meet receptive skills such as reading and listening, productive skills—especially

speaking—continue to lag. Several studies have highlighted that students often struggle with fluency, lexical variety, and pragmatic appropriateness when expressing themselves orally in German (Wisniewski, 2020; Hu & Lu, 2021). These deficiencies are partly due to the absence of assessment instruments that authentically capture the functional use of language in natural contexts, particularly for A2-level learners.

In many educational settings, speaking tests are either integrated within general course evaluations or rely on rubrics that fail to differentiate between controlled, classroom-based speaking and actual communicative performance. This mismatch results in evaluations that do not reflect learners' abilities in real-life communication scenarios (Xi, 2017; Callies & Götz, 2015). Furthermore, research has shown that speaking evaluation tools often emphasize grammatical accuracy without sufficiently addressing fluency or sociopragmatic appropriateness, which are key aspects of communicative competence (Council of Europe, 2020; Biber et al., 1998).

Prior efforts in the development of speaking evaluation tools have focused on English or other widely taught languages (Afia & Laili, 2023; Belz & Vyatkina, 2008), with limited emphasis on German as a foreign language, especially in Indonesian higher education. Although learner corpora such as MERLIN and FOLK provide insights into the linguistic performance of L2 German learners, these resources have not been adequately utilized in the design of localized evaluation instruments tailored to Indonesian learners' contexts (Boyd et al., 2014; Moehrs & Meliss, 2017). Studies by Lemmenmeier-Batinic (2020) and Weber (2018) emphasize the importance of utilizing learner data to design more context-sensitive assessments, yet these insights remain underapplied in Southeast Asian contexts.

The gap this study seeks to fill lies in the lack of corpus-informed evaluation instruments specifically targeting A2-level speaking competence in German for Indonesian university students. No existing tool to date integrates corpus-based insights with localized pedagogical needs and CEFR-aligned benchmarks. This creates a gap not only in practice but also in the empirical research literature on L2 German speaking evaluation (Weiss & Meurers, 2019; Schmidt, 2014; Fellbaum & Geyken, 2005). Additional works by Evert (2008), Boyd et al. (2014), and Timpe-Laughlin et al. (2015) further emphasize the importance of leveraging multi-word expressions, learner performance data, and intercultural pragmatics in developing effective assessments. However, most tools remain focused on advanced learners or are applied in Western educational contexts (Leńko-Szymańska & Boulton, 2015; Sharwood Smith, 2013).

The novelty of this study lies in its emphasis on integrating spoken corpus data—especially authentic informal speech patterns—into the construction of an evaluative tool tailored for early-stage learners. While most previous research focuses on written corpora or advanced learners, this study applies those principles to assess oral competence at the foundational level. The use of a corpus-based approach ensures that lexical items and discourse patterns included in the evaluation instrument reflect actual usage in natural contexts (Furko & Kecskes, 2021; Götz, 2013). Furthermore, designing rubrics based on empirical learner performance offers a more transparent and pedagogically valid basis for assessment (Tracy-Ventura & Huensch, 2022).

The state of the art in language assessment shows a growing trend toward data-driven and performance-based evaluations. However, as Sharwood Smith (2013) and Timpe-Laughlin et al. (2015) argue, these innovations have yet to permeate evaluation design in less commonly taught languages such as German in non-European contexts. This research therefore positions itself as a pioneering effort to bridge corpus linguistics, applied linguistics, and CEFR-aligned assessment in

a localized Indonesian setting. It is informed by interdisciplinary research on communicative competence, corpus-based lexicography, and test validation frameworks (Taguchi & Roever, 2017; Fulcher & Davidson, 2007; Tono, 2012; Alderson, 2000; Norris et al., 2002).

The focus of this study is threefold: (1) to identify speaking competence indicators appropriate for A2-level German learners in the Indonesian context, (2) to design an evaluation instrument that aligns with those indicators and incorporates data from spoken German corpora, and (3) to validate the instrument's effectiveness and reliability through expert judgment and empirical testing. The central research questions are: (1) What specifications of A2-level speaking competence are relevant as a basis for instrument development? (2) How is the evaluation instrument constructed using corpus-informed linguistic data? (3) To what extent is the developed instrument valid and reliable for assessing German-speaking performance in an educational setting?

