Mapping the literature landscape of artificial intelligence and machine learning in physics education research

Amir Bralin
Department of Physics and Astronomy, Purdue University.

Amogh Sirnoorkar
Center for Advancing the Teaching and Learning of STEM (CATALYST), Purdue University.

Yiyuan Zhang
Department of Computer Science, Purdue University.

N. Sanjay Rebello
Department of Curriculum and Instruction, and Department of Physics and Astronomy, Purdue University.

This paper examines the topical areas where machine learning (ML) and artificial intelligence (AI) techniques have been applied in physics education research (PER), based on a systematic review of studies published in the Physical Review Physics Education Research journal and Physics Education Research Conference proceedings. Following the PRISMA guidelines, 79 articles were analyzed, revealing six major themes: (i) development and analysis of assessments, (ii) student success, (iii) student engagement, (iv) theoretical and methodological overview, (v) admissions, and (vi) other areas like literature reviews. The most common theme was using AI/ML for developing, evaluating, and characterizing student responses on physics assessments. Other prominent areas included predicting student success, analyzing collaboration patterns, and providing methodological guidance. The study scope was limited to these two publication venues and classified articles primarily based on abstracts.

2024 PERC Proceedings edited by Ryan, Pawl, and Zwolak; Peer-reviewed, doi.org/10.1119/perc.2024.pr.Bralin

I. INTRODUCTION

Advances in the fields of Artificial Intelligence (AI) and Machine Learning (ML) have led to their widespread adoption in physics education research (PER). This is evident from the recent call for papers [1] and published studies employing emerging technologies and computational techniques to analyze educational data. With the rise of Generative AI, the interest in these technologies has only increased. This growing interest manifests itself also in the increasing attention of researchers and educators to AI-related topics at national conferences. For example, the 2024 Summer Meeting of the American Association of Physics Teachers (AAPT) hosted a total of seven sessions dedicated to leveraging AI for physics learning compared to only one in 2023 [2].

Recent studies have explored how to leverage AI for qualitative analysis of visual-based (image and video) data [3], the feasibility of ML-based physics assessment [4], how to deploy AI to generate feedback on student responses [5], ML techniques to accurately predict student performance [6], the extent of AI usage by students in a physics class [7], the potential of AI as a personal assistant in teacher preparation [8]. Nevertheless, given this growing interest in AI and ML, there is relatively less emphasis on systematically documenting the body of literature surrounding their use in PER.

This work is a step toward the systematic review of studies in PER that use AI and/or ML in a meaningful way. First, we provide our view of AI and ML which dictates our selection criteria for relevant articles. Then, we briefly analyze each article in our review list and synthesize broad topical areas common to most of them. Given the space constraints of this conference paper, its target audience, and scope, we only focus on studies published in American Physical Society's "Physical Review Physics Education Research" (PRPER) and the Physics Education Research Conference (PERC) Proceedings. In doing so, we address the following research question: *What are the broad topical areas of studies published in PRPER and PERC Proceedings employing AI and/or ML?*

To address the question above, we adopt the latest version of the "Preferred Reporting Items for Systematic Reviews and Meta-Analyses" (PRISMA) statement [9] which provides a 27-item checklist ensuring systematic collection and analysis of manuscripts for literature reviews. This review resulted in the total of 79 articles (64 from PRPER and 15 from PERC) which we analyzed and classified into six following themes (in increasing order of associated articles) or topical areas in which researchers employ AI and ML in PER: (i) development and analysis of assessments, (ii) student success, (iii) student engagement, (iv) theoretical and methodological overview, (v) admissions, and (vi) others (including literature review, evaluation of pedagogical transformations, etc.).

In the next section, we provide a brief overview of AI/ML before discussing our approaches to selecting and analyzing papers. We then discuss the results, limitations of this work, and our intended follow-up on this article.

II. ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING: AN OVERVIEW

Artificial Intelligence (AI) and Machine Learning (ML) are two rapidly evolving fields that have gained significant attention in recent years. Collectively, we refer to them both as AI/ML from here on. In this section, we provide our view of what constitutes AI/ML, setting the stage for our exploration of their utilization in PER.

AI refers to computer systems that can mimic intelligent human behavior or perform tasks that typically require human intelligence. These tasks include learning from experience, adapting to new information, and making decisions based on that information. AI encompasses a wide range of computational techniques or algorithms, including ML, natural language processing (NLP), and clustering. We will discuss these terms further below. Applications of AI include: development of autonomous vehicles such as self-driving cars and virtual personal assistants like those used by Google and Amazon. In addition, Generative AI has become a popular collocation in the media. Its meaning emphasizes the ability of AI systems to generate digital content: fast, of high quality, and with minimal input. [10]

ML is a field related to AI that involves data to simulate learning, progressively enhancing their accuracy over time by learning from data. An ML algorithm comprises three primary components. First, there is a decision process where the algorithm, based on input data, makes predictions or classifications, estimating patterns within the data. Second, an error function evaluates the accuracy of these predictions by comparing them to known examples. Finally, there is a model optimization process where the algorithm adjusts its internal parameters, iteratively refining its predictions to minimize the discrepancy between the model's estimates and the known examples, thus enhancing its accuracy over time. For example, in email filtering, a ML algorithm can be used to predict spams based on various features such as the email content, sender information, and subject line [11]. Given this broad definition of ML, we assign every statistical algorithm such as linear or logistic regression under its umbrella.

