2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #51217

# Incorporating Physical Tools, Simulator, and Machine Learning for Teaching Parallax Distance-Based Measurements in High School Astronomy

**Ryan Parham, College Station High School, College Station ISD**

Ryan Parham is a high school science teacher at College Station High School in College Station, Texas, where he teaches physics and astronomy. He has ten years of experience teaching secondary science, including courses in physics, astronomy, earth and space science, and chemistry. He has been with College Station ISD for six years. He holds a Bachelor of Science degree from Texas A&M University. His instructional interests include hands-on learning, simulation, and data-driven approaches to improving student understanding of complex scientific concepts.

**Dr. Sheng-Jen Hsieh, Texas A&M University**

Dr. Sheng-Jen (”Tony”) Hsieh is a Professor in the Department of Engineering Technology and Industrial Distribution and a member of the Graduate Faculty at Texas A&M University, College Station, TX. His research interests include automation, robotics, cyber-manufacturing and Industry 4.0; optical/infrared imaging and instrumentation; micro/nano manufacturing; and design of technology for engineering education. He is also the Director of the Rockwell Automation Laboratory at Texas A&M University, a state-of-the-art facility for education and research in the areas of automation, robotics, and Industry 4.0 systems. He was named Honorary International Chair Professor for National Taipei University of Technology in Taipei, Taiwan, for 2015-21. Dr. Hsieh received his Ph.D. in Industrial Engineering from Texas Tech University, Lubbock, TX.

# Incorporating Physical Tools, Simulator, and Machine Learning for Teaching Parallax-based Distance Estimation Measurements in High School Astronomy

Abstract

Understanding how astronomers measure stellar distances is critical for high school students' comprehension of space, yet traditional lessons often rely on simplified tools that do not reflect how these methods are used in modern research. This project introduces a novel high school astronomy instructional approach that combines physical tools, a pixel-based simulation, and a machine learning extension to better align classroom practice with contemporary scientific methods. While prior instructional models have improved conceptual understanding through hands-on activities and geometric calculations, they often stop short of incorporating digital analysis or computational thinking. This study addresses that gap by guiding students through manual parallax measurements, simulated image-based calculations, and algorithmic distance prediction using real and simulated data. The lesson unfolds over seven days and includes a physical lab, a browser-based simulation, and a teacher-guided machine learning activity. Evaluation results suggest group-level increases in accuracy in distance estimation, a shift toward image-based reasoning in students' models, and greater confidence with both traditional and computational tools.

## 1 Introduction

Astronomy presents a unique instructional challenge: its core phenomena distant stars, vast galaxies, and the immense scale of space are inherently abstract and inaccessible to direct experience [1]. One of the most foundational skills in astronomy is estimating stellar distances, a process formalized in the Cosmic Distance Ladder [2]. At the base of this ladder lies geometric parallax, which provides the first rungs for determining distances to nearby stars.

Classroom instruction on parallax often relies on simplified analogues, such as protractors and fixed baselines. While these models effectively introduce geometric principles, they diverge significantly from how modern astronomy applies parallax—namely, through pixel-based image analysis and computational modeling [3, 4]. For instance, the European Space Agency's Gaia mission measures stellar centroids across multiple epochs and computes sub-pixel shifts to derive parallax angles and distances [2]. This disconnect risks limiting students' understanding of the concept's real-world relevance and may inadvertently misrepresent the scientific process.

To address this gap, we propose a hybrid instructional approach that combines three components: (1) a student-designed physical measurement lab, (2) a custom-built web-based pixel simulation, and (3) a machine learning extension. These components are scaffolded to align with Texas Essential Knowledge and Skills (TEKS) standards related to modeling, data analysis, and computational reasoning. Beyond meeting curricular requirements, this integrated approach introduces students to the tools and workflows they are likely to encounter in college-level science and engineering environments.

This approach addresses several important gaps in high school astronomy instruction by having the students learn methods for taking data in space and their associated limitations. The simulation, along with the machine learning integration afterwards, provides context for ensuing classroom discussions on studying galaxies, stars, and our own solar system. By engaging students in authentic practices, collecting physical data, measuring pixel displacement, and analyzing predictions from a neural network, this study cultivates both conceptual mastery and future-ready skills. The layered design encourages students to experience astronomy as a dynamic, data-driven science, rather than a static collection of formulas.

## 2 Existing Work

Prior efforts to teach astronomical parallax in secondary and undergraduate classrooms have largely centered on hands-on activities and geometric demonstrations. For example, Ladd and Nottis [5] reported significant learning gains from a hybrid lab involving both physical observation and use of the WorldWide Telescope (WWT) interface. Pössel [6] developed a classic in-class activity using protractors and fixed baselines to reinforce spatial reasoning and introduce trigonometric parallax. These approaches are effective in supporting conceptual understanding of geometric relationships, but they depart from the digital and computational workflows employed in contemporary astronomy.

Recent work in the field underscores the increasing integration of pixel-based image analysis and machine learning. Leung and Bovy [7] applied deep learning models to refine parallax estimates using Gaia DR2 data, while Shen and Luo [8] explored regression models for predicting stellar parameters. Despite the widespread use of such methods in professional research, they remain almost entirely absent from K–12 astronomy instruction. This disconnect is a missed opportunity to build students' data literacy and introduce them to emerging tools in scientific analysis; a concern noted by Fluke and Jacobs [9] and echoed in calls for early AI education in schools [10].

