

Structured Generative AI Co-Tutoring for Learning Outcomes and Motivation in Information Systems

¹Mei Ratnasari*, ²Amalyanda Azhari, ³Diyah Trinovita, ⁴Widi Andewi, ⁵Eva Nurchurifiani

^{1,3}Information System, STMIK Kalirejo Lampung, Indonesia

²Computer Science, STMIK Kalirejo Lampung, Indonesia

⁴Information System, Institut Bakti Nusantara, Indonesia

⁵English Department, STKIP PGRI Bandar Lampung, Indonesia

^{1,2,3}Jl. Jend. Sudirman No.569, Kali Rejo, Kec. Kalirejo, Kabupaten Lampung Tengah, Lampung

⁴Jl. Wisma Rini No. 09, Pringsewu Selatan, Kec. Pringsewu, Kabupaten Pringsewu, Lampung

⁵Jl. Khairil Anwar Nomor 79, Kec. Tanjung Karang Pusat, Kota Bandar Lampung, Lampung

*e-mail: meiratnasarialfian@gmail.com

(received: 17 August 2026, revised: 28 August 2026, accepted: 29 August 2026)

Abstract

Generative AI (GenAI) offers interactive explanations, feedback, and personalized support in higher education, yet its educational value depends on how such support is structured. Evidence remains limited regarding whether structured GenAI co-tutoring can simultaneously improve independent learning outcomes and learning motivation in concept-oriented Information Systems courses. This study examined the effectiveness of structured GenAI co-tutoring compared with conventional instruction in an undergraduate Information Systems Concepts course in Indonesia. A quasi-experimental nonequivalent control group pretest–posttest design involved 60 students from two intact classes, with 30 students in each condition. Learning outcomes were measured using parallel forms of the Information Systems Concepts Achievement Test, while learning motivation was assessed using the EVC Light scale. Data collection comprised pre-intervention assessment, the instructional intervention, and post-intervention assessment; achievement tests were completed without GenAI access to capture independently demonstrated learning. ANCOVA was used to compare posttest outcomes while controlling for corresponding pretest scores. The experimental group achieved a significantly higher adjusted ISCAT-B score than the control group, with an adjusted mean difference of 7.89 points, $F(1, 57) = 27.65$, $p < .001$, partial $\eta^2 = .327$. Learning motivation was also significantly higher in the experimental group, with an adjusted EVC Index difference of 0.053, $F(1, 57) = 13.53$, $p < .001$, partial $\eta^2 = .192$. The study extends evidence on structured GenAI support by examining independently assessed conceptual learning alongside expectancy–value–cost motivation in Information Systems education. In practice, lecturers can adopt a sequence of independent attempts, guided AI feedback, verification, and revision, followed by assessment without AI assistance.

Keywords: conceptual learning, Generative AI, information systems education, learning motivation, learning outcomes

1 Introduction

Generative artificial intelligence (GenAI), particularly large language models (LLMs) such as ChatGPT, has rapidly expanded the possibilities for technology-supported learning in higher education. Unlike conventional digital resources that primarily provide static information, GenAI can engage students in interactive dialogue, generate context-sensitive explanations and examples, respond to follow-up questions, and provide immediate feedback. These capabilities allow GenAI to function not only as a content-generation tool but also as a potential virtual tutor or co-tutor that supports students during independent learning [1]. Recent evidence has indicated that appropriately integrated GenAI can contribute to students' learning outcomes and affective-motivational experiences [2]. This potential is particularly relevant in university courses that require students to understand abstract concepts, connect theoretical knowledge with real-world situations, and independently monitor their understanding beyond scheduled classroom instruction.

However, access to GenAI does not necessarily lead to improved learning outcomes or stronger learning motivation. The educational benefits of GenAI appear to depend considerably on how the technology is incorporated into learning activities. Although experimental evidence generally points to positive effects of ChatGPT-supported instruction, considerable variation has been reported across intervention designs, learning tasks, and measured outcomes [3]. Moreover, unrestricted reliance on GenAI may encourage students to delegate cognitive work to the technology rather than actively process, evaluate, and verify information. For example, GenAI could improve performance while assistance was available but could disadvantage subsequent independent performance when pedagogical safeguards were absent [4]. Similarly, motivational benefits should not be assumed simply because learners interact with an AI system, as the effects of educational chatbots on motivation vary according to instructional design and the type of support provided [5]. The key issue is whether GenAI is structured to support meaningful cognitive engagement rather than replace it.

This issue is particularly important in computing-related education, where GenAI has increasingly been used to provide explanations, feedback, hints, debugging support, and personalized assistance. A growing body of research has examined GenAI-supported learning in programming and related computing contexts [6]. For example, a ChatGPT-based intelligent tutoring system has shown potential for providing progressively structured support in Python learning [7], while an LLM-based programming environment has been associated with improvements in both students' motivation and programming ability [8]. Nevertheless, findings remain mixed. ChatGPT-supported programming has been shown to influence students' learning behaviors and perceptions without producing significant differences in programming performance [9]. However, most controlled studies focus on programming tasks such as code generation, debugging, and procedural problem solving. These settings differ from concept-oriented Information Systems learning, which requires students to understand relationships among people, processes, data, and technology and apply them to organizational contexts.

Consequently, a more specific gap remains concerning the use of structured GenAI-supported co-tutoring in foundational Information Systems education. Existing research demonstrates that GenAI can support academic and computing tasks [4], [8], [9], [10], but it remains less clear whether a deliberately structured co-tutoring approach can simultaneously improve students' learning outcomes and learning motivation in a concept-oriented Information Systems course compared with conventional instruction. Immediate task performance and positive perceptions of AI do not necessarily reflect deeper learning, sustained motivation, or improved academic achievement. Research linking motivational outcomes with measurable learning outcomes in educational chatbot use remains limited [5]. This gap is especially relevant in Indonesian higher education, where a recent study documented students' use of GenAI for academic tasks alongside concerns about overdependence, critical thinking, and academic integrity [11]. However, the study relied on self-reported experiences and did not establish whether GenAI improves independently demonstrated learning. For Information Systems education, this distinction matters because students must understand and apply relationships among people, processes, data, and technology rather than rely on AI-generated answers. Further research is needed to examine whether structured GenAI co-tutoring can support both learning outcomes and motivation in regular undergraduate instruction. This study addressed this gap by examining structured GenAI co-tutoring in an undergraduate Information Systems Concepts course in Indonesia. The intervention used guided prompting, concept exploration, GenAI-supported feedback, and reflection, positioning GenAI as learning support rather than an answer generator. Students in the experimental group were compared with students receiving conventional instruction on the same course content. The study examined whether this structured approach was associated with differences in both learning outcomes and learning motivation. Accordingly, the following research questions were addressed:

- 1 To what extent do students receiving structured GenAI-supported co-tutoring differ in their learning outcomes from students receiving conventional instruction?
- 2 To what extent do students receiving structured GenAI-supported co-tutoring differ in their learning motivation from students receiving conventional instruction?

