https://jsras.rcc.edu.ly/
Vol.3 No.2 (2026),197-202
Article history:
Received: 1 .Sep .2026
Accepted: 20.Sep . 2026
Published: 20 Oct .2026

Journal of Sustainable Research in Applied Sciences

PhET supported instruction and achievement in geometrical optics among Libyan ninth grade students

Mohamed Abugharsa[1],* and Ali M. Abdulshahed[2]
[1] Department of Information Systems, Faculty of Information Technology, Misurata University, Libya
[2] Department of Electrical and Electronic Engineering, Faculty of Engineering, Misurata University, Libya
* Corresponding author: m.abugharsa@it.misuratau.edu.ly

Index Terms
PhET; geometrical optics; physics education; simulation-based learning; Libya; ninth grade.

Abstract
This descriptive reassessment examines archived achievement summaries from a two-class quasi-experimental study of mirror and image formation in Libya. Thirty ninth-grade students received either guided PhET-supported instruction or conventional instruction over one academic semester, with 15 students per class. A specialist-reviewed 12-item test was administered before and after instruction. Only aggregate transformed-score summaries were available; individual scores, item responses, and the transformation rule are unavailable. Recorded means increased from 17.87 to 49.00 in the PhET class and from 16.53 to 20.87 in the comparison class. Subtraction of the displayed means gives gains of 31.13 and 4.34 points and a gain contrast of 26.79 points. The post-test gap is 28.13 points. Results are descriptive because class and intervention are inseparable. An internal audit identifies an unusually large standardized gap and opposing implied pre/post correlations. These unresolved features prevent validation of the magnitude of the recorded differences or attribution to PhET. Replication requires multiple classes and transparent individual and item-level records.

التدريس المدعوم بمحاكاة (PhET) والتحصيل في البصريات الهندسية لدى طلبة الصف التاسع في ليبيا

محمد أبوغرسة *[1]، علي عبد الشاهد [2]
[1]جامعة مصراتة ، كلية تقنية المعلومات، ليبيا
[2] جامعة مصراتة ,قسم الهندسة الكهربائية والإلكترونية، كلية الهندسة

الكلمات المفتاحية
البصريات الهندسية؛ تعليم الفيزياء؛ التعلمPhET بالمحاكاة؛ ليبيا؛ الصف التاسع

الملخص
تقدم هذه الدراسة إعادة فحص وصفية لملخصات التحصيل المحفوظة من دراسة شبه تجريبية عن المرايا وتكوّن الصور في ليبيا. شارك ثلاثون طالبًا من الصف التاسع في فصلين، ضم كل منهما خمسة عشر طالبًا، وتلقوا أو التدريس المعتاد خلال فصل دراسي واحد. طُبق اختبار من اثني عشر PhET تدريسًا موجهًا مدعومًا بمحاكاة بندًا راجعه مختصون قبل التدريس وبعده. لم تتوفر للمراجعة إلا ملخصات الدرجات المحوّلة؛ أما الدرجات الفردية واستجابات البنود ومعادلة التحويل فغير متاحة. ارتفع المتوسط المسجل لفصل المحاكاة من 17.87 إلى 49.00، وللفصل المقارن من 16.53 إلى 20.87. ويعطي طرح المتوسطات المعروضة تحسنًا قدره 31.13 و4.34 نقطة، وفرقًا بين التحسنين قدره 26.79 نقطة، بينما بلغ الفرق البعدي 28.13 نقطة. النتائج وصفية لتعذر فصل التدخل عن خصائص الفصل. وأظهر التدقيق فجوة معيارية كبيرة بصورة غير معتادة وارتباطين ضمنيين متعاكسين بين القياسين. تمنع هذه المشكلات غير المحسومة التحقق من حجم الفروق المسجلة أو نسبتها إلى المحاكاة. ويلزم0تكرار الدراسة بفصول متعددة ودرجات فردية وبيانات بنود موثقة .

1. INTRODUCTION

Learning geometrical optics requires students to connect ray diagrams with image position, orientation, and the behavior of an actual optical system. Difficulties with these connections can persist after instruction. Goldberg and McDermott (1987), for example, documented problems in applying ray diagrams to real images formed by converging lenses and concave mirrors. This makes mirror and image formation a useful topic for examining how students engage with visual representations.

