JoTaLP: Journal of Teaching and Learning Physics 10, 2 (2025): 103-110
Website: http://journal.uinsgd.ac.id/index.php/jtlp/index
ISSN 2580-3107 **(online) ISSN** 2528-5505 **(print)**

[Research Article]

PATTERNS OF MISCONCEPTION CHANGE IN STELLAR EVOLUTION UNDER DIFFERENTIATED INSTRUCTION

***Hilda Permata*[1]*, Winny Liliawati*[1] *and Judhistira Aria Utama*[1]**

[1]*Magister of Physics Education, Faculty of Mathematics and Science Education, Indonesia University of Education, Bandung, Indonesia*
E-mail: winny@upi.edu

DOI: http://dx.doi.org/10.15575/jotalp.v10i2.49489
Received: 13 August 2025 ; Accepted: 25 August 2025 ; Published: 31 August 2025

ABSTRACT

This study maps changes in misconceptions about stellar evolution by emphasising three student-level categories: positive, stable, and negative change. Misconceptions in astronomy often persist in definitional concepts and causal relations. A quasi-experimental pretest–posttest design with experimental and control classes was employed. Data were collected using a fifteen-item, four-tier diagnostic; an item was considered understood when both the answer and the written reason were correct. We analysed each student's total understanding score to classify individual change and then computed class-level proportions. Results show the experimental class was dominated by positive change 62.9%, followed by stable 25.7% and negative 11.4%; the control class showed positive 48.6%, stable 24.3%, and negative 27.0%. These findings indicate that differentiated instruction is more effective in reducing misconceptions at the individual level.

Keywords: misconceptions, stellar evolution, four-tier diagnostic test, differentiated instruction, conceptual change

How to cite: Permata, H., Liliawati, Winny., and Utama, J. A. (2025) Patterns of Misconception Change in Stellar Evolution under Differentiated Instruction, *Journal of Teaching and Learning Physics 10* (2), 103-110. DOI: http://dx.doi.org/10.15575/jotalp.v10i2.49489

1. INTRODUCTION

Understanding in astronomy especially stellar evolution often suffers from persistent misconceptions, both on concise definitional items and on those requiring causal reasoning across quantities such as mass, fusion rate, luminosity, and stellar lifetime (Bitzenbauer et al., 2023; Salimpour et al., 2024; Ubben et al., 2022). Such misconceptions may be accompanied by high confidence in incorrect ideas, making them resistant to change without targeted instructional support (Guerra-Reyes et al., 2024; Herder & Rau, 2022; Keppens et al., 2025; Kersting et al., 2024; Ubben & Bitzenbauer, 2022).

A sharp lens on these difficulties is the four-tier diagnostic test, which evaluates both the selected answer and the written reason; an item is scored “understood” only when both are correct. This approach enables fine-grained item- and topic-level mapping, while also allowing analysis of individual shifts (not-understood → understood and the reverse) rather than relying solely on class averages (Astuti et al., 2023; Çelikkanlı & Kızılcık, 2022; Istiyono et al., 2023; Rusilowati et al., 2024).

This study focuses on patterning misconception change before and after differentiated instruction, emphasising three individual-level categories: positive change (increase in total understood score), stable (no change), and negative change (decrease). The research questions are: (1) What are the item- and topic-level patterns of change in the experimental and control classes? and (2) What are the class-level distributions of positive, stable, and negative change? Accordingly, the objectives are to map per-item/per-topic changes and to describe the category distribution in each class as a basis for actionable instructional recommendations.

The study is grounded in the framework of conceptual change, which highlights claim-reason coherence as an indicator of understanding (Paçacı & Çetin-Dindar, 2024), and in differentiated instruction, which provides adaptive support aligned with learners’ readiness and profiles (Langelaan et al., 2024). The working hypothesis is that the experimental class will show a higher proportion of positive change and a lower proportion of negative change than the control class.

2. METHOD

This study was organised rationally, empirically, and systematically as a quasi-experimental pretest-posttest design with a control class. It was conducted in February 2025 in the Earth and Space Science class, Bandung, Indonesia. Two comparable classes participated: one served as the experimental class (differentiated instruction) and one as the control class (conventional instruction). Sampling was non-random based on class availability. The paired sample analyzed comprised N=33 (experimental) and N=37 (control), namely students who completed both the pretest and posttest.

