2026 ASEE ANNUAL CONFERENCE June 21–24, 2026 | Charlotte Convention Center, NC
Host Sponsor UNC Charlotte
Paper ID #53351

NSF ITEST: Pulse oximeters as a concrete anchor for illuminating bias in healthcare AI

Dr. Kathryn Jessen Eller, Brown Center for Biomedical Informatics, Brown University and The Concord Consortium

Kathryn Jessen Eller has a PhD in Oceanography from UCONN and postdoctoral training from the Marine Biological Laboratory and Tufts University. She is a nature and ocean enthusiast, teacher advocate, and currently the secretary for the Rhode Island Science Teachers Association. Her research interests include developing engaging courses and experiences co-designed with teachers and their students to increase student data, AI and ocean literacy.

Brandy Jackson, Scoutlier by Aecern
Dr. Michael Cassidy, TERC
Leo Anthony Celi, Massachusetts Institute of Technology

NSF ITEST: Pulse oximeters as a concrete anchor for illuminating bias in healthcare AI

Purpose

With healthcare systems rapidly adopting artificial intelligence (AI) for the triage and diagnosis of medical conditions [1], [2], engineering students need greater awareness of how technical design choices can create systematic error in healthcare data and, in turn, influence downstream AI models and clinical decisions. Healthcare is personally and socially meaningful and connected to students' lives and possible careers. Therefore, healthcare-based AI and data science learning can support engagement for students with varied prior knowledge and aspirations, as reflected in teacher perspectives from prior implementations of this curriculum [3], [4].

This paper examines a pulse oximetry activity as a concrete, engineering-valid anchor for helping high school students reason about bias in healthcare AI. The activity is briefly described in an earlier publication of the unit plan [5]. Here, we report early, triangulated evidence of student engagement and bias-oriented reasoning. Our research question is: *How do high school students and teachers engage with and reason about a pulse oximetry activity as an anchor for understanding and mitigating bias in healthcare AI?*

Background

Pulse oximeters estimate blood oxygen saturation (SpO_2) using light absorption through the skin [6]. Substantial evidence shows that these devices can produce systematically inaccurate readings for individuals with darker skin pigmentation, increasing the risk of missed hypoxemia [7], [8], [9] [10]. Recent regulatory efforts underscore the engineering nature of the issue. For example, the FDA recommends evaluating pulse oximeter performance across the range of skin pigmentations using more representative study populations and standardized skin tone assessment as part of device validation [11]. Because pulse oximeters are familiar, widely used, and embedded in downstream clinical and algorithmic systems, they provide a compelling context for connecting measurement error to data quality, AI outputs, and health outcomes. The pulse oximeter activity described here leverages this real-world issue to make bias concrete.

Program Context and Instructional Design

The pulse oximetry activity is part of an NSF-funded data science and AI in healthcare program, that includes a semester-long high school course and two-day datathon [12], [5]. The datathon helps students engage with timely issues in healthcare AI, including bias and fallibility. At the 2023 datathon, teams were challenged to develop a pulse oximetry correction factor [13], later published by mentors [14]. Pulse oximetry is introduced in Lesson 6. In Phase 1 (Spring 2023–Spring 2024), it functioned as a conceptual anchor for healthcare AI bias, but teacher feedback indicated that students struggled to connect abstract discussions to concrete mechanisms. In response, Phase 2 (2024) redesigned the lesson as a hands-on investigation in which students collected, visualized, and evaluated their own pulse oximetry data.

A student interview from the 2024 Datathon illustrates the intended learning:
I was talking to him [clinician mentor] about a pulse oximeter and how the melanin is different, and target's different than the lighter skin, and how oxygen is shown differently." This student's

explanation shifts toward mechanistic reasoning (how sensing works and why error arises), aligning with an engineering framing of bias as a design and validation challenge.

The pulse oximetry experiment (expanded from [5]) follows a streamlined 5E-inspired sequence [15] emphasizing phenomenon-driven inquiry and data/AI literacy:

- **Engage:** Students use a pulse oximeter and discuss what it measures and how it works.
- **Explore (data collection):** Students collect pulse and oxygen readings before and after brief exercise, record variables, and contribute to a shared class dataset.
- **Explain (data practices):** Students use CODAP [16] to visualize and analyze their data.
- **Elaborate/Evaluate (sensemaking and ethics):** Facilitated prompts connect student's observed patterns to clinical risk, AI impacts, and possible mitigations.