The proposed lesson builds on and responds to these developments by integrating two significant innovations not commonly seen in classroom-based parallax instruction. First, students transition from physical measurement to a browser-based simulation that uses pixel displacement to calculate distance. Second, they engage with a supervised machine learning model, implemented via a neural network (MLPRegressor), that analyzes and predicts object distance using real and simulated data. This scaffolded integration introduces students to core concepts in computational modeling, without requiring programming experience.

To better contextualize the strengths and limitations of prior instructional models, four key examples from the literature are summarized and directly compared to the proposed lesson. These include: Ladd and Nottis [5], who combine physical and virtual labs; Pössel [6], who use classic protractor-based instruction; Newland [11], who introduce web-based data science platforms; and Nora's Guide to the Galaxy [12], a tutorial using pixel-shift methods with external software.

Prior efforts to teach parallax-based distance estimation estimation at the secondary and undergraduate levels have emphasized hands-on geometric demonstrations and guided software exploration. For example, Ladd and Nottis [5] combined physical measurement with visualization tools and reported measurable gains in conceptual understanding, while Pössel [6] developed structured protractor-based activities to reinforce spatial reasoning and trigonometric relationships. These approaches are effective at establishing geometric foundations but typically stop short of integrating pixel-based image analysis or computational modeling.

More recent educational resources incorporate digital tools and data-science platforms that expose students to authentic datasets. Newland's data-science platform offers broad functionality but can diffuse focus when instructors wish to target parallax specifically, and multimedia tutorials on pixel-shift techniques often require external software and manual counting rather than an embedded, streamlined simulation [11].

The present study builds on these prior works by sequencing experiences from physical geometry to pixel-based simulation and then to supervised machine learning. Unlike fixed-protocol lessons, students design their own physical procedures to promote ownership. The custom web simulation narrows attention to parallax-based distance estimation estimation, reducing cognitive load. Finally, integrating a teacher-guided machine learning activity introduces students to algorithmic prediction and model evaluation, thereby extending classroom practice toward contemporary research workflows.

Prior efforts to teach parallax-based distance estimation estimation at the secondary and undergraduate levels have emphasized hands-on geometric demonstrations and guided software exploration. For example, Ladd and Nottis [5] combined physical measurement with visualization tools and reported measurable gains in conceptual understanding, while Pössel [6] developed structured protractor-based activities to reinforce spatial reasoning and trigonometric relationships. These approaches are effective at establishing geometric foundations but typically stop short of integrating pixel-based image analysis or computational modeling.

More recent educational resources incorporate digital tools and data-science platforms that expose students to authentic datasets. Newland's data-science platform offers broad functionality [11] but can diffuse focus when instructors wish to target parallax specifically, and multimedia tutorials on pixel-shift techniques often require external software and manual counting rather than an embedded, streamlined simulation.

The present study builds on these prior works by sequencing experiences from physical geometry to pixel-based simulation and then to supervised machine learning. Unlike fixed-protocol lessons, students design their own physical procedures to promote ownership. The custom web simulation narrows attention to parallax-based distance estimation estimation, reducing cognitive load. Finally, integrating a teacher-guided machine learning activity introduces students to algorithmic prediction and model evaluation, thereby extending classroom practice toward contemporary research workflows.

## 3 Objectives

This work addresses a critical instructional gap by extending students' engagement beyond manual measurement into algorithmic prediction and data-informed analysis. In contrast to traditional parallax lessons that end with protractor-based calculations, the proposed curriculum integrates a hands-on physical lab, a pixel-based simulation modeled after professional workflows, and a supervised machine learning extension.

Together, these components aim to provide students with a more authentic understanding of how distances are measured in modern astronomy. The instructional sequence is designed to scaffold both conceptual understanding and data literacy while aligning with computational and modeling standards outlined in the Texas Essential Knowledge and Skills (TEKS).

## 4 Instructional Components

This study is structured around three core instructional components presented in sequence: a physical parallax lab, a browser-based pixel simulation, and a machine learning extension. Each component builds upon previous learning to deepen students' understanding of distance estimation while introducing increasingly sophisticated scientific tools.

### 4.1 Physical Tools

The first component is a hands-on investigation into angular parallax and trigonometric reasoning. Students work individually to design a method for estimating the distance to a fixed object located in front of the school. Using simple tools, such as a protractor or compass, measuring tape (or calibrated pacing), and a calculator, students apply geometric principles to construct their own experimental protocol. Before collecting data, students complete a brief conceptual pre-lab addressing the relationship between apparent motion and distance, as well as the limitations of parallax on Earth.

Students then execute their procedures outdoors by measuring angular displacement from two fixed observation points equidistant from a central reference. Using their recorded angles and baseline distances, they calculate the object's estimated position through basic trigonometry (Figure 1). Afterwards, they compute percent error based on the actual distance and submit both a diagram of their setup and a written analysis of their results. This activity provides an experiential foundation for understanding how geometric parallax functions and highlights the constraints of manual measurement techniques.

