Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 29 Number 11, July 2024 ISSN 1531-7714

Discovering Educational Data Mining: An Introduction

Zachary K. Collier, *University of Connecticut*
Joshua Sukumar, *University of Connecticut*
Roghayeh Barmaki *University of Delaware*

This article introduces researchers in the science concerned with developing and studying research methods, measurement, and evaluation (RMME) to the educational data mining (EDM) community. It assumes that the audience is familiar with traditional priorities of statistical analyses, such as accurately estimating model parameters and inferences from those models. Instead, this article focuses on data mining's adoption of statistics and machine learning to produce cutting-edge methods in educational contexts. It answers three questions: (1) What are the primary interests of EDM and RMME researchers? (2) What is their discipline-specific vocabulary? and (3) What are the similarities and differences in how the EDM and RMME communities analyze similar types of data?

Keywords: Educational Measurement and Evaluation, Educational Data Mining, Disciplinary Research Comparison, Causal Inference, Prediction.

Introduction

Until recently, students from colleges of education interested in data analysis took research methods, measurement, and evaluation (RMME) courses. These courses are no longer the "only game in town" (Friedman, 1998). Computer science courses, notably data mining, may be more attractive for developing a young researcher's analytic talent. For example, educational data mining (EDM) and learning analytics are promising alternatives to making inference and drawing conclusions about a population based on a sample of educational data (R. S. Baker & Inventado, 2014; Slater et al., 2017). RMME programs typically focus on three statistics-dependent subject areas (1) Educational Measurement, (2) Statistics, and (3) Program Evaluation (Randall et al., 2021). These programs go by different names, including but not limited to (1) Educational Statistics and Research Methods, (2) Research, Educational Measurement, and Psychometrics, and (3) Research and Evaluation Methodology.

Although RMME programs are typically housed in colleges of education, and most EDM instructors are computer scientists, both communities share a common use of advanced computational methods, statistical techniques, and machine learning algorithms to analyze large-scale educational datasets. Despite this shared interest, these communities often do not take courses together or attend the same conferences. However, advancements of EDM techniques are being supported by nationally recognized initiatives. For example, the Georgia Institute of Technology leads the National Artificial Intelligence (AI) Institute for Adult Learning and Online Education (Aialoe, 2024), with the goal of enhancing adult online education via AI. Furthermore, the National Science Foundation (NSF) supported CIRCLS, or the Center for Integrative Research in Computing and Learning Sciences, to support the community and explore future learning technologies. Within the CIRCLS community, the largest proportion of members contribute to "AI Approaches and Technologies" out of all the expertise areas currently listed on their website (Center for Integrative Research in Computing and Learning Sciences, 2022). Potentially, these initiatives demonstrate the relevance and timeliness of RMME research and exemplify significant national interest in the growth and application of EDM, especially related to future learning tools available to students. Additionally, a policy report from the Office of Educational Technology further emphasizes the growing importance of AI tools in enhancing educational outcomes (Office of Educational Technology, 2023). Furthermore, recent NSF awards (e.g., AI Institute awards of 2112635[1] and 2229612[2]) underscore the national investment and validation of research areas, not only highlighting the significance of AI in education, but also indicate a national trend toward data-driven educational strategies, which this article explores.

Purpose

This article contributes to the field by providing a more nuanced and comprehensive understanding of the current state of educational data science. It goes beyond a simple summary of the literature and offers insights that can inform future research and practice in the field. Specifically, we identify gaps between RMME and EDM methodologies, as well as potential areas for innovation and improvement in methodological approaches.

The selection of topics covered in this article will be somewhat informal; rather than attempting to encompass every study ever conducted by EDM researchers, or every tool ever created and used by a single research group, we will focus on delineating the primary differences and similarities across the fields. Therefore, some niche areas of a community may be excluded. We nonetheless hope that this review will provide useful information to researchers new to EDM with a more nuanced and comprehensive understanding of the current state of the field.

Areas of Educational Data Science

Figure 1 is an original illustration developed by the authors to display the overlap of different disciplines in educational data science. Sweeping over an extensive area, educational data science "is an umbrella for a fleet of new computational techniques being used to identify new forms of data, measures, descriptives, predictions, and experiments in education" (McFarland et al., 2021). Our manuscript extends from existing research, refining the definition of educational data science as a specialized application of data science within education fields (e.g., RMME and EDM). This entails working with data collected from educational environments/settings to effectively address and solve pertinent educational challenges.