The lesson emphasizes core engineering practices, including experimental design, measurement under uncertainty, user validation, and analysis of system limitations. This design choice is intentional: teachers have reported that hands-on activities using personal data are especially engaging for students, particularly when students can see immediate results [4]. Teachers also noted that these moments help students move beyond viewing AI as "*just a computer*" and toward recognizing how AI can affect society [4]. We also draw on a context-based learning lens that connects students to healthcare data and practice, emphasizes hands-on disciplinary work, and supports use of disciplinary language around measurement error, bias, and uncertainty [3]. In the revised Lesson 6 (Fall 2024–Spring 2025), we strengthened this approach by centering the pulse oximeter investigation as the primary data-generating experience rather than a supplemental reflection.

Research Design and Data Sources

This study uses a mixed-methods, design-based research approach, drawing on multiple complementary data sources to examine engagement, reasoning, and feasibility across phases. Table 1 summarizes data sources by phase and analytic contribution. Trace data are interpreted as participation indicators, not learning outcomes. Data sources include: (1) Teacher-written implementation notes and focus group notes (Phase I), (2) Teacher implementation notes, survey (likert scale and open responses), and interviews (Phase II), (3) Student interview excerpts, (4) Student artifacts (e.g., explanations, visualizations), and (5) Platform trace indicators (participation, completion, time-on-task). Qualitative data were coded using four prioritized themes aligned to instructional goals. These were (1) Engagement, (2) Conceptual/mechanistic understanding, (3) Bias-oriented reasoning and mitigation, and (4) Challenges and needs.

Table 1. Summary of Data Sources and Analytic Contribution

Phase	Lesson 6 Context	Data source	Sample size	Primary Analytic Contribution
PHASE I Spring 2023- Spring 2024	Pulse oximetry as a conceptual anchor	Teacher implementation notes	n= 5 teachers	- Engagement - Feasibility - Challenges

	Task 3 prompts, no full experiment	Teacher focus group notes	1 focus group (5 teachers)	- Cross-teacher patterns - Implementation constraints - Recommended pulse oximetry experiment
		Student artifacts (Task 3 responses on the platform)	69 students (5 classes)	- Bias-oriented reasoning - Conceptual Understanding
		Student interview excerpts	1 focus group session of 5 students (excerpt)	- Engagement - Bias awareness
		Platform trace data	n= 69 students enrolled across 5 classes	- Participation indicators - Completion rates (0.50-0.85)
PHASE 2 Fall 2024- Spring 2025	Full pulse oximeter experiment (revised Lesson 6)	Teacher implementation notes (vignette)	n= 1 teacher	- Enactment detail of the revised experiment sequence - Feasibility and practical constraints
		Teacher implementation interview	n= 4 teachers	- Engagement - Bias-oriented reasoning - Implementation context
		Teacher implementation survey	n= 4 teachers	- Engagement - Reasoning mitigation
		Student interview excerpts	1 focus group (4 students)	- Mechanistic understanding - Bias-oriented reasoning
Contextual (Datathon)	Pulse oximetry referenced in capstone work	Datathon student interviews	Spring 2024 n= 1 student Spring 2025 n= 1 focus group (5 students)	- Illustrative mechanistic reasoning

Results
As shown in Table 1, evidence is strongest for student engagement and implementation feasibility, with converging but more moderate evidence for conceptual/mechanistic understanding and bias-oriented reasoning.

Engagement*:* Across teacher surveys and interviews (n = 4), teacher implementation notes, student interviews, and platform trace indicators, the pulse oximetry experiment was consistently engaging. All teachers rated student engagement and the activity's usefulness as a real-world anchor for bias in healthcare AI at the highest levels. Teachers described students as "*thoroughly engaged*," working collaboratively to collect measurements and consolidate class data for visualization. In one classroom implementation vignette, the teacher documented that students "*really enjoyed the hands-on nature of the pulse oximeter experiments*," and captured a student comment: **"***I like the Hands-On with the pulse oximeter, this is a fun class***"**. Students echoed this sentiment, describing the experience of moving from reading about bias to seeing it "*translated to real life*," which students described as "*impactful*".

Conceptual and Mechanistic Understanding*:* Student artifacts and interviews indicate emerging mechanistic understanding of pulse oximetry. Students articulated that the device estimates oxygen indirectly using light and skin tone can affect absorption. One student explained that light may not "*penetrate… through the skin*" in the same way and therefore may not read oxygen "*properly*" for students with darker skin tones. Teachers reported that students used visualizations to reason about variability and uncertainty rather than seeking single correct values. Student artifacts reflect attention to measurement assumptions and system limitations.