2. Literature Review

2.1 Pedagogical Design of Structured GenAI Co-Tutoring

GenAI can support digital learning through on-demand explanations, examples, feedback, and personalized assistance. Systematic reviews have identified key benefits such as continuous learning support, clarification of difficult concepts, self-assessment, engagement, learner autonomy, accessibility, personalization, and interactivity [12]. However, GenAI also presents risks related to inaccurate information, bias, privacy, academic integrity, overdependence, and digital inequality [13]. Recent research further highlights its potential to support critical thinking, reflective learning, autonomy, and prompt literacy, while warning that excessive reliance may reduce students' cognitive and metacognitive effort [14]. Therefore, the educational value of GenAI depends on how it is integrated into learning activities that require students to evaluate, question, and regulate AI-generated support.

Structured GenAI support appears more effective when students remain responsible for the learning process. In guided settings, students attempt problems independently before receiving hints, which has been associated with stronger knowledge construction, self-regulated learning, and higher-order thinking skills than unrestricted ChatGPT use [15]. Structured activities such as generating practice questions, checking answers, receiving feedback, evaluating AI outputs, and engaging in dialogue can also keep students actively involved in learning [16]. By contrast, GenAI may improve performance while assistance is available but weaken subsequent independent performance when adequate safeguards are absent [4].

Evidence from computing education shows that the effects of GenAI depend on how it is used. In programming courses, GenAI use has been linked to students' learning practices and performance, but experimental findings remain mixed [17]. Scaffolded AI tutors can improve exercise performance and intrinsic motivation yet gains in conceptual knowledge or code comprehension are not always observed [18]. Similarly, structured ChatGPT activities have been associated with self-regulated learning, knowledge development, engagement, and motivation [19]. However, most existing studies focus on programming, science, or task-specific contexts. Evidence remains limited on whether structured GenAI co-tutoring can support conceptual learning in foundational Information Systems courses while reducing the risks of unrestricted AI assistance.

2.2 Learning Outcomes and Motivation in GenAI-Supported Learning

GenAI can improve learning outcomes, but its effects depend on how students use it. Meta-analytic evidence shows positive effects on academic performance and higher-order thinking skills, although results vary across intervention designs [2]. Higher-level learning is more likely when students use GenAI to build and extend knowledge rather than simply complete tasks [20]. Learning performance is also associated with factors such as self-efficacy, creativity, fairness and ethics, and trust [21]. These findings indicate that GenAI supports learning most effectively when students engage with it actively and use it to strengthen, rather than replace, their cognitive processes.

Learning motivation is another important outcome of GenAI-supported learning. Personalized feedback, individualized assistance, and greater control over learning can strengthen students' competence and engagement [8]. Studies of intentionally designed GenAI environments have reported improvements in autonomous motivation and learning performance [22]. Meta-analytic evidence also shows positive effects on motivation and behavioral, cognitive, and emotional engagement, although these effects vary across subjects, learning strategies, and instructional contexts [23]. However, excessive reliance on GenAI may reduce engagement. These findings indicate that motivational benefits depend on how GenAI activities are designed and used, rather than on access to the technology alone.

Learning outcomes and motivation should be examined together because stronger AI-assisted performance does not always reflect deeper learning. Frequent GenAI use may increase confidence and perceived efficiency while also increasing technological dependence [24]. Experimental evidence shows that AI-supported conditions can improve task performance without producing greater gains in conceptual knowledge, while scaffolded support may be more effective for motivation [18]. These findings show that performance, learning, and motivation can respond differently depending on how

GenAI is structured. Evidence is still limited on whether structured GenAI co-tutoring can improve both learning outcomes and motivation in foundational, concept-oriented Information Systems courses.

Based on the literature reviewed, the conceptual logic of this study is summarized in Figure 1. Structured GenAI co-tutoring is defined as guided learning support that includes an initial independent attempt, guided prompting, AI-generated explanations and feedback, evaluation or verification of AI responses, and revision or reflection [15], [16]. This structure is expected to sustain students' active cognitive engagement and self-regulated learning, which may contribute to stronger independently demonstrated conceptual learning outcomes [20]. It is also expected to support learning motivation because structured assistance may strengthen students' expectancy and perceived value while reducing perceived cost in learning activities [22], [25]. At the same time, the framework assumes that the educational value of GenAI depends on instructional guidance, because unstructured use may increase technological dependence and weaken deeper learning [4], [24].

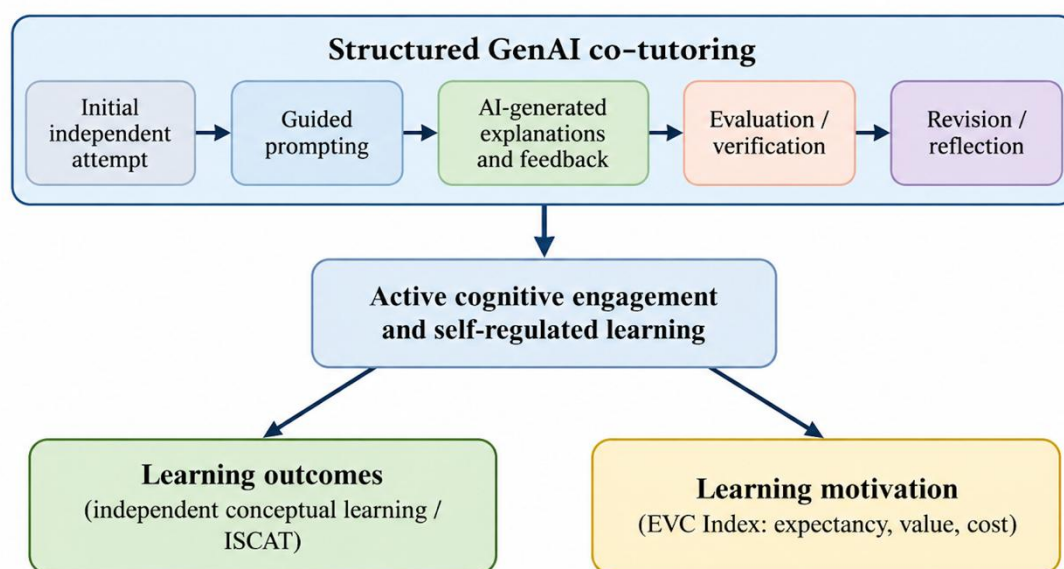


Figure 1 Conceptual framework of structured GenAI co-tutoring, learning outcomes, and learning motivation

