

Does it feel like it works? Three-dimensional modeling and thermodynamics conceptual understanding in teacher education

Janine Refil¹ , Georis Aplacador¹ , Rosemarian Arendain¹ , Nelson Ross¹ , Erdee Cajurao^{1*} 

¹ College of Education, Dr. Emilio B. Espinosa Sr. Memorial State College of Agriculture and Technology, Mandaon, Masbate, PHILIPPINES

*Corresponding Author: eccajurao@debesmscat.edu.ph

Citation: Refil, J., Aplacador, G., Arendain, R., Ross, N., & Cajurao, E. (2027). Does it feel like it works? Three-dimensional modeling and thermodynamics conceptual understanding in teacher education. *International Journal of Changes in Education*, 4(1), Article em112. <https://doi.org/10.29333/ijce/19474>

ARTICLE INFO

Received: 25 Jul. 2026

Accepted: 15 Sep. 2026

ABSTRACT

Three-dimensional (3D) modeling is promoted as a way to make abstract thermodynamic processes tangible, yet whether favorable perceptions accompany stronger learning is unclear. This descriptive-correlational and predictive study examined the association between perceived effectiveness of 3D modeling and thermodynamics conceptual understanding among 28 second-year pre-service science teachers at a Philippine state college. Each participant completed a 16-item questionnaire covering conceptual understanding, engagement, motivation and ease of learning, plus the 19-item thermodynamics conceptual test. Participants rated 3D modeling favorably (mean = 4.19, standard deviation = 0.46) and scored 14.82 of 19 (78.00%). No dimension was significantly associated with test scores ($r = -.113$ to $.164$, all $p > .05$), and backward regression retained no predictor, $F(4, 23) = 0.51$, $p = .726$, $R^2 = .0855$. Given the small sample and wide confidence intervals, these findings are inconclusive rather than evidence of no association. Larger studies are needed for more precise estimates.

Keywords: three-dimensional modeling, perceived effectiveness, conceptual understanding, thermodynamics, pre-service science teachers, statistical power

INTRODUCTION

Science education worldwide asks learners to reason about phenomena they cannot see. Energy transfer, molecular motion and system-level change are not directly observable, so learners must build and hold internal representations of them. Where instruction does not support that representational work, conceptual understanding tends to stall regardless of how much content is delivered. The instructional question is therefore not only what to teach but what to make visible.

Thermodynamics is where this difficulty is sharpest. Brundage et al. (2024a) administered a validated 78-item instrument to more than 1,000 students across 12 introductory and advanced physics classes at four institutions and found persistent difficulty with change in internal energy, work and heat transfer among both introductory and advanced students. Using the same instrument, Brundage et al. (2024b) found that entropy and the second law produce equally durable difficulties. Introductory students frequently import the properties of energy into entropy and reason as though entropy were conserved. Advanced students expect entropy increases even in processes where entropy does not change, and both groups struggle to separate the entropy of a system

from that of its surroundings. These are not gaps that repetition closes, and they persist in advanced coursework. Cognitive load theory offers one account of why: when instructional materials leave the whole burden of visualizing an invisible process to the learner, working memory is overtaxed and conceptual understanding suffers (Sweller, 1988; Sweller et al., 1998).

Three-dimensional (3D) modeling is one response to that burden. By rendering spatial and structural relationships externally, 3D models offload part of the visualization work and, in principle, free cognitive resources for reasoning (Teplá et al., 2022). The term covers many implementations, from computer-generated interactive models and simulations to printed physical artifacts and models that students build themselves, and these are not interchangeable. We later specify the implementation used in the course studied here, which was a student-constructed physical diorama.

The evidence on 3D modeling supports three claims of differing strength. First, 3D modeling can support conceptual understanding. Aslan et al. (2024) reported gains among primary school science students, and Abdo et al. (2024) traced the mechanism through think-aloud interviews: undergraduates who anchored their reasoning in 3D models showed deeper understanding when solving science problems,

Table 1. Theoretical basis and sample item for each perceived effectiveness dimension

Dimension	Theoretical basis	Items	Sample item
Conceptual understanding	Cognitive theory of multimedia learning (Mayer, 1997, 2001)	4	3D modeling helps me visualize abstract thermodynamics topics.
Engagement	Constructivism (Piaget, 1954)	4	3D modeling makes thermodynamics learning more interactive for me.
Motivation	Expectancy-value theory (Eccles & Wigfield, 2002)	4	3D modeling motivates me to study thermodynamics.
Ease of learning	Cognitive load theory (Sweller, 1988; Sweller et al., 1998)	4	3D modeling simplifies difficult thermodynamics problems.

which locates the benefit in how learners coordinate two-dimensional and 3D representations. Second, 3D modeling can improve engagement, motivation and spatial reasoning. Usembayeva et al. (2024) found it supported physics concept interpretation across institutions, Tejera et al. (2025) concluded in a teacher education review that it consistently supports spatial reasoning and hands-on engagement, and Karaismailoğlu and Yıldırım (2024) found improved spatial ability among pre-service teachers whether digital or concrete materials were used. Related work reports the same direction of benefit in other subject areas, with 3D printing improving elementary students' attitudes toward STEM (Abu Khurma et al., 2023) and supporting chemistry instruction across STEM disciplines (Trujillo-Cayado et al., 2024). Third, and most relevant here, the evidence for measured achievement is weaker than the first two claims suggest. Kwon (2023) synthesized 26 classroom studies and placed the average effect at 0.759, a medium rather than large magnitude, and Tejera et al. (2025) and Fokides and Lagopati (2024) both cautioned that the literature rests on small, single-institution samples of uneven design quality.

That third claim is reinforced by the wider evidence on manipulable representations. This literature is relevant here because 3D modeling belongs to that family and inherits its central puzzle: tools that reliably improve how learning feels do not reliably improve what learning produces. Wörner et al. (2022) reviewed 42 experimental studies and found that combining real and virtual experiments generally promoted conceptual understanding better than either alone, though no sequence proved superior. Mulwijik and Lazonder (2023) found no overall advantage for physical over virtual investigation across 35 studies ($g = -0.14$, 95% confidence interval [CI] $[-0.33, 0.06]$), with the advantage depending on learner age and on whether touch conveyed conceptually relevant information. Tseng et al. (2023) tested this on heat and temperature specifically: physical and virtual manipulatives were equally effective for changes of state, while virtual manipulatives outperformed physical ones for thermal expansion and contraction. Manipulable representations, then, influence engagement and subjective ease of learning reliably, while their effect on measured understanding is plausible but inconsistent.