Bias-oriented Reasoning and Mitigation: Students demonstrated bias-oriented reasoning about who is affected by measurement limitations and why this matters. Several students expressed surprise at learning that a common medical device could be biased, with one noting, *"I had no idea that this actually could be biased towards people of color".* Teacher interviews and surveys reinforce these findings. They described student learning in terms of real-world relevance, awareness of device limitations, and caution in interpreting healthcare data. Teachers emphasized that students learned not to "*blindly trust healthcare devices*," and to consider "*what it could mean for them personally.*" Student-proposed mitigations commonly included awareness of uncertainty, confirmatory testing, and improved validation or recalibration across users.

Challenges and Needs: Teachers identified practical constraints shaping implementation. Small class sizes and limited demographic diversity can make patterns difficult to observe. Teachers also emphasized the need for consistent approaches to recording skin tone and sufficient time for visualization and discussion.

Overall, results are best interpreted as design and feasibility evidence rather than outcome claims.

Discussion
The findings reinforce why pulse oximetry functions as a strong "concrete anchor" for bias in healthcare AI through an engineering lens. Rather than framing bias as an abstract ethical concern, the activity situates bias as a measurement validity problem rooted in sensing assumptions, calibration, and uncertainty. This supports prior work emphasizing the need for

students to reason about how upstream measurement and data collection decisions shape downstream algorithmic behavior [14], [17].

Students' mitigation ideas, though often conceptual, closely mirror real-world engineering and regulatory responses. Recent FDA draft guidance emphasizes representative clinical data, standardized skin tone assessment, and clearer labeling to ensure pulse oximeters perform comparably across skin pigmentation [11]. The alignment between student reasoning and these regulatory priorities underscores the authenticity of the learning opportunity.

Engagement emerged as a consistent strength. Teachers and students highlighted the value of embodied, context-rich data experiences, aligning with prior research on making complex AI issues accessible through real-world phenomena [14]. At the same time, implementation challenges highlight the importance of instructional scaffolding. Small datasets and classroom constraints mirror the same issues that motivate regulatory concern about unrepresentative data, reinforcing the relevance of these discussions for engineering education.

Conclusion

This study provides preliminary, triangulated design evidence that a pulse oximetry investigation can support student engagement and bias-oriented reasoning in healthcare AI. Across multiple data sources, the activity was feasible in real classrooms and prompted mechanistic reasoning about measurement bias and mitigation. By centering an engineered sensing system with documented limitations, the activity positions bias as an engineering problem of validation, calibration, and uncertainty. While claims are limited by sample size and classroom variability, results suggest that pulse oximetry is a promising, replicable anchor phenomenon for integrating technical and societal dimensions of AI in engineering education.

Future plans

Future work will expand implementations across additional classrooms and refine scaffolds for data collection and visualization. Planned improvements include standardized approaches to recording skin tone and rubrics for evaluating mechanistic explanations. Upcoming datathon implementations will enable larger datasets and examination of how classroom learning transfers to multigenerational problem-solving contexts.

Acknowledgment / NSF disclaimer

This work is supported by the National Science Foundation Innovative Technology Experiences for Students and Teachers (Award # 2517189). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

References

[1] S. Tyler, M. Olis, N. Aust, L. Patel, L. Simon, C. Triantafyllidis, V. Patel, D. W. Lee, B. Ginsberg, H. Ahmad & R. J. Jacobs, “Use of artificial intelligence in triage in hospital emergency departments: A scoping review,” *Cureus*, vol. 16, no. 5, Art. no. e59906, 2024, doi: 10.7759/cureus.59906.

[2] Faiyazuddin et al., 2025 M. Faiyazuddin, S. J. Q. Rahman, G. Anand, R. K. Siddiqui, R. Mehta, M. N. Khatib & S. Gaidhane, “The impact of artificial intelligence on healthcare: A comprehensive review of advancements in diagnostics, treatment, and operational efficiency,” *Health Science Reports*, 2025.

[3] K. Shreeve, A. Perry, M. Cassidy, K. Jessen Eller, B. Price, B. Jackson, L. A. Celi, I. Lourentzou, and L. Hendrik, “Incorporating teacher effect when modeling student engagement in smart STEM classrooms: A cluster analysis,” *Smart Learning Environments*, vol. 12, no. 1, Art. no. 58, 2025.

[4] A. Bopardikar, M. Cassidy, and K. Jessen Eller, “Exploring teachers’ perspectives on enacting context-based learning of artificial intelligence (AI) and data science to support students’ engagement and learning,” in *Proc. 2025 Int. Soc. Learn. Sci. Conf.*, 2025. Available: https://repository.isls.org/handle/1/11581.