3 Research Method

3.1 Research Design

This study employed a quantitative quasi-experimental nonequivalent control group pretest–posttest design to examine the effects of structured GenAI-supported co-tutoring on students' learning outcomes and learning motivation. A quasi-experimental design was appropriate because the study involved two intact classes and individual random assignment was not feasible, while the inclusion of a comparison group and pretest measures allowed baseline differences between the groups to be examined before the intervention [26], [27]. One class was assigned as the experimental group and received structured GenAI-supported co-tutoring, whereas the other class served as the control group and received conventional instruction covering the same course content and learning objectives. Both groups completed the Information Systems Concepts Achievement Test and the EVC Light scale before and after the instructional period. Parallel forms of the achievement test were used at pretest and posttest to reduce recall effects, while EVC Light was administered at both measurement points to capture changes in motivation within the computing-course context [25]. The primary between-group comparisons were conducted using ANCOVA, with posttest scores treated as the dependent variables, instructional condition as the fixed factor, and the corresponding pretest scores entered as covariates. This approach allows post-intervention group differences to be estimated while statistically adjusting for baseline variation and is generally preferable to relying solely on raw gain-score comparisons when baseline and follow-up measures are available [28]. However, with only one class per condition,

instructional condition was confounded with class membership, and unmeasured class differences could have contributed to the observed outcomes.

3.2 Participants

The study was conducted in the undergraduate Information Systems program at STMIK Kalirejo, a higher education institution in Lampung, Indonesia. The research was embedded in the Information Systems Concepts course during the second semester of the 2025–2026 academic year. This course was selected because it introduces foundational concepts concerning systems, information, organizational processes, data, technology, and the role of information systems in organizational activities. The instructional context was suitable for examining whether structured GenAI-supported co-tutoring could assist students in understanding and applying concept-oriented course content. The study involved 60 undergraduate Information Systems students drawn from two intact classes. Because the institutional timetable and existing class organization did not permit individual random assignment, the intact classes were retained throughout the study, consistent with the logic of nonequivalent-group quasi-experimental research [26], [27].

Participants were selected using purposive cluster sampling. The two classes were chosen because they were comparable in enrollment size and course scheduling and were enrolled in the same Information Systems Concepts course. One class, comprising 30 students, served as the experimental group and received structured GenAI-supported co-tutoring, while the other class, also comprising 30 students, served as the control group and received conventional instruction. Using intact classes reduced disruption to regular teaching and preserved the naturally occurring instructional setting, although equivalence between groups could not be assumed. Purposive class selection may have introduced differences in prior knowledge, learning habits, and GenAI experience that were not fully captured by the pretest measures. For this reason, both groups completed pre-intervention measures of learning outcomes and learning motivation, and the corresponding baseline scores were considered in the subsequent statistical analyses.

Eligibility criteria were established before data collection to maintain a reasonably homogeneous course-based sample. Participants were required to be actively enrolled in the Information Systems program, officially registered for the Information Systems Concepts course, not repeating the course, and willing to participate in the study. All eligible students in the two selected classes were invited to participate, resulting in a total sample of 60 students. Table 1 summarizes the participant characteristics available for the two instructional conditions.

Table 1 Demographic and academic characteristics of participants

Characteristic	Experimental Group (n = 30)	Control Group (n = 30)	Total (N = 60)
Age, years, M (SD)	19.43 (0.82)	19.37 (0.85)	19.40 (0.83)
Gender, n (%)			
Male	18 (60.0%)	17 (56.7%)	35 (58.3%)
Female	12 (40.0%)	13 (43.3%)	25 (41.7%)
Prior GenAI experience, n (%)			
Never/rarely	5 (16.7%)	6 (20.0%)	11 (18.3%)
Occasionally	14 (46.7%)	13 (43.3%)	27 (45.0%)
Frequently	11 (36.7%)	11 (36.7%)	22 (36.7%)
Active student status	30 (100%)	30 (100%)	60 (100%)
Repeating the course	0 (0%)	0 (0%)	0 (0%)
Study program	Information Systems	Information Systems	Information Systems
Course	Information Systems Concepts	Information Systems Concepts	Information Systems Concepts

3.3 Instructional Conditions

The two groups followed the same course learning objectives and content but differed in the type of instructional support provided during the intervention. In the experimental group, GenAI was integrated as a structured co-tutor rather than used as an unrestricted answer generator. Each learning cycle began with an independent initial attempt in which students examined a concept, question, or case before consulting GenAI. Students then used lecturer-prepared guided prompts to request clarification, examples, comparisons, or hints related to the target concept. This was followed by concept exploration and practice activities in which students interacted with GenAI-generated feedback. Importantly, students were required to evaluate and verify the responses before revising their initial answers and completing a brief reflection on what they had learned. This sequence was designed to maintain students' cognitive involvement while allowing GenAI to provide timely instructional support. The approach was informed by evidence that guidance-based ChatGPT use, in which learners attempt problems before receiving hints rather than direct solutions, can better support knowledge construction, SRL, and HOTS than unrestricted ChatGPT use [15]. Meanwhile, structured GenAI learning activities can position students as active participants who plan, evaluate, and reflect on their learning [16].

The control group addressed the same course topics and learning objectives without GenAI-supported co-tutoring. Learning activities consisted of lecturer explanations, classroom discussion, assigned learning resources, and written practice. To maintain comparability between conditions, the instructional content and targeted learning outcomes were aligned across both groups, while access to structured GenAI assistance constituted the principal instructional difference. An intervention protocol was prepared for each meeting to specify the topic, learning objective, experimental-group activity, corresponding control-group activity, and instructional duration, thereby supporting consistent implementation and replicability of the instructional conditions.