Physics Education Technology (PhET) simulations provide interactive environments in which learners can vary parameters and inspect the resulting representations. Their educational design aims to connect observable phenomena with scientific models (Perkins et al., 2006; Wieman et al., 2008). However, the presence of a simulation does not by itself establish learning. Research on PhET and guided inquiry emphasizes the roles of task design, questioning, and instructional context (Banda & Nzabahimana, 2021; Srisawasdi & Panjaburee, 2015).

Relevant evidence exists in geometrical optics and in Arab educational settings. Ndihokubwayo et al. (2020) compared usual instruction with PhET- or video-supported optics teaching in Rwandan secondary schools. Uwamahoro et al. (2021) examined conceptual understanding of geometric optics, including persistent difficulties with lens images. In Morocco, Ben Ouahi et al. (2021) studied PhET-supported learning of Ohm's law, while Ben Ouahi et al. (2022) investigated teachers' views and implementation barriers. These studies establish relevant precedents, although they do not directly answer questions about ninth-grade mirror instruction in Libya.

The present article examines archived pre-test and post-test summaries for two Libyan classes taught over one academic semester. Its contribution is limited to documenting the recorded classroom pattern and assessing the internal consistency of the available statistics. Because

there is one class per condition and the scoring records cannot be reconstructed, the study cannot estimate an independent intervention effect or establish the validity of the unusually large recorded differences.

2. Literature Review

2.1 Evidence on PhET and physics learning

The literature includes favorable findings across several physics topics, but the designs and outcomes differ. Banda and Nzabahimana (2023) studied oscillations and waves among Malawian secondary students. Almadrones and Tadifa (2024) examined PhET-supported general physics instruction in the Philippines. Batuyong and Antonio (2018) investigated simulation activities in electromagnetism. These studies provide a rationale for investigating simulations, rather than a basis for assuming a particular result in Libyan optics classes.

Evidence is not uniformly strong. Yunzal and Casinillo (2020) described only slight improvement in an electrodynamics study. Fadillah et al. (2026) synthesized 47 effect sizes from 20 studies and reported an overall standardized difference of 0.83. Their synthesis supports examining contextual variation, but its pooled estimate cannot validate an individual dataset or justify an extreme local estimate. Differences in topics, instruments, and designs must be considered when comparing findings.

2.2 Guidance and learning with representations

Simulation-based inquiry can involve prediction, manipulation of variables, and comparison of representations. Srisawasdi and Panjaburee (2015) examined inquiry learning with formative assessment, while Ceberio et al. (2016) investigated interactive simulations in university physics problem solving. Earlier work also examined simulation use before laboratory inquiry and comparisons between physical and virtual experimentation (Zacharia & Anderson, 2003; Zacharia & Olympiou, 2011). These contributions concern broader instructional arrangements; they should not be treated as direct evidence that a particular mechanism operated in the present classes.

2.3 Prior work on optics and the regional context

Research on image formation shows why achievement in optics needs conceptual assessment. Goldberg and McDermott (1987) found that ray-tracing procedures did not necessarily imply an ability to explain an actual lens or mirror system. Ndihokubwayo et al. (2020) assessed geometric optics in 136 students across six Rwandan secondary schools, comparing usual instruction, PhET-supported instruction, and video-supported instruction. Their PhET condition used teacher presentation without student manipulation, an important difference from the intended small-group activities described here. Uwamahoro et al. (2021) also found that conceptual outcomes differed across optics topics and resources. A total achievement score alone therefore cannot identify which misconceptions changed.

Regional studies address different outcomes and designs. Ben Ouahi et al. (2021) examined Ohm's law achievement among 182 Moroccan middle-school students in six classes, three per condition. They reported higher post-test performance in the simulation condition. This provides a closer regional achievement comparison, although its electrical content and design limit transfer to Libyan optics instruction. Ben Ouahi et al. (2022), by contrast, surveyed 114 science teachers about use, perceived value, and barriers; those perceptions are implementation evidence rather than achievement measurements. The present contribution is a local descriptive record, not a claim that optics or Arab simulation research is absent.