NLP is another field related to AI. It merges computational linguistics with statistical and machine learning techniques. It enables computers to comprehend, generate, and interact with human language in both written and spoken forms. NLP algorithms typically involve several steps. First, text is preprocessed through tasks like tokenization (breaking it into words or phrases) and removing stop words (commonly used words that add little meaning). Next, features are extracted, turning the text into numerical data for machine learning algorithms. The model is then trained using these features, and its performance is evaluated using metrics like accuracy. Finally, the trained model is deployed to make predictions on new text data. NLP applications are found in various fields, such as sentiment analysis, aimed at classifying the emotional intent of the text, and machine translation, for automating translation between different languages [12].

A sub-field of NLP which has made great progress in the recent years is language modeling. In essence, a language model is a statistical model that is trained to predict the probability of a word or sequence of words occurring in a sentence. This predictive ability enables computers to generate coherent and contextually relevant text based on a given input. For instance, consider the following scenario: we are provided with the beginning of the sentence "The dog sat on the..." In this case, a language model would analyze the contextual information within the sentence to determine the probabilities of the next word. It would predict that the most probable next word is "floor" or "chair", based on the provided context. Large language models (LLMs), such as OpenAI's GPT-4, have gained immense popularity due to their remarkable ability to generate human-like text. These models are trained on vast amounts of text data, learning the intricate patterns and relationships among words and phrases. As a result, they can generate text that is not only grammatically correct but also contextually relevant and coherent [13].

Finally, there is a family of ML algorithms that does not fit well within the learning framework given above. That is, in some cases, there are no a priori known examples to compare the predictions with. Then the algorithm will only look for some patterns within the data. Such an approach is called unsupervised learning. It includes clustering algorithms, network analysis, and topic modeling, among others. We refer to them all collectively as simply clustering.

III. METHODS

To guide our literature review, we used the twenty seven items of the PRISMA statement [9], such as: the objective of the literature review, databases used in searching for the manuscripts, keywords used in the database search, and inclusion/exclusion criteria. The statement also provides a template for the flow diagram that summarizes the details of identification, screening, and selection of the manuscripts targeted by the review. Figure 1 provides our adaptation of this template. The item checklist of the PRISMA statement was greatly simplified when we decided to focus only on two sources of literature: the PRPER journal and PERC Proceedings. Nevertheless, to provide in detail the checklist completed for our review would not fit within the constraints of this paper. Instead, the main steps are described below.

PRPER is the main scientific journal for publishing PER in the U.S. In addition, it is "the only fully open access journal for physics education research" as stated on its webpage [14]. It was established in 2005 under the name "Physical Review Special Topics – Physics Education Research." In 2016, it was renamed as PRPER. Thus, among our references, the earlier journal articles have extra acronym "ST" due to the old name. We refer to all articles published in this journal as that of PRPER. For our literature review, we included all articles published from its inception date (Jan, 2005) to the time of writing this paper (Apr, 2024).

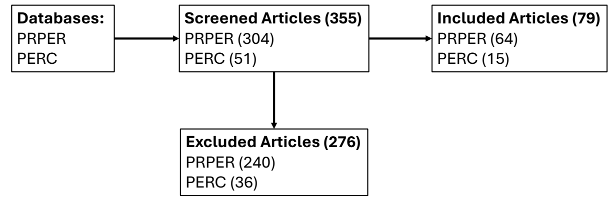

FIG. 1. The PRISMA flow diagram adapted and simplified for our review. Databases refer to the sources of literature used for our article search. The screened articles are the ones which resulted from our keyword search. They were included/excluded according to the outlined inclusion/exclusion criteria. The number of articles at each stage is given in the parentheses.

PERC is the annual conference dedicated to PER in the U.S. Works presented here are often expanded and published in journals. Until 2013, PERC was published by the American Institute of Physics (AIP). Therefore, some of the earlier PERC references are cited as AIP Conference Proceedings. We included PERC papers published between 2003 and 2023 in our review.

We initiated the data collection, i.e., screening the articles, by searching for keywords "artificial intelligence" and "machine learning." In addition, we also searched for the abbreviated keywords "AI" and "ML" because they might be used in the title of an article, while the full term might be reserved for its full text. During this process, we found that there was a significant number of articles that could be identified by the keywords "Natural Language Processing" and "clustering," but wouldn't be identified otherwise. This may be due to treating each of these families of algorithms separately from the AI/ML context. All in all, with these four keywords and their abbreviated versions we searched for articles published in the relevant databases of the PRPER journal [15] and the PERC proceedings [16]. This yielded a total of 304 articles in the former and 51 articles in the latter.