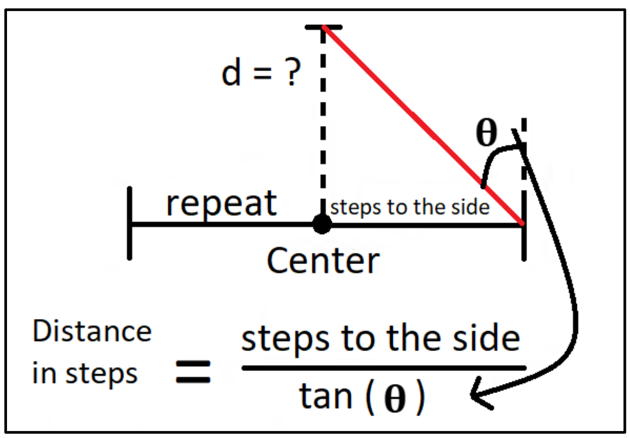



Figure 1. Formula for determining distance

### 4.2 Simulation

After completing the physical parallax lab, students transition to a browser-based simulation designed to replicate astronomical measurement through pixel-based analysis. The simulation, developed using JavaScript and the Three.js library, provides an interactive 3D environment where students emulate photographing an object from two viewpoints to observe parallax displacement. The simulation is hosted on a publicly accessible server and designed as a fully client-side application, allowing for seamless classroom use, offline accessibility, and compatibility with a variety of devices and browsers.

The custom-built simulation is intentionally streamlined to support focused, concept-driven learning. Its interface allows students to manipulate virtual camera positions, capture two-angle images, and calculate parallax-based distance estimation estimations using pixel displacement—all within a simplified 3D environment. Designed for accessibility and minimal cognitive load, the simulation removes extraneous controls while offering just enough interactivity to mirror real-world astronomical workflows. This balance enables students to engage meaningfully with parallax analysis without being distracted or overwhelmed by unfamiliar software features. A visual overview of this interface is shown in Figure 2, and Table 1 summarizes each tool's function.

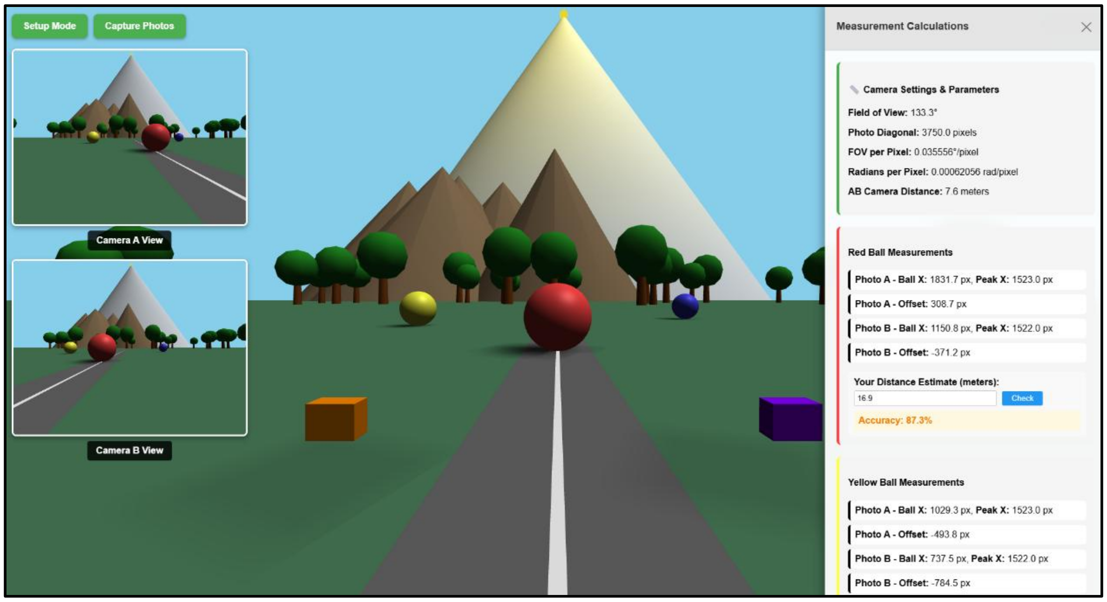
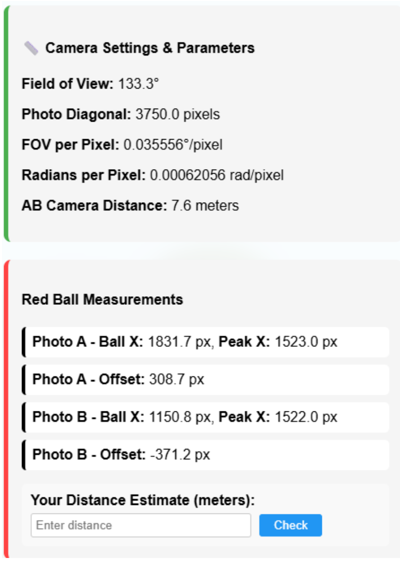
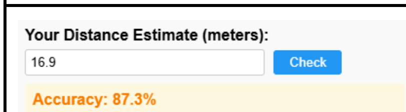



Figure 2. Custom simulation created for teaching parallax showing interface layout used during classroom instruction.

Table 1. Student-Facing Simulation Tools and Functions

| User Interface | Label | Function |
|---|---|---|
|   | Setup mode | Used to move the objects and the cameras. |
|   | Capture Photos | Used to generate two images, one from each camera at the same moment. For the images to be considered usable, the desired object must be visible in both images. |
|   | Calculate Results | Generates the needed information to calculate the distance of each object. |