The distinction matters most for pre-service science teachers, who must learn thermodynamics themselves and later decide whether to use tools such as 3D modeling with their own students. López et al. (2025) analyzed 49 learning activities designed by pre-service physics, chemistry and biology teachers and found that measured technological pedagogical content knowledge correlated with the integration of content-specific technologies, and that such

activities carried significantly higher cognitive demand. Pre-service teachers' judgments about a tool therefore propagate into future classrooms. Two recent studies show why those judgments should not be taken at face value. Wekerle et al. (2024) found that interactive technology-enhanced activities strongly predicted students' perceptions of their own learning while the expected advantage for measured cognitive outcomes did not materialize, and Noroozi et al. (2025) found perceived learning unrelated to actual success level among 284 university students.

These questions carry particular weight in the Philippines, where Filipino learners scored among the lowest performing participants in science literacy in the 2022 Programme for International Student Assessment (PISA), with minimal improvement since 2018 (Acido & Caballes, 2024), and where access to instructional technology and hands-on materials distinguishes higher performing students from consistently low-proficiency peers (Bernardo et al., 2023). Philippine classroom studies report favorable results for manipulable representations. Amil and Buensuceso (2025) found significant pretest to posttest gains for both virtual simulations and 3D models in grade 8 science, and Ugaldi et al. (2025) reported significant gains and favorable perceptions among grade 6 students using 3D-printed models, alongside practical concerns about fragility and supply. Teacher education programs in such settings adopt instructional technology under real resource constraints, which makes the difference between a tool that is liked and a tool that works a budgeting question rather than an academic one.

This study is guided by five theories, each supporting one or more dimensions of perceived effectiveness rather than introduced as a free-standing framing. The cognitive theory of multimedia learning holds that learners understand complex ideas better when information arrives through both visual and verbal channels (Mayer, 1997, 2001), which underpins the conceptual understanding dimension. Cognitive load theory holds that learning improves when materials reduce extraneous load (Sweller, 1988; Sweller et al., 1998), which underpins ease of learning. Constructivism holds that learners build knowledge by acting on their environment (Piaget, 1954), which underpins engagement. The technology acceptance model holds that perceived usefulness and perceived ease of use govern acceptance of a technology (Davis, 1989), and it underpins the perceived effectiveness construct as a whole. Expectancy-value theory holds that motivation depends on expectancy of success and task value (Eccles & Wigfield, 2002), which underpins the motivation dimension.

Table 1 maps each dimension to its theoretical basis. Together the framework predicts a chain: favorable perception

supports engagement and motivation, which support learning behavior, which may support conceptual understanding. This study tests the final link in that chain, which is the one the literature has least often examined.

The gap is therefore specific. Studies have measured perceptions of 3D modeling and studies have measured learning outcomes, but few have measured both in the same sample and tested whether the first is associated with the second, and fewer still have done so among pre-service science teachers in a Philippine state college. This study addressed that gap by examining three questions among second-year pre-service science teachers: the level of perceived effectiveness of 3D modeling across conceptual understanding, engagement, motivation and ease of learning; their level of thermodynamics conceptual understanding; and whether perceived effectiveness is associated with and predicts that understanding. Because the design is correlational, it establishes association rather than causal effect, and the analysis reports the precision of its estimates alongside the point estimates.

MATERIALS AND METHODS

Research Design

This study used a descriptive-correlational and predictive design, which suits quantitative research that describes a population, tests associations among variables and determines predictive value within a single non-experimental dataset (Creswell & Creswell, 2018). The descriptive component characterized the level of perceived effectiveness of 3D modeling and the level of thermodynamics conceptual understanding. The correlational component tested whether the two were associated. The predictive component used backward multiple regression to determine whether perceived effectiveness, overall or by dimension, predicted conceptual understanding. No variable was manipulated and no comparison group was used, so the design supports statements about association and not about causal effect. Data collection took place immediately after respondents completed the thermodynamics course in which 3D modeling was used, so both measures referred to the same completed instructional period.

The Three-Dimensional Modeling Intervention

This section describes the 3D modeling component of the thermodynamics course in which participants were enrolled. It is reported so that readers can judge exactly what participants were rating when they completed the perceived effectiveness questionnaire.

The models used in this course were 3D physical dioramas rather than computer-generated models, screen-based simulations or commercially printed artifacts. The students designed and built the dioramas themselves for the thermodynamics course, so no modeling software or external supplier was involved at any stage. Because the models were student-constructed, the act of building each diorama formed part of the instructional experience that participants were later asked to rate.

3D modeling activities were scheduled twice each week across the thermodynamics course in which the cohort was enrolled. The course itself was delivered primarily through traditional lecture-discussion together with teacher-guided instruction. The modeling activities supplemented that primary approach rather than replacing any part of it.

Because 3D modeling supplemented rather than replaced the primary instruction, conceptual understanding measured by the thermodynamics conceptual test reflects the combined effect of everything the cohort received during the semester. The study therefore examines whether students' judgments about the modeling component are associated with their conceptual understanding. It does not attribute that conceptual understanding to 3D modeling by itself.

One further point of scope deserves statement here rather than in the limitations alone. Much of the literature reviewed concerns computer-generated models, virtual manipulatives or 3D printed artifacts produced outside the classroom. Student-built physical dioramas share the manipulable and spatial character that this literature identifies as the active ingredient, but they differ in how the models are produced and in who produces them. Comparisons drawn between the present findings and that literature should be read with this difference in view.

Participants and Sampling

Participants were second-year pre-service science teachers enrolled in a thermodynamics course at a Philippine state college during the first semester of academic year 2025 to 2026. Total enumeration was used because the entire second-year cohort constituted the population of theoretical interest and the population was small and clearly bounded. The class comprised 28 students and all 28 participated, completing both the perceived effectiveness questionnaire and the conceptual test. The resulting sample size limits the precision of the regression estimates, which is addressed in the discussion and the limitations.

Instrumentation

Perceived effectiveness questionnaire

Perceived effectiveness was measured with a researcher-made 16-item questionnaire covering four dimensions, with four items per dimension: conceptual understanding, engagement, motivation and ease of learning. Each dimension was operationalized as the mean (M) of its four items on a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree), and the overall score was the M of the four dimension Ms. All items were worded positively, so no reverse scoring was applied. Higher scores indicate more favorable judgments of 3D modeling for learning thermodynamics.

Item content was derived from the theoretical framework, with each dimension anchored to a specific theory rather than to the framework in general. **Table 1** gives the mapping and a sample item for each dimension. The complete instrument appears in **Appendix A**.

M scores were interpreted against a five-band scale: 1.00 to 1.80 not effective, 1.81 to 2.60 less effective, 2.61 to 3.40 moderately effective, 3.41 to 4.20 effective and 4.21 to 5.00 very effective.

