Journal Singular: Pedagogical Language, Literature, and Cultural Studies
Volume 03 No 02
http://dx.doi.org/10.63011/js.v3i2.65

Academic Integrity and Generative AI Use among Indonesian Students of German: Perceptions, Practices, and Pedagogical Implications

Alamsyah[1]*, Syarifah Fatimah[2], Muftihaturahmah Burhamzah[3]

[1] Universitas Negeri Makassar, Indonesia
[2] Universitas Negeri Makassar, Indonesia
[3] Universitas Negeri Makassar, Indonesia

Corresponding author:
E-mail: der_alamsyah@unm.ac.id

Abstract
Generative artificial intelligence (GenAI) has blurred the boundary between legitimate language support and academic misconduct, particularly in foreign language programmes, where the linguistic form of a text is the assessed competence. This study examined how Indo-nesian university students of German perceive, practise and justify GenAI use in their coursework. An explanatory sequential mixed-methods design combined a survey of 264 students from four Ger-man education programmes with interviews of 14 students. Students accepted language-support uses and rejected wholesale substitution, but three tasks, most notably translating their own Indonesian drafts into German, were judged problematic yet widely practised. In a PLS-SEM model based on the Theory of Planned Behaviour and neu-tralization theory, attitude, subjective norms, perceived behavioural control and neutralization explained 53.8% of the variance in the intention to use GenAI questionably, and intention predicted reported questionable practice. Interviews showed that students separated their own content from German form and cited limited proficiency as a necessity. German programmes should therefore introduce task-specific AI rules, proficiency-appropriate tasks, process-based as-sessment and AI literacy within language courses.

Keywords: *Academic integrity, Generative AI, German as a foreign language, Neutralization theory, Theory of planned behaviour*

Introduction

German as a foreign language in Indonesia is a relevant setting for examining this problem. Indonesian university students usually learn German as a third language after Indonesian, often alongside a regional language, and English. Most begin as beginners and work towards CEFR A2–B1 during their studies. At these levels, the gap between what learners can write unaided and what a chatbot produces in fluent German is large, which makes outsourcing tempting. Indone-sian higher education has adopted GenAI rapidly (Kharis et al., 2024), while institutional guid-ance on acceptable use remains uneven. Evidence on how these students perceive and use GenAI is needed for fair assessment policies, for the education of future German teachers and for testing whether integrity findings from English-medium settings transfer to other languages.

How to cite:
Alamsyah., Fatimah, S., & Burhamzah, M. (2026). Academic Integrity and Generative AI Use among Indonesian Students of German: Perceptions, Practices, and Pedagogical Implications. *Journal Singular: Pedagogical Language, Literature, and Cultural Studies.* Nexus Publishing. Pages 68-77. doi: 10.63011/js.v3i2.65

This ambiguity is sharper in foreign language education than in most disciplines, because the linguistic form of a text is the very competence being taught and assessed. When a chatbot corrects grammar, rephrases a sentence or translates a paragraph, it performs the skill that the assignment is meant to develop (Barrot, 2023; Köhncke, 2025; Kohnke et al., 2023). Machine translation had already blurred the line between legitimate support and substitution (Ducar & Schocket, 2018); GenAI extends this problem to whole texts. Moreover, learners need AI literacy, particularly the ability to evaluate AI output critically (Ng et al., 2021), but learners at lower proficiency levels are least able to judge whether fluent AI-generated text is correct or reflects what they meant.

Since the public release of ChatGPT in November 2022, generative artificial intelligence (GenAI) has become a routine study tool in higher education. Large language models can draft, translate, paraphrase and correct text on demand, which challenges established notions of authorship, originality and fair assessment (Kasneci et al., 2023). Conventional plagiarism detection and assessment design are poorly equipped for machine-generated text (Cotton et al., 2024). (Chan, 2025) describes this problem as "AI-giarism" and shows that students clearly reject submitting AI-generated content as their own work but are ambivalent about subtler uses such as paraphrasing or language polishing. Academic integrity, understood as a commitment to values such as honesty, fairness and responsibility rather than only the avoidance of misconduct (Bretag, 2016; International Center for Academic, 2021), has therefore become a spectrum of practices whose acceptability depends on task, purpose and disclosure.