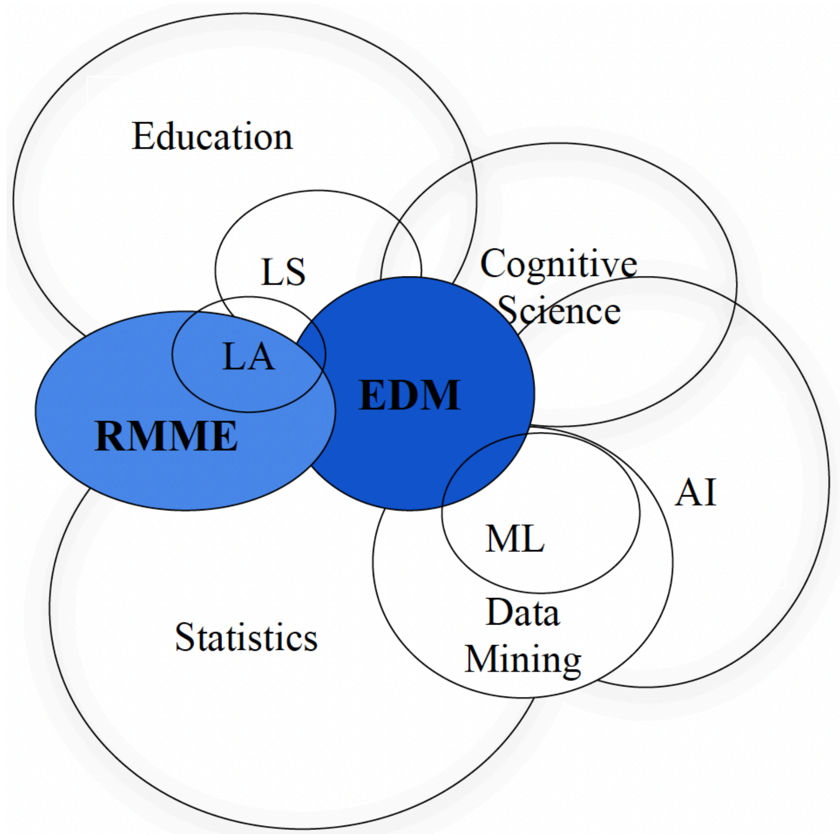
While previous research considers educational data science as a fusion of four to seven fields (Piety et al., 2014), Figure 1 illustrates our expanded conceptualization, encompassing a combination of 10 communities within educational data science:

- *Education*: Education is a broad field encompassing various disciplines, theories, and practices aimed at facilitating learning and development in individuals and communities (Brady et al., 2023). It includes educational psychology, curriculum development, pedagogy, instructional design, and educational leadership, among others. Education seeks to understand how people learn, develop effective teaching strategies, and create supportive learning environments. It plays a critical role in shaping individuals' knowledge, skills, attitudes, and behaviors, contributing to personal growth and societal progress.

- *Statistics*: In Figure 1, statistics is shown as a stand-alone field, blended into other areas of data science, such as RMME and data mining. There is overlap in that there is a shared use of statistics in all data science fields; the difference is how these concepts are applied and the overall goal of the research within that field. The field of statistics itself dates back hundreds of

[1] https://www.nsf.gov/awardsearch/showAward?AWD_ID=2112635&HistoricalAwards=false
[2] https://www.nsf.gov/awardsearch/showAward?AWD_ID=2229612&HistoricalAwards=false

Figure 1. Diagram of Fields that Support Educational Data Science



Abbreviations: AI (Artificial Intelligence), ML (Machine Learning), RMME (Research Methods, Measurement, and Evaluation), EDM (Educational Data Mining), LA (Learning Analytics), LS (Learning Science).

Note: The main emphasis is on the overlapping regions, emphasizing that the circle sizes do not accurately depict the boundaries or significance of respected disciplines.

years and studying "statistics" can be seen as establishing a wide breadth of knowledge in its broad range of theories and tests (Stigler, 1986).