The primary data were responses to a 15-item four-tier diagnostic test covering three topics: stellar birth (S1-S6), stellar lifetime (S7-S11), and stellar death (S12-S15). An item was coded only as understood (1) only when the answer (Tier-1) was correct and the written reason (Tier-3) matched the rubric; otherwise, it was not understood (0). Materials included the test booklet, answer sheets, the rubric for scoring, and data-processing tools (spreadsheets and scripts) to ensure consistency.

The procedure began with a pretest to map initial understanding. The experimental class then received differentiated instruction (multi-representation, flexible grouping, guided-question scaffolding, and formative checks), while the control class received conventional instruction on the same content. After the intervention, all participants took a posttest using the identical instrument. Answer sheets were anonymized, written reasons were scored with the rubric, and data were entered into the spreadsheet.

Let Δ denote each student's change score. For each student, we computed the total "understood" score on the pretest (U_{pre}) and on the posttest (U_{post}) using the same scoring rubric stated earlier $(range\ 0-15)$. The change index was defined as $\Delta = U_{post} - U_{pre}$, Students were classified into three mutually exclusive categories: positive change ($\Delta > 0$), stable ($\Delta = 0$), and negative change ($\Delta < 0$). For each class (experimental and control) we report the frequency (n) and percentage (%) in each category; the same classification is then summarized by topic according to the test blueprint (Items 1–6: stellar birth; 7–11: stellar lifetime; 12–15: stellar death) and complemented with item-level narrative where relevant. Paired summaries include only complete pre–post cases; students missing either test are excluded from paired counts but retained in item-level descriptives with the corresponding N explicitly stated. All analyses use an identical coding scheme across classes to ensure comparability, and all procedures comply with educational research ethics (informed participation, confidentiality, academic use of data).

3. RESULT AND DISCUSSION

Paired pretest-posttest analysis shows the experimental class dominated by positive change (62.9%), followed by stable (25.7%) and negative (11.4%). The control class presents positive 48.6%, stable 24.3%, and negative 27.0%. This pattern indicates that differentiated instruction is more effective at promoting individual improvement while limiting regress (see Table 1; Figure 1). Conceptually, the pattern aligns with a conceptual change view—correction requires reasons that coherently support the claim—and with differentiated instruction principles that provide adaptive support. [19] This overall shift is in line with recent work emphasizing immersive/visual designs in astronomy education (Kersting et al., 2024).

Table 1. Distribution of change categories (paired) by class

Class	N Paired	Positif (%)	Tetap (%)	Negatif (%)
Experiment	33	69,7	15,2	15,2
Control	37	48,6	24,3	27,0

Table 1 reports counts and percentages for each category in both classes. The experimental class leads in positive change, whereas the control class shows a higher share of negative change. A stable portion near one quarter in both classes suggests items that are relatively resistant to the intervention.

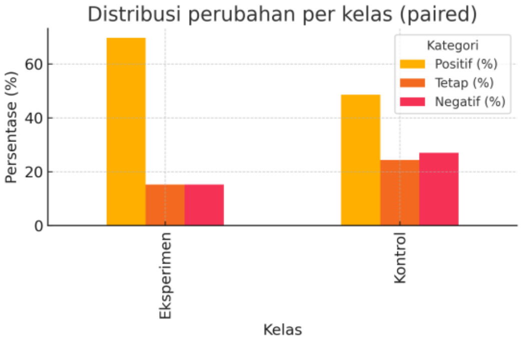

FIgure 1. Comparison of change-category percentages by class

Figure 1 reinforces Table 1 visually: the experimental bar for positive greatly exceeds negative, while in the control class the gap between positive and negative is smaller. This highlights a qualitative difference in individual shifts across approaches.

The experimental class shows a dominant positive proportion (62.9%), with stable near one quarter (25.7%) and the smallest negative share (11.4%). Conversely, the control class features a lower positive rate (48.6%) and a higher negative rate (27.0%). The gap implies differentiated instruction not only raises the likelihood of improvement but also reduces the probability of post-instruction regress.