<table>
<tr><th>User Interface</th><th>Label</th><th>Function</th></tr>
<tr><td>
</td><td>Camera Settings & Parameters</td><td>This screen appears when the calculate results button has been pressed. It shows the settings that are both needed to calculate the distance, and the background ones already used in the other generated information.</td></tr>
<tr><td>
</td><td>Radians per Pixel, AB Camera Distance, Offsets for each image.</td><td>These are the required pieces of information needed to calculate the distance of the object. Information on all three objects is generated simultaneously, even if a student is only working on one object. The students for this lab will be using Google Sheets with the calculations built in, and they just need to paste in the required information.</td></tr>
<tr><td colspan="2">$$\frac{\text{(AB Camera Distance)}}{\tan\left(\left(\text{(Photo A-Offset)} - \text{(Photo B-Offset)}\right) * \text{(Radians per Pixel)}\right)}$$</td><td>This is the formula required based on the provided information and background calculations already completed by the program, if not using a Google Sheets.</td></tr>
<tr><td>
</td><td>Your Distance Estimate (meters)</td><td>Students type in their answer and click the Check button. The simulation then compares their answer to the hidden correct answer. While the student can guess repeatedly until they find the correct distance, this information is not useful for the Machine Learning aspect of the lab and is discouraged.</td></tr>
</table>

The simulation experience follows a four-phase workflow: (1) selecting the setup, (2) capturing images, (3) calculating pixel displacement, and (4) comparing results to known values (see Table 2). Students begin by configuring the virtual scene in Setup Mode, using a top-down view to drag and drop one of three colored target spheres and adjust the baseline distance between Camera A and Camera B. Once configured, students capture two simulated images, one from each camera's perspective, by clicking the "Capture Photos" button.

Behind the scenes, the application positions the virtual camera at each marker, renders the 3D scene, and generates two static HTML5 canvas images. To calculate parallax displacement, the simulation uses built-in projection functions to convert each object's 3D world coordinates into 2D screen coordinates. This yields precise pixel positions for both views, allowing for accurate measurement of pixel shift between images. Students record these values in an integrated Google Sheet that automates the distance calculation process using radians-per-pixel conversions.

To discourage guessing and preserve data validity for the machine learning phase, students are prompted to record their calculated distances before checking against hidden true values. This structured interaction mirrors scientific inquiry and supports reflective reasoning.

Table 2. Summary of Simulation Workflow

| Phase | Student Task | Input | Output |
|---|---|---|---|
| Setup | Select camera angle & baseline | Angle, distance | Configured Scene |
| Capture | Identify object pixel shift between views | Pixel Measurements | Observed Parallax shift |
| Calculate | Enter pixel data into Google Sheet calculator | Pixel Values | Estimated object distance |
| Compare | Compare hidden known with calculated distance. | Student data | Accuracy |

## 4.3 Machine Learning

The final instructional phase introduces students to supervised machine learning through a guided neural network activity. After completing the simulation, students export a dataset consisting of five measurements, each capturing pixel offset and baseline distance for different object positions. These data are then analyzed using a teacher-led Google Colab notebook that demonstrates regression-based modeling.

The notebook implements a feedforward neural network using the MLPRegressor class from the scikit-learn library. The neural network model architecture includes two hidden layers of 100 neurons each, using ReLU activation and the Adam optimizer (See Figure 3). Input features include the AB camera distance and the angular offset (converted from pixels to radians), and the output is a predicted object distance. This machine learning model was chosen over simpler alternatives, such as linear regression, due to its ability to capture complex, non-linear relationships between pixel offset and distance—an approach shown to improve accuracy in parallax estimation by Leung and Bovy [7].

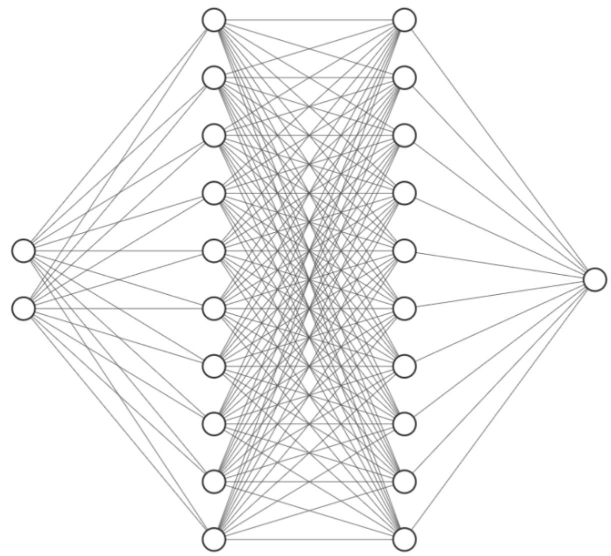


Figure 3. ANN Topology 2 x 100 x 100 x 1

## 5 Implementation and Evaluation Methodology

### 5.1 Participants

A lesson plan utilizing the three components was delivered and evaluated within three intact sections of a high-school astronomy elective course at a suburban public high school. Each section enrolls 25–35 juniors and seniors (ages 16–18). A total of 68 students participated in the evaluation.

The lesson is embedded within a broader unit on stellar measurement and observational methods. All participants had completed prerequisite units on basic geometry, measurement, and scientific inquiry. Instruction takes place in the school's computer lab (with reliable internet and individual workstations) and on school grounds suitable for outdoor measurements. The instructor followed a detailed pacing guide and used identical materials and environments for all sections.

### 5.2 Implementation

The lesson is structured over seven 50-minute class periods and includes three integrated instructional phases:

1. **Lecture and Physical Parallax Lab (Days 1–3):** Students engage in traditional parallax measurement using physical tools (e.g., protractors).
2. **Pixel-Based Computer Simulation (Days 4–5):** Students use a custom-built web simulation to estimate distances via pixel displacement.
3. **Machine Learning Extension (Days 6–7):** Students compare their calculated data to machine-generated predictions using a neural network.

The materials used can be found in Appendix A and a breakdown of the lesson by minute can also be found in Appendixes B and C.