- *Learning Analytics*: Learning Analytics involves considering and analyzing data at the micro or 'learning' level to uncover insights into teaching methods and academic interventions that are most likely to enhance the learning of specific content for individual learners (Romero & Ventura, 2020; Romero & Ventura, 2010, 2017). It focuses on leveraging data to understand and improve the learning process, often in real-time, to make informed decisions that benefit learners and educators (Rodrigues et al., 2018). Learning analytics applies various data analytic techniques from other disciplines, including RMME and EDM.
- *Learning Science*: Learning science is dedicated to the systematic investigation of learning and teaching, encompassing human development and educational technology (Sommerhoff et al, 2018). It involves applying research findings to design innovative educational interventions. Drawing on expertise from developmental psychology, educational psychology, cognitive science, educational technology, and special education, this field scrutinizes significant educational and developmental challenges (Packer & Maddox, 2016).
- *Cognitive Science*: Cognitive science represents an interdisciplinary exploration of the mind and intelligence, encompassing philosophy, psychology, artificial intelligence, neuroscience, linguistics, and anthropology (Thagard, 2013). At its core, cognitive science posits that

understanding thinking processes is most effectively achieved by examining the representational structures within the mind and the computational procedures that manipulate these structures.

- *Research Methods, Measurement, and Evaluation*: RMME involves the development, evaluation, and application of methods and tools for assessing knowledge, skills, abilities, and other educational outcomes. It encompasses a range of techniques such as testing, assessment design, psychometrics, and data analysis to inform educational practices, policy decisions, and improve the effectiveness of teaching and learning processes. (Randall et al., 2021).
- *Educational Data Mining*: EDM encompasses various techniques for mining largescale educational data, including the discovery of patterns and trends in educational contexts. By using computational methods to analyze educational data, EDM can provide insights into how different factors, such as student background, learning styles, and instructional methods can influence learning outcomes (Rodrigues et al., 2018). These insights can then be used to develop more personalized and effective learning experiences for students.
- *Artificial Intelligence/Machine Learning/Data Mining*: Within the broader domain of AI, machine learning (ML) is a subfield that has garnered significant attention. In EDM, ML plays a crucial role as it enables computers to learn and produce behaviors not explicitly programmed. While AI encompasses general problem-solving capabilities, ML, in the educational context, focuses on tasks such as understanding student behavior and predicting learning outcomes. It is important to note the distinction between ML and data mining; while often used interchangeably, they serve different purposes. Data mining, being broader, involves both data management and analyses, aiming to discover patterns and relationships (Shu & Ye, 2023). ML, on the other hand, is directed at developing predictive models for making accurate predictions on new educational data (Sarker, 2021).

What are the Primary Interests of EDM and RMME Researchers?

The primary interests for RMME researchers typically entail (1) testing the relationships between independent and dependent variables (e.g., linear) and testing the significance of those relationships, 2) measuring learning and related constructs, 3) evaluating the effectiveness of an educational program or intervention, and 4) developing and refining assessment techniques and designing robust evaluation frameworks. In contrast, EDM researchers primarily focus on the development and application of prediction, clustering, and relationship mining methods (Baker et al., 2010).

RMME goals benefit the EDM community, and vice versa, fostering mutual enrichment and progress. For instance, program evaluation is an important aspect of the RMME field that employs diverse qualitative and quantitative data collection methods, such as surveys, focus groups, and interviews, to assess the effectiveness of educational programs. By drawing upon the expertise of program evaluators, EDM researchers can ensure that their insights obtained from large datasets are relevant and applicable to real-world educational contexts. In turn, program evaluators can utilize EDM techniques to analyze data from large-scale programs and uncover patterns that may not be immediately evident through traditional RMME methods. The integration of program evaluation and EDM methods can lead to a more comprehensive understanding of educational outcomes and the effectiveness of educational programs, benefiting both communities.

What are their Approaches to Data Analysis?

Hypothesis Testing

Data mining and statistics are two fields that involve analyzing data, but they differ in their approaches and goals. Traditional statistical methods, such as regression analysis, have been used by both data miners and statisticians for decades. However, data mining algorithms often link variables and determine the model's functional form, which can help researchers determine distributions and discover patterns. In contrast, statisticians typically start with models based on hypotheses and assumptions about the data distribution. They may also use non-parametric (distribution free) approaches, such as X^2, to analyze correlations among observed events.

Data miners utilize various tools such as association mining and cluster analysis to uncover patterns and relationships within datasets. While association mining doesn't adhere to traditional hypothesis testing, it remains a pivotal tool for exploratory data analysis. For instance, it can shed light on educational data patterns like the correlation between student absences and academic performance. Researchers might employ association rule mining techniques to establish rules capturing the connection between student attendance patterns and academic achievements, such as the indication that "students with frequent absences are more likely to have lower grades." Subsequently, they evaluate the confidence[3] and lift for each rule to gauge its strength and significance, with higher confidence values indicating stronger predictive power, while lift measures the degree of dependence beyond random chance. To refine the analysis, thresholds are applied to filter out rules with low confidence or lift values, prioritizing those most relevant for further exploration. By interpreting these association rules, researchers gain valuable insights into the relationship between student absences and academic performance, potentially formulating hypotheses that can be subjected to traditional statistical testing methods.