2.4 Research questions and exploratory expectations

The research questions are: (1) What pre-test and post-test means and standard deviations are recorded for each class? (2) What within-class changes and between-class contrasts follow from those displayed means? (3) Are the archived statistics internally consistent enough to support interpretation of their magnitude?

Two directional expectations guide description: H1, both classes will have higher recorded post-test than pre-test means; and H2, the PhET class will have a larger recorded mean gain than the comparison class. These are exploratory expectations formulated for this reassessment, not preregistered confirmatory hypotheses. They are assessed descriptively; no null-hypothesis test is used to infer an intervention effect.

3. Methodology

3.1 Design and unit of comparison

The original classroom study used a nonrandomized pre-test/post-test comparison involving two intact ninth-grade classes. One received guided PhET-supported instruction and the other conventional instruction. The present reassessment uses only the surviving aggregate summaries. It is therefore reported as a descriptive analysis of a two-class quasi-experiment.

Instruction was allocated at class level, giving two clusters and only one cluster per condition. Students in a class share instructional and peer environments. Treating all 30 students as independent intervention replications would ignore this structure. Class and condition are inseparable, and teacher-related or scheduling influences cannot be disentangled from the instructional package. Neither an individual-level t test nor a multilevel model can identify a separate intervention effect from these two class summaries. Accordingly, no p values, intervention confidence intervals, or inferential effect-size claims are reported.

3.2 Participants and setting

The archived account identifies 30 ninth-grade students in a public school in Misurata Province, Libya, during the 2025/2026 academic year. Each intact class contained 15 students. Instruction extended over one academic semester. The available materials do not specify the number or length of lessons, exact testing dates, attendance, or whether the same teacher taught both

classes. Equivalence in instructional time, teacher input, and test administration therefore cannot be verified.

3.3 Content and instructional conditions

The unit addressed mirrors and image formation, including image position and orientation, real and virtual images, and ray diagrams. The archived description states that the PhET class used guided small-group activities: students varied object position, recorded observations on worksheets, discussed explanations, and received teacher clarification. The comparison class studied the same content using teacher explanation, board work, and standard lesson materials without PhET.

This description identifies the intended instructional sequence, not verified implementation fidelity. No observation records, interviews, interaction logs, or systematic fidelity measures are available. The simulation version, worksheet text, and student exposure time are not preserved in the materials examined. Software use, worksheets, group discussion, and teacher facilitation form a combined package whose components cannot be evaluated separately.

3.4 Achievement measure and data availability

The original account describes a 12-item multiple-choice achievement test reviewed by subject specialists for content relevance and grade suitability. Such review addresses content relevance but does not establish score reliability or construct validity. The available record contains group means, standard deviations, sample sizes, and earlier test statistics for transformed scores. It does not contain de-identified individual pre/post scores, item responses, a scoring key, the scale range, or the transformation rule.

Consequently, the scores cannot be interpreted as counts out of 12, percentages correct, or validated mastery levels. Reliability, item characteristics, ceiling effects, score distributions, missingness, and pre/post matching cannot be independently checked. The original mean and standard deviation summaries are retained as recorded; their accuracy is not established by repeating the calculations. Individual data and the transformation rule are unavailable, so a full reanalysis cannot be performed.

3.5 Descriptive calculations and consistency audit

For each class g, the mean gain G is calculated from the displayed post-test and pre-test means M using (1). The descriptive gain contrast is then calculated using (2), where E denotes the PhET class and C the comparison class. Both calculations use the rounded means in Table 1 and assume that the recorded scores share a comparable metric.

$$G_{\mathrm{g}} = M_{\mathrm{g,post}} - M_{\mathrm{g,pre}} \tag{1}$$

$$D_{\mathrm{G}} = G_{\mathrm{E}} - G_{\mathrm{C}} \tag{2}$$

The standardized post-test gap is inspected only as a scale diagnostic. The pooled within-class standard deviation is calculated using (3), with n denoting class size and s^2 each recorded post-test variance. Division of the post-test mean gap by this pooled standard deviation gives the descriptive ratio R in (4). This ratio is not an adjusted or validated intervention effect.