### 5.3 Data Collection

To evaluate learning gains, students complete three identical assessments at designated points as shown in Table 3. Note that because the lesson implementation and evaluation activities took place over seven class periods, the number of students who completed the evaluation instruments varied from 62 to 68.

Table 3. Data Collection

| Assessment | Timing | Instrument |
|---|---|---|
| Conceptual Model Prompt (n=68) | Pre-instruction (Day 1) | "Draw a model illustrating how astronomers determine distances to nearby stars." |
| Mid-point Check (n=62) | Day 4 | Same model prompt |
| Post-instruction (n=64) | Day 7 | Same model prompt |
| Procedural Accuracy | Day 3 and Day 5 | Percent-error from physical lab calculations and simulation accuracy |
| Self-Efficacy Survey | Day 1 (n=63) and Day 7 (n=62) | 7-point Likert items on confidence with each method |

### 5.4 Rubric-Based Evaluation

Student understanding of parallax measurement and astronomical distance estimation was assessed using a combination of written responses, opinion surveys, lab submissions, and teacher observations. At three points in the lesson, before, during, and after instruction, students respond to an open-ended prompt: *"*Develop a model on how astronomers determine the distances to nearby stars." Each response included a brief written explanation and a visual model incorporating diagrams, arrows, calculations, and annotations. Responses were evaluated using a structured rubric (Table 4) that scored both the distance estimation method and the quality of the model. Base scores were assigned based on whether students used an incorrect method, a protractor-based method, or an image-analysis approach. These scores were then modified by a quality multiplier that reflects clarity, accuracy, and completeness. This system captures both conceptual understanding and the effectiveness of students' scientific communication.

In addition to model evaluations, students completed self-efficacy surveys to track their confidence in using different measurement strategies. Items are rated on a 7-point Likert scale and include statements such as: "I can explain how to calculate distances using photographs," and "I have mastered parallax-based distance estimation estimation measurements as described in this class." These surveys provide insight into students' self-perceived growth and comfort with both traditional and computational tools.

Student worksheets, group reflections, and lab data submissions serve as additional evidence of engagement and procedural understanding. Throughout the instructional sequence, the teacher also collects informal observational notes related to collaboration, questioning, and the emergence of misconceptions. Together, these mixed methods provide a comprehensive and triangulated view of student learning, allowing for analysis of shifts in reasoning from geometric calculations to image-based or algorithmic models and identifying where deeper support may be needed.

Table 4. Model Grading Rubric

| **Method used to calculate distance** | **Max Points** | **High Quality Model**<br>**100% of Max Points** | **Medium Quality Model**<br>**66% of Max Points** | **Low Quality Model**<br>**33% of Max Points** | **Points Earned** |
|---|---|---|---|---|---|
| Incorrect Method | 6 points | Clearly understandable model with high efficiency use of space. Contains the expected arrows, explanations, and mathematical calculations that completely encapsulate and expand upon their two-sentence summary. | Understandable model that fairly accurately clarifies their two-sentence summary. Fair use of the space provided but lacks quality or contains a few inaccurate statements. | Model does not match the two-sentence summary of their topic, or it is incomplete, or it has several highly inaccurate statements that take away from the understanding of the model. | ___/6 |
| Protractor-Based Method | 12 points | | | | ___/12 |
| Image Analysis Method | 18 points | | | | ___/18 |

## 5.5 Statistical Analysis

Due to absences and logistical limitations during data collection, student responses could not be unambiguously matched across pre-, mid-, and post-assessment timepoints. Survey instruments were recorded in order of submission rather than by unique student identifier, and some students did not complete all instruments. As a result, repeated measures or paired analyses could not be validly conducted.

To avoid invalid within-subject assumptions, analyses were conducted at the group level, treating each timepoint as an independent sample. Descriptive statistics (n, mean, standard deviation, and 95% confidence intervals) were calculated for each measure. Pre- and post-survey comparisons were analyzed using Welch's independent samples t-tests to account for unequal sample sizes and potential heterogeneity of variance. For objective test scores (pre, mid, post), pairwise Welch t-tests were conducted. Effect sizes were calculated using Hedges' g to provide a bias-corrected estimate appropriate for unequal sample sizes. Statistical significance was evaluated at $\alpha = .05$.

# 6 Results

Group-level descriptive statistics and inferential comparisons are presented in Table 5.

Table 5. Group-Level Descriptive Statistics and Independent-Sample Comparisons

| **Measure** | **Time point** | **n** | **Mean** | **SD** | **95% CI** | **Comparison** | **Welch t** | **P** | **Hedges' g** |
|---|---|---|---|---|---|---|---|---|---|
| Q1 Self-Efficacy | Pre | 63 | 2.14 | 1.45 | [1.78, 2.51] | Pre vs Post | 12.16 | < .001 | 2.16 |
| | Post | 62 | 5.03 | 1.37 | [4.62, 5.44] | | | | |
| Q2 Self-Efficacy | Pre | 63 | 2.89 | 1.50 | [2.51, 3.27] | Pre vs Post | 8.63 | < .001 | 1.53 |
| | Post | 62 | 5.05 | 1.29 | [4.72, 5.37] | | | | |
| Q3 Self-Efficacy | Pre | 63 | 3.17 | 1.58 | [2.78, 3.57] | Pre vs Post | 7.27 | < .001 | 1.29 |
| | Post | 62 | 5.13 | 1.42 | [4.77, 5.49] | | | | |
| Q4 Self-Efficacy | Pre | 63 | 2.62 | 1.34 | [2.30, 2.94] | Pre vs Post | 9.73 | < .001 | 1.72 |
| | Post | 62 | 5.26 | 1.19 | | | | | |
| Test Score | Pre | 68 | 4.50 | 1.57 | [4.12, 4.88] | Pre vs Mid | −18.21 | < .001 | 3.21 |
| | Mid | 62 | 10.11 | 1.91 | [9.63, 10.60] | Pre vs Post | −25.08 | < .001 | 4.42 |
| | Post | 64 | 15.09 | 3.02 | [14.34, 15.85] | Mid vs Post | −11.11 | < .001 | 1.95 |

Substantial increases were observed across all four self-efficacy items from pre- to post-assessment. For Question 1, mean scores increased from 2.14 ($SD = 1.45$) to 5.03 ($SD = 1.37$), a statistically significant difference, $t = 12.16$, $p < .001$, with a very large effect size ($g = 2.16$).