For decades, RMME researchers, particularly psychometricians, have applied clustering algorithms to detect test collusion, which is cheating on tests (Sinharay, 2017). Clustering groups based on shared characteristics or patterns, allow for meaningful insights to be drawn from the data without the need for a pre-defined hypothesis. Typically, EDM and RMME researchers use clustering algorithms for exploratory/unsupervised data analysis. Exploratory data analysis is a method that is employed to discover patterns, relationships, and potential hypotheses about the data, while confirmatory data analysis is focused on validating a specific hypothesis by testing it against empirical data (Marcoulides, 1993). Clustering can be employed in both contexts. For EDM researchers, clustering is as a data-driven approach to gain insights from the data without drawing any causal or correlational conclusions. Clustering is also one of the core topics in a multivariate statistics course, taught by RMME and statistics professors.

Predictive Modeling

Predictive modeling encompasses a spectrum of approaches, ranging from interpretable models to "black box" models, a concept of interest to both EDM and RMME researchers. While black box models pose challenges by obscuring the understanding of variable interactions, they are not the sole representation of predictive modeling. Specifically, black box models employ complex functions that exceed human comprehension regarding variable relationships. RMME researchers, often collaborating with education researchers, prioritize transparency in understanding variable relationships to develop new methodologies and insights into learning (Sawyer, 2005). Hence, many RMME researchers prefer models that offer clear explanations of how variables are interconnected. For instance, linear models, where a few variables are weighted and combined, are commonly used. In contrast, EDM researchers tend to explore models that may offer higher accuracy but are not inherently interpretable in a straightforward manner. For example, weights in deep neural networks (DNNs) represent the parameters learned during the training process, which adjust the strength of connections between neurons in different layers of the network to optimize the model's performance on a given task (Collier & Leite, 2020; Collier et al., 2022). Despite the challenges in interpreting individual weights, EDM has developed various techniques and tools to improve the interpretability of DNNs. These include visualization methods, feature attribution techniques, and model explanation frameworks designed to shed light on the inner workings of DNNs and enhance our understanding of their decision-making processes (Adadi & Berrada, 2018).

In the realm of model interpretability, both communities share a vested interest, prompting a burgeoning field known as *explainable AI* or "XAI". XAI constitutes a subset of AI dedicated to crafting machine learning models capable of offering

[3] In the context of association mining and association rule learning, "confidence" refers to a measure of the reliability or strength of an association rule. It is not the same as "confidence intervals" in traditional statistical analysis.

transparent and interpretable explanations for their decision-making processes. This is particularly relevant when considering algorithms such as neural networks-based or deep learning-based models, which often result in black box models. These models, characterized by their complex and opaque decision-making mechanisms, underscore the need for XAI to shed light on the inner workings of AI systems. In essence, XAI strives to cultivate AI systems that provide clear and understandable reasons for the predictions or decisions they make.

Differences in Classification Frameworks

Table 1 lists some analytic methods and then classifies them as being either "unsupervised" or "supervised" learning approaches. It is important to note these distinctions because both EDM and RMME communities use these methods. However, the classification of "unsupervised" or "supervised" learning is mainly emphasized in EDM journals (e.g., Journal of Educational Data Mining) but not as frequently in RMME journals (e.g., Journal of Education and Behavioral Statistics). Supervised learning has clear labels and bases its results off a predetermined attribute, and the algorithm stops once an acceptable level of performance is achieved (Berry et al., 2019). An example of a supervised learning approach is linear regression because there is a dependent variable of interest that is based on the results of predetermined independent variables. The stopping point for linear regression is when the model finds the relationship between the independent and dependent variables. On the other hand, unsupervised learning operates in the absence of a target attribute, involves pattern recognition, and inherent groupings are identified via the algorithm which can then be used for supervised learning processes (Berry et al., 2019).

An example of an unsupervised learning technique used by both RMME and EDM researchers is principal component analysis or PCA. PCA is used primarily as a data reduction technique to take a large multi-dimensional data space and reduce it down to something that is of a smaller dimension that still captures most of the relevant variability. There is no dependent variable of interest when conducting the PCA. The results can be fed into a supervised statistical procedure such as multivariate regression, but the PCA itself has no dependent variable.