Theoretical Background

TPB explains intention but not how students reconcile questionable behaviour with a self-image as honest learners. Neutralization theory (Sykes & Matza, 1957) addresses this by describing justifications that temporarily neutralise guilt, such as *denial of responsibility* ("everyone uses it"), *denial of injury* ("no one is harmed") and *condemnation of the condemners* ("the lecturers use AI too"). In language learning, a typical neutralization may be describing AI use as "only correcting grammar" when the tool rewrites most of the text. Combining both theories allows this study to explain both what drives questionable use and how students justify it.

Two theories guide the explanatory part of this study. The Theory of Planned Behaviour (Ajzen, 1991) holds that behaviour follows from intention, which is shaped by *attitude* (the evaluation of the behaviour, for example that AI saves time), *subjective norms* (perceived expectations of peers and lecturers) and *perceived behavioural control* (the perceived ease of the behaviour, including the risk of detection). In a survey of 702 academics, (Bin-Nashwan et al., 2023) found that time saving, self-efficacy and stress increased ChatGPT use, whereas peer influence and academic integrity decreased it.

Previous Studies and Research Gap

Table 1. Previous studies and remaining research gaps

Study	Context and sample	Main focus	What remains open
(Bin-Nashwan et al., 2023)	702 academics, international	Drivers of ChatGPT use; integrity as predictor	Academics, not students; no language tasks
(Stöhr et al., 2024)	University students, Sweden	Usage and attitudes across groups	Foreign languages not examined separately
(Chan, 2025)	393 students, Hong Kong	Perceptions of AI-giarism scenarios	Scenarios not tied to language-skill tasks

Study	Context and sample	Main focus	What remains open
(Katanneck & Suner Muñoz, 2024)	19 students of German	Self-regulated learning with ChatGPT feedback	Learning gains, not integrity
(Köhncke, 2025)	17 B2 learners of German	ChatGPT and four language skills	Small pilot; integrity not addressed
(Amin et al., 2026)	56 students, one Indonesian university	Usefulness and ethical concerns of ChatGPT	Single site; no discipline-specific analysis

Four gaps emerge. First, integrity research treats students as one group or focuses on English, although foreign language learning is where GenAI most directly performs the assessed skill; studies on German have examined AI for feedback and skill gains but not integrity (Katanneck & Suner Muñoz, 2024; Köhncke, 2025; Kohnke et al., 2023). Second, existing instruments ask about AI use in general (Chan, 2025) and do not distinguish language-specific practices such as translating an Indonesian draft into German. Third, Indonesian studies report perceptions and general ethical concerns (Amin et al., 2026; Kharis et al., 2024) but rarely relate beliefs to practices or explain the gap between them. Fourth, few studies derive concrete implications for task design, assessment and teacher education. This study addresses these gaps with four research questions:

1. RQ1: How do Indonesian students of German judge the acceptability of different GenAI practices in German language tasks?
2. RQ2: Which GenAI practices do they report using, and how often?
3. RQ3: To what extent do attitudes, subjective norms, perceived behavioural control and neutralization predict their intention to use GenAI in ways that conflict with integrity norms?
4. RQ4: How do students justify practices that they themselves consider problematic?

Method

Research Design

The study used an explanatory sequential mixed-methods design (Creswell & Plano Clark, 2018). A survey (Phase 1) answered RQ1–RQ3, and semi-structured interviews (Phase 2) answered RQ4 by explaining the survey patterns. The strands were integrated through a joint display (Fetters et al., 2013). The structural model tested the following hypotheses: attitude (H1), subjective norms (H2) and perceived behavioural control (H3) positively predict the intention to use GenAI in ways that conflict with integrity norms (Ajzen, 1991); neutralization tendency positively predicts this intention (H4; (Haines et al., 1986); intention positively predicts reported questionable practice (H5); and perceived behavioural control alsopredicts practice directly (H6).

Participants

Participants were undergraduate students in German language education programmes at Indonesian state universities, recruited through the Indonesian Germanists Association (IGV) network, with Universitas Negeri Makassar as the coordinating site. Eligible students were in semesters 2–8, had completed at least one German skills course and consented voluntarily. An a priori power analysis in G*Power (Faul et al., 2007) for four predictors ($f^2 = 0.15$, $\alpha = .05$, power = .95) indicated a minimum of 129 respondents. For Phase 2, 14 survey respondents were

selected through maximum variation purposive sampling based on their reported use and on the consistency between their judgements and practices. Interviewing stopped when no new justification types emerged in two consecutive interviews.