Thermodynamics conceptual test

Thermodynamics conceptual understanding was measured with the thermodynamics conceptual test, a researcher-made 19-item multiple-choice instrument with four options per item. Each correct answer scored 1 and each incorrect answer scored 0, giving a possible range of 0 to 19. The final instrument covers seven content areas: heat engines and efficiency (4 items), thermodynamic systems (4 items), heat transfer mechanisms (3 items), the laws of thermodynamics and internal energy (2 items), entropy and the second law (2 items), heat and temperature and their units (2 items) and real-world applications (2 items). Items were written to require interpretation and application rather than recall alone. Fourteen items ask students to classify a system, identify a mechanism or interpret a described situation. Two require computation, one of a change in internal energy from heat and work and one of work output from efficiency and heat input. The remaining three items test definitional knowledge of laws and units. The complete instrument appears in **Appendix B**.

Validation and reliability

Content validity of both instruments was established through expert evaluation alongside a review of related literature. Experts in physics education, educational research and measurement reviewed the initial questionnaire and test items for clarity, relevance and alignment with the study variables and with thermodynamics content, and their feedback was used to revise the instruments before pilot testing. The number of experts, their individual qualifications and any quantitative content validity index were not documented at the time, which is noted in the limitations.

A pilot test was conducted with 30 third-year pre-service science teachers at the same institution who had already completed the thermodynamics course and who did not take part in the main study. The initial thermodynamics conceptual test comprised 40 multiple-choice items. Items were retained if the difficulty index fell between .30 and .80 and the discrimination index was at least .20. Twenty-one items failed one or both criteria and were removed, leaving the final 19 items. Across the retained items the M difficulty index was .62 and the M discrimination index was .41, indicating a moderately difficult test that separated higher from lower scoring pilot participants. Distractor analysis beyond the difficulty and discrimination indices was not conducted. Several perceived effectiveness items were reworded for clarity on the basis of the same pilot, and the 16-item structure was retained.

Internal consistency was computed on the main study sample. The perceived effectiveness questionnaire returned an overall Cronbach's alpha of .86, with dimension coefficients of .912 for conceptual understanding, .889 for engagement, .872 for motivation and .872 for ease of learning, all above the conventional threshold of .70 (Field, 2013). The thermodynamics conceptual test returned a Kuder-Richardson formula 20 coefficient of .79, which suits dichotomously scored multiple-choice items and also exceeds .70. Because both instruments were researcher-made, these estimates rest on a single administration rather than an established psychometric record, and no factor analysis was conducted to

examine the four-dimension structure. Both points are carried into the limitations.

Data Collection

Data were collected face to face in a single class period. Permission was obtained from the institutional authorities and the course instructor beforehand, and participants gave informed consent. The researchers distributed printed copies of the questionnaire followed by the conceptual test, gave standard instructions and remained present to answer procedural questions while participants worked independently. All materials were collected immediately afterward. All 28 members of the population participated and no response was incomplete, so no case was excluded.

Data Analysis

Analyses were conducted in JASP version 0.18.3 (JASP Team, 2024). Ms and standard deviations (SDs) characterized perceived effectiveness by dimension and overall, and performance on the thermodynamics conceptual test. Normality was assessed with the Shapiro-Wilk test before inferential analysis, and because the data were normally distributed, Pearson product-moment correlations were used to examine associations between each perceived effectiveness dimension and test score, reported with 95% CIs.

Backward multiple regression then tested whether the four dimensions predicted test score. JASP has no automated backward elimination routine in its standard linear regression module, so the procedure was performed manually. All four predictors were entered simultaneously. The predictor with the highest p value above the removal criterion of $p > .05$ was removed and the model re-estimated, and the cycle repeated until no predictor met the retention criterion. R^2 , adjusted R^2 and predictor p values were recorded at each step.

Regression assumptions were assessed before interpretation. Normality of residuals was tested with the Shapiro-Wilk test, which was non-significant ($p = .342$). Linearity was examined through scatterplots of residuals against predicted values, which showed no systematic pattern. Independence of errors was checked with the Durbin-Watson statistic, which ranged from 1.82 to 2.11 across the five models and fell within the acceptable range of 1.5 to 2.5. Homoscedasticity was assessed by visual inspection of residual plots, which showed no funnel-shaped spread. Multicollinearity was examined using the variance inflation factor, with values from 1.87 to 2.93, below the conventional threshold of 5.0. No assumption was violated.

Because the sample was small, the precision of the estimates is reported directly through CIs for every correlation and regression coefficient and through adjusted R^2 at each elimination step. A sensitivity calculation following Buckley (2024) was also conducted. It used Cohen's (1988) f^2 , the number of predictors and the sample size to establish the sample size a future study would need to detect associations of the observed and of conventional medium magnitude with power of .80. This calculation is reported as guidance for replication and is not used to interpret the present null result.

Table 2. Descriptive statistics of perceived effectiveness of 3D modeling (N = 28)

Dimension	M	SD	Interpretation
Conceptual understanding	4.21	0.48	Very effective
Engagement	4.15	0.52	Effective
Motivation	4.18	0.44	Effective
Ease of learning	4.41	0.41	Very effective
Overall perceived effectiveness	4.19	0.46	Effective

Note. 1.00-1.80: Not effective; 1.81-2.60: Less effective; 2.61-3.40: Moderately effective; 3.41-4.20: Effective; & 4.21-5.00: Very effective

Ethical Considerations

Approval to conduct the study was secured from the dean of the college of education of the participating state college, which is the body that authorizes undergraduate research at that institution. The institution does not issue ethical clearance certificates or protocol numbers for undergraduate research, so none is available for this study. Participation was voluntary and preceded by written informed consent describing the purpose, procedures, voluntary nature and right to withdraw without penalty. Confidentiality was maintained by assigning participant codes in place of personal identifiers, and all data were stored securely and used solely for academic purposes. No deception was involved and the study posed no more than minimal risk.

RESULTS

Results are reported in the order of the study's three questions: the level of perceived effectiveness of 3D modeling, the level of thermodynamics conceptual understanding, and the association and predictive value between the two.

Level of Perceived Effectiveness of Three-Dimensional Modeling

Table 3. Descriptive statistics and score distribution for the thermodynamics conceptual test (N = 28)

Measure	Value
Mean score (of 19 items)	14.82
Standard deviation (items)	1.81
Mean percentage	78.00%
Minimum score	11.00 (57.89%)
Maximum score	18.00 (94.73%)
Needs improvement (below 60%)	1 student (3.6%)
Satisfactory (60% to 89%)	25 students (89.3%)
Outstanding (90% or above)	2 students (7.1%)

Note. Bands follow the institutional criterion-referenced scale. Possible score range is 0 to 19

Table 2 presents the descriptive statistics for the four perceived effectiveness dimensions and the overall composite.