3.1 Item and Topic-Wise Patterns

This section elaborates on item-wise and topic-wise performance. A summary of pre-post per cent understood per item appears in Table 2, followed by per-class visualisations in Figures 2-3. Topic means are reported in Table 3 and visualized in Figures 4-5. (Herder & Rau, 2022; Keppens et al., 2025).

Table 2. Persentation of understood per item (Pre–Post) and Delta (percentage points, pp)

Class	Item	Delta (pp)	Post (%)	Pre (%)
Experiment	S1	14,3	20,0	5,7
Experiment	S10	28,6	88,6	60,0
Experiment	S11	0,0	77,1	77,1
Experiment	S12	37,1	77,1	40,0
Experiment	S13	45,7	68,6	22,9
Experiment	S14	17,1	45,7	28,6
Experiment	S15	11,4	82,9	71,4
Experiment	S2	0,0	0,0	0,0
Experiment	S3	5,7	68,6	62,9
Experiment	S4	14,3	57,1	42,9
Experiment	S5	5,7	80,0	74,3
Experiment	S6	-8,6	37,1	45,7
Experiment	S7	11,4	42,9	31,4
Experiment	S8	0,0	0,0	0,0
Experiment	S9	2,9	51,4	48,6
Control	S1	2,1	20,5	18,4
Control	S10	-6,8	59,0	65,8
Control	S11	-19,8	53,8	73,7
Control	S12	32,4	69,2	36,8
Control	S13	37,9	61,5	23,7
Control	S14	17,5	35,9	18,4
Control	S15	13,9	71,8	57,9
Control	S2	0,0	0,0	0,0
Control	S3	14,2	61,5	47,4
Control	S4	19,4	61,5	42,1
Control	S5	-4,4	66,7	71,1
Control	S6	-13,9	28,2	42,1
Control	S7	11,7	53,8	42,1
Control	S8	0,0	0,0	0,0
Control	S9	6,2	64,1	57,9

The largest gains generally occur on end-of-life items (S12-S13) and the causal indicator S10, while a consistent decline appears on S6. The definitional item S2 remains low in both classes, signaling a need for explicit terminology support.

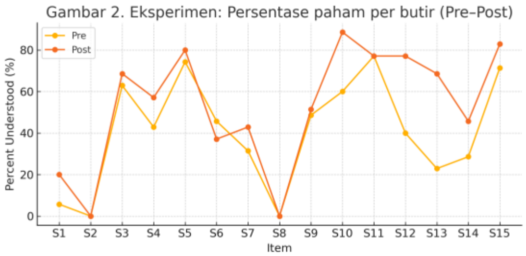

Figure 2. Experimental: presentation of understood per item (Pre–Post)

The posttest curve generally lies above the pretest across most items, with clear jumps at S10 and S12-S15, but a drop at S6. This indicates that causal and end-of-life indicators are the most responsive to the intervention.

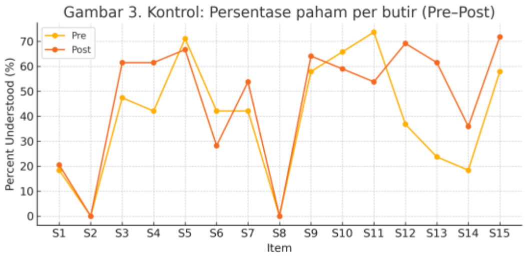

Figure 3. Control: percent understood per item (Pre–Post)

Moderate increases appear at S3-S4 and S12-S15, but declines occur at S10-S11 and S6. The S2 line remains flat at 0%, reinforcing that terminology obstacles seldom resolve without explicit supports (Ubben & Bitzenbauer, 2022).

3.1.1 Stellar birth (S1-S6)

The experimental class improves on S1, S3-S5 and declines on S6; the control class improves on S3-S4 and declines on S6. S2 stays at 0% understood in both classes. The pattern is consistent with fragile definitional understanding and representational load on S6.

3.1.2 Stellar lifetime (S7-S11)

Causal-mechanistic indicators S10-S11 stand out. The experimental class shows a strong S10 jump and a stable-to-slight increase on S11, while the control class declines on both items. This suggests better concept-bridging from mass → fusion rate → luminosity → lifetime under differentiation.