Instruments

The questionnaire comprised six sections (Table 2). Items were translated from English into Indonesian and back-translated by an independent translator (Brislin, 1970). Sections B and C used the same 12 German learning tasks, ranging from spelling checks to generating a complete assignment, so that each student's acceptability judgement could be compared with their reported practice for every task.

Table 2. Structure of the questionnaire

Section	Construct	Example item	Scale	Basis
A	Background	Semester, CEFR level, GenAI tools used	Categorical	Researcher-developed
B	Acceptability of GenAI practices (RQ1)	"Using GenAI to translate my Indonesian draft into German and submitting it as my own text"	5-point, not misconduct at all to serious misconduct	Adapted from (Chan, 2025)
C	Reported GenAI practices (RQ2)	Same 12 tasks as Section B	5-point, never to very often	Researcher-developed
D	Attitude, subjective norms, perceived behavioural control, intention (RQ3)	"Most of my classmates use GenAI to write German assignments"	7-point agreement	(Ajzen, 1991; Bin-Nashwan et al., 2023)
E	Neutralization tendency (RQ3)	"Using GenAI is acceptable if the lecturer does not explain the rules clearly"	7-point agreement	(Haines et al., 1986)
F	Social desirability (control)	Short-form items	True/false	(Reynolds, 1982)

Five experts in German language education, academic integrity, measurement and school teaching assessed content validity, and items were retained at I-CVI ≥ .78 (Polit & Beck, 2006). A pilot with students outside the main sample tested clarity and completion time. The interview guide asked participants to describe a recent German assignment completed with GenAI, how they decided what was acceptable and how they would explain their use to a lecturer.

Data Collection

After ethics approval, the anonymous online survey was distributed over four weeks by programme coordinators and class representatives rather than by the researchers. Students who wished to be interviewed registered through a separate form, so that interview contact details could not be linked to survey responses. Interviews lasted 30–45 minutes, were conducted in Indonesian by a researcher who did not teach the participant, and were recorded and transcribed verbatim. To reduce socially desirable responding and common method bias (Podsakoff

et al., 2003), the survey was fully anonymous, avoided the word "cheating", used different scale formats for predictors and outcomes, and included a social desirability control.

Data Analysis

Descriptive statistics answered RQ1 and RQ2. A task was defined in advance as a "gap task" if it was rated at least 3.00 for seriousness and used at least *sometimes* by at least 40% of students. Group differences were tested with Kruskal–Wallis tests. RQ3 was tested with PLS-SEM in SmartPLS 4, which suits prediction-oriented models with newly adapted constructs and non-normal data (Hair et al., 2013; Sarstedt et al., 2014). The measurement model was evaluated through loadings, composite reliability, average variance extracted and the heterotrait–monotrait ratio (HTMT < .85; (Fornell & Larcker, 1981; Henseler et al., 2015). Paths were tested with 10,000 bootstrap subsamples.

Interviews were analysed through thematic analysis (Braun & Clarke, 2006), combining deductive codes from neutralization theory and TPB with inductive codes for language-specific justifications. Two researchers independently coded 20% of the transcripts and refined the codebook before coding the remainder. Trustworthiness was supported through member checking, triangulation with survey data, an audit trail and reflexive memos on the researchers' role as German lecturers (Lincoln & Guba, 1985).

Ethical Considerations

The study was approved by the research ethics committee of the coordinating university (approval number 1578/UN36/LP2MP/2025). Participants gave informed consent and could withdraw at any time. Because the study asked about behaviour that may breach academic rules, no identifying data were collected in the survey, transcripts were pseudonymised, and data were never shared with lecturers or used in disciplinary processes. Participation had no effect on grades.

Of 291 returned questionnaires from four programmes, 264 complete responses were retained (90.7%). Most respondents were female (74.2%), were in semesters 2–5 (71.6%) and reported A2 (42.4%) or B1 (33.7%) proficiency. Nearly all (90.2%) had used GenAI for German coursework in the current semester, most often ChatGPT (91.3%), Google Translate (76.9%) and DeepL (63.3%).