[5] K. Jessen Eller, A. Larson, E. Price, B. Jackson, L. Hendrik, M. Cassidy, D. Rodemacher, and L. A. Celi, ‘Increasing data literacy in high school students through data science, AI, and healthcare’, *The Science Teacher* (Special Issue on Data Science in the Science Classroom) [March/April Issue, accepted 2026].

[6] K. D. Torp, P. Modi, E. J. Pollard & L. V. Simon, “Pulse oximetry,” in *StatPearls* [Internet]. Treasure Island, FL, USA: StatPearls Publishing, 2023.

[7] A. I. Wong, M. Charpignon, H. Kim, C. Josef, A. de Hond, J. J. Fojas, A. Tabaie, X. Liu, E. Mireles-Cabodevila, L. Carvalho, R. Kamaleswaran, R. W. Madushani, L. Adhikari, A. Holder, E. W. Steyerberg, T. G. Buchman, M. E. Lough & L. A. Celi, “Analysis of discrepancies between pulse oximetry and arterial oxygen saturation measurements by race and ethnicity and association with organ dysfunction and mortality,” *JAMA Network Open*, vol. 4, no. 11, Art. no. e2131674, 2021, doi: 10.1001/jamanetworkopen.2021.31674.

[8] L. A. Celi, J. Cellini, M.-L. Charpignon, E. C. Dee, F. Dernoncourt, R. Eber, W. G. Mitchell, L. Moukheiber, J. Schirmer, J. Situ, J. Paguio, J. Park, J. Wawira Gichoya, and S. Yao, ‘Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review’, *PLOS Digital Health*, 1(3) (2022), e0000022. doi:10.1371/journal.pdig.0000022.

[9] E. R. Gottlieb, J. Ziegler, K. Morley, B. Rush, and L. A. Celi, ‘Assessment of racial and ethnic differences in oxygen supplementation among patients in the intensive care unit’, *JAMA Internal Medicine*, 182(8) (2022), pp. 849–858. doi:10.1001/jamainternmed.2022.2587.

[10] A Fawzy, T. D. Wu, K. Wang, M. L. Robinson, J. Farha, A. Bradke, S. H. Golden, Y. Xu, and B. T. Garibaldi, ‘Racial and ethnic discrepancy in pulse oximetry and delayed identification of treatment eligibility among patients with COVID-19’, *JAMA Internal Medicine*, 182(7) (2022), pp. 730–738. doi:10.1001/jamainternmed.2022.1906.

[11] U.S. Food and Drug Administration, ‘Pulse Oximeters for Medical Purposes – Non-Clinical and Clinical Performance Testing, Labeling, and Premarket Submission Recommendations’ (Draft Guidance for Industry and Food and Drug Administration Staff, January 2025). Available: https://www.fda.gov/news-events/press-announcements/fda-proposes-updated-recommendations-help-improve-performance-pulse-oximeters-across-skin-tones?utm_source=chatgpt.com. [Accessed Jan. 19, 2026].

[12] K. Jessen Eller, “Data science, AI, and you in healthcare: teaching students about medical data bias” (Concord Consortium, 2024). Available: https://concord.org/newsletter/2025-winter/data-science-ai-and-you-in-healthcare-teaching-students-about-medical-data-bias/ [Accessed Oct. 15, 2025].

[13] CriticalDatathon, ‘MIT Critical Datathon 2023’ (GitHub Pages, 2023). Available: https://criticaldatathon.github.io/. [Accessed Jan. 18, 2026].

[14] J. Matos, T. Struja, J. Gallifant, M.-L. Charpigno, J.S. Cardoso, and L. A. Celi, ‘Shining light on dark skin: pulse oximetry correction models’, in *Proceedings of the 2023 IEEE 7th Portuguese Meeting on Bioengineering (ENBENG)* (Porto, Portugal, 2023), pp. 211–214.

[15] R. W. Bybee, J. A. Taylor, A. Gardner, P. Van Scotter, J. C. Powell, A. Westbrook, and N. Landes, *The BSCS 5E instructional model: Origins and effectiveness*. Colorado Springs, CO, USA: BSCS, 2006.

[16] Concord Consortium, *Common Online Data Analysis Platform (CODAP)* [computer software] (Concord, MA: The Concord Consortium, 2014). Available: https://codap.concord.org/. [Accessed Jan. 19, 2026].

[17] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, ‘Dissecting racial bias in an algorithm used to manage the health of populations’, *Science*, 366(6464) (2019), pp. 447–453.