3.1.3 Stellar death (S12-S15)

Both classes improve on S12-S15, but jumps are larger in the experimental class (e.g., S12 +36.4 pp; S13 +48.5 pp). Clarifying mass-based evolutionary pathways effectively counters the common generalization that "all stars end as black holes."

Table 3. Mean percent understood by topic (Pre–Post) and Delta (pp)

Class	Topic	Pre-mean (%)	Post-mean (%)	Delta mean (pp)
Experiment	Stellar birth (S1-S6)	38,6	43,8	5,2
Experiment	Stellar lifetime (S7-S11)	43,4	52,0	8,6
Experiment	Stellar death (S12-S15)	40,7	68,6	27,9
Control	Stellar birth (S1-S6)	36,8	39,7	2,9
Control	Stellar lifetime (S7-S11)	47,9	46,2	-1,7
Control	Stellar death (S12-S15)	34,2	59,6	25,4

Topic means sharpen the picture: stellar birth rises only slightly (hampered by S2 and S6); stellar lifetime increases, especially in the experimental class; stellar death shows the largest gains in both classes, with higher levels in the experimental class.

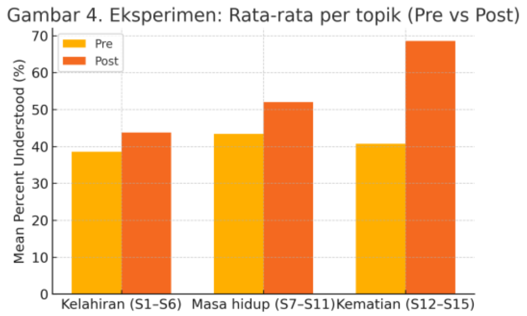

Figure 4. Experimental: topic means (Pre vs Post)

The largest increase is in stellar death, followed by stellar lifetime. Stellar birth shows the smallest rise, consistent with very low S2 performance and the S6 decline.

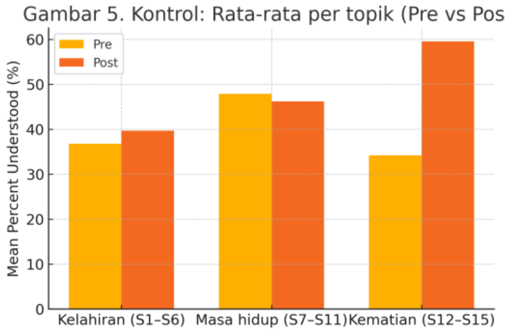

Figure 5. Control: topic means (Pre vs Post)

A similar ordering appears in the control class, but the magnitudes are smaller than in the experimental class. The biggest between-class gap is in stellar death.

Paired pretest-posttest analysis shows the experimental class dominated by positive change (62.9%), followed by stable (25.7%) and negative (11.4%), whereas the control class shows positive 48.6%, stable 24.3%, and negative 27.0%. The strongest gains appear on causal-mechanistic indicators (S10-S11) and on the end-of-life topic (S12-S15), while the concise definitional item S2 remains persistently low and S6 declines in both classes. At the topic level, mean understanding increases are largest on stellar death, followed by stellar lifetime, while stellar birth shows the smallest rise. At the individual level, claim-reason coherence improves in the experimental class, indicating that differentiated instruction not only lifts scores but also improves the quality of conceptual justification. This overall shift is in line with recent work emphasizing immersive/visual designs in astronomy education (Kersting et al., 2024).

Table 4. Consistency triangle (problem–objectives–conclusions)

Component	Formulation in this study
Research problem	Persistent misconceptions in stellar evolution; problematic causal and definitional items.
Research objectives	Map change patterns per item and per topic; describe proportions of change (positive, stable, negative).
Conclusions	Experimental: positive 62.9%, stable 25.7%, negative 11.4%; Control: positive 48.6%, stable 24.3%, negative 27.0%; S10 and S12–S13 improve; S2 persists; S6 declines.

Table 5. Key findings by topic and instructional implications

Topic	Δ mean Experimental (pp)	Δ mean Control (pp)	Instructional implications
Stellar birth (S1–S6)	5.2	2.9	Small rise; strengthen terminology & representation (S2, S6).
Stellar lifetime (S7–S11)	8.6	-1.7	Moderate rise; bridge mass→fusion rate→luminosity→lifetime.
Stellar death (S12–S15)	27.9	25.4	Largest rise; clarify mass-based evolutionary pathways.