3.4 Research Instruments

Two instruments were used to measure the outcomes specified in the research questions: the Information Systems Concepts Achievement Test (ISCAT) for learning outcomes and the EVC Light scale for learning motivation. ISCAT was developed as a curriculum-aligned achievement measure because RQ1 concerned students' demonstrated mastery of the Information Systems Concepts course rather than their perceptions of learning. The test comprised four constructed-response items and one integrated case-analysis task covering systems and information fundamentals, information system components, organizational applications, decision support, and ethical or governance issues. Two parallel forms were prepared: Form A for the pretest and Form B for the posttest. Both forms followed the same content blueprint, targeted equivalent learning objectives and cognitive demands, and used the same scoring criteria, while differing in wording and case contexts to reduce recall effects. Responses were scored on a 0–100 scale using an analytic rubric addressing conceptual accuracy, completeness, conceptual integration, and application or justification. Evidence of content validity was obtained through expert review of the alignment among test items, course learning objectives, and intended cognitive demands, consistent with established principles of educational test validation [29], [30]. A subset of responses was independently scored by two raters, and inter-rater reliability was examined using the ICC, with the specific ICC model and confidence interval reported [31].

Learning motivation for RQ2 was measured using EVC Light [25], a brief instrument specifically designed and validated for undergraduate computing courses. The scale contains six items representing three theoretically distinct dimensions: expectancy, value, and cost, with two items assigned to each dimension. Expectancy represents students' beliefs about their capability to learn and succeed in the course, value reflects the perceived importance and usefulness of learning activities, and cost represents the perceived effort, time, or sacrifice associated with engagement. Responses are recorded on a five-point Likert scale. The original validation involved 245 undergraduate students from four universities and provided evidence for the three-factor structure through exploratory and confirmatory factor analyses as well as reliability analyses. Following the original scoring framework, expectancy, value, and cost indices were calculated, together with the overall EVC Index representing students' overall motivational level. The EVC Index served as the primary motivation outcome for

<http://sistemasi.ftik.unisi.ac.id>

RQ2, while the three component scores were retained for supplementary interpretation of how the intervention influenced particular aspects of motivation.

For use in the present Indonesian Information Systems context, the EVC Light items were translated and contextually adapted to refer to the Information Systems Concepts course without changing the underlying meaning of the constructs. The translated version was reviewed for linguistic and conceptual equivalence and piloted with students outside the main sample before administration. Internal consistency was examined for the three dimensions and the overall scale before inferential analyses were conducted. Both ISCAT and EVC Light were administered before and after the instructional period, enabling baseline-adjusted comparisons of learning outcomes and learning motivation between the experimental and control groups.

3.5 Data Collection

Data collection was conducted in three sequential phases: pre-intervention assessment, instructional intervention, and post-intervention assessment. Before the intervention, participants in both groups completed ISCAT Form A followed by EVC Light. ISCAT Form A established baseline learning outcomes, whereas EVC Light measured students' initial expectancy, value, cost, and overall motivation toward learning in the Information Systems Concepts course. Both assessments were administered under comparable conditions using the same instructions. Importantly, students completed ISCAT without access to GenAI or other external assistance so that the scores represented independently demonstrated knowledge rather than AI-assisted task performance. The experimental group then participated in structured GenAI-supported co-tutoring, incorporating independent initial attempts, guided prompting, concept exploration, practice with GenAI feedback, evaluation and verification of AI responses, and reflection or revision. The control group studied the same course content and targeted the same learning outcomes through lecturer explanations, classroom discussion, assigned learning resources, and written practice without GenAI-supported co-tutoring. Attendance and implementation records were maintained throughout the intervention to document exposure to the assigned instructional condition.

After the instructional period, both groups completed ISCAT Form B followed by EVC Light. ISCAT Form B was administered without GenAI under the same assessment conditions to measure learning that students could demonstrate independently. Prior studies show that AI-supported task performance does not always translate into stronger conceptual learning once AI assistance is removed [4], [18]. The pretest–posttest procedure assessed independent learning outcomes through ISCAT and changes in learning motivation through EVC Light for baseline-adjusted group comparisons

3.6 Data Analysis

Data analysis was conducted using descriptive and inferential statistics. Means, standard deviations, and score distributions were first calculated for all pretest and posttest measures. RQ1 was addressed using ANCOVA, with the ISCAT-B posttest total score as the dependent variable, instructional condition (structured GenAI-supported co-tutoring vs. conventional instruction) as the fixed factor, and the ISCAT-A pretest total score as the covariate. RQ2 was examined using a separate ANCOVA in which the posttest EVC Index served as the dependent variable, instructional condition as the fixed factor, and the pretest EVC Index as the covariate. Baseline-adjusted analysis was selected because it enables post-intervention group differences to be estimated while accounting for variation in corresponding pre-intervention scores and is preferable to relying solely on raw change scores when baseline and follow-up measurements are available.

Before interpreting each ANCOVA, assumptions concerning linearity between the covariate and outcome, homogeneity of regression slopes, homogeneity of variance, normality of residuals, and influential observations were examined. Homogeneity of regression slopes was assessed by testing the interaction between instructional condition and the corresponding pretest score. Statistical significance was evaluated at an alpha level of .05. For each primary analysis, adjusted posttest means, standard errors, adjusted mean differences, 95% confidence intervals, F statistics, p values, and partial η^2 were reported to provide information about both statistical and practical significance [32]. Expectancy, value, and cost scores from EVC Light were not treated as separate primary

<http://sistemasi.ftik.unisi.ac.id>

outcomes; when examined to clarify the pattern underlying the overall EVC Index, they were reported as secondary analyses. Because these analyses involved multiple statistical tests, Holm's sequential procedure was applied to control the familywise Type I error rate. Given the use of intact, nonrandomized classes, baseline adjustment was interpreted cautiously and was not assumed to eliminate all potential pre-existing group differences. The individual-level ANCOVA did not account for within-class dependence. With one class per condition, class effects could not be separated from instructional effects. Consequently, the adjusted group differences were interpreted as associations rather than definitive causal effects.

3.7 Ethical Consideration

Ethical approval for the study was obtained from the institutional research ethics authority at STMIK Kalirejo (No. 005/LPPM/STMIKKalirejo/XVI/2026). Before data collection, all participants received information about the study procedures and provided informed consent. Participation was voluntary, and students were informed that they could withdraw without academic penalty. Participant identities were anonymized during data processing and reporting, and research data were handled confidentially. GenAI use in the experimental group was limited to the designated instructional activities and was framed as learning support rather than a substitute for students' academic work.