Most similar to "supervised" and "unsupervised" learning distinctions, RMME published works primarily focus on "parametric" and "non-parametric" distinctions. Non-parametric statistical models have no assumptions about the shape or parameters of the sample's population distribution, whereas parametric models make assumptions about the sample's population distribution. Supervised learning models are not, by default, parametric models. For example, PCA is an unsupervised learning technique that does not require normality for the extraction of components. A PCA only seeks to optimally describe data by using (sparse) data points in a (high-dimensional) space. Therefore, PCA is both an unsupervised learning approach and a non-parametric model.

Software. As the two communities evolve, both EDM and RMME researchers increasingly employ

Table 1. Methods and Examples of Supervised & Unsupervised Learning

Unsupervised Learning	
Method	**Example**
Clustering	K-Means Clustering
Dimensionality	
Reduction	Principal Components Analysis
Unsupervised Learning	
Method	**Example**
Regression	Linear Regression
Classification	Logistic Regression
Deep Learning	Neural Networks

statistical methods like PCA, Random Forest, and Bayes Network Classifiers. Nonetheless, the specific tools they utilize to implement these methods often vary. For instance, Hussain et al. (2018) utilized Weka, a widely used suite of machine learning software tools known as the "Waikato Environment for Knowledge Analysis." Conversely, Suk and Han (2023) employed R to conduct PCA. R is a programming language and open-source software environment specifically designed for statistical computing and data analysis (R Core Team, 2022). This variability in software[4] choice may reflect differences in preferences, expertise, and specific requirements of each research community.

EDM tools are closely integrated with educational research and practice, aiming to inform instructional design, student assessment, and learning interventions (Slater et al., 2017).RMME tools contribute to the methodological aspects of educational research, helping researchers design rigorous studies, select appropriate measurement instruments, and analyze data effectively. It is important to note that there is some overlap in the types of analyses conducted by EDM and RMME researchers. However, the tools they use differ in terms of purpose, functionality, specialization, and integration with educational research. For a comprehensive overview of software commonly utilized by EDM researchers, refer to Slater et al. (2017), which provides a review of 40 tools frequently employed for data mining and analytics in the field of education.

Assessments and Skill Mastery

Psychometrics. Within the field of RMME, researchers delve into measurement theory courses that underpin their understanding of latent variables influencing observable variables. In these courses, RMME researchers explore classical test theory, item response theory, and generalizability theory.

RMME professionals specializing in psychometrics employ a systematic approach to assess whether a test taker has mastered a particular skill. The utilization of assessments and tests tailored for this purpose involves the following key components (Crocker & Algina, 1986):

- *Test Development*: Psychometricians engage in developing tests meticulously crafted to accurately measure the specific skill under consideration. This process entails creating questions or tasks aligned with the content and objectives of the skill being assessed.
- *Reliability*: Psychometric assessments strive for reliability, ensuring consistent results over time. Achieving this requires rigorous testing and statistical analysis to validate that the test reliably measures the intended skill.
- *Validity*: Validity holds paramount importance in psychometrics. A test is considered valid if it accurately measures the targeted skill. Various validity types, including content, construct, and criterion-related, undergo examination to ensure the test's appropriateness.
- *Scoring*: Psychometricians develop scoring methods to precisely quantify a test taker's performance. This may involve assigning numerical scores or categorizing performance levels based on predefined criteria.
- *Norms and Benchmarking*: Test results are compared to established norms or benchmarks, aiding in understanding a test taker's performance relative to a broader group that has undergone the same assessment.
- *Item Analysis*: Individual test items undergo meticulous analysis to gauge their effectiveness in measuring the targeted skill. This iterative process contributes to refining and improving the overall test over time.
- *Feedback and Reporting*: Test results, coupled with constructive feedback on strengths and weaknesses, are provided to test takers.

By adhering to these psychometric principles, practitioners in RMME strive to ensure that assessments are not only rigorous but also fair, accurate, and meaningful in determining whether a test

[4] In the field of EDM, R and Python are extensively utilized. Weka, on the other hand, may be perceived as an older and less potent tool in comparison to R and Python

taker has genuinely mastered a specific skill. This comprehensive exploration contributes to the ongoing discourse on robust skill assessment methodologies within the realm of RMME.