Participants rated 3D modeling favorably, with an overall M of 4.19 (SD = 0.46), which falls in the effective band and sits 0.02 points below the threshold for very effective. Ease of learning received the highest rating (M = 4.41), followed by conceptual understanding (M = 4.21), motivation (M = 4.18) and engagement (M = 4.15). All four dimensions exceeded 4.00, so respondents judged 3D modeling beneficial across every measured aspect of their learning experience. **Figure 1** shows these results against the scale thresholds.

Level of Thermodynamics Conceptual Understanding

Table 3 summarizes performance on the thermodynamics conceptual test.

Participants achieved an M of 14.82 of 19 items, or 78.00%, which falls in the satisfactory band of the institutional criterion-referenced scale. Scores ranged from 11.00 (57.89%, needs improvement) to 18.00 (94.73%, outstanding), and 25 of 28 students fell within the satisfactory band. The SD of 1.81 items indicates limited variability in test scores across the sample. Limited variability is a property of the score

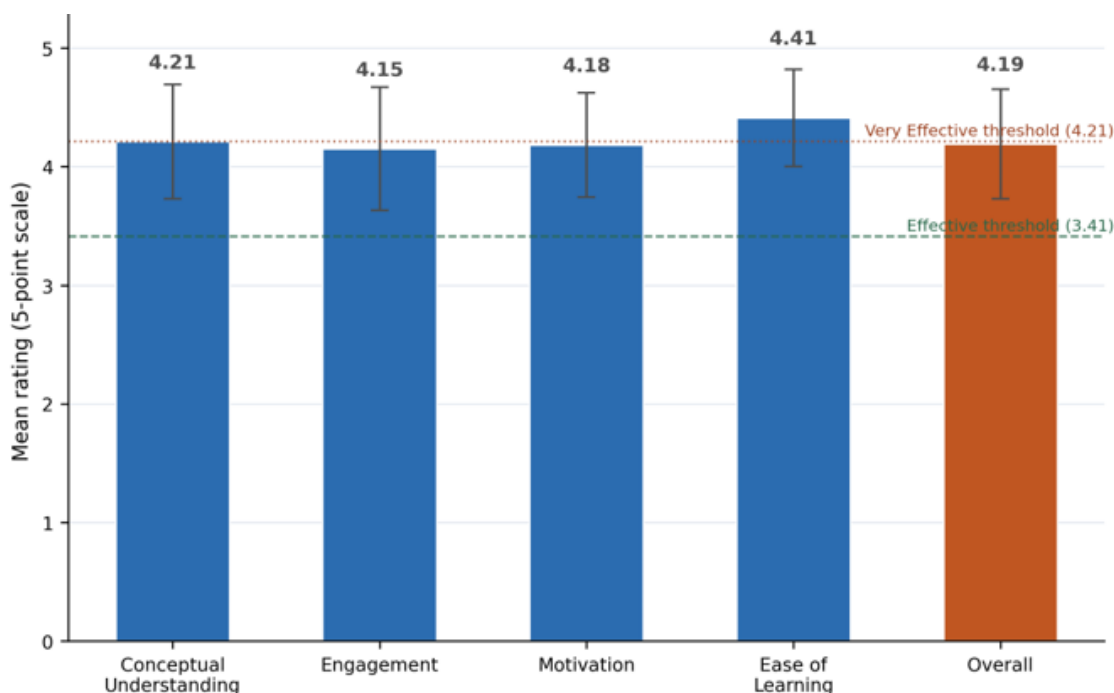


Figure 1. Average perceived effectiveness of 3D modeling by dimension (the authors' own elaboration)

Table 4. Pearson correlations between perceived effectiveness dimensions and thermodynamics conceptual test score (N = 28)

Dimension	r	95% CI	p
Conceptual understanding	-.113	[-.48, .27]	.566
Engagement	-.005	[-.38, .37]	.980
Motivation	.164	[-.21, .51]	.404
Ease of learning	.101	[-.27, .46]	.609

distribution and does not by itself establish that participants held equivalent conceptual understanding, since students may reach similar totals through different patterns of correct and incorrect responses.

Association Between Perceived Effectiveness and Conceptual Understanding

Table 4 presents Pearson correlations between each perceived effectiveness dimension and thermodynamics conceptual test score, with 95% CIs.

No dimension was significantly associated with test score. Conceptual understanding showed a small negative correlation ($r = -.113$, $p = .566$) and engagement essentially none ($r = -.005$, $p = .980$). Motivation and ease of learning showed small positive correlations ($r = .164$, $p = .404$ and $r = .101$, $p = .609$). Every CI was wide and included zero. The interval for motivation, $[-.21, .51]$, is the most informative of the four: it is consistent with anything from a small negative association to a moderate positive one, which is the level of precision a sample of 28 supports.

Perceived Effectiveness as a Predictor of Conceptual Understanding

A backward multiple regression tested whether the four dimensions, entered together, predicted test score. Table 5 presents the initial model with unstandardized coefficients and 95% CIs.

The initial model was not statistically significant, $F(4, 23) = 0.514$, $p = .726$. It accounted for 8.55% of the variance in test score, and the adjusted R^2 of $-.07$ indicates that the four predictors together explained less variance than the number of parameters would be expected to absorb by chance. No

predictor reached significance individually and every coefficient interval spanned zero by a wide margin. Ease of learning carried the highest p value ($p = .793$) and was removed first. Elimination continued through three further steps, removing conceptual understanding, then engagement, then motivation, until no predictor remained. Table 6 summarizes model fit at each step and Figure 2 shows the same sequence with the corresponding p values.

R^2 stayed low throughout, from .0855 with four predictors to .0270 with motivation alone, and adjusted R^2 was at or below zero at every step except the two-predictor model, where it reached .007. The overall model p value never approached significance. R^2 was nearly identical between Step 1 and Step 3 despite the removal of two predictors, so ease of learning and conceptual understanding added almost nothing to explanatory power. These results indicate that perceived effectiveness of 3D modeling, whether measured as conceptual understanding, engagement, motivation or ease of learning, was not significantly associated with thermodynamics conceptual understanding in this sample.