For Question 2, mean scores increased from 2.89 ($SD = 1.50$) to 5.05 ($SD = 1.29$), $t = 8.63$, $p < .001$, $g = 1.53$.

For Question 3, mean scores increased from 3.17 ($SD = 1.58$) to 5.13 ($SD = 1.42$), $t = 7.27$, $p < .001$, $g = 1.29$.

For Question 4, mean scores increased from 2.62 ($SD = 1.34$) to 5.26 ($SD = 1.19$), $t = 9.73$, $p < .001$, $g = 1.72$.

Objective test scores increased significantly across timepoints. Pre-test scores ($M = 4.50$, $SD = 1.57$) increased to 10.11 ($SD = 1.91$) at mid-assessment and to 15.09 ($SD = 3.02$) at post-assessment.

Pairwise comparisons revealed statistically significant differences between all timepoints: pre vs. mid ($t = -18.21$, $p < .001$, $g = 3.21$), pre vs. post ($t = -25.08$, $p < .001$, $g = 4.42$), and mid vs. post ($t = -11.11$, $p < .001$, $g = 1.95$). These effect sizes indicate extremely large improvements in group-level objective performance over the duration of the intervention.

Figures 4 to 8 presents results from the pre-, mid-, and post-instruction assessments and the self-efficacy surveys. Data from the student worksheets, group reflections, lab data submissions, and teacher notes are still being reviewed.

These results show a clear overall group-level trend of improvement in learners' understanding of how astronomers determine distances to nearby stars as the lesson progresses from the Physical Parallax Lab to the Pixel-Based Computer Simulation to the Machine Learning Extension. Because assessment data could not be reliably matched to individual students across timepoints, these findings represent changes at the cohort level rather than within-person gains.

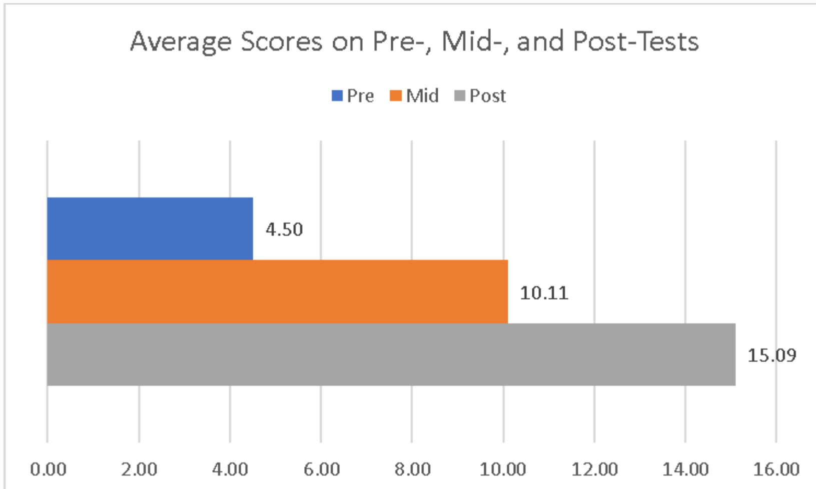



Figure 4. Average scores on pre-, mid, and post-tests administered on Day 1 (n=68), Day 4 (n=62), and Day 7 (n=64), respectively


Figure 5. Average scores on pre- and post- self-efficacy survey question "I have mastered parallax-based distance estimation estimation measurements as described in this class."

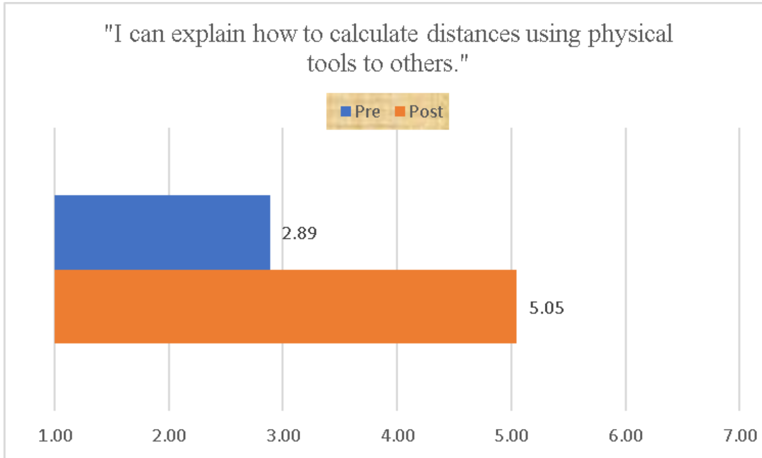



Figure 6. Average scores on pre- and post- self-efficacy survey question
“I can explain how to calculate distances using physical tools to others.”