Table 5 presents the topic-wise distribution of change categories aligned with the test blueprint (Items 1–6: stellar birth; 7–11: stellar lifetime; 12–15: stellar death). This view clarifies which conceptual clusters contribute most to positive change and where stable/negative responses persist, providing context for the item-level narratives (S1–S15).

Table 6. Priority items for follow-up and recommendations

Item	Δ Experimental (pp)	Δ Control (pp)	Status	Brief recommendation
S2	0.0	0.0	Persistent	Strengthen terminology; mini-glossary & contrasting examples
S6	-8.6	-13.9	Decline	Audit wording & prerequisite sequence; use alternative representations
S10	28.6	-6.8	Increase	Sustain multi-representation & scaffold causal reasoning
S12	37.1	32.4	Increase	Use mass-based evolutionary pathway maps & case examples
S13	45.7	37.9	Increase	Use mass-based evolutionary pathway maps & case examples

Table 6 lists priority items for instructional follow-up. Items were flagged based on the persistence of non-understanding, recurring claim–reason inconsistency, or signs of regress (Δ<0). The table is intended as an actionable bridge to classroom planning, indicating where targeted reinforcement and representational supports are most needed.

Instruction that adapts to learners' readiness, interests, and profiles—a hallmark of differentiated instruction—is better suited to causal concepts that require linking quantities (mass, fusion rate, luminosity, lifetime). In contrast, definitional concepts require explicit terminology supports (mini-glossaries, contrasting examples, and practice in testing concise statements) to trigger conceptual change. From an assessment standpoint, the four-tier diagnostic test is effective because it enables reporting change categories (positive-stable-negative) at the individual level and item-wise mapping; this practice is worth adopting in other astronomy/physics topics. Generalization is strongest for comparable Earth and Space Science contexts; cross-institution/curriculum replications will strengthen explanatory power, especially for re-evaluating definitional items (e.g., S2) and items with high representational load (e.g., S6). The four-tier diagnostic has shown utility across science contexts (Astuti et al., 2023; Kiray & Simsek, 2021). Reported astronomy mental models further contextualize persistent patterns (Ubben & Bitzenbauer, 2022). Conceptual-change activities can strengthen these gains (Cardinot, 2024). Building representational competencies remains a key lever for lasting understanding (Herder & Rau, 2022).

4. CONCLUSION

Differentiated instruction that accommodates multiple intelligences improved students' conceptual understanding and reduced the persistence of key misconceptions compared with conventional teaching. Improvements were most consistent on causal–mechanistic indicators (e.g., stellar lifetime and end states), whereas definitional items remained the most resistant to change. Overall, the experimental class exhibited a predominance of positive change with minimal regress, indicating the approach's feasibility and instructional value in Earth and Space Science contexts.

5. REFERENCES

Salimpour, S., Fitzgerald, M., & Hollow, R. (2024). Examining the mismatch between the intended astronomy curriculum content, astronomical literacy, and the astronomical universe. *Physical Review Physics Education Research, 20*(1), 010135. https://doi.org/10.1103/PhysRevPhysEducRes.20.010135

Bitzenbauer, P., Navarrete, S., Hennig, F., Ubben, M. S., & Veith, J. M. (2023). A cross-age study on secondary school students' views of stars. *Physical Review Physics Education Research, 19*(2), 020165. https://doi.org/10.1103/PhysRevPhysEducRes.19.020165

Ubben, M. S., Hartmann, J., & Pusch, A. (2022). "Holes in the atmosphere of the universe": An empirical qualitative study on mental models of students regarding black holes. *Astronomy Education Journal, 2*(1), 029ra. https://doi.org/10.32374/AEJ.2022.2.1.029ra

Herder, T., & Rau, M. A. (2022). Representational-competency supports in an educational video game for undergraduate astronomy. *Computers & Education, 190,* 104602. https://doi.org/10.1016/j.compedu.2022.104602

Çelikkanlı, N. Ö., & Kızılcık, H. Ş. (2022). A review of studies about four-tier diagnostic tests in physics education. *Journal of Turkish Science Education, 19*(4), 1291–1311. https://doi.org/10.36681/tused.2022.175