DISCUSSION

Favorable Perceptions Alongside Unrelated Conceptual Understanding

Participants judged 3D modeling positively on every measured dimension, which is consistent with the literature linking 3D modeling and printing to favorable student and teacher attitudes (Thyssen & Meier, 2023; Ugaldi et al., 2025; Usembayeva et al., 2024). Higher perceived effectiveness was not significantly associated with higher scores on the thermodynamics conceptual test. The pattern echoes Wekerle et al. (2024), who found that interactive technology-enhanced activities strongly predicted students' perceptions of learning without a matching advantage in measured cognitive outcomes, and Noroozi et al. (2025), who found perceived learning unrelated to actual success level. Under the technology acceptance model, perceived usefulness and perceived ease of use are theorized to drive acceptance and engagement rather than to guarantee learning outcomes

Table 5. Initial backward regression model predicting thermodynamics conceptual test score from all four perceived effectiveness dimensions (step 1)

Predictor	B	SE	95% CI	t	p
Intercept	85.49	21.00	[42.05, 128.93]	4.071	< .001
Conceptual understanding	-8.67	8.82	[-26.92, 9.58]	-0.982	.336
Engagement	2.07	6.90	[-12.20, 16.34]	0.300	.767
Motivation	3.60	5.26	[-7.28, 14.48]	0.684	.501
Ease of learning	1.37	5.14	[-9.26, 12.00]	0.266	.793

Note. Model fit: $R = .292$, $R^2 = .0855$, Adjusted $R^2 = -.074$, $F(4, 23) = 0.514$, & $p = .726$; CIs were computed as B plus or minus 2.069 times SE, with $df = 23$; & Removal criterion was $p > .05$

Table 6. Summary of backward elimination steps

Step	Predictors retained	R^2	Adjusted R^2	F	df	p	Predictor removed
1	CU, E, M, EoL	.0855	-.074	0.514	4, 23	.726	Ease of learning
2	CU, E, M	.0827	-.032	0.715	3, 24	.553	Conceptual understanding
3	E, M	.0803	.007	1.092	2, 25	.351	Engagement
4	M	.0270	-.010	0.720	1, 26	.404	Motivation
5	None	n/a	n/a	n/a	n/a	n/a	Final model, no predictors remain

Note. CU: Conceptual understanding; E: Engagement; M: Motivation; & EoL: Ease of learning

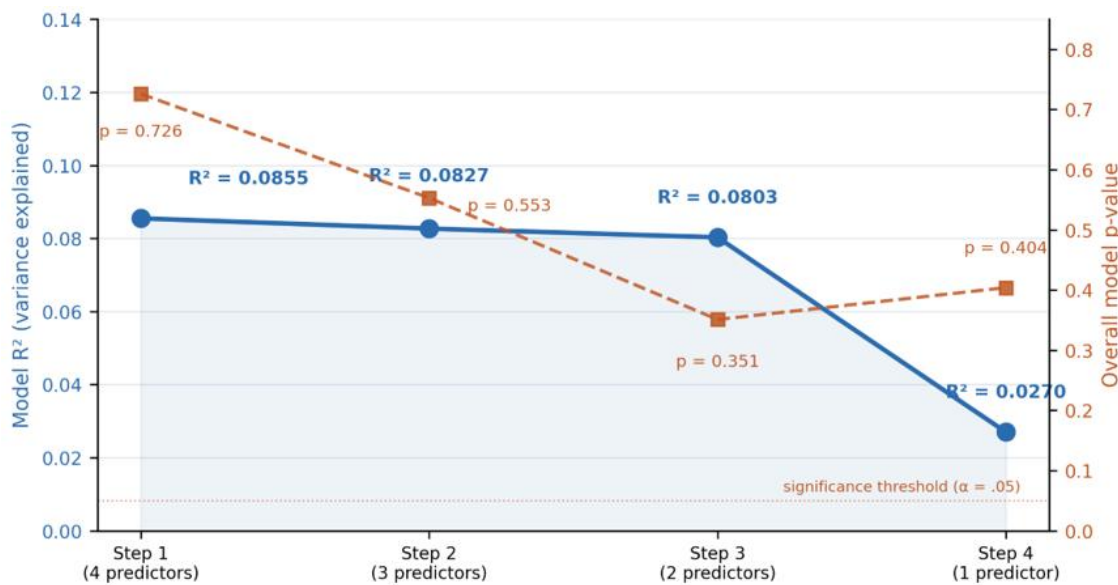


Figure 2. Model R² and overall model p value across the four backward elimination steps (the authors' own elaboration)

(Davis, 1989), and the present result is consistent with that separation.

Ease of learning received the highest rating of any dimension. Participants found 3D modeling subjectively easier to learn from, and that subjective ease was not associated with higher scores on a conceptual test. Expectancy-value theory offers a complementary reading, since positive perceptions are theorized to shape engagement and persistence more directly than knowledge, which depends additionally on prior preparation and assessment alignment (Eccles & Wigfield, 2002). This is also why the result sits alongside Amil and Buensuceso (2025) rather than contradicting them. Their within-individual pretest to posttest design tests a change in learning over time, whereas the present design tests an association between two measures taken on one occasion. The two designs answer different questions.

Satisfactory Scores With Limited Variability

Average performance was satisfactory but not exceptional, at 78.00%, with most students in the satisfactory rather than the outstanding band. This is consistent with reports that Filipino science learners continue to underperform relative to international benchmarks even when instructional resources are available (Acido & Caballes, 2024; Bernardo et al., 2023). The limited variability in test scores may also have attenuated the observed associations, since a correlation requires meaningful variation in both variables to express itself numerically. With 25 of 28 participants inside a single band, the spread available for a linear association to emerge was small. No formal test of range restriction was conducted, so this remains an interpretation rather than a demonstrated property of the data. The narrow spread may reflect a single cohort taught by one instructor. Sampling across multiple sections would widen the range of observed scores and is recommended for replication.

Interpreting the Non-Significant Association

The absence of a statistically significant association does not establish that perceived effectiveness of 3D modeling is

unrelated to thermodynamics conceptual understanding. The clearest evidence for that reading is the imprecision of the estimates themselves. Every correlation CI was wide and crossed zero, and the widest, [-.48, .27] for conceptual understanding, spans almost the entire range of small to moderate associations in both directions. The regression coefficients tell the same story. The interval for conceptual understanding, [-26.92, 9.58] percentage points per scale unit, is compatible with a substantial negative association, with no association and with a modest positive one. Estimates this imprecise cannot distinguish a true null from a real association of the size this literature typically reports.

The adjusted R² reinforces the point. At -.07 for the initial model, it indicates that four predictors estimated from 28 observations absorbed more degrees of freedom than they explained variance. This is a characteristic signature of a model specified more richly than the available data can support, and it is a reason to treat the point estimates as unstable rather than as evidence of absence.