Bayesian Knowledge Tracing. Bayesian Knowledge Tracing or BKT (see Figure 2) is an AI algorithm that lets EDM researchers infer a student's current knowledge state to predict if they have learned a skill (student skill acquisition) (Slater & Baker, 2018). In a recent article (Bulut et al., 2023), the authors highlight several similarities between IRT and BKT. Both approaches aim to understand and predict learners' knowledge and performance based on their responses to educational tasks. Additionally, both involve estimating parameters that describe learners' abilities and item characteristics.

There are four key parameters of BKT, each in the [0-1] range:

- Pr(known) or $p(L_0)$: the probability that the student already knew a skill
- Pr(will learn) or p(T): the probability that the student will learn a skill on the next practice opportunity
- Pr(slip) or p(S): the probability that the student will answer incorrectly despite knowing a skill
- Pr(guess) or p(G): the probability that the student will answer correctly despite not knowing a skill.

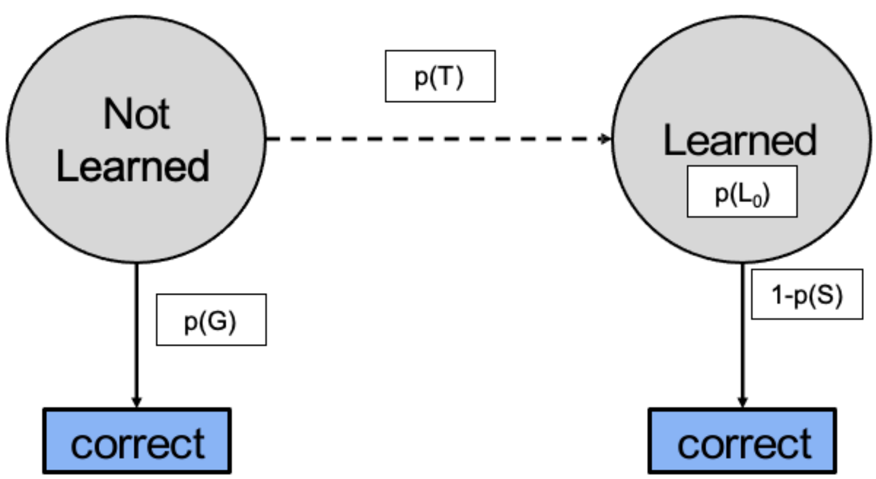
After each student response, the BKT algorithm calculates the probability of the student having learned a particular skill, denoted as P(learned), based on the current values of its parameters. The formula for calculating P(learned) varies depending on whether the student's answer was correct or incorrect. In this way, the BKT algorithm updates the student's knowledge state over time, allowing for personalized and adaptive learning experiences.

While both psychometric approaches and Bayesian Knowledge Tracing offer valuable insights into modeling student learning and assessing their mastery of skills, they each come with their own set of strengths and limitations.

What is their discipline-specific vocabulary?

Table 2 presents some common terminology used in data mining. This list is not exhaustive but highlights terms that have slightly different and even completely different meanings than terms familiar to RMME researchers (presented in Table 3). For example, "generalizability" in RMME measures how well one's experimental findings from a sample extend to the population. Additionally, "generalizability theory" is a framework for estimating measurement reliability and understanding the various sources of measurement errors (Marcoulides, 1993). In data mining, the understanding is similarly sample dependent, yet more

Figure 2. Bayesian Knowledge Tracing (diagram reproduced from (Slater & Baker, 2018)



Note: The BKT framework estimates a binary latent variable. That is, mastery of a skill or topic.

Table 2. Key Terminology for EDM Researchers

Term	Meaning
Generalizability	Model's ability to adapt properly to new, previously unseen data
Recursive Model	A neural network that processes the variables in a hierarchical structure
Learning	Training models (estimating their parameters) based on existing data
Test Error	The difference between the predictions and the observed values on a test data set
Regressor	Algorithm that predicts a continuous outcome based on the value of predictor variables
Classifier	Algorithm used to assign a class label to a data input
Knowledge Discovery	Process of identifying novel pattern or knowledge in data
Supervised Learning	Model with clear labels and bases its results off a predetermined attribute
Unsupervised Learning	Model operates in the absence of a target attribute
Sampling Methods	Probabilistic methods seeking to avoid selection bias in training data

Note. The terminology in the table above was acquired from several sources (Collier et al., 2022; Fayyad et al., 1996; Kushwah et al., 2021)