Next, we performed data selection by including only those studies which employ AI/ML as their method for data analysis. This criterion excluded the studies that (a) write about AI/ML but don't use it, (b) use AI/ML for transcribing qualitative data (e.g. the popular software OtterAI), (c) mention AI/ML only in the references. Another inclusion criterion was to be a peer-reviewed article. In the case of PRPER, this excluded opinion articles, calls for papers, editorials, and comments. In the case of PERC Proceedings, this excluded plenary and invited papers. As a result, the total number of articles eligible for our review dropped down to 79 (64 for PRPER and 15 for PERC). Figure 2 breaks down this number into columns based on the publication year of each article.

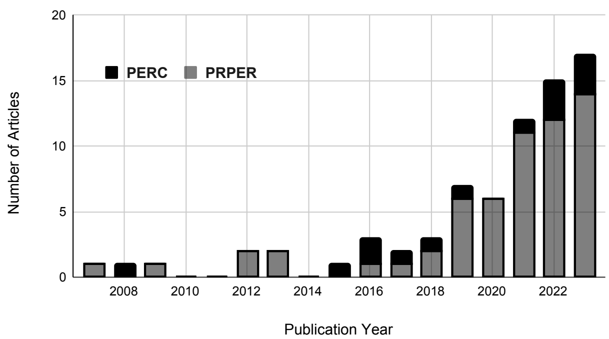

FIG. 2. The number of articles selected for our review versus the year of publication. Articles published in PRPER are shown in gray. Articles published in PERC are shown in solid black.

The last several years have witnessed the rise of Generative AI and the popularization of ML among researchers. While PERC Proceedings are still catching up in this regard, the dramatic growth of PRPER articles on the chart seems to reflect this trend.

As the last step in our review process, the first two authors divided all articles among themselves and described how the study in each subset of articles employed AI/ML. This was based primarily on the article's abstract. However, in those cases which required additional information, the rest of the text was referred to as well. The remaining two authors then randomly selected 15 articles from the entire set and cross-checked whether the provided descriptions were indeed correct. The disagreements (which were relatively few) were discussed and resolved. Based on the descriptions, we then identified six common themes.

IV. RESULTS

In this section, we report the broad topical areas that have emerged as the result of our classification of all articles under review. We refer to them simply as *themes*. Even though they are relatively self-contained, they are not strictly mutually exclusive. The reader may notice that some articles are highlighted under several themes below. We did not try to make our classification exact. The purpose of these themes is to provide the reader with a ten-thousand foot view of the broad areas in which researchers in PER have employed AI/ML.

1. Development and analysis of assessments. This major common theme among the articles that we have reviewed was about leveraging AI/ML in the context of physics assessments. Because this theme was much larger in proportion than the others, we divided it further into three sub-themes: (i) characterization of student responses, (ii) development and evaluation of physics assessments with Generative AI, and (iii) analysis of standardized assessments.

In the first sub-theme, studies have used diverse AI/ML techniques for analyzing student responses to various kinds of physics problems [17–25]. They could be: open-ended [26], Likert-style surveys [27], or multiple-choice questions [28, 29]. In addition, studies under this sub-theme used AI/ML to assess: student reasoning [30, 31], epistemic framing [32], and the quality of mental models [33], based on student responses.

In the second sub-theme, studies have mainly focused on the accuracy of Generative-AI responses to typical physics course assessments [34–37] and kinematic graphs [38]. In addition, Generative AI was used for: generating synthetic student data [39], grading assessments [40], and developing physics tasks for prospective teachers [41].

In the third sub-theme, studies leveraged AI/ML to analyze student responses specifically to standardized assessments: Force Concept Inventory (FCI) [28, 42–45], Force and Motion Conceptual Evaluation (FMCE) [46, 47], and Test of Understanding Graphs in Kinematics (TUG-K) [48].

2. Student success. Studies in this theme have mainly used AI/ML to develop predictive models that facilitate student success in class [49–56]. These physics classes could be: hybrid [57], flipped [58], or remote [59]. Other studies in this theme used AI/ML for comparing asynchronous online with traditional lecture methods [60], supporting student agency in physics laboratory [61], unpacking alternate world views of students [62], and exploring students' choice of courses [63]. Still others tried to establish a relation (or a correlation) between student success and: high school preparation [46, 47, 64], formative practice [65], prior knowledge [51]. In addition, some studies leveraged AI/ML to identify the demographic and academic predictors of student success [66, 67].

3. Student engagement. Exploring student collaboration and engagement within and beyond academic environments corresponds to the third major theme in our list. A predominant section of these studies focuses on social network analysis and clustering algorithms. Thrust areas of these studies are diverse: student interaction and collaboration in online and hybrid classrooms [57], student collaboration in sequential courses [68], student interaction in forum logs [69, 70], exploration of gender inequities in physics labs [71], examination of social interactions in labs [72, 73], student emotions in learning [74, 75], effect of social belonging on students' performance [49], examination of perception of learning environment with respect to students' academic grades [50], students' group roles in physics laboratories [76], and the impact of informal physics learning on student identity [77].