METHODS

This study adopts a Research and Development (R& D) approach using the ADDIE instructional model—Analysis, Design, Development, Implementation, and Evaluation—to create a valid and reliable speaking evaluation instrument aligned with CEFR A2 standards. The ADDIE model, widely applied in instructional design, offers a systematic structure for developing educational assessments that are learner-centered and empirically grounded (Branch, 2009; Molenda, 2003; Almelhi, 2021).

In the Analysis phase, data were collected through interviews and classroom observations involving German language instructors and A2-level students from three Indonesian teacher-training institutions (UNJ, Unnes, Unesa). Additionally, existing curricula and speaking assessments were reviewed to identify pedagogical gaps. Corpus-based analysis was conducted using Sketch Engine to extract commonly used spoken expressions from authentic German corpora such as FOLK and DGD2.

The Design phase focused on developing a blueprint for the evaluation instrument. Test constructs and rubrics were created based on CEFR descriptors for A2 speaking, incorporating key indicators such as fluency, accuracy, vocabulary range, and pragmatic appropriateness. Rubrics were designed to reflect authentic speech performance by integrating corpus-derived expressions.

During the Development phase, a prototype of the instrument—including task instructions, sample prompts, and a scoring rubric—was constructed and reviewed by experts in German language education and applied linguistics. A limited trial was conducted with 30 students to test usability and reliability. Feedback informed revisions to improve clarity, validity, and task alignment.

The Implementation and Evaluation phases focused on the refinement and expert review of the instrument. Expert validation was conducted to examine the clarity, relevance, and alignment of the instrument with CEFR A2 descriptors. Iterative feedback from language education professionals guided the improvement of item prompts, rubric formulation, and the linguistic appropriateness of the assessment tasks. The instrument underwent multiple rounds of revisions to ensure theoretical alignment, empirical clarity, and pedagogical applicability.

RESULTS AND DISCUSSION

The developed speaking evaluation instrument consists of three core components: (1) a task-based test format featuring six speaking prompts grounded in authentic communicative scenarios; (2) a four-band holistic rubric evaluating fluency, accuracy, vocabulary range, and pragmatic appropriateness; and (3) a reference glossary of corpus-derived German expressions including informal and colloquial items relevant to CEFR A2-level learners.

Table 1. Sample Rubric of Speaking Evaluation Instrument (A2 Level)

Criterion	Score 1: Strongly Disagree	Score 2: Disagree	Score 3: Neutral	Score 4: Agree	Score 5: Strongly Agree
Fluency	Constant pauses and breakdowns	Noticeable hesitations, limited flow	Adequate fluency, occasional hesitation	Generally fluent speech	Smooth, natural flow throughout
Accuracy	Frequent grammatical errors that distort meaning	Some errors that hinder understanding	Understandable with several errors	Minor errors, meaning preserved	Grammatically accurate and clear
Vocabulary Range	Extremely limited vocabulary, repetitive use	Limited vocabulary, lacks variation	Basic vocabulary for familiar topics	Good vocabulary range with some variation	Wide and appropriate vocabulary including informal expressions
Pragmatic Appropriateness	Inappropriate responses, no awareness of context	Minimal awareness of appropriateness	Generally contextually appropriate	Appropriate language with minor slips	Fully appropriate and context-sensitive use of language

Scoring Scale Explanation (Likert-Based):
Each performance descriptor is rated using a 5-point Likert scale:

- 1 = Strongly Disagree (Does not meet the criteria at all)
- 2 = Disagree (Meets the criteria poorly)
- 3 = Neutral (Meets the criteria adequately)
- 4 = Agree (Meets the criteria well)
- 5 = Strongly Agree (Exceeds the criteria excellently)

Evaluators assign scores for each of the four criteria (Fluency, Accuracy, Vocabulary Range, and Pragmatic Appropriateness), with a total maximum score of 20. The cumulative scores are interpreted as follows:

- 17–20: Excellent – Exceeds A2 level expectations
- 13–16: Good – Meets A2 level expectations
- 9–12: Basic – Approaching expectations
- 5–8: Limited – Needs significant improvement
- 1–4: Very Limited – Major support required

Initial expert validation indicated high content relevance, with a mean score of 92% agreement on the alignment between test items and CEFR descriptors. The inter-rater reliability of the scoring rubric, as measured using Cohen's Kappa, was 0.84, indicating strong consistency among independent raters.Pre- and post-test results from 300 students revealed statistically

significant improvements in all four assessed domains. The mean speaking score increased from 62.7 (SD = 8.5) to 74.9 (SD = 7.1), representing a gain of approximately 19.4%. A paired-sample t-test confirmed this difference as statistically significant ($t(299) = 14.73$, $p < 0.001$), supporting the instrument's effectiveness.

Qualitative feedback from both instructors and learners emphasized the value of including informal expressions, which were reported to increase learners' engagement and perceived communicative relevance. Many students noted that the tasks felt more realistic and aligned with conversational German as used in social media, travel, and everyday interactions.

These findings align with prior studies that support the use of corpus-informed materials in L2 speaking assessment (Götz, 2013; Wisniewski, 2020; Tracy-Ventura & Huensch, 2022). The instrument's ability to capture pragmatic nuances while remaining accessible to A2 learners highlights its novelty and pedagogical utility in Indonesian tertiary education contexts. It offers a model for localized test development that incorporates global standards while addressing contextual language use.

CONCLUSION

This study has successfully developed a corpus-informed evaluation instrument tailored to assess A2-level German speaking competence within Indonesian higher education settings. Grounded in the ADDIE development model and CEFR framework, the instrument was constructed through a systematic process involving needs analysis, corpus-based task design, expert validation, and iterative refinement. Its structure incorporates authentic speaking tasks and a rubric that addresses fluency, accuracy, lexical range, and pragmatic appropriateness—dimensions often underrepresented in traditional assessments.

The use of spoken German corpora provided a foundation for integrating real-world language use, particularly informal and colloquial expressions, enhancing the instrument's communicative and cultural relevance. Expert review confirmed the content validity of the instrument, while reliability metrics demonstrated its consistency across evaluators. The Likert-based rubric allows for nuanced, performance-based assessment and offers clear interpretability.The results of this study contribute to the broader field of applied linguistics and language testing by offering a localized, empirically grounded model for evaluating L2 speaking proficiency. It addresses a clear gap in German language education in Indonesia, particularly regarding informal communicative competence at the A2 level. Future research may explore the adaptation of this model for other CEFR levels or for other foreign languages within similar educational contexts.

REFERENCES

Afia, N., & Laili, M. (2023). Developing a Corpus-Based English Vocabulary Dictionary using The ADDIE Model. Nusantara Educational Review, 1(1), 56–62. https://doi.org/10.5678/ner.v1i1.123

Alderson, J. C. (2000). Assessing reading. Cambridge University Press. https://doi.org/10.1017/CBO9780511732935

Almelhi, A. M. (2021). The effect of instructional design based on the ADDIE model on developing EFL learners' writing performance. English Language Teaching, 14(2), 20–30. https://doi.org/10.5539/elt.v14n2p20

Belz, J. A., & Vyatkina, N. (2008). Learner corpus analysis and the development of L2 pragmatic competence in networked intercultural language study. Canadian Modern Language Review, 64(4), 567–600. https://doi.org/10.3138/cmlr.64.4.567

Biber, D., Conrad, S., & Reppen, R. (1998). Corpus linguistics: Investigating language structure and use. Cambridge University Press. https://doi.org/10.1017/CBO9780511804489

Boyd, A., et al. (2014). The MERLIN corpus: Learner language and the CEFR. LREC 2014, 1281–1288.

Branch, R. M. (2009). Instructional design: The ADDIE approach. Springer. https://doi.org/10.1007/978-0-387-09506-6

Callies, M., & Götz, S. (Eds.). (2015). Learner corpora in language testing and assessment. John Benjamins. https://doi.org/10.1075/scl.70

Council of Europe. (2020). CEFR companion volume. Council of Europe Publishing.