$$s_{\mathrm{p}} = \sqrt{\frac{(n_{\mathrm{E}}-1){s_{\mathrm{E}}}^2 + (n_{\mathrm{C}}-1){s_{\mathrm{C}}}^2}{n_{\mathrm{E}}+n_{\mathrm{C}}-2}} \tag{3}$$

$$\mathrm{R} = \frac{M_{\mathrm{E,post}} - M_{\mathrm{C,post}}}{s_{\mathrm{p}}} \tag{4}$$

To examine the paired summaries, the standard deviation of individual changes sΔ is reconstructed algebraically from the earlier paired t statistic using (5), where ΔM is that record's mean change and n = 15. The implied pre/post correlation r is then obtained from the variance identity in (6). The earlier t statistics are retained solely as audit inputs; they are not used for significance testing or intervention inference.

$$s_{\Delta} = \frac{\sqrt{n}\,|\Delta \mathrm{M}|}{|\mathrm{t}|} \tag{5}$$

$$\mathrm{r} = \frac{{s_{\mathrm{pre}}}^2 + {s_{\mathrm{post}}}^2 - {s_{\Delta}}^2}{2 s_{\mathrm{pre}} s_{\mathrm{post}}} \tag{6}$$

All derived values are approximate because the inputs are rounded. As a sensitivity check, each displayed two-decimal SD, paired mean difference, and absolute t value was allowed to vary by ±0.005, assuming conventional rounding to nearest. Conservative interval arithmetic applied to (5) and (6) bounds the implied correlations. These are rounding bounds, not confidence intervals, and they do not account for transcription, matching, or scoring errors. No individual scores are reconstructed or imputed.

3.6 Ethics and reporting

This reassessment reports only aggregate summaries and includes no student identifiers. The materials available do not document a formal ethics approval reference, school authorization, parental permission, or student assent procedures, so those procedures cannot be verified from the archive. No retrospective approval or consent claim is made. Any further use of student records must follow applicable institutional requirements and appropriate privacy safeguards.

4. RESULTS

4.1 Recorded achievement summaries

Table 1 presents the recorded sample sizes, means, and standard deviations. The pre-test difference was 1.34 points in favor of the PhET class; this does not establish baseline equivalence. At post-test the recorded mean gap was 28.13 points. These are comparisons between the two observed classes, conditional on the accuracy and comparability of the archived metric.

TABLE 1. Recorded pre-test and post-test summaries

Class	Test	n	Mean	SD
PhET	Pre-test	15	17.87	2.42
PhET	Post-test	15	49.00	2.85
Comparison	Pre-test	15	16.53	2.26
Comparison	Post-test	15	20.87	2.26

Note. SD = standard deviation. Values are reproduced from the archived summaries on an undocumented transformed metric. Its range and relationship to the 12-item test cannot be verified.

4.2 Changes derived from displayed means

Table 2 gives the arithmetic contrasts from (1) and (2). The PhET class's recorded mean gain was 31.13 points and the comparison class's was 4.34 points, giving a descriptive gain contrast of 26.79 points. The direction accords with H1 and H2 in these summaries; no hypothesis is accepted or rejected statistically.

TABLE 2. descriptive contrasts from the displayed means

Quantity	Calculation	Points
PhET mean gain	49.00 − 17.87	31.13
Comparison mean gain	20.87 − 16.53	4.34
Pre-test gap	17.87 − 16.53	1.34
Post-test gap	49.00 − 20.87	28.13
Gain contrast	31.13 − 4.34	26.79

Note. Gaps and the gain contrast are PhET minus comparison. All values use the displayed rounded means and are descriptive, without inferential uncertainty estimates.

The earlier paired summary reports a comparison-class gain of 4.33 points, yielding the previously stated gain contrast of 26.80 points. Subtraction of the displayed means gives 4.34 and 26.79 instead. The 0.01-point differences may reflect rounding from unreported precision, but this cannot be checked. The revised descriptive tables consistently use displayed-mean arithmetic; the audit below explicitly uses the earlier paired summary's own 4.33-point input.