Table 3. Key Terminology for RMME Researchers

Term	Meaning
Generalizability	Measures how well one's experimental findings from a sample extend to the population
Measurement	The process of assessing an individual's ability, trait, or attribute through the use of tests or assessments
Test Error	The difference between a test score and a student's actual knowledge and ability
Regression	A set of statistical procedures relating independent variables to a dependent variable
Non-parametric Models	Models with no assumptions about the shape of the population distribution
Parametric Models	Models that make assumptions about the distribution of the population
Quasi-experimental	Intervention seeking to estimate causal effects without randomization of participants
Experimental	Intervention where participants are randomly assigned
Reliability	The degree to which a test or measurement procedure produces consistent and stable results over repeated administrations or under different conditions
Validity	The extent to which a test or assessment measures what it is intended to measure

Note. The terminology in the table above was acquired from several sources (Grimm et al., 2016; Marcoulides, 1993; Mueller & Hancock, 2018)

focus is on the model's ability to adapt (or generalize) to new data (Glavatskikh et al., 2019).

What are the similarities and differences in how the EDM and RMME communities analyze similar types of data?

Disclaimer: Finding two separate studies from distinct communities that are chronologically and systematically comparable is a challenging task. Following extensive research, we have selected McArdle et al. (2013) and Yanagiura et al. (2023) as our reference points. This comparison underscores the broader educational endeavor of leveraging data analysis to support student success, albeit through distinct analytical frameworks.

The comparison between the RMME (McArdle et al., 2013) and EDM (Yanagiura et al., 2023) studies sheds light on how the EDM and RMME communities approach similar research goals with varying methodologies and objectives. McArdle et al. (2013) employ multilevel multivariate analysis to predict GPA for first-year college students, focusing on hierarchical data structures and variable relationships across different levels. In contrast, Yanagiura et al. (2023) utilize machine learning algorithms to forecast first-term college GPA, emphasizing algorithmic fairness and the relevance of non-academic skills in predictive analytics. While McArdle et al. (2013) delve into statistical models such as multilevel analysis, Yanagiura et al. (2023) address algorithmic fairness and predictive analytics, highlighting the diverse approaches within the educational research landscape.

Both studies utilize empirical datasets but pursue distinct methodological objectives. McArdle et al. (2013) aim to demonstrate the practical utility and limitations of multilevel models in standard validation studies, offering insights for future research directions. Conversely, Yanagiura et al. (2023) seek to provide insights into developing more equitable early warning systems in higher education.

In terminology and concepts, McArdle et al. (2013) introduce educational measurement terms such as predictive validity, multiple linear regression, multilevel models, variance components, and random coefficients. Predictive validity assesses a measure's ability to predict future outcomes such as college grades. Multiple linear regression examines the relationship between one outcome variable and multiple predictors. Multilevel models account for hierarchical data structures, estimating individual and group effects and variances. Variance components explain the proportion of total variance in an outcome variable due to different sources, while random coefficients indicate variations in regression coefficients across groups or contexts. Conversely, Yanagiura et al. (2023) explore EDM terms including Early Warning Systems (EWS), algorithmic fairness, classification parity, and calibration. EWS predicts students' college success risks and intervenes early. Algorithmic fairness ensures models do not discriminate against specific groups, with classification parity and calibration ensuring equal prediction accuracy and risk assessment across protected groups, respectively.

Despite methodological differences, both investigations highlight grade point average (GPA), student performance, and academic achievement, emphasizing their shared goal of enhancing educational outcomes. McArdle et al. (2013) seek to identify factors affecting academic performance through statistical analysis, while Yanagiura et al. (2023) examine the efficacy and fairness of AI models in GPA prediction, addressing the ethical implications of early warning systems and the potential insights from non-academic data.

In terms of variables and features, McArdle et al. (2013) analyze college students who are also National Collegiate Athletic Association (NCAA) student-athletes, considering variables such as high school academic records, college characteristics, and student demographics. High school variables include core GPA, ACT or SAT scores and core units taken. College variables encompass freshman GPA, credits, quality points, graduation rate, cost, and public/private status, while student demographics cover gender and ethnicity. Conversely, Yanagiura et al. (2023) focus on different variables, including demographic and pre-college academic data such as gender, age, department, entrance examination type, high school rank, GPA, and achievement test scores. They also examine non-academic skills through the PROG test, which measures interpersonal, task execution, and self-control skills. Their study aims to predict first-term GPA, classifying GPAs below 2.0 as low-performing.

Conclusion

This article aims to introduce RMME researchers to the EDM community and to provide a synthesis of perspectives between them. It recognizes that each of these communities has its own unique language, practices, and perspectives on educational data, which makes bridging them a challenging task. While this article did not seek to (and cannot) cover all aspects of these fields, it provides a fundamental understanding of their concepts, methodologies, and terminologies.