Results

Acceptability and Practice (RQ1, RQ2)

Students distinguished clearly between support and substitution (Table 3). Language-support uses such as vocabulary lookup and grammar checking were rated as acceptable and used most often, whereas generating complete assignments and answering tests with AI were rated as serious misconduct and used rarely.

Table 3. Acceptability and frequency of GenAI practices

Task	Seriousness M (SD)	Frequency M (SD)	Users (%)	Gap task
T11 Generating and submitting a complete assignment	4.47 (0.81)	1.41 (0.79)	9.8	
T10 Answering online tests with AI	4.21 (0.92)	1.72 (0.98)	18.2	
T12 Having AI write a reflection or journal	3.64 (1.06)	1.98 (1.07)	27.3	

Task	Seriousness M (SD)	Frequency M (SD)	Users (%)	Gap task
T7 Translating own Indonesian draft into German	3.38 (1.09)	2.96 (1.21)	54.9	Yes
T9 Having AI write a presentation script	3.29 (1.10)	2.48 (1.19)	40.5	Yes
T8 Adapting an AI-generated model text	3.12 (1.03)	2.54 (1.15)	42.4	Yes
T6 Having AI rewrite a whole own text	2.91 (1.05)	2.71 (1.17)	47.0	
T5 Paraphrasing own German sentences	2.47 (1.01)	2.95 (1.12)	55.3	
T4 Generating ideas or an outline	2.06 (0.94)	2.88 (1.14)	52.7	
T1 Checking spelling and grammar	1.42 (0.71)	3.61 (1.02)	78.4	
T3 Asking AI to explain a grammar rule	1.38 (0.66)	3.27 (1.08)	66.3	
T2 Looking up words and example sentences	1.31 (0.62)	3.84 (0.95)	83.7	

Note. Seriousness: 1 = not misconduct at all, 5 = serious misconduct. Frequency: 1 = never, 5 = very often. Users = students reporting the practice at least sometimes.

Three gap tasks were identified: translating one's own Indonesian draft (T7), generating a presentation script (T9) and adapting a model text (T8). In all three, GenAI produces the German text while the student supplies the content or structure. Students in semesters 6–8 rated T7 as less serious than students in semesters 2–3 ($H(2) = 7.84$, $p = .020$), and lower self-reported proficiency was associated with more frequent use of T7 ($\rho = -.21$, $p < .001$).

Predictors of Questionable Use (RQ3)

After one perceived behavioural control item with a low loading (.612) was removed, all constructs met the thresholds for reliability and validity (loadings .709–.912; composite reliability .863–.925; AVE .581–.804; all HTMT < .85). Full collinearity VIFs were below 3.3. Table 4 presents the structural model.

Table 4. Structural model and hypothesis testing

H	Path	β	t	p	f^2	Decision
H1	ATT → INT	.384	6.52	< .001	.21	Supported
H2	SN → INT	.196	3.31	.001	.06	Supported
H3	PBC → INT	.142	2.41	.016	.03	Supported
H4	NEUT → INT	.231	4.08	< .001	.08	Supported
H5	INT → QPB	.452	7.86	< .001	.24	Supported
H6	PBC → QPB	.087	1.39	.164	.01	Not supported

Note. ATT = attitude; SN = subjective norms; PBC = perceived behavioural control; NEUT = neutralization; INT = intention; QPB = questionable practice (frequency of T7–T12). Social desirability was controlled ($\beta = -.11$, $p = .041$).

The four predictors explained 53.8% of the variance in intention, and intention explained 26.8% of the variance in questionable practice. Attitude was the strongest predictor, and neutralization added explanatory power beyond the three TPB constructs. All four predictors affected questionable practice indirectly through intention (indirect $\beta = .064$–$.174$, all $p < .05$).

Justifications (RQ4)

Interviews with 14 students (nine with large judgement–practice gaps) yielded five themes (Table 5).

Table 5. Themes in students' justifications

Theme	Link to theory	n	Illustrative excerpt
"It's only correction"	Denial of responsibility	11	"The ideas are mine. ChatGPT only puts them into correct German." (P04)
Proficiency gap as necessity	Language-specific justification	10	"At A2 I cannot write what the task asks." (P09)
Unclear rules	Condemnation of the condemners	9	"No lecturer ever said translation with AI is forbidden." (P02)
"Everyone does it"	Subjective norms, denial of injury	8	"In our class group people share the prompts." (P12)
Lines students still hold	Integrity values	13	"Tests and the whole assignment, no. That is really cheating." (P07)

The most frequent theme separated *content*, claimed as the student's own, from *German form*, treated as a technical service. The proficiency-gap theme was voiced mainly by A1–A2 students and mirrors the quantitative association between lower proficiency and translation use. Integrating both strands, questionable use concentrated on tasks in which GenAI produces the German language, and it was sustained by justifications rather than by ignorance of the rules.