Istiyono, E., Dwandaru, W. S. B., Fenditasari, K., Ayub, M. R. S. S. N., & Saepuzaman, D. (2023). The development of a four-tier diagnostic test based on modern test theory in physics education. *European Journal of Educational Research, 12*(1), 371–385. https://doi.org/10.12973/eu-jer.12.1.371

Astuti, I. A. D., Zulherman, & Yustika, G. P. (2023). Android-based four-tier physics test app to identify misconception profiles. *International Journal of Evaluation and*

Research in Education, 12(3), 1356–1363. https://doi.org/10.11591/ijere.v12i3.25536

Rusilowati, A., Supriyadi, S., & Cahyono, A. N. (2024). Development of a four-tier e-diagnostic test on the topic of momentum to measure and reduce student misconception. *Jurnal Pendidikan Fisika Indonesia, 20*(2). https://doi.org/10.15294/jpfi.v20i2.1297

Paçacı, C., & Çetin-Dindar, A. (2024). Effectiveness of conceptual change strategies in science: A meta-analysis. *Journal of Research in Science Teaching, 61*(10), 1801–1836. https://doi.org/10.1002/tea.21887

Langelaan, B. N., Gaikhorst, L., Smets, W., & Oostdam, R. J. (2024). Differentiating instruction: Understanding the key elements for successful teacher preparation and development. *Teaching and Teacher Education, 140,* 104464. https://doi.org/10.1016/j.tate.2023.104464

Kersting, M., Bondell, J., Steier, R., & Myers, M. (2024). Virtual reality in astronomy education: Reflecting on design principles. *International Journal of Science Education, Part B, 14*(2), 157–176. https://doi.org/10.1080/21548455.2023.2238871

Keppens, W., De Cock, M., Van Winckel, H., Van Dooren, W., & Sermeus, J. (2025). Exploring student estimates of astronomical scales: Impact of question formulation and visualization. *Physical Review Physics Education Research, 21,* 010159. https://doi.org/10.1103/PhysRevPhysEducRes.21.010159

Guerra-Reyes, F., Guerra-Dávila, E., Naranjo-Toro, M., Basantes-Andrade, A., & Guevara-Betancourt, S. (2024). Misconceptions in the learning of natural sciences: A systematic review. *Education Sciences, 14*(5), 497. https://doi.org/10.3390/educsci14050497

Utami, A., Sujarwo, S., Fauziyah, P. Y., Mustadi, A., Hidayat, R., & Rofiki, I. (2024). Bibliometric analysis of research developments on differentiated instruction. *European Journal of Educational Research, 13*(3), 1421–1439. https://doi.org/10.12973/eu-jer.13.3.1421

Azizah, S. N., Akhsan, H., Muslim, M., & Ariska, M. (2022). Analysis of college students' misconceptions in astronomy using a four-tier test. *Journal of Physics: Conference Series, 2165,* 012004. https://doi.org/10.1088/1742-6596/2165/1/012004

Kiray, S. A., & Simsek, S. (2021). Four-tier diagnostic test on density misconceptions. *International Journal of Science and Mathematics Education, 19*(5), 935–955. https://doi.org/10.1007/s10763-020-10087-5

Astuti, I. A. D., Bhakti, Y. B., Prasetya, R., & Rahmawati, Y. (2023). Four tier-relativity diagnostic test (4T-RDT) to identify student misconception. *JIPF: Jurnal Ilmu Pendidikan Fisika, 8*(1), 75–84. https://doi.org/10.26737/jipf.v8i1.3668

Ubben, M. S., & Bitzenbauer, P. (2022). Two cognitive dimensions of students' mental models in science. *Education Sciences, 12*(3), 163. https://doi.org/10.3390/educsci12030163

Cardinot, A. (2024). Non-digital educational games to support conceptual change in astronomy education. *Astronomy Education Journal, 4*(1). https://doi.org/10.32374/AEJ.AECON.2023.111aept

Herder, T., & Rau, M. A. (2022). The role of representational competencies for learning from an educational video game for astronomy. *Frontiers in Education, 7,* 919645. https://doi.org/10.3389/feduc.2022.919645