4 Results

Data from 60 participants were included in the analysis, with 30 students in the structured GenAI-supported co-tutoring group and 30 in the conventional-instruction group. Descriptive statistics indicated that the two groups had similar scores before the intervention. For learning outcomes, the experimental group obtained a mean ISCAT-A pretest score of 51.16 (SD = 6.09), while the control group obtained a mean of 51.10 (SD = 8.54). Following the instructional period, the experimental group increased to 74.80 (SD = 6.74), corresponding to an unadjusted mean gain of 23.64 points. The control group also improved, reaching a posttest mean of 66.88 (SD = 7.66), with an unadjusted gain of 15.78 points. For learning motivation, the experimental group increased from a pretest EVC Index of 0.699 (SD = 0.078) to 0.769 (SD = 0.062), whereas the control group increased from 0.691 (SD = 0.074) to 0.713 (SD = 0.071). Thus, both groups showed positive changes across the instructional period, although descriptively larger improvements were observed in the experimental group for both outcomes. Descriptive statistics for learning outcomes and learning motivation can be seen in Table 2.

Table 2 Descriptive statistics for learning outcomes and learning motivation

Outcome	Group	Pretest M (SD)	Posttest M (SD)	Unadjusted Gain
ISCAT	Experimental	51.16 (6.09)	74.80 (6.74)	23.64
	Control	51.10 (8.54)	66.88 (7.66)	15.78
EVC Index	Experimental	0.699 (0.078)	0.769 (0.062)	0.070
	Control	0.691 (0.074)	0.713 (0.071)	0.022

Before conducting the primary ANCOVA models, the relevant assumptions were examined. For ISCAT scores, the interaction between instructional condition and pretest score was not statistically significant, $F(1, 56) = 0.01$, $p = .921$, supporting the homogeneity-of-regression-slopes assumption. Levene's test was also nonsignificant, $F(1, 58) = 0.64$, $p = .425$, indicating that the homogeneity-of-variance assumption was satisfied. Examination of the model residuals did not indicate a substantial departure from normality, $p = .254$, and no influential cases were identified. The same assumptions were supported for the EVC Index. The interaction between instructional condition and baseline EVC Index was nonsignificant, $F(1, 56) = 0.03$, $p = .866$, and Levene's test was also nonsignificant, $F(1, 58) = 0.18$, $p = .677$. Residuals were approximately normally distributed ($p = .230$), and inspection of influential observations did not indicate cases that materially affected the model estimates. These results supported proceeding with the planned baseline-adjusted analyses. Assumption checks for the primary ANCOVA models can be seen in Table 3.

Table 3 Assumption checks for the primary ANCOVA models

Assumption	ISCAT Model	EVC Index Model	Interpretation
Group $\times$ pretest interaction	$F(1, 56) = 0.01, p = .921$	$F(1, 56) = 0.03, p = .866$	Homogeneity of regression slopes supported
Levene's test	$F(1, 58) = 0.64, p = .425$	$F(1, 58) = 0.18, p = .677$	Homogeneity of variance supported
Residual normality	$p = .254$	$p = .230$	No substantial violation detected

4.1 RQ1: Learning Outcomes

RQ1 examined whether students receiving structured GenAI-supported co-tutoring differed in their learning outcomes from students receiving conventional instruction. Descriptive results showed improvement in both groups, although the increase was greater among students in the experimental condition. As presented in Table 4, the experimental group increased from an ISCAT-A pretest mean of 51.16 (SD = 6.09) to an ISCAT-B posttest mean of 74.80 (SD = 6.74), representing an unadjusted gain of 23.64 points. The control group increased from 51.10 (SD = 8.54) to 66.88 (SD = 7.66), corresponding to a gain of 15.78 points. The difference between the two unadjusted gains was 7.86 points in favor of the experimental group. The similar pretest means also indicated that, at the descriptive level, the groups entered the intervention with comparable average levels of conceptual knowledge.

Table 4 Descriptive and baseline-adjusted results for learning outcomes

Indicator	Experimental Group	Control Group	Between-Group Comparison
ISCAT-A pretest, M (SD)	51.16 (6.09)	51.10 (8.54)	0.06
ISCAT-B posttest, M (SD)	74.80 (6.74)	66.88 (7.66)	7.92
Unadjusted gain	23.64	15.78	7.86
Adjusted posttest M (SE)	74.79 (1.06)	66.90 (1.06)	7.89
95% CI for adjusted difference			[4.89, 10.89]
ANCOVA instructional-condition effect			$F(1, 57) = 27.65$
Statistical significance			$p < .001$
Effect size			partial $\eta^2 = .327$

To determine whether the observed posttest difference persisted after accounting for baseline achievement, an ANCOVA was conducted with ISCAT-B posttest score as the dependent variable, instructional condition as the fixed factor, and ISCAT-A pretest score as the covariate. The pretest score was significantly associated with posttest performance, $F(1, 57) = 19.84, p < .001$, partial $\eta^2 = .258$, showing that initial conceptual knowledge contributed to subsequent achievement. After this baseline variation was statistically controlled, instructional condition remained significant, $F(1, 57) = 27.65, p < .001$, partial $\eta^2 = .327$. The adjusted posttest mean was 74.79 (SE = 1.06) for the experimental group and 66.90 (SE = 1.06) for the control group. This produced an adjusted mean difference of 7.89 points, with a 95% CI of [4.89, 10.89]. Because the confidence interval remained entirely above zero, the adjusted difference consistently favored the structured GenAI-supported co-tutoring condition.

The descriptive and adjusted analyses showed the same pattern. Both groups improved during the instructional period, but the experimental group demonstrated a substantially greater increase and maintained a higher posttest score after baseline achievement was controlled. The experimental group gained 23.64 points compared with 15.78 points in the control group, and the baseline-adjusted difference remained 7.89 points. The partial η^2 of .327 further indicated a substantial association between instructional condition and posttest learning outcomes. Importantly, ISCAT-B was

<http://sistemasi.ftik.unisi.ac.id>

administered without access to GenAI, so the posttest scores reflected knowledge that students could demonstrate independently rather than performance produced while receiving AI assistance. Thus, RQ1 showed that students receiving structured GenAI-supported co-tutoring achieved significantly higher independent learning outcomes than students receiving conventional instruction after controlling for pre-intervention achievement.