4. Theoretical and methodological overview. Our next theme corresponds to the studies that provide theoretical and practical guidelines for employing AI/ML in PER. They elaborate on how to: make reliable claims with machine coding [78], leverage computational methods for qualitative research [79], evaluate statistical models [80], perform social network analysis [81, 82], and analyze multiple-choice questions [83].

TABLE I. Summary of the emergent themes concerning the studies in PER which pertain to the use of AI/ML. Each theme is given a name and comes with a bried description at the corresponding row.

Theme	Description
1. Development and analysis of assessments	Exploring the affordances of leveraging AI/ML in developing assessments and analyzing student responses (including standardized assessments).
2. Student success	Development of predictive models that facilitate student success in physics courses.
3. Student engagement	Investigations of student engagement patterns within and beyond academic environments.
4. Theoretical and methodological overview	Discussion of AI/ML theoretical and methodological approaches relevant to PER.
5. Admissions	Unpacking factors that influence college admissions.
6. Others	Literature synthesis, mapping pedagogical practices in classrooms, and evaluating the effectiveness of course transformations.

5. Admissions. The penultimate theme on our list includes the studies that have used AI/ML for predicting factors that influence university admissions. They have explored application features that may drive the admission decisions [84, 85], predictive factors for students' grades upon admission [86] and factors that predict doctoral degree completion [87, 88].

6. Others. Our final theme captures all other topical areas surrounding the use of AI/ML in PER which have been relatively less explored. The objective to club these articles under the single umbrella is primarily for the feasibility to characterize the existing literature for the reader. The topical areas of studies in this category include: synthesis of literature [89], mapping pedagogical practices in classrooms [90–93], and evaluating the effectiveness of course transformations (or interventions) [94, 95].

Table I provides the summary of each theme along with its brief description.

V. DISCUSSION AND CONCLUSION

We presented six broad emergent themes from manuscripts published in Physical Review Physics Education Research journal and proceedings of the Physics Education Research Conference which leverage AI/ML in their research. The six themes include: (i) development and analysis of assessments, (ii) student success, (iii) student engagement, (iv) theoretical and methodological overview, (v) admissions, and (vi) others (including literature review, evaluation of pedagogical transformations, etc.). Note that these themes are not mutually exclusive. That is, a given article may belong to more than one theme at the same time. Their purpose is to provide a broad overview of the research landscape surrounding the use of AI/ML in PER.

The main limitation of our study comes from our sources of literature: PRPER and PERC. Therefore, it does not include studies published in other sources: other journals and conferences relevant to PER. Since both PRPER and PERC are primarily US-based, adding more international sources of literature will be important for our future work. Furthermore, the papers were classified based primarily on the abstracts. The full text of the paper was only referred to if the information in the abstract was incomplete or ambiguous. The validation of our classification was only done only on about one-sixth of the papers which yielded an 85% agreement between raters. Finally, the classification was done using our working definition of AI/ML. While this definition was carefully constructed, it is likely that some papers may not have been included in our study. Nevertheless we are confident that we have included most of the works related to AI/ML in these two sources.

Our future work will address the aforementioned limitations. Specifically, we will include a larger corpus of literature including other journals where PER researchers are likely to publish. We will continue to refine our definition of AI/ML and hence modify our inclusion and exclusion criteria. We will also expand our analysis to focus more on the technical aspects of the papers i.e. the algorithm and methods used by the authors. Finally, with a greater number of documents to analyze, one may use quantitative (NLP) methods such as topic analysis to investigate the emerging themes in the field.

We have seen that the number of articles published in PER (in the two sources we selected) has increased dramatically in the last few years. This trend is likely to continue so long as the advances in AI/ML continue to occur. Conducting a literature review like this one in the future could result in meaningful comparisons between the state of the field (with regard to AI/ML usage) then versus the state of the field today. How will the emergent themes that we have identified evolve? Will any new themes emerge? These questions could be asked then.

Last but not the least comment worth mentioning here is that of ethics. What are the consequences of employing AI/ML in education? This question did not make up a separate theme in our review analysis because this development is so recent. But as Generative AI is used in education for longer periods of time, one may well be in a position to ask it in the future.

The limitations of this work notwithstanding, this work is one of the earliest attempts to characterize the state of AI/ML in PER and can potentially inform its future evolution.

VI. ACKNOWLEDGEMENTS

This work was supported in part by U.S. National Science Foundation grant 2300645.