Evert, S. (2008). A lexicographic evaluation of German adjective-noun collocations. LREC Workshop, 20–27.

Fellbaum, C., & Geyken, A. (2005). Corpus-based studies of German idioms and light verbs. International Journal of Lexicography, 19(4), 349–360. https://doi.org/10.1093/ijl/ecm030

Fulcher, G., & Davidson, F. (2007). Language testing and assessment: An advanced resource book. Routledge.

Furko, P., & Kecskes, I. (2021). The pragmatics of English as a lingua franca in intercultural communication. Intercultural Pragmatics, 18(4), 451–476. https://doi.org/10.1515/ip-2021-0009

Götz, S. (2013). Fluency in native and nonnative English speech. John Benjamins. https://doi.org/10.1075/scl.57

Hu, X., & Lu, X. (2021). Predicting CEFR levels in L2 oral speech. Language Testing, 38(3), 1–25. https://doi.org/10.1177/0265532220980423

Leńko-Szymańska, A., & Boulton, A. (Eds.). (2015). Multiple affordances of language corpora for data-driven learning. John Benjamins. https://doi.org/10.1075/scl.69

Lemmenmeier-Batinić, D. (2020). Lexical Explorer: Extending access to the Database for Spoken German for user-specific purposes. Corpora, 15(1), 55–76. https://doi.org/10.3366/cor.2020.0186

Moehrs, C., & Meliss, M. (2017). LeGeDe project: Spoken German lexical resources. Lexikographie und Wörterbuchforschung, 1(1), 45–60.

Molenda, M. (2003). In search of the elusive ADDIE model. Performance Improvement, 42(5), 34–36. https://doi.org/10.1002/pfi.4930420508

Norris, J. M., Brown, J. D., Hudson, T., & Yoshioka, J. (2002). Designing second language performance assessments. University of Hawai'i Press.

Schmidt, T. (2014). The Database for Spoken German – DGD2. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14) (pp. 1451–1457).

Sharwood Smith, M. (2013). Rethinking the role of consciousness in second language learning. In VanPatten, B. & Benati, A. (Eds.), Key terms in second language acquisition (pp. 27–29). Bloomsbury.

Taguchi, N., & Roever, C. (2017). Second language pragmatics. Oxford University Press.

Timpe-Laughlin, V., Wain, J., & Schmidgall, J. (2015). Defining and operationalizing the construct of interactional competence. TOEFL Research Report Series, 2015(2), 1–57. https://doi.org/10.1002/ets2.12059

Tono, Y. (2012). Designing and implementing a corpus-based dictionary of collocations for learners of Japanese. International Journal of Lexicography, 25(4), 396–433. https://doi.org/10.1093/ijl/ecs017

Tracy-Ventura, N., & Huensch, A. (2022). Measuring fluency across tasks: What matters and what doesn't. Language Testing, 39(2), 211–235. https://doi.org/10.1177/02655322211035121

Weber, T. (2018). Grammatik und Lernerkorpora: Eine korpusorientierte Untersuchung von Präpositionalphrasen im deutschen MERLIN-Korpus. In E. Fuß et al. (Eds.), Grammar and Corpora 2016 (pp. 415–424). Heidelberg University Publishing. https://doi.org/10.17885/heiup.361.c4714

Weiss, Z., & Meurers, D. (2019). Broad linguistic modeling is beneficial for German L2 proficiency assessment. In Widening the scope of learner corpus research (pp. 419–435). Presses Universitaires de Louvain.

Wisniewski, K. (2020). SLA developmental stages in the CEFR-related learner corpus MERLIN: Inversion and verb-end structures in German A2 and B1 learner texts. International Journal of Learner Corpus Research, 6(1), 1–17. https://doi.org/10.1075/ijlcr.18008.wis

Xi, X. (2017). Validity and validation in language testing. In E. Shohamy, I. G. Or, & S. May (Eds.), Language testing and assessment (pp. 1–14). Springer. https://doi.org/10.1007/978-3-319-02261-1_1