Discussion

A Language-Specific Grey Zone (RQ1, RQ2)

Students judged GenAI use along a continuum, as (Chan, 2025) found, but the con-tested middle ground in foreign language learning was cross-lingual meaning transfer rather than paraphrasing in general. This extends the earlier problem of machine trans-lation (Ducar & Schocket, 2018). Students' distinction between their content and German form fits general notions of authorship, which locate it in ideas, but not language courses, where form is the assessed competence (Barrot, 2023). This mismatch is the central finding for German programmes. The association between lower proficiency and translation use further suggests that tasks set above learners' level push them towards outsourcing, which bypasses the cognitive work through which writing develops. The lower seriousness ratings of senior students may indicate normalisation, as observed in earlier cheating research (Diekhoff et al., 1996), but longitudinal data are needed to confirm this.

Drivers and Justifications (RQ3, RQ4)

The model supports TPB as a framework for GenAI-related integrity behaviour (Ajzen, 1991). In contrast to (Bin-Nashwan et al., 2023), who found that peer influence reduced ChatGPT use among academics, subjective norms increased questionable inten-tion among students, plausibly because visible peer use in class groups normalises rather than restrains it. Perceived behavioural control did not predict behaviour directly, prob-ably because free and undetectable tools leave almost no student without control. Neu-tralization explained intention beyond TPB,

consistent with (Haines et al., 1986). The interviews show which justifications are at work: most correspond to (Sykes & Matza, 1957) techniques, but the proficiency-gap justification extends the typology because students did not deny the norm but argued that the task made compliance impossible. As in the Swedish surveys (Malmström et al., 2023), unclear rules also played a role, while students still held firm boundaries around tests and complete assignments.

Pedagogical Implications

The findings suggest that clear, task-level guidance is more promising than general bans. German programmes can (1) state for each assignment which AI uses are allowed, (2) set writing and speaking tasks at learners' actual CEFR level, (3) use process-based assessment such as in-class drafts and revision logs (Cotton et al., 2024) and (4) teach students to evaluate AI-generated German as part of AI literacy (Ng et al., 2021), in-cluding in teacher education.

Limitations

Self-reports of sensitive behaviour probably underestimate questionable use. The cross-sectional design does not permit causal conclusions or confirmation of the normali-sation interpretation. Programmes were recruited through a professional network, and proficiency was self-reported. Future studies could use longitudinal designs, analyse submitted texts and compare German with other foreign languages.

Conclusion

Indonesian students of German accepted GenAI as language support and rejected wholesale substitution, but they widely used it for tasks they considered problematic, above all translating their own Indonesian drafts into German. Attitude, subjective norms, perceived behavioural control and neutralization explained their intention to use GenAI in these ways, and students justified such use by separating content from German form and by citing their limited proficiency. For German programmes, these findings point to task-specific AI rules, proficiency-appropriate tasks, process-based assessment and AI literacy embedded in language courses and teacher education.

References

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50(2), 179-211. https://doi.org/10.1016/0749-5978(91)90020-T

Amin, M., Mustaqim, B., Pratama, W., & Sibuea, A. M. (2026). Assessing student perspectives on ChatGPT in higher education: A quantitative analysis. IAES International Journal of Artificial Intelligence, 15(2). https://ijai.iaescore.com/index.php/IJAI/article/view/29514

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing, 57, 100745. https://doi.org/10.1016/j.asw.2023.100745

Bin-Nashwan, S. A., Sadallah, M., & Bouteraa, M. (2023). Use of ChatGPT in academia: Academic integrity hangs in the balance. Technology in Society, 75, 102370. https://doi.org/10.1016/j.techsoc.2023.102370

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

Bretag (Ed.). (2016). Handbook of academic integrity. Springer. https://doi.org/10.1007/978-981-287-098-8.