4.2 RQ2: Learning Motivation

RQ2 examined whether students receiving structured GenAI-supported co-tutoring differed in their learning motivation from students receiving conventional instruction. Descriptive results indicated increases in the EVC Index in both instructional conditions, although the improvement was greater in the experimental group. As shown in Table 5, the experimental group increased from a pretest mean of 0.699 (SD = 0.078) to a posttest mean of 0.769 (SD = 0.062), representing an unadjusted increase of 0.070. The control group showed a smaller increase, from 0.691 (SD = 0.074) at pretest to 0.713 (SD = 0.071) at posttest, corresponding to an increase of 0.022. The difference between the two unadjusted changes was 0.048 in favor of the experimental group. The relatively similar pretest means suggested that the groups entered the instructional period with comparable average levels of learning motivation, whereas the posttest values showed a clearer difference between the two conditions.

Table 5 Descriptive and baseline-adjusted results for learning motivation

Indicator	Experimental Group	Control Group	Between-Group Comparison
EVC Index pretest, M (SD)	0.699 (0.078)	0.691 (0.074)	0.008
EVC Index posttest, M (SD)	0.769 (0.062)	0.713 (0.071)	0.056
Unadjusted change	0.070	0.022	0.048
Adjusted posttest M (SE)	0.767 (0.010)	0.715 (0.010)	0.053
95% CI for adjusted difference			[0.024, 0.082]
ANCOVA instructional-condition effect			F(1, 57) = 13.53
Statistical significance			p < .001
Effect size			partial η^2 = .192

To determine whether the post-intervention motivational difference remained after accounting for students' initial motivation, an ANCOVA was conducted with the posttest EVC Index as the dependent variable, instructional condition as the fixed factor, and the pretest EVC Index as the covariate. Baseline motivation was significantly associated with posttest motivation, $F(1, 57) = 33.12$, $p < .001$, partial $\eta^2 = .367$, indicating that students' initial motivational levels contributed substantially to variation in their post-intervention scores. After controlling for this baseline variation, instructional condition remained statistically significant, $F(1, 57) = 13.53$, $p < .001$, partial $\eta^2 = .192$. The adjusted posttest EVC Index was 0.767 (SE = 0.010) for the experimental group and 0.715 (SE = 0.010) for the control group, yielding an adjusted mean difference of 0.053. The 95% CI for this difference ranged from 0.024 to 0.082 and remained entirely above zero, indicating that the higher adjusted motivational score in the experimental group was statistically reliable after baseline motivation was considered.

The descriptive and baseline-adjusted results consistently favored the structured GenAI-supported co-tutoring condition. Although learning motivation increased in both groups, the increase was approximately three times larger in the experimental group at the descriptive level (0.070 versus 0.022). The adjusted difference of 0.053 also remained significant after controlling for pre-intervention motivation, while the partial η^2 of .192 indicated a notable instructional-condition effect on posttest EVC Index scores. Because the EVC Index integrates expectancy, value, and cost into the overall motivational indicator

<http://sistemasi.ftik.unisi.ac.id>

specified for the primary analysis, the result reflects a broader difference in students' motivation toward learning the course rather than a change in only one motivational component. Thus, RQ2 showed that students receiving structured GenAI-supported co-tutoring demonstrated significantly higher post-intervention learning motivation than students receiving conventional instruction after controlling for their initial motivational levels.

5 Discussion

This study examined whether structured GenAI co-tutoring was associated with differences in learning outcomes and learning motivation in an undergraduate Information Systems Concepts course. After baseline adjustment, the experimental group obtained higher scores on both outcomes. The adjusted ISCAT-B difference was 7.89 points (partial $\eta^2 = .327$), while the adjusted EVC Index difference was 0.053 (partial $\eta^2 = .192$). The achievement result is particularly relevant because ISCAT-B was completed without GenAI access. Thus, the difference reflects knowledge students could demonstrate independently after instruction rather than performance produced while AI assistance was available.

The higher learning outcome is consistent with evidence that GenAI can support learning when integrated into purposeful instructional activities. Previous studies have reported positive effects on academic performance, knowledge construction, self-regulated learning, and higher-order thinking, particularly when students receive guided support rather than unrestricted assistance [2], [15]. GenAI use focused on knowledge construction and augmentation has also been associated with higher-level learning than procedural use [20]. In the present study, students first developed their own responses and then used guided prompts, feedback, verification, and revision to refine their understanding. This structure may have supported deeper engagement with course concepts rather than simple retrieval of AI-generated answers. The improvement in both groups also suggests a role for regular instruction and practice, while the larger gains in the experimental group may reflect the combined contribution of GenAI assistance, guided feedback, and revision activities.

This pattern differs from evidence showing that stronger AI-assisted performance does not always transfer to independent learning. Unrestricted GenAI support may improve performance while the tool is available but weaken subsequent performance after its removal [4]. Similarly, AI-supported programming can improve exercise scores without producing comparable gains in conceptual knowledge or code comprehension [18]. In the present study, the experimental group still outperformed the control group when GenAI was unavailable during the posttest. This may reflect the instructional sequence, which required students to attempt tasks independently, verify AI responses, and revise their work rather than simply accept generated output. Such a design is consistent with research emphasizing reflective engagement and learner autonomy in productive GenAI use [14].

Learning motivation was also higher in the experimental group. This result is consistent with evidence that GenAI can improve motivation and engagement, although the effects vary across learning strategies and contexts [22]. Intentionally designed AI-supported environments have also been associated with stronger autonomous motivation [8]. In the present course, students received explanations, examples, and feedback while remaining responsible for their initial responses and evaluations. This type of support may have made difficult concepts easier to manage while preserving students' active role in learning. Personalized tutoring and responsive assistance have likewise been identified as factors that can strengthen engagement [23].

The motivational result also shows the importance of instructional structure. GenAI can increase confidence but may also increase technological dependence [24]. Scaffolded support appears more effective for intrinsic motivation than unrestricted AI access, even when both conditions improve task performance [18]. In this study, GenAI was introduced only after

<http://sistemasi.ftik.unisi.ac.id>

students had attempted the task independently and was followed by evaluation and revision. This structure supports planning, monitoring, feedback, and reflection rather than replacing these processes [16]. It may also help explain why the experimental group showed advantages in both achievement and motivation. The motivational difference may also reflect students' prior experience with GenAI, the interactive nature of the activities, and opportunities to receive immediate feedback. These factors suggest that motivation may be associated with the overall learning experience rather than GenAI access alone.