[1] C. Henderson, Editorial: Call for papers for Focused Collection of Physical Review Physics Education Research: AI tools in physics teaching and PER, Phys. Rev. Phys. Educ. Res. **19**, 020003 (2023).
[2] https://www.aapt.org/Conferences/index.cfm.
[3] G.-G. Lee and X. Zhai, Realizing Visual Question Answering for Education: GPT-4V as a Multimodal AI (2024), arXiv:2405.07163.
[4] X. Zhai, AI and Machine Learning for Next Generation Science Assessments (2024), arXiv:2405.06660.
[5] T. Wan and Z. Chen, Exploring generative ai assisted feedback writing for students' written responses to a physics conceptual question with prompt engineering and few-shot learning, Phys. Rev. Phys. Educ. Res. **20**, 010152 (2024).
[6] J. Pace, J. Hansen, and J. Stewart, Exploring techniques to improve machine learning's identification of at-risk students in physics classes, Phys. Rev. Phys. Educ. Res. **20**, 010149 (2024).
[7] G. Kortemeyer and W. Bauer, Cheat sites and artificial intelligence usage in online introductory physics courses: What is the extent and what effect does it have on assessments?, Phys. Rev. Phys. Educ. Res. **20**, 010145 (2024).
[8] B. Gregorcic, G. Polverini, and A. Sarlah, ChatGPT as a tool for honing teachers' Socratic dialogue skills, Physics Education **59**, 045005 (2024).
[9] M. L. Rethlefsen, S. Kirtley, S. Waffenschmidt, A. P. Ayala, D. Moher, M. J. Page, and J. B. Koffel, PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews, Sys. Rev. **10**, 1 (2021).
[10] https://www.nvidia.com/en-us/glossary/generative-ai/.
[11] https://www.ibm.com/topics/machine-learning.
[12] https://www.deeplearning.ai/resources/natural-language-processing.
[13] https://developers.google.com/machine-learning/resources/intro-llms.
[14] https://journals.aps.org/prper/about.
[15] https://journals.aps.org/search.
[16] https://www.per-central.org/perc/.
[17] J. Campbell, K. Ansell, and T. Stelzer, Evaluating IBM's Watson natural language processing artificial intelligence as a short-answer categorization tool for physics education research, Phys. Rev. Phys. Educ. Res. **20**, 010116 (2024).
[18] C. M. Nakamura, S. K. Murphy, M. G. Christel, S. M. Stevens, and D. A. Zollman, Automated analysis of short responses in an interactive synthetic tutoring system for introductory physics, Phys. Rev. Phys. Educ. Res. **12**, 010122 (2016).
[19] A. Bralin, J. W. Morphew, C. M. Rebello, and N. S. Rebello, Analysis of student essays in an introductory physics course using natural language processing, in *Physics Education Research Conference Proceedings* (2023).
[20] J. M. Geiger, L. M. Goodhew, and T. O. B. Odden, Developing a natural language processing approach for analyzing student ideas in calculus-based introductory physics, in *Physics Education Research Conference Proceedings* (2022).
[21] J. Campbell, K. Ansell, and T. Stelzer, Using IBM's Watson to automatically evaluate student short answer responses, in *Physics Education Research Conference Proceedings* (2022).
[22] R. Fussell, A. Mazrui, and N. G. Holmes, Machine learning for automated content analysis: characteristics of training data impact reliability, in *Physics Education Research Conference Proceedings* (2022).
[23] J. Munsell, N. S. Rebello, and C. M. Rebello, Using natural language processing to predict student problem solving performance, in *Physics Education Research Conference Proceedings* (2021).
[24] N. S. Rebello, M. H. Nguyen, Y. Wang, T. Zu, J. Hutson, and L. Loschky, Machine learning predicts responses to conceptual tasks using eye movements, in *Physics Education Research Conference Proceedings* (2018).
[25] C. M. Nakamura, S. K. Murphy, M. Christel, S. M. Stevens, and D. A. Zollman, Finding meaningful search features for automated analysis of short responses to conceptual questions, in *AIP Conference Proceedings*, Vol. 1413 (2012) pp. 283–286.
[26] J. Wilson, B. Pollard, J. M. Aiken, M. D. Caballero, and H. J. Lewandowski, Classification of open-ended responses to a research-based assessment using natural language processing, Phys. Rev. Phys. Educ. Res. **18**, 010141 (2022).
[27] R. P. Dalka, D. Sachmpazidi, C. Henderson, and J. P. Zwolak, Network analysis approach to Likert-style surveys, Phys. Rev. Phys. Educ. Res. **18**, 020113 (2022).
[28] E. Brewe, J. Bruun, and I. G. Bearden, Using module analysis for multiple choice responses: A new method applied to Force Concept Inventory data, Phys. Rev. Phys. Educ. Res. **12**, 020131 (2016).
[29] A. Pawl, R. E. Teodorescu, and J. D. Peterson, Assessing class-wide consistency and randomness in responses to true or false questions administered online, Phys. Rev. ST Phys. Educ. Res. **9**, 020102 (2013).
[30] O. R. Battaglia, B. Di Paola, and C. Fazio, Unsupervised quantitative methods to analyze student reasoning lines: Theoretical aspects and examples, Phys. Rev. Phys. Educ. Res. **15**, 020112 (2019).
[31] R. P. Springuel, M. C. Wittmann, and J. R. Thompson, Applying clustering to statistical analysis of student reasoning about two-dimensional kinematics, Phys. Rev. ST Phys. Educ. Res. **3**, 020107 (2007).
[32] M. Bodin, Mapping university students' epistemic framing of computational physics using network analysis, Phys. Rev. ST Phys. Educ. Res. **8**, 010115 (2012).
[33] C. Fazio, O. R. Battaglia, and B. Di Paola, Investigating the quality of mental models deployed by undergraduate engineering students in creating explanations: The case of thermally activated phenomena, Phys. Rev. ST Phys. Educ. Res. **9**, 020101 (2013).
[34] G. Kortemeyer, Could an artificial-intelligence agent pass an introductory physics course?, Phys. Rev. Phys. Educ. Res. **19**, 010132 (2023).
[35] M. N. Dahlkemper, S. Z. Lahme, and P. Klein, How do physics students evaluate artificial intelligence responses on comprehension questions? A study on the perceived scientific accuracy and linguistic quality of ChatGPT, Phys. Rev. Phys. Educ. Res. **19**, 010142 (2023).
[36] S. Wheeler and R. E. Scherr, ChatGPT reflects student misconceptions in physics, in *Physics Education Research Conference Proceedings* (2023).
[37] D. Zollman, A. Sirnoorkar, and J. T. Laverty, Analyzing AI and student responses through the lens of sensemaking and mechanistic reasoning, in *Physics Education Research Conference*