A sensitivity calculation indicates the scale a more precise study would need. For an association of the magnitude observed here, expressed as Cohen's (1988) $f^2 = .09$ with four predictors, detecting it with power of .80 would require approximately 133 participants. For a conventional medium association of $f^2 = .15$ (Cohen, 1988), approximately 85 would suffice. These figures are offered as design guidance for replication. They are not evidence about what the present study found, since a power calculation derived from an observed effect adds little beyond what the CIs already show. The defensible summary is that the association between perceived effectiveness and conceptual understanding remains inconclusive in this sample. The study did not find evidence of an association, and it was not precise enough to rule one out.

A Possible Mismatch Between Tool and Assessment

The theoretical framework predicts that reducing extraneous load through spatial and structural representations should support conceptual understanding, which should in

turn show up in measured performance. The present findings did not confirm the final link. One possible explanation is that 3D modeling may primarily support visualization of abstract processes, whereas the present assessment emphasized conceptual classification and application in a single-answer multiple-choice format. The study did not measure cognitive load, spatial ability or visualization directly, so this reading is an interpretation drawn from the literature and not a finding of the present data.

Three studies make that interpretation plausible. Abdo et al. (2024) showed that the benefit students derive from 3D models lies in how they coordinate two-dimensional and 3D representations while working through a problem, a process a single-answer item is poorly positioned to capture. Teplá et al. (2022) proposed that 3D models mainly influence how learners process spatial and structural information, a skill closer to spatial visualization than to the conceptual classification the present test emphasizes. Karaismailoğlu and Yıldırım (2024) located the gain from 3D modeling among pre-service teachers on a spatial ability measure, which is the kind of outcome the present test did not assess.

The wider literature on manipulable representations points the same way, and one finding in it bears directly on the present intervention. Muilwijk and Lazonder (2023) found no overall conceptual knowledge advantage for physical over virtual investigation, with any advantage depending on whether touch conveyed conceptually relevant information. A diorama renders the spatial arrangement of a thermodynamic system tangible, but the quantities that the conceptual test asks about, such as entropy change and efficiency, are not themselves things a learner can touch. Wörner et al. (2022) reached a compatible conclusion, namely that the function a representation serves for the learning task matters more than its format. Kwon (2023) placed the average effect of 3D printing in education at a medium rather than a large magnitude, and associations of that size are precisely the ones a sample of 28 cannot estimate with useful precision. The null finding may therefore reflect, at least in part, the combination of limited statistical precision and possible misalignment between the assessment and the skills that 3D modeling is intended to support. Wekerle et al. (2024) found the same asymmetry in multilevel modeling, with interactive activities benefiting perceived learning more consistently than measured cognitive outcomes, so the pattern observed here may not be peculiar to this sample.

Practical Implications for Teacher Education

Positive student perceptions of an instructional tool should not, on their own, be treated as evidence that the tool improves conceptual understanding. Pre-service science teachers who adopt 3D modeling in their own classrooms should pair it with explicit problem-solving practice and formative assessment aligned to their intended learning outcomes, rather than assume that favorable engagement will independently produce stronger conceptual results. López et al. (2025) support this recommendation: pre-service science teachers' technological pedagogical content knowledge was associated with the integration of content-specific technologies and with activities of higher cognitive demand. Teacher education programs are therefore well placed to

model the pairing of engagement-focused tools with outcome-focused assessment design during pre-service training itself.

Resourcing matters alongside pedagogy. Ugaldi et al. (2025) found that Filipino learners using 3D-printed models reported favorable perceptions alongside practical concerns such as model fragility and insufficient supply. Maintaining and replacing 3D resources is part of what determines whether a promising tool sees sustained classroom use, and programs adopting it should budget for these constraints alongside pedagogical training.

Limitations

Several limitations qualify how these findings should be used. The sample of 28 participants from a single institution limited the precision of every estimate reported here and restricts transferability to other pre-service teacher populations. Both instruments were researcher-made. Internal consistency was satisfactory: an overall Cronbach's alpha of .86 for the perceived effectiveness questionnaire and a Kuder-Richardson formula 20 coefficient of .79 for the conceptual test. Content validity was established through expert review and a 30-participant pilot with item analysis. Two elements of that psychometric record are nonetheless incomplete: the number and qualifications of the content experts and any quantitative content validity index were not documented at the time, and no factor analysis was conducted to confirm the four-dimension structure of the questionnaire, which a sample of 28 could not support in any case. Both instruments therefore rest on a single administration rather than an established psychometric record.

The four-predictor regression structure, though theoretically motivated, was more complex than a sample of this size can reliably support, as the negative adjusted R^2 shows. A single-predictor model using the overall perceived effectiveness composite would have offered a more stable, if less granular, test of the same question. The design relied on one self-reported perception measure and one conceptual test administered on the same occasion, which cannot rule out common method bias. Because 3D modeling was one component of a course delivered through several instructional methods, the study cannot isolate its contribution to the conceptual understanding measured at the end. The intervention itself also bounds transferability in a way worth stating explicitly. Participants rated student-built physical dioramas, whereas much of the comparison literature examines computer-generated models or printed artifacts supplied to learners ready-made. Findings from this cohort should therefore not be read as evidence about digital modeling tools without further study. Finally, both variables were measured once and only for thermodynamics, so the findings cannot speak to other physics topics or to the same cohort at a later point.

CONCLUSION

Pre-service science teachers in this study judged 3D modeling favorably on every dimension measured, and those judgments were not significantly associated with their conceptual understanding of thermodynamics. Each of the

four perceived effectiveness dimensions dropped out of the regression model in turn, and no dimension showed a significant correlation with test score.

How useful an instructional tool feels to a learner and what that learner can demonstrate on a conceptual assessment are separate outcomes. Neither should be read as evidence for the other when a teacher education program decides what to adopt. The non-significant relationship itself should be interpreted cautiously, given the small sample and the limited precision of the estimates. This study did not find evidence of an association, and it was not precise enough to rule one out.

Two directions follow for future research. Larger samples drawn from multiple sections and institutions would produce the precision this design lacked. Pairing a perception measure with an assessment built around the spatial and visualization skills that 3D modeling is meant to support, rather than conceptual classification alone, would test whether the alignment between tool and assessment is what the present design missed.

Author contributions: JR: conceptualization, data curation, resources, writing – review & editing; GA: conceptualization, data curation, resources, writing – review & editing; RA: conceptualization, writing – review & editing; NR: conceptualization, data curation, resources, writing – review & editing; EC: conceptualization, formal analysis, supervision, writing – original draft, writing – review & editing. All authors agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Acknowledgments: The authors would like to thank the second-year pre-service science teachers who participated in this study, and the faculty of the college of education for granting access to the thermodynamics course.

Ethical statement: The authors stated that approval to conduct the study was secured from the dean of the college of education of the participating state college, which is the body that authorizes undergraduate research at that institution. The institution does not issue ethical clearance certificates or protocol numbers for undergraduate research. Participation was voluntary and all participants provided written informed consent prior to data collection.