4.3 Internal consistency and magnitude

Equations (3) and (4) give a pooled post-test SD of approximately 2.572 and a standardized gap of 10.94 within-class SDs. This exceptional ratio is a reason to investigate the measurement and records. It is not evidence of exceptional instructional effectiveness. A common positive linear transformation of all scores would preserve both this standardized ratio and the pre/post correlations; the fact that scores were transformed does not by itself explain the unusual pattern.

Table 3 shows the audit based on (5) and (6). The earlier paired summaries imply a pre/post correlation of approximately −0.787 in the PhET class and +0.879 in the comparison class. These are algebraic implications of the archived summaries, not correlations calculated from individual observations.

TABLE 3. Audit of the earlier paired summaries

Class	|ΔM|	|t|	Implied sΔ	Implied r
PhET	31.13	24.19	4.984	−0.787
Comparison	4.33	15.08	1.112	+0.879

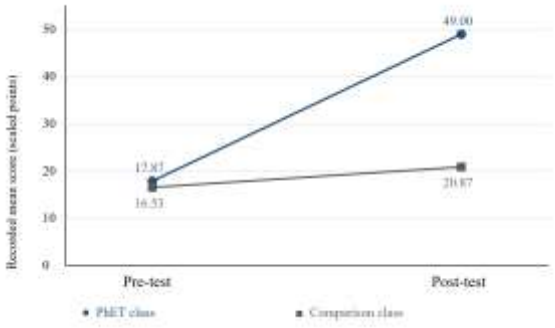
Note. n = 15 per class. Pre/post SD inputs are those in Table 1. Absolute paired mean differences and t values are archived audit inputs, not valid tests of the intervention. Derived values are approximate.

The opposing implied correlations are mathematically possible but unusual. Conditional on accurate paired summaries, the negative value indicates a negative linear association between individual pre-test and post-test scores in the PhET class. It does not show that every initially higher-scoring student subsequently scored lower. Pre/post matching, transcription, scoring, and consistency of the records and metric need checking. No explanation can be selected without the underlying data.

The rounding sensitivity check gives conservative correlation bounds of [−0.796, −0.779] for the PhET class and [0.870, 0.888] for the comparison class. The opposite signs therefore persist under the stated rounding assumption; two-decimal rounding alone cannot explain them. This calculation narrows one possible explanation but does not validate the archive or establish the cause of the pattern. The data-provenance concern remains unresolved.

4.4 Visual summary

Fig. 1 provides a visual summary of the four recorded means. The plot has no inferential error bars and contributes no independent evidence beyond Table 1.



Figer. 1. Recorded mean scores before and after instruction

5. DISCUSSION

5.1 Interpretation within the design

The archived summaries record increases in both classes, with a larger increase in the PhET class. This answers the descriptive questions about direction, but neither the source of the difference nor its magnitude is validated. With one class in each condition, the instructional package cannot be separated from class composition, classroom conditions, or teacher-related influences. The inference is further restricted by the undocumented scoring metric and the audit findings.

The 26.79-point gain contrast is an arithmetic comparison, not a causal difference-in-differences estimate. Baseline adjustment would not remove complete class-condition confounding, even if individual scores were recovered. The very large standardized gap and unusual implied correlations make stronger claims inappropriate until the records can be checked.

5.2 Hypotheses about possible learning processes

Several mechanisms could be investigated in a future study. One hypothesis is that manipulating object position while inspecting image changes helps students connect ray diagrams with spatial relationships. Another is that prediction and explanation prompts encourage students to compare their initial ideas with observations. A third is that peer discussion creates opportunities to articulate and revise explanations. Research on representations and inquiry provides a rationale for examining such processes (Ceberio et al., 2016; Srisawasdi & Panjaburee, 2015).

None of these mechanisms was measured here. There are no observation, interview, interaction-log, or fidelity data showing that students used the intended processes or that those processes accounted for their scores. Worksheets, teacher attention, and group work could also matter. The recorded differences therefore do not establish reduced representational ambiguity, peer correction, or any other proposed mediator.