Proceedings (2023).
[38] G. Polverini and B. Gregorcic, Performance of ChatGPT on the test of understanding graphs in kinematics, Phys. Rev. Phys. Educ. Res. **20**, 010109 (2024).
[39] F. Kieser, P. Wulff, J. Kuhn, and S. Küchemann, Educational data augmentation in physics education research using ChatGPT, Phys. Rev. Phys. Educ. Res. **19**, 020150 (2023).
[40] G. Kortemeyer, Toward AI grading of student problem solutions in introductory physics: A feasibility study, Phys. Rev. Phys. Educ. Res. **19**, 020163 (2023).
[41] S. Küchemann, S. Steinert, N. Revenga, M. Schweinberger, Y. Dinc, K. E. Avila, and J. Kuhn, Can ChatGPT support prospective teachers in physics task development?, Phys. Rev. Phys. Educ. Res. **19**, 020128 (2023).
[42] J. Stewart, M. Miller, C. Audo, and G. Stewart, Using cluster analysis to identify patterns in students' responses to contextually different conceptual problems, Phys. Rev. ST Phys. Educ. Res. **8**, 020112 (2012).
[43] J. Stewart, B. Drury, J. Wells, A. Adair, R. Henderson, Y. Ma, A. Pérez-Lemonche, and D. Pritchard, Examining the relation of correct knowledge and misconceptions using the nominal response model, Phys. Rev. Phys. Educ. Res. **17**, 010122 (2021).
[44] Y. Li and C. Singh, Inclusive learning environments can improve student learning and motivational beliefs, Phys. Rev. Phys. Educ. Res. **18**, 020147 (2022).
[45] S. M. Stoen, M. A. McDaniel, R. F. Frey, K. M. Hynes, and M. J. Cahill, Force Concept Inventory: More than just conceptual understanding, Phys. Rev. Phys. Educ. Res. **16**, 010105 (2020).
[46] D. Hewagallage, E. Christman, and J. Stewart, Examining the relation of high school preparation and college achievement to conceptual understanding, Phys. Rev. Phys. Educ. Res. **18**, 010149 (2022).
[47] J. Stewart, G. L. Cochran, R. Henderson, C. Zabriskie, S. DeVore, P. Miller, G. Stewart, and L. Michaluk, Mediational effect of prior preparation on performance differences of students underrepresented in physics, Phys. Rev. Phys. Educ. Res. **17**, 010107 (2021).
[48] P. Klein, S. Becker, S. Küchemann, and J. Kuhn, Test of understanding graphs in kinematics: Item objectives confirmed by clustering eye movement transitions, Phys. Rev. Phys. Educ. Res. **17**, 013102 (2021).
[49] J. D. Edwards, L. Laguerre, R. S. Barthelemy, C. De Grandi, and R. F. Frey, Relating students' social belonging and course performance across multiple assessment types in a calculus-based introductory physics 1 course, Phys. Rev. Phys. Educ. Res. **18**, 020150 (2022).
[50] S. Cwik and C. Singh, How perception of learning environment predicts male and female students' grades and motivational outcomes in algebra-based introductory physics courses, Phys. Rev. Phys. Educ. Res. **17**, 020143 (2021).
[51] E. W. Burkholder, G. Murillo-Gonzalez, and C. Wieman, Importance of math prerequisites for performance in introductory physics, Phys. Rev. Phys. Educ. Res. **17**, 010108 (2021).
[52] E. Altermatt, R. Teodorescu, and E. R. Iverson, Teaching strategies predict instructors' perceptions of their effectiveness in engaging students in introductory physics for life sciences courses, Phys. Rev. Phys. Educ. Res. **17**, 020133 (2021).
[53] J. Yang, S. DeVore, D. Hewagallage, P. Miller, Q. X. Ryan, and J. Stewart, Using machine learning to identify the most at-risk students in physics classes, Phys. Rev. Phys. Educ. Res. **16**, 020130 (2020).