AI statement: No generative AI tool was used to produce the data, the analysis or the interpretation reported in this manuscript. The authors take full responsibility for the content of the submission.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

- Abdo, S. N., Hsu, J. L., Kapetanakis, C., Newman, D. L., Wright, L. K., & Bailey, J. (2024). An exploration of spatial visualization skills: Investigating students' use of 3D models in science problems during think-aloud interviews. *Journal of Chemical Education*, 101(9), 3624-3634. <https://doi.org/10.1021/acs.jchemed.3c01355>
- Abu Khurma, O., Ali, N., & Swe Khine, M. (2023). Exploring the impact of 3D printing integration on STEM attitudes in elementary schools. *Contemporary Educational Technology*, 15(4), Article ep458. <https://doi.org/10.30935/cedtech/13568>
- Acido, J. V., & Caballes, D. G. (2024). Assessing educational progress: A comparative analysis of PISA results (2018 vs. 2022) and HDI correlation in the Philippines. *World Journal of Advanced Research and Reviews*, 21(1), 462-474. <https://doi.org/10.30574/wjarr.2024.21.1.0020>
- Amil, A. V., & Buensuceso, D. B. (2025). A comparison of virtual simulation and 3D model as tools in teaching and learning rates of reaction in science 8. *International Journal of Research and Innovation in Social Science*, 9(7), 3120-3140. <https://doi.org/10.47772/IJRISS.2025.907000255>
- Aslan, A., Avcı, S. G., & Gökçü, M. Ş. (2024). Using 3-dimensional models as teaching tools in science education for primary school students. *International Journal of 3D Printing Technologies and Digital Industry*, 8(2), 237-254. <https://doi.org/10.46519/ij3dptdi.1473140>
- Bernardo, A. B. I., Cordel, M. O., II, Calleja, M. O., Teves, J. M. M., Yap, S. A., & Chua, U. C. (2023). Profiling low-proficiency science students in the Philippines using machine learning. *Humanities and Social Sciences Communications*, 10, Article 192. <https://doi.org/10.1057/s41599-023-01705-y>
- Brundage, M. J., Meltzer, D. E., & Singh, C. (2024a). Investigating introductory and advanced students' difficulties with change in internal energy, work, and heat transfer using a validated instrument. *Physical Review Physics Education Research*, 20(1), Article 010115. <https://doi.org/10.1103/PhysRevPhysEducRes.20.010115>
- Brundage, M. J., Meltzer, D. E., & Singh, C. (2024b). Investigating introductory and advanced students' difficulties with entropy and the second law of thermodynamics using a validated instrument. *Physical Review Physics Education Research*, 20(1), Article 020110. <https://doi.org/10.1103/PhysRevPhysEducRes.20.020110>
- Buckley, J. (2024). Conducting power analyses to determine sample sizes in quantitative research: A primer for technology education researchers using common statistical tests. *Journal of Technology Education*, 35(2), 81-109. <https://doi.org/10.21061/jte.v35i2.a.5>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340. <https://doi.org/10.2307/249008>
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53(1), 109-132. <https://doi.org/10.1146/annurev.psych.53.100901.135153>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE.

- Fokides, E., & Lagopati, G. (2024). The utilization of 3D printers by elementary-aged learners: A scoping review. *Journal of Information Technology Education: Innovations in Practice*, 23, Article 006. <https://doi.org/10.28945/5288>
- JASP Team. (2024). JASP (version 0.18.3). *JASP*. <https://jasp-stats.org/>
- Karaismailoğlu, F., & Yıldırım, M. (2024). The effect of 3D modeling performed using Tinkercad or concrete materials in the context of the flipped classroom on pre-service teachers' spatial abilities. *Research in Science & Technological Education*, 42(4), 1264-1283. <https://doi.org/10.1080/02635143.2023.2223134>
- Kwon, H. (2023). Meta analysis on effects of using 3D printing in South Korea K-12 classrooms. *Education and Information Technologies*, 28(9), 11733-11758. <https://doi.org/10.1007/s10639-023-11644-5>
- López, A., Meneses, A., & Montenegro, M. (2025). Pre-service teachers' TPAC and its relationship to technology integration in science learning activities. *Journal of Science Education and Technology*, 34(2), 436-451. <https://doi.org/10.1007/s10956-025-10201-8>
- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32(1), 1-19. https://doi.org/10.1207/s15326985ep3201_1
- Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139164603>
- Muilwijk, S. E., & Lazonder, A. W. (2023). Learning from physical and virtual investigation: A meta-analysis of conceptual knowledge acquisition. *Frontiers in Education*, 8, Article 1163024. <https://doi.org/10.3389/feduc.2023.1163024>
- Noroozi, O., Alqassab, M., Taghizadeh Kerman, N., Banihashem, S. K., & Panadero, E. (2025). Does perception mean learning? Insights from an online peer feedback setting. *Assessment & Evaluation in Higher Education*, 50(1), 83-97. <https://doi.org/10.1080/02602938.2024.2345669>
- Piaget, J. (1954). *The construction of reality in the child*. Basic Books. <https://doi.org/10.1037/11168-000>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251-296. <https://doi.org/10.1023/A:1022193728205>
- Tejera, M., Galiç, S., & Lavicza, Z. (2025). 3D modelling and printing in teacher education: A systematic literature review. *Journal for STEM Education Research*, 9(1), 1-32. <https://doi.org/10.1007/s41979-025-00147-2>
- Teplá, M., Teplý, P., & Šmejkal, P. (2022). Influence of 3D models and animations on students in natural subjects. *International Journal of STEM Education*, 9, Article 65. <https://doi.org/10.1186/s40594-022-00382-8>
- Thyssen, C., & Meier, M. (2023). 3D printing as an element of teaching: Perceptions and perspectives of teachers at German schools. *Frontiers in Education*, 8, Article 1233337. <https://doi.org/10.3389/feduc.2023.1233337>
- Trujillo-Cayado, L. A., Santos, J., Cordobés, F., & Ramos-Payán, M. (2024). Influence of the use of 3D printing technology for teaching chemistry in STEM disciplines. *Computer Applications in Engineering Education*, 32(4), Article e22738. <https://doi.org/10.1002/cae.22738>
- Tseng, Y.-K., Lin, F.-S., Tarng, W., Lu, Y.-L., & Wang, T.-L. (2023). Comparing the effects of physical, virtual, and hybrid labs on primary school students' conceptual learning of heat and temperature. *Journal of Baltic Science Education*, 22(1), 153-166. <https://doi.org/10.33225/jbse/23.22.153>
- Ugaldi, F., Dulnuan, A., Canuto, P. P., Hiteg, D., & Hiteg, N. (2025). Impact of 3D-printed models on elementary students' space science learning: Mixed methods and classroom action research study. *STEM Education*, 5(6), 1102-1131. <https://doi.org/10.3934/steme.2025047>
- Usembayeva, I., Kurbanbekov, B., Ramankulov, S., Batyrbekova, A., Kelesbayev, K., & Akhanova, A. (2024). 3D modeling and printing in physics education: The importance of STEM technology for interpreting physics concepts. *Qubahan Academic Journal*, 4(3), 45-58. <https://doi.org/10.48161/qaj.v4n3a727>
- Wekerle, C., Daumiller, M., Janke, S., Dickhäuser, O., Dresel, M., & Kollar, I. (2024). Putting ICAP to the test: How technology-enhanced learning activities are related to cognitive and affective-motivational learning outcomes in higher education. *Scientific Reports*, 14, Article 16289. <https://doi.org/10.1038/s41598-024-66069-y>
- Wörner, S., Kuhn, J., & Scheiter, K. (2022). The best of two worlds: A systematic review on combining real and virtual experiments in science education. *Review of Educational Research*, 92(6), 911-952. <https://doi.org/10.3102/00346543221079417>