These findings have practical implications for instructional design. GenAI can be introduced after an initial independent attempt so that it supports rather than replaces students' reasoning. AI-generated responses can then be evaluated through verification, comparison, and revision activities. Independent assessment is also important when the goal is conceptual understanding. In this study, administering ISCAT-B without GenAI helped distinguish independent learning from AI-assisted performance. These practices are consistent with research emphasizing guided interaction rather than unrestricted answer generation [15], [16].

The study contributes to the GenAI literature by extending structured co-tutoring beyond the programming-oriented settings that dominate computing education research. Rather than focusing on coding, debugging, or task completion, the intervention addressed conceptual learning about systems, information, organizational processes, data, and technology. It also examined independent achievement and learning motivation within the same course-based design. The results suggest that the educational value of GenAI may depend less on access to the technology than on how interaction with it is organized. For concept-oriented courses, structured co-tutoring that preserves initial reasoning, requires evaluation of AI output, and separates assisted learning from independent assessment provides a useful direction for instructional design and further research.

6 Conclusion

This study examined structured GenAI co-tutoring in an undergraduate Information Systems Concepts course by comparing students' learning outcomes and learning motivation with those receiving conventional instruction. After baseline adjustment, the experimental group achieved a 7.89-point higher ISCAT-B score than the control group (partial $\eta^2 = .327$) and a 0.053 higher EVC Index (partial $\eta^2 = .192$). The higher ISCAT-B performance is particularly relevant because the assessment was completed without GenAI, indicating that the difference reflected knowledge students could demonstrate independently after instruction. These results suggest that GenAI may be more educationally useful when its use is organized around initial independent attempts, guided assistance, evaluation of AI output, and revision rather than unrestricted answer generation. For concept-oriented Information Systems courses, this approach offers a practical way to use GenAI while preserving students' responsibility for reasoning and maintaining independent assessment of learning. The findings should be interpreted within the scope of the study. Participants came from two intact classes in one institution, individual random assignment was not possible, and outcomes were measured immediately after the intervention. With only one class per condition, class characteristics could not be separated from the intervention, and baseline adjustment could not eliminate unmeasured differences. The single-institution sample also limits the generalizability of the findings to other educational settings. The intervention also combined several elements, including guided prompting, feedback, verification, reflection, and revision, so the contribution of each component could not be determined separately. In addition, the primary motivational analysis relied on the overall EVC Index and did not establish which motivational dimension contributed most strongly to the observed difference. Future research should validate these findings using larger samples from multiple institutions and several classes per condition, with class-level random assignment where feasible, incorporate delayed posttests to examine the retention of conceptual learning and motivation, and compare individual components of structured GenAI co-tutoring. Further studies could also examine

expectancy, value, and cost separately and test whether similar patterns emerge in other concept-oriented computing courses.

Acknowledgement

The authors thank the Directorate General of Higher Education, Ministry of Education and Culture of the Republic of Indonesia, for providing financial support for this research under Contract No. 94/DST/C/AL.04.02/2026, dated April 9, 2026.

References

- [1] A. Al-Abri, "Exploring ChatGPT as a Virtual Tutor: A Multi-Dimensional Analysis of Large Language Models in Academic Support," *Educ. Inf. Technol. (Dordr.)*, Vol. 30, no. 12, pp. 17447–17482, Aug. 2025, DOI: 10.1007/s10639-025-13484-x.
- [2] R. Deng, M. Jiang, X. Yu, Y. Lu, and S. Liu, "Does ChatGPT Enhance Student Learning? A Systematic Review and Meta-Analysis of Experimental Studies," *Comput. Educ.*, Vol. 227, p. 105224, Apr. 2025, DOI: 10.1016/j.compedu.2024.105224.
- [3] W. Andewi, W. Waziana, D. Wibisono, K. A. Putra, T. Hastomo, and I. B. Oktarin, "From Prompting to Proficiency: A Mixed-Methods Analysis of Prompting with ChatGPT Versus Lecturer Interaction in an EFL Classroom," *Journal of Studies in the English Language*, Vol. 20, No. 2, pp. 210–238, 2025, Accessed: Sep. 01, 2025. [Online]. Available: <https://so04.tci-thaijo.org/index.php/jsel/article/view/282318>
- [4] H. Bastani, O. Bastani, A. Sungu, H. Ge, Ö. Kabakcı, and R. Mariman, "Generative AI without Guardrails can Harm Learning: Evidence from High School maThematics," *Proceedings of the National Academy of Sciences*, Vol. 122, No. 26, Jul. 2025, DOI: 10.1073/pnas.2422633122.
- [5] W. Huang, J. Jiang, R. B. King, and L. K. Fryer, "Chatbots and Student Motivation: A Scoping Review," *International Journal of Educational Technology in Higher Education*, Vol. 22, No. 1, p. 26, Apr. 2025, DOI: 10.1186/s41239-025-00524-2.
- [6] F. J. Agbo, C. Olivia, G. Oguibe, I. T. Sanusi, and G. Sani, "Computing Education using Generative Artificial Intelligence Tools: A Systematic Literature Review," *Computers and Education Open*, Vol. 9, p. 100266, Dec. 2025, DOI: 10.1016/j.caeo.2025.100266.
- [7] A. C. M. Yang, J.-Y. Lin, C.-Y. Lin, and H. Ogata, "Enhancing Python Learning with PyTutor: Efficacy of a ChatGPT-based Intelligent Tutoring System in Programming Education," *Computers and Education: Artificial Intelligence*, Vol. 7, p. 100309, Dec. 2024, DOI: 10.1016/j.caeai.2024.100309.
- [8] S. Choi and H. Kim, "The Impact of a Large Language Model-based Programming Learning Environment on Students' Motivation and Programming Ability," *Educ. Inf. Technol. (Dordr.)*, Vol. 30, No. 6, pp. 8109–8138, Apr. 2025, DOI: 10.1007/s10639-024-13107-x.
- [9] D. Sun, A. Boudouaia, C. Zhu, and Y. Li, "Would ChatGPT-Facilitated Programming Mode Impact College Students' Programming Behaviors, Performances, and Perceptions? An Empirical Study," *International Journal of Educational Technology in Higher Education*, Vol. 21, No. 1, p. 14, Feb. 2024, DOI: 10.1186/s41239-024-00446-5.
- [10] M. Gerlich, "From Offloading to Engagement: An Experimental Study on Structured Prompting and Critical Reasoning with Generative AI," *Data (Basel)*, Vol. 10, No. 11, p. 172, Oct. 2025, DOI: 10.3390/data10110172.
- [11] M. D. H. Rahiem, "Generative AI in Higher Education in Indonesia: Patterns of use and Learning Impact," *Social Sciences & Humanities Open*, Vol. 13, p. 102672, Jun. 2026, DOI: 10.1016/j.ssaho.2026.102672.
- [12] H. Crompton and D. Burke, "The Educational Affordances and Challenges of ChatGPT: State of the Field," *TechTrends*, Vol. 68, No. 2, pp. 380–392, Mar. 2024, DOI: 10.1007/s11528-024-00939-0.
- [13] N. Wang, X. Wang, and Y.-S. Su, "Critical Analysis of the Technological Affordances, Challenges and Future Directions of Generative AI in Education: A Systematic Review," *Asia Pacific Journal of Education*, Vol. 44, No. 1, pp. 139–155, Jan. 2024, DOI: 10.1080/02188791.2024.2305156.