[54] C. Zabriskie, J. Yang, S. DeVore, and J. Stewart, Using machine learning to predict physics course outcomes, Phys. Rev. Phys. Educ. Res. **15**, 020120 (2019).
[55] N. T. Young, G. Allen, J. M. Aiken, R. Henderson, and M. D. Caballero, Identifying features predictive of faculty integrating computation into physics courses, Phys. Rev. Phys. Educ. Res. **15**, 010114 (2019).
[56] J. W. T. Pond and J. J. Chini, Exploring student learning profiles in algebra-based studio physics: A person-centered approach, Phys. Rev. Phys. Educ. Res. **13**, 010119 (2017).
[57] J. Pulgar, D. Ramírez, and C. Candia, Enhanced social connectivity in hybrid classrooms versus academic centrality in online settings, Phys. Rev. Phys. Educ. Res. **19**, 020155 (2023).
[58] G. Kortemeyer, W. Bauer, and W. Fisher, Hybrid teaching: A tale of two populations, Phys. Rev. Phys. Educ. Res. **18**, 020130 (2022).
[59] I. Marzoli, A. Colantonio, C. Fazio, M. Giliberti, U. Scotti di Uccio, and I. Testa, Effects of emergency remote instruction during the COVID-19 pandemic on university physics students in Italy, Phys. Rev. Phys. Educ. Res. **17**, 020130 (2021).
[60] G. Kortemeyer, C. Kortemeyer, and W. Bauer, Taking introductory physics in studio, lecture, or online format: What difference does it make in subsequent courses, and for whom?, Phys. Rev. Phys. Educ. Res. **19**, 020148 (2023).
[61] N. G. Holmes, B. Keep, and C. E. Wieman, Developing scientific decision making by structuring and supporting student agency, Phys. Rev. Phys. Educ. Res. **16**, 010109 (2020).
[62] P. Eaton, K. Vavruska, and S. Willoughby, Exploring the preinstruction and postinstruction non-Newtonian world views as measured by the Force Concept Inventory, Phys. Rev. Phys. Educ. Res. **15**, 010123 (2019).
[63] J. M. Aiken, R. Henderson, and M. D. Caballero, Modeling student pathways in a physics bachelor's degree program, Phys. Rev. Phys. Educ. Res. **15**, 010128 (2019).
[64] J. Stewart, J. Hansen, and E. Burkholder, Visualizing and predicting the path to an undergraduate physics degree at two different institutions, Phys. Rev. Phys. Educ. Res. **18**, 020117 (2022).
[65] W. K. Black, R. L. Matz, M. Mills, and A. E. Evrard, Quantifying grade benefits of study: Practice makes better, Phys. Rev. Phys. Educ. Res. **19**, 010139 (2023).
[66] Z. Richards and A. M. Kelly, Predicting community college astronomy performance through logistic regression, Phys. Rev. Phys. Educ. Res. **19**, 010119 (2023).
[67] A. B. Simmons and A. F. Heckler, Grades, grade component weighting, and demographic disparities in introductory physics, Phys. Rev. Phys. Educ. Res. **16**, 020125 (2020).
[68] D. L. Vargas, A. M. Bridgeman, D. R. Schmidt, P. B. Kohl, B. R. Wilcox, and L. D. Carr, Correlation between student collaboration network centrality and academic performance, Phys. Rev. Phys. Educ. Res. **14**, 020112 (2018).
[69] A. Traxler, A. Gavrin, and R. Lindell, Networks identify productive forum discussions, Phys. Rev. Phys. Educ. Res. **14**, 020107 (2018).
[70] A. L. Traxler, A. Gavrin, and R. Lindell, Coursenetworking and community: Linking online discussion networks and course success, in *Physics Education Research Conference Proceedings* (2016).
[71] M. Dew, E. Hunt, V. Perera, J. Perry, G. Ponti, and A. Loveridge, Group dynamics in inquiry-based labs: Gender