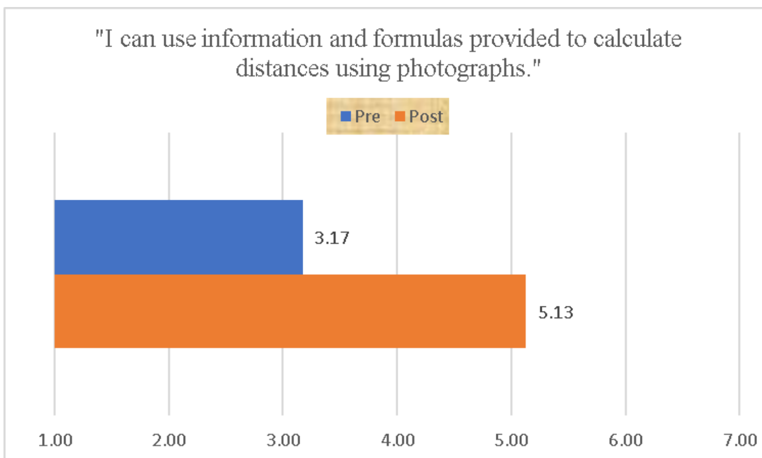



Figure 7. Average scores on pre- and post- self-efficacy survey question
“I can use information and formulas provided to calculate distances using photographs.”

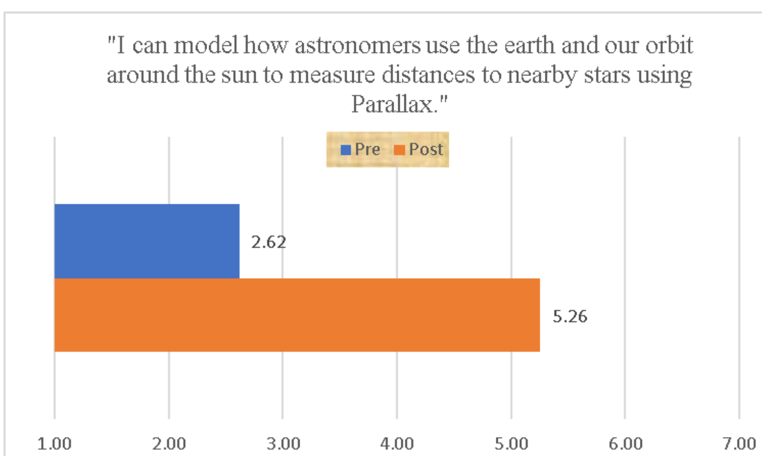



Figure 8. Average scores on pre- and post- self-efficacy survey question
“I can model how astronomers use the earth and our orbit around the sun

to measure distances to nearby stars using parallax."

## 7 Potential Impacts, Conclusions, and Future Directions

This study did not include a separate lecture-only control group, and individual assessment records could not be unambiguously matched across timepoints. Because of these constraints, we interpret the reported statistical results as indicators of group-level change rather than definitive within-student or strictly causal effects. Future research should include matched longitudinal data and, where possible, comparison/control conditions to strengthen causal claims.

Results suggest that the scaffolded sequence is associated with shifts from physical to image-based reasoning and increases in group-level measures of conceptual understanding and confidence. While these associations are consistent with the instructional design, caution is warranted: without matched within-student comparisons and a control condition, the evidence supports cohort-level change rather than causal attribution to any single component.

The use of simulation and machine learning is intended not to replace traditional instruction, but to extend it into domains that reflect tools used in research settings. Given that the current analysis reports group-level outcomes, these findings should be seen as evidence that the blended approach can align classroom practice with research workflows and support cohort-level gains in accuracy and confidence; further work is needed to isolate component effects.

This study also represents a model for how advanced tools can be introduced in a high school context without overwhelming students or requiring major curriculum overhauls. By embedding machine learning within a guided, teacher-led framework, students are exposed to powerful analytical methods while still receiving the structure they need to succeed. Importantly, the entire sequence remains rooted in TEKS standards, ensuring that the activity is both innovative and instructionally aligned.

In the future, we plan to further evaluate the extent to which students' abilities to explain and model parallax-based distance estimation change at the individual level, and to examine attitudes toward computational tools. Critical next steps include collecting matched student identifiers (with appropriate consent) to enable within-subject analyses (e.g., paired t-tests or mixed-effects models), implementing comparison groups across instructors or schools, and refining the rubric to ensure inter-rater reliability.

Ultimately, this project offers a replicable strategy for integrating AI and simulation into science instruction in a way that is accessible, engaging, and educationally meaningful. Connecting traditional astronomical concepts with modern analytical tools prepares students not only to understand the cosmos, but to navigate the data-rich scientific challenges of the future.

**Acknowledgement**

This material is based upon work supported by the National Science Foundation under Award No. 2206953. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

The authors would like to express gratitude to Dr. Bart Taylor and Mr. Cheng Niu for their invaluable assistance and insightful suggestions for improvement during the process of working on this paper.

## References


[1] Taasoobshirazi, G., & Sneha, S. (in press). Assessing misconceptions in astronomy: The use of ordered multiple-choice items. *Contemporary Mathematics and Science Education*.

[2] Luri, X., Brown, A. G. A., Sarro, L. M., Arenou, F., Bailer-Jones, C. A. L., Castro-Ginard, A., ... & Smart, R. L. (2018). Gaia Data Release 2: Using parallax for stellar distance estimation. *Astronomy & Astrophysics, 616*, A9.