- [14] Y. Qian, "Pedagogical Applications of Generative AI in Higher Education: A Systematic Review of the Field," *TechTrends*, Vol. 69, No. 5, pp. 1105–1120, Sep. 2025, DOI: 10.1007/s11528-025-01100-1.
- [15] H.-Y. Lee, P.-H. Chen, W.-S. Wang, Y.-M. Huang, and T.-T. Wu, "Empowering ChatGPT with Guidance Mechanism in Blended Learning: Effect of Self-Regulated Learning, Higher-Order Thinking Skills, and Knowledge Construction," *International Journal of Educational Technology in Higher Education*, Vol. 21, No. 1, p. 16, Mar. 2024, DOI: 10.1186/s41239-024-00447-4.
- [16] T. K. F. Chiu, "A Classification Tool to Foster Self-Regulated Learning with Generative Artificial Intelligence by Applying Self-Determination Theory: A Case of ChatGPT," *Educational Technology Research and Development*, Vol. 72, No. 4, pp. 2401–2416, Aug. 2024, DOI: 10.1007/s11423-024-10366-w.
- [17] M. Lepp and J. Kaimre, "Does Generative AI Help in Learning Programming: Students' Perceptions, Reported use and Relation to Performance," *Computers in Human Behavior Reports*, Vol. 18, p. 100642, May 2025, DOI: 10.1016/j.chbr.2025.100642.
- [18] P. Bassner, B. Lenk-Ostendorf, R. Beinstingel, T. Wasner, and S. Krusche, "Less Stress, Better Scores, Same Learning: The Dissociation of Performance and Learning in AI-Supported Programming Education," *Computers and Education: Artificial Intelligence*, Vol. 10, p. 100537, Jun. 2026, DOI: 10.1016/j.caeai.2025.100537.
- [19] D. T. K. Ng, C. W. Tan, and J. K. L. Leung, "Empowering Student Self-Regulated Learning and Science Education through ChatGP: A Pioneering Pilot Study," *British Journal of Educational Technology*, Vol. 55, No. 4, pp. 1328–1353, Jul. 2024, DOI: 10.1111/bjet.13454.
- [20] J. L. Pallant, J. Blijlevens, A. Campbell, and R. Jopp, "Mastering Knowledge: The Impact of Generative AI on Student Learning Outcomes," *Studies in Higher Education*, Vol. 51, No. 4, pp. 714–735, Apr. 2026, DOI: 10.1080/03075079.2025.2487570.
- [21] Febriyanti, E. Nurchurifiani, T. Hastomo, A. Wicaksono, and S. S. Nasution, "AI Literacy of Pre-Service English Teachers: Proficiency, Strategies, and Barriers," *JEELS (Journal of English Education and Linguistics Studies)*, Vol. 13, No. 1, pp. 155–186, Apr. 2026, DOI: 10.30762/jeels.v13i1.6293.
- [22] Q. Xia, W. Li, Y. Yang, X. Weng, and T. K. F. Chiu, "A Systematic Review and Meta-Analysis of the Effectiveness of Generative Artificial Intelligence (GenAI) on Students' Motivation and Engagement," *Computers and Education: Artificial Intelligence*, Vol. 9, p. 100455, Dec. 2025, DOI: 10.1016/j.caeai.2025.100455.
- [23] Y. M. E. Heung and T. K. F. Chiu, "How ChatGPT Impacts Student Engagement from a Systematic Review and Meta-Analysis Study," *Computers and Education: Artificial Intelligence*, Vol. 8, p. 100361, Jun. 2025, DOI: 10.1016/j.caeai.2025.100361.
- [24] L. Zhang and J. Xu, "The Paradox of Self-Efficacy and Technological Dependence: Unraveling Generative AI's Impact on University Students' Task Completion," *Internet High. Educ.*, Vol. 65, p. 100978, Apr. 2025, DOI: 10.1016/j.iheduc.2024.100978.
- [25] P. Schoeffel, V. F. C. Ramos, C. Cechinel, and R. S. Wazlawick, "The Expectancy-Value-Cost Light Scale to Measure Motivation of Students in Computing Courses," *Informatics in Education*, Jun. 2021, DOI: 10.15388/infedu.2022.04.
- [26] W. R. Shadish, T. D. Cook, and D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Houghton Mifflin, 2002.
- [27] J. W. Creswell and J. D. Creswell, *Research design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage publications, 2018.
- [28] A. J. Vickers and D. G. Altman, "Analysing Controlled Trials with Baseline and Follow up Measurements," *BMJ*, Vol. 323, No. 7321, pp. 1123–1124, Nov. 2001, DOI: 10.1136/bmj.323.7321.1123.
- [29] A. P. A. & N. C. on M. in E. American Educational Research Association, *Standards for Educational and Psychological Testing*. American Educational Research Association, 2014.
- [30] S. Sireci and M. Faulkner-Bond, "Validity Evidence based on Test Content," *Psicothema*, Vol. 1, No. 26, pp. 100–107, Feb. 2014, DOI: 10.7334/psicothema2013.256.

- [31] T. K. Koo and M. Y. Li, “A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research,” *J. Chiropr. Med.*, Vol. 15, No. 2, pp. 155–163, Jun. 2016, DOI: 10.1016/j.jcm.2016.02.012.
- [32] D. Lakens, “Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for t-tests and ANOVAs,” *Front. Psychol.*, Vol. 4, 2013, DOI: 10.3389/fpsyg.2013.00863.