inequities and the efficacy of partner agreements, Phys. Rev. Phys. Educ. Res. **20**, 010121 (2024).

[72] M. Sundstrom, D. G. Wu, C. Walsh, A. B. Heim, and N. G. Holmes, Examining the effects of lab instruction and gender composition on intergroup interaction networks in introductory physics labs, Phys. Rev. Phys. Educ. Res. **18**, 010102 (2022).

[73] A. R. Warren, Network analysis of social interactions in laboratories, in *AIP Conference Proceedings*, Vol. 1064 (2008) pp. 219–222.

[74] L. Timonen, K. Juuti, and S. Harmoinen, Learning enhancing emotions predict student retention: Multilevel emotions of finnish university physics students in and outside learning situations, Phys. Rev. Phys. Educ. Res. **18**, 020126 (2022).

[75] P. Kelley, A. Gavrin, and R. S. Lindell, Text mining online discussions in an introductory physics course, in *Physics Education Research Conference Proceedings* (2017).

[76] K. N. Quinn, M. M. Kelley, K. L. McGill, E. M. Smith, Z. Whipps, and N. G. Holmes, Group roles in unstructured labs show inequitable gender divide, Phys. Rev. Phys. Educ. Res. **16**, 010129 (2020).

[77] Z. Hazari, R. Dou, G. Sonnert, and P. M. Sadler, Examining the relationship between informal science experiences and physics identity: Unrealized possibilities, Phys. Rev. Phys. Educ. Res. **18**, 010107 (2022).

[78] R. K. Fussell, E. M. Stump, and N. G. Holmes, Method to assess the trustworthiness of machine coding at scale, Phys. Rev. Phys. Educ. Res. **20**, 010113 (2024).

[79] P. Tschisgale, P. Wulff, and M. Kubsch, Integrating artificial intelligence-based methods into qualitative research in physics education research: A case for computational grounded theory, Phys. Rev. Phys. Educ. Res. **19**, 020123 (2023).

[80] J. M. Aiken, R. De Bin, H. J. Lewandowski, and M. D. Caballero, Framework for evaluating statistical models in physics education research, Phys. Rev. Phys. Educ. Res. **17**, 020104 (2021).

[81] R. Dou and J. P. Zwolak, Practitioner's guide to social network analysis: Examining physics anxiety in an active-learning setting, Phys. Rev. Phys. Educ. Res. **15**, 020105 (2019).

[82] R. P. Dalka, D. Sachmpazidi, C. Henderson, and J. P. Zwolak, Network analysis approach to Likert-style surveys, Phys. Rev. Phys. Educ. Res. **18**, 020113 (2022).

[83] L. Ding and R. Beichner, Approaches to data analysis of multiple-choice questions, Phys. Rev. ST Phys. Educ. Res. **5**, 020103 (2009).

[84] N. T. Young, N. Verboncoeur, D. C. Lam, and M. D. Caballero, Rubric-based holistic review represents a change from traditional graduate admissions approaches in physics, Phys. Rev. Phys. Educ. Res. **19**, 010134 (2023).

[85] J. Doyle and G. Potvin, In search of distinct graduate admission strategies in physics: An exploratory study using topological data analysis, in *Physics Education Research Conference Proceedings* (2015).

[86] M. Verostek, C. W. Miller, and B. Zwickl, Analyzing admissions metrics as predictors of graduate GPA and whether graduate GPA mediates Ph.D. completion, Phys. Rev. Phys. Educ. Res. **17**, 020115 (2021).

[87] N. J. Mikkelsen, N. T. Young, and M. D. Caballero, Investigating institutional influence on graduate program admissions by modeling physics graduate record examination cutoff scores, Phys. Rev. Phys. Educ. Res. **17**, 010109 (2021).

[88] N. N. Young and M. D. Caballero, Using machine learning to understand physics graduate school admissions, in *Physics Education Research Conference Proceedings* (2019).

[89] T. O. B. Odden, A. Marin, and M. D. Caballero, Thematic analysis of 18 years of physics education research conference proceedings using natural language processing, Phys. Rev. Phys. Educ. Res. **16**, 010142 (2020).

[90] K. Commeford, E. Brewe, and A. Traxler, Characterizing active learning environments in physics using latent profile analysis, Phys. Rev. Phys. Educ. Res. **18**, 010113 (2022).

[91] G. Kestin and K. Miller, Harnessing active engagement in educational videos: Enhanced visuals and embedded questions, Phys. Rev. Phys. Educ. Res. **18**, 010148 (2022).

[92] T. Wan, C. M. Doty, A. A. Geraets, C. A. Nix, E. K. H. Saitta, and J. J. Chini, Evaluating the impact of a classroom simulator training on graduate teaching assistants' instructional practices and undergraduate student learning, Phys. Rev. Phys. Educ. Res. **17**, 010146 (2021).

[93] T. Wan, C. M. Doty, A. A. Geraets, E. K. Saitta, and J. J. Chini, Characterizing graduate teaching assistants' teaching practices in physics "mini-studios", in *Physics Education Research Conference Proceedings* (2019).

[94] A. Werth, B. Pollard, R. Hobbs, and H. J. Lewandowski, Investigating changes in student views of measurement uncertainty in an introductory physics lab course using clustering algorithms, Phys. Rev. Phys. Educ. Res. **19**, 020146 (2023).

[95] Z. Felker and Z. Chen, Reducing procrastination on introductory physics online homework for college students using a planning prompt intervention, Phys. Rev. Phys. Educ. Res. **19**, 010123 (2023).