[3] Shen, X. J., Hou, L. G., Liu, H. L., & Gao, X. Y. (2025). parallax-based distance estimation estimations to Galactic H II regions: Nearby spiral structure. Astronomy & Astrophysics, 696, A67.

[4] Andriantsaralaza, M., Ramstedt, S., Vlemmings, W. H. T., & De Beck, E. (2022). Distance estimates for AGB stars from parallax measurements. *Astronomy & Astrophysics, 667*, A74.

[5] Ladd, N., & Nottis, K. E. (2019). A hybrid hands-on and computer simulation laboratory activity for the teaching of astronomical parallax. *Journal of Astronomy & Earth Sciences Education, 6*(2), 31–44.

[6] Pössel, M. (2017). Parallax: Reaching the stars with geometry. *Science in School, 39*, 40–44.

[7] Leung, H. W., & Bovy, J. (2019). Simultaneous calibration of spectro-photometric distances and the Gaia DR2 parallax zero-point offset with deep learning. *Monthly Notices of the Royal Astronomical Society, 489*(2), 2079–2096.

[8] Shen, Y. Y., & Luo, A. L. (2024). Distance and stellar parameter estimations of solar-like stars from the LAMOST spectroscopic survey. *Astronomy & Astrophysics, 691*, A218.

[9] Fluke, C. J., & Jacobs, C. (2020). Surveying the reach and maturity of machine learning and artificial intelligence in astronomy. *WIREs Data Mining and Knowledge Discovery, 10*, e1349.

[10] Touretzky, D. S., Gardner-McCune, C., Martin, F., & Seehorn, D. (2019). Envisioning AI for K–12: What should every child know about AI? In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 9795–9799).

[11] Newland, J. (2025). Using data science in high school astronomy. *arXiv preprint* arXiv:2501.04856.

[12] Nora's Guide to the Galaxy. (2024, August 26). *Putting parallax to the test* [Video]. YouTube. https://youtu.be/-bNyZZ85Zos

[13] Elyiv, A. A., Melnyk, O. V., Vavilova, I. B., Dobrycheva, D. V., & Karachentseva, V. E. (2020). Machine-learning computation of distance modulus for local galaxies. *Astronomy & Astrophysics, 635*, A124.

[14] Meher, S. K., & Panda, G. (2021). Deep learning in astronomy: A tutorial perspective. *The European Physical Journal Special Topics, 230*(10), 2285–2317.


## Appendix A

Materials Required

- Protractors and measuring tapes
- Desktop or Chromebook computers with internet access
- Custom web-based simulation interface
- Student packet (lab worksheets, data collection forms, reflection prompts)

- Google Sheets (for simulation calculations)
- Teacher slide deck for instruction and guided discussion
- Google Colab notebook for machine-learning demonstration

**Appendix B**

Students progress through a structured seven-day sequence designed to scaffold their understanding of parallax measurement and machine-learning applications:

**Day 1: Pre-Assessment & Introduction**

- Students complete a written pre-test and opinion survey to establish baseline conceptual understanding and confidence.
- The teacher introduces the Cosmic Distance Ladder via direct instruction and guided discussion.

**Day 2: Physical Parallax Lab**

- Alone, students design a procedure for a hands-on lab measuring an object's distance using protractors and measuring tapes.

**Day 3: Data Collection and Reflection**

- Individuals record angles and compute distances using trigonometric formulas.
- Students compile their physical-lab measurements and reflect on the process in small groups.
- The teacher leads a class discussion highlighting the limitations of traditional tools and transitioning to modern pixel-based methods.

**Day 4: Mid-Unit Assessment & ML Overview**

- Students complete a mid-unit assessment on parallax measurement concepts.
- The teacher introduces machine learning fundamentals and its role in modern astronomy, preparing students for the simulation.

**Day 5: Pixel-Based Simulation**

- Students use a web simulation to mimic photographing a fixed object from two positions (analogous to Earth's orbit).
- They identify the object's position shift in pixels and calculate distances using a provided Google Sheet.

**Day 6: ML Integration and Analysis**

- The simulation outputs are compared to predictions generated by a pre-trained machine learning model (MLPRegressor).
- Students analyze discrepancies between manual and ML-predicted results, evaluating potential sources of error and model performance.

**Day 7: Post-Assessment & Synthesis**

- Students complete a post-test and final opinion survey.
- A class-wide synthesis discussion helps consolidate understanding and encourages reflection on the use of modern tools in astronomy.

**Appendix C**

Schedule of Student Activities and Data Collection

| Day | Time (min) | Activity | Assessment or collected data |
|---|---|---|---|
| 1 | 50 | Intro (5), Pre-Test (10), Survey (5), Notes (20), Practice Problems (10) | Pre-Test, Opinion Survey |
| 2 | 50 | Teacher Instructions (10), Student Creation (30), Approval/Advice (10) | Student Observations, Lab Procedures |
| 3 | 50 | Outdoor Lab (20), Data Gathering (15), Sketch (10), Calc Distance (10), Percent Error (5) | Lab Packet, Reflection |
| 4 | 50 | Mid-Test (15), ML Notes (30), Conclusions (5) | Mid-Test |
| 5 | 50 | Instructions (10), Sim Demo (5), Sheets Demo (5), Data Gathering (35), Conclusion Qs (5) | Simulation Data Sheet |
| 6 | 50 | Instructions (10), Sim Demo (5), Sheets Demo (5), Data Gathering (30), Conclusion Qs (10) | ML Results, Student Notes |
| 7 | 50 | Q&A (10), Post-Test (30), Final Survey (10) | Post-Test, Final Opinion Survey |