5.3 Relationship to previous research

Prior PhET studies motivate further investigation but cannot verify the present archive. The recorded direction resembles favorable findings in some studies, including Banda and Nzabahimana (2023), while the extreme magnitude requires scrutiny. The standardized gap of 10.94 here is an unaudited ratio from two class summaries, not an effect estimate directly comparable with the pooled standardized result in Fadillah et al. (2026). Differences in design, scoring, and analysis prevent an interpretation that the present teaching package outperformed interventions in the wider literature. Yunzal and Casinillo (2020) also illustrate that reported outcomes are not uniformly large.

The optics literature emphasizes conceptual assessment (Goldberg & McDermott, 1987; Ndihokubwayo et al., 2020; Uwamahoro et al., 2021). Regional achievement research and teacher surveys likewise provide distinct kinds of evidence (Ben Ouahi et al., 2021; Ben Ouahi et al., 2022). These studies inform the design of future Libyan classroom research; they do not resolve this report's missing scoring information or class-condition confounding.

5.4 Implications for further classroom investigation

The descriptive record identifies questions for further classroom investigation. A future study could test the hypothesis that guided simulation activities improve students' coordination of ray diagrams and image properties when content, lesson time, worksheets, and assessment conditions are comparable across replicated classes. Observations, task logs, and explanation assessments would be needed to examine whether guidance and peer discussion contribute. These are hypotheses for testing, not mechanisms demonstrated by the present study. No conclusions about cost-effectiveness, laboratory substitution, or institutional investment follow from this archive.

5.5 Limitations

The main limitations concern design and data provenance. Two intact classes in one school provide no replication of either instructional condition. The small sample and undocumented lesson exposure further restrict generalization. Missing individual and item-level records prevent checks of matching, distributions, reliability, missing data, and scoring. The scale transformation is unknown, so absolute score meaning and comparability cannot be established.

The unusual standardized gap and opposing implied correlations remain unexplained. The rounding check excludes ordinary two-decimal rounding as the sole explanation under its stated assumptions; it does not rule out transcription, matching, or scoring problems. Removing inferential tests does not validate the summaries. There was no delayed assessment, and learning processes and fidelity were not measured. Ethics and consent documentation are also unavailable in the examined materials. The findings remain a provisional description of recorded class summaries.

6. CONCLUSION

The archived summaries show larger recorded end-of-instruction mean gains in the PhET class than in the comparison class. However, complete class-condition confounding, unavailable individual records and scoring rules, and unresolved statistical anomalies prevent validation of the magnitude or attribution of the differences to PhET. Ordinary rounding alone does not explain the opposing implied correlations under the assumptions examined. The report supports a need for a documented replication, not a conclusion about instructional effectiveness.

7. RECOMMENDATIONS

The immediate priority is a verifiable data record. Future studies should preserve de-identified linked pre/post scores, item responses, scoring keys, transformation rules, and an analysis log, alongside documented ethics and consent procedures. If original records are recovered, the extreme gap, student matching, and implied correlations should be checked before further inference.

Replication should include multiple classes per condition and, where feasible, random assignment of classes. Sample planning and analysis should account for clustering, with adequate replication across teachers and schools. Transparent instruments, equivalent instructional exposure, and a delayed post-test would strengthen assessment of achievement and retention. Observations or task logs would allow hypotheses about guidance and peer explanation to be tested. Classroom adoption decisions should draw on this broader evidence rather than the magnitude reported in the present archive.

8. REFERENCES

Almadrones, R. D. G., & Tadifa, F. G. (2024). Physics educational technology (PhET) simulations in teaching general physics 1. *International Journal of Instruction, 17*(3), 635–650. https://doi.org/10.29333/iji.2024.17335 a

Banda, H. J., & Nzabahimana, J. (2021). Effect of integrating physics education technology simulations on students' conceptual understanding in physics: A review of literature. *Physical Review Physics Education Research, 17*(2), Article 023108. https://doi.org/10.1103/PhysRevPhysEducRes.17.023 10 8

Banda, H. J., & Nzabahimana, J. (2023). The impact of Physics Education Technology (PhET) interactive simulation-based learning on motivation and academic achievement among Malawian physics students. *Journal of Science Education and Technology, 32*(1), 127–141. https://doi.org/10.1007/s10956-022-10010- 3