APPENDIX A: PERCEIVED EFFECTIVENESS OF THREE-DIMENSIONAL MODELING QUESTIONNAIRE

Directions given to participants: Please read each statement carefully and indicate your level of agreement by selecting the number that best corresponds to your response. Use the scale provided below.

- 5 – Strongly agree
- 4 – Agree
- 3 – Neutral
- 2 – Disagree
- 1 – Strongly disagree

All 16 items are positively worded, so no item is reverse scored. Each dimension score is the mean of its four items and the overall score is the mean of the four dimension means.

Conceptual Understanding

1. 3D modeling improves my understanding of thermodynamics concepts.
2. 3D modeling helps me visualize abstract thermodynamics topics.
3. Using 3D models enhances my comprehension of heat transfer, energy, and entropy.
4. 3D modeling strengthens my ability to explain thermodynamics concepts clearly.

Engagement

5. I am more engaged in physics lessons when 3D modeling is used.
6. 3D modeling increases my participation during thermodynamics classes.
7. I pay more attention to thermodynamics discussions with 3D models.
8. 3D modeling makes thermodynamics learning more interactive for me.

Motivation

9. 3D modeling motivates me to study thermodynamics.
10. I feel more confident in my thermodynamics performance when using 3D models.
11. 3D modeling encourages me to perform better in thermodynamics tasks.
12. I am more interested in learning thermodynamics because of 3D modeling.

Ease of Learning

13. 3D modeling makes learning thermodynamics easier for me.
14. I understand thermodynamics lessons faster when 3D models are used.
15. 3D modeling simplifies difficult thermodynamics problems.
16. Using 3D models reduces my confusion in thermodynamics topics.

APPENDIX B: THERMODYNAMICS CONCEPTUAL TEST

Directions given to participants: Read each question carefully and encircle the letter of the best answer. Choose only one answer for each item.

Each correct answer scores 1 and each incorrect answer scores 0, giving a possible range of 0 to 19. Every item offers four options.

1. Which law states that energy cannot be created nor destroyed?
 - A. Zeroth law of thermodynamics
 - B. First law of thermodynamics
 - C. Second law of thermodynamics
 - D. Third law of thermodynamics
2. Which SI unit is used to measure heat energy?
 - A. Watt
 - B. Joule
 - C. Kelvin
 - D. Pascal
3. Which process transfers heat through direct contact?
 - A. Radiation
 - B. Convection
 - C. Conduction
 - D. Expansion
4. The degree of randomness or disorder in a thermodynamic system is measured by:
 - A. Energy
 - B. Temperature
 - C. Entropy
 - D. Pressure
5. A machine that converts thermal energy into mechanical motion is called a:
 - A. Refrigerator
 - B. Heat engine
 - C. Thermometer
 - D. Insulator
6. Temperature in the International System of Units is measured in:
 - A. Celsius
 - B. Kelvin
 - C. Fahrenheit
 - D. Joule
7. A perfectly insulated container that does not exchange energy or matter is classified as:
 - A. Open
 - B. Closed
 - C. Isolated
 - D. Thermal
8. The working substance inside most heat engines is usually:
 - A. Fuel
 - B. Gas
 - C. Water
 - D. Oil

9. An open system is best described as a system where:
 - A. Only energy enters the system
 - B. Only matter enters the system
 - C. Both matter and energy enter and leave the system
 - D. Nothing enters or leaves the system
10. When entropy increases in a system, it means that:
 - A. The system becomes more ordered
 - B. The system becomes more disordered
 - C. The temperature decreases
 - D. Energy is destroyed
11. A system absorbs 100 J of heat and performs 40 J of work. What is the change in internal energy?
 - A. 60 J
 - B. 140 J
 - C. 100 J
 - D. 40 J
12. A heat engine operates at 25% efficiency with 400 J heat input. What is the work output?
 - A. 100 J
 - B. 200 J
 - C. 300 J
 - D. 400 J
13. Which situation is an example of heat transfer by radiation?
 - A. Boiling water
 - B. Sun heating Earth
 - C. Touching metal
 - D. Air rising
14. A heat engine loses more heat than it converts. What does this indicate?
 - A. High efficiency
 - B. Low efficiency
 - C. No work
 - D. Perfect system
15. Which thermodynamic system does not exchange energy or matter with the surroundings?
 - A. Open
 - B. Closed
 - C. Isolated
 - D. Thermal
16. Which method is best to reduce heat loss in houses?
 - A. Use metal walls
 - B. Add insulation
 - C. Open windows
 - D. Increase airflow
17. Which thermodynamic system prevents both matter and energy transfer?
 - A. Open
 - B. Closed
 - C. Isolated
 - D. Leaking system

18. A candle heats water to produce steam that spins a small fan. What is the main energy source in this process?

- A. Electricity
- B. Heat
- C. Sound
- D. Light

19. Which situation demonstrates an application of thermodynamics in real life?

- A. Designing an efficient engine
- B. Measuring temperature
- C. Observing heat
- D. Reading a thermometer

Note. Content coverage of the final 19-item instrument: Heat engines and efficiency (items 5, 8, 12, 14); thermodynamic systems (items 7, 9, 15, 17); heat transfer mechanisms (items 3, 13, 16); laws of thermodynamics and internal energy (items 1, 11); entropy and the second law (items 4, 10); heat, temperature and units (items 2, 6); real-world applications (items 18, 19).