Brislin, R. W. (1970). Back-translation for cross-cultural research. Journal of Cross-Cultural Psychology, 1(3), 185-216. https://doi.org/10.1177/135910457000100301

Chan, C. K. Y. (2025). Students' perceptions of 'AI-giarism': Investigating changes in understandings of academic misconduct. Education and Information Technologies. https://doi.org/10.1007/s10639-024-13151-7

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, 61(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Creswell, J. W., & Plano Clark, V. L. (2018). Designing and conducting mixed methods research (3rd ed.). SAGE.

Diekhoff, G. M., LaBeff, E. E., Clark, R. E., Williams, L. E., Francis, B., & Haines, V. J. (1996). College cheating: Ten years later. Research in Higher Education, 37(4), 487-502. https://doi.org/10.1007/BF01730111

Ducar, C., & Schocket, D. H. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google Translate. Foreign Language Annals, 51(4), 779-795. https://doi.org/10.1111/flan.12366

Fetters, M. D., Curry, L. A., & Creswell, J. W. (2013). Achieving integration in mixed methods designs: Principles and practices. Health Services Research, 48(6 Pt 2), 2134-2156. https://doi.org/10.1111/1475-6773.12117

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39-50. https://doi.org/10.1177/002224378101800104

Haines, V. J., Diekhoff, G. M., LaBeff, E. E., & Clark, R. E. (1986). College cheating: Immaturity, lack of commitment, and the neutralizing attitude. Research in Higher Education, 25(4), 342-354. https://doi.org/10.1007/BF00992130

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2013). Partial least squares structural equation modeling: Rigorous applications, better results and higher acceptance. Long Range Planning, 46(1-2), 1-12. https://doi.org/10.1016/j.lrp.2013.01.001

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. Journal of the Academy of Marketing Science, 43(1), 115-135. https://doi.org/10.1007/s11747-014-0403-8

International Center for Academic, I. (2021). The fundamental values of academic integrity.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T.,…Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Katanneck, S., & Suner Muñoz, F. (2024). Selbstreguliertes Lernen mit Feedback von ChatGPT: Die Rolle individueller Einstellungen und Wahrnehmungen beim selbstregulierten Lernen mit KI-basiertem Feedback. KONTEXTE: Internationales Journal zur Professionalisierung in Deutsch als Fremdsprache, 2(2). https://doi.org/10.24403/jp.1394575

Kharis, M., Wijaya, D. N., Yafie, E., Waluyo, B., & Resultay, R. G. (2024). The impact of AI usage on transforming student learning culture in Indonesia in alignment with SDG 4. Journal of Ecohumanism, 3(8), 1651-1661.

Köhncke, C. (2025). The impact of using ChatGPT in foreign language acquisition: A pilot study EuroCALL 2025 Conference Proceedings,

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. RELC Journal, 54(2), 537-550. https://doi.org/10.1177/00336882231162868

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic inquiry. SAGE.

Malmström, H., Stöhr, C., & Ou, A. W. (2023). Chatbots and other AI for learning: A survey of use and views among university students in Sweden [2023:1](Chalmers Studies in Communication and Learning in Higher Education, Issue.

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, 100041. https://doi.org/10.1016/j.caeai.2021.100041

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. Journal of Applied Psychology, 88(5), 879-903. https://doi.org/10.1037/0021-9010.88.5.879

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. Research in Nursing & Health, 29(5), 489-497. https://doi.org/10.1002/nur.20147

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. Journal of Clinical Psychology, 38(1), 119-125. https://doi.org/10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I

Sarstedt, M., Ringle, C. M., Smith, D., Reams, R., & Hair, J. F. (2014). Partial least squares structural equation modeling (PLS-SEM): A useful tool for family business researchers. Journal of Family Business Strategy, 5(1), 105-115. https://doi.org/10.1016/j.jfbs.2014.01.002

Stöhr, C., Ou, A. W., & Malmström, H. (2024). Perceptions and usage of AI chatbots among students in higher education across genders, academic levels and fields of study. Computers and Education: Artificial Intelligence, 7, 100259. https://doi.org/10.1016/j.caeai.2024.100259

Sykes, G. M., & Matza, D. (1957). Techniques of neutralization: A theory of delinquency. American Sociological Review, 22(6), 664-670. https://doi.org/10.2307/2089195