Batuyong, C. T., & Antonio, V. V. (2018). Exploring the effect of PhET® interactive simulation-based activities on students' performance and learning experiences in electromagnetism. *Asia Pacific Journal of Multidisciplinary Research, 6*(2), 121–131. http://www.apjmr.com/wp- content/uploads/2018/02/APJMR-2017.6.2.15.pd f

Ben Ouahi, M., Ait Hou, M., Bliya, A., Hassouni, T., & Al Ibrahmi, E. M. (2021). The effect of using computer simulation on students' performance in teaching and learning physics: Are there any gender and area gaps? *Education Research International, 2021*, Article 6646017. https://doi.org/10.1155/2021/664601 7

Ben Ouahi, M., Lamri, D., Hassouni, T., & Al Ibrahmi, E. M. (2022). Science teachers' views on the use and effectiveness of interactive simulations in science teaching and learning. *International Journal of Instruction, 15*(1), 277–292. https://doi.org/10.29333/iji.2022.15116 a

Ceberio, M., Almudí, J. M., & Franco, Á. (2016). Design and application of interactive simulations in problem-solving in university-level physics education. *Journal of Science Education and Technology, 25*(4), 590–609. https://doi.org/10.1007/s10956-016-9615- 7

Fadillah, M. A., Alawyah, K., Syafrijon, S., & Usmeldi, U. (2026). Meta-analysis of the effectiveness of PhET simulations in physics education. *International Journal of Online and Biomedical Engineering (iJOE), 22*(2), 21–37. https://doi.org/10.3991/ijoe.v22i02.5900 7

Goldberg, F. M., & McDermott, L. C. (1987). An investigation of student understanding of the real image formed by a converging lens or concave mirror. *American Journal of Physics, 55*(2), 108–119. https://doi.org/10.1119/1.1525 4

Ndihokubwayo, K., Uwamahoro, J., & Ndayambaje, I. (2020). Effectiveness of PhET simulations and YouTube videos to improve the learning of optics in Rwandan secondary schools. *African Journal of Research in Mathematics, Science and Technology Education, 24*(2), 253–265. https://doi.org/10.1080/18117295.2020.181804 2

Perkins, K., Adams, W., Dubson, M., Finkelstein, N., Reid, S., Wieman, C., & LeMaster, R. (2006). PhET: Interactive simulations for teaching and learning physics. *The Physics Teacher, 44*(1), 18–23. https://doi.org/10.1119/1.215075 4

Srisawasdi, N., & Panjaburee, P. (2015). Exploring effectiveness of simulation-based inquiry learning in science with integration of formative assessment. *Journal of Computers in Education, 2*(3), 323–352. https://doi.org/10.1007/s40692-015-0037- y

Uwamahoro, J., Ndihokubwayo, K., Ralph, M., & Ndayambaje, I. (2021). Physics students' conceptual understanding of geometric optics: Revisited analysis. *Journal of Science Education and Technology, 30*(5), 706–718. https://doi.org/10.1007/s10956-021-09913- 4

Wieman, C. E., Adams, W. K., & Perkins, K. K. (2008). PhET: Simulations that enhance learning. *Science, 322*(5902), 682–683. https://doi.org/10.1126/science.116194 8

Yunzal, A. N., Jr., & Casinillo, L. F. (2020). Effect of Physics Education Technology (PhET) simulations: Evidence from STEM students' performance. *Journal of Education Research and Evaluation, 4*(3), 221–226. https://doi.org/10.23887/jere.v4i3.2745 0

Zacharia, Z., & Anderson, O. R. (2003). The effects of an interactive computer-based simulation prior to performing a laboratory inquiry-based experiment on students' conceptual understanding of physics. *American Journal of Physics, 71*(6), 618–629. https://doi.org/10.1119/1.156642 7

Zacharia, Z. C., & Olympiou, G. (2011). Physical versus virtual manipulative experimentation in physics learning. *Learning and Instruction, 21*(3), 317–331. https://doi.org/10.1016/j.learninstruc.2010.03.00 1