Future Research

Additionally, we propose several directions for future research in RMME and EDM. One potential avenue is the exploration of advanced modeling techniques, such as deep learning, to improve knowledge tracing models (Bulut et al., 2023). Another promising direction involves the integration of models, specifically combining BKT with IRT to enhance accuracy and interoperability. Additionally, it is crucial to study how educational context and teaching methods impact learners' knowledge states and model performance. Personalized learning can be advanced by developing models that provide individualized interventions and support based on learner data.

Furthermore, it is important to engage educators in the development, implementation, and decision-making processes of AI in education. This collaboration will help ensure the models are practical and effective in real-world settings. Emphasis should also be placed on research and development focused on adapting AI models to diverse educational contexts and enhancing trust and safety (Fenu et al., 2022). Developing education-specific guidelines and guardrails for responsible AI use is another critical area. Finally, the application of automated machine learning should be further developed to reduce the computational burden of training deep learning models in educational research (Collier et al., 2022).

Through our investigation, we have identified key findings that highlight the potential benefits of connecting researchers in these fields to improve educational outcomes. EDM can provide valuable insights into large-scale data analysis, while RMME can contribute to the development and validation of assessments and measurement tools and ensure that the insights obtained from large datasets are relevant and applicable to real-world educational contexts. By exploring the intersection of these fields, we can work towards a more data-informed and effective educational system.

References

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE access*, *6*, 52138–52160.

Aialoe website [Accessed: July 16, 2024]. (n.d.). https://aialoe.org/

Baker, R. (2015). *Big data and education*. New York: Teachers College, Columbia University.

Baker, R., et al. (2010). Data mining for education. *International encyclopedia of education*, *7*(3), 112–118.

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics* (pp. 61–75). Springer.

Berry, M. W., Mohamed, A., & Yap, B. W. (2019). *Supervised and unsupervised learning for data science*. Springer.

Brady, A. C., Griffin, M. M., Lewis, A. R., Fong, C. J., & Robinson, D. H. (2023). How scientific is educational psychology research? the increasing trend of squeezing causality and recommendations from non-intervention studies. *Educational Psychology Review*, *35*(1), 37.

Bulut, O., Shin, J., Yildirim-Erbasli, S. N., Gorgun, G., & Pardos, Z. A. (2023). An introduction to Bayesian knowledge tracing with pybkt. *Psych*, *5*(3), 770–786.

Center for Innovative Research in Cyberlearning. (2022). *Partnerships for Change: Existing Partnership Trends in the CIRCLS Community*. for-change-existing-partnership-trends-in-the-circls-community

Collier, Z. K., & Leite, W. L. (2020). A tutorial on artificial neural networks in propensity score analysis. *The Journal of Experimental Education*, 1–18.

Collier, Z. K., Zhang, H., & Liu, L. (2022). Explained: Artificial intelligence for propensity score estimation in multilevel educational settings. *Practical Assessment, Research, and Evaluation*, *27*(1), 3.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. ERIC.

da Silva, M. A., Liu, R., Huggins-Manley, A. C., & Bazán, J. L. (2019). Incorporating the q-matrix into multidimensional item response theory models. *Educational and Psychological Measurement*, *79*(4), 665–687.

Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., et al. (1996). Knowledge discovery and data mining: Towards a unifying framework. *KDD*, *96*, 82–88.

Fenu, G., Galici, R., & Marras, M. (2022). Experts' view on challenges and needs for fairness in artificial intelligence for education. *International Conference on Artificial Intelligence in Education*, 243–255.

Friedman, J. H. (1998). Data mining and statistics: What's the connection? *Computing science and statistics*, *29*(1), 3–9.

Glavatskikh, M., Leguy, J., Hunault, G., Cauchy, T., & Da Mota, B. (2019). Dataset's chemical diversity limits the generalizability of machine learning predictions. *Journal of cheminformatics*, *11*(1), 1–15.

Granville, V. (2014). *Developing analytic talent: Becoming a data scientist*. John Wiley & Sons.

Grimm, K. J., Mazza, G. L., & Mazzocco, M. M. (2016). Advances in methods for assessing longitudinal change. *Educational Psychologist*, *51*(3-4), 342–353.

Hussain, S., Dahan, N. A., Ba-Alwib, F. M., & Ribata, N. (2018). Educational data mining and analysis of students' academic performance using weka. *Indonesian Journal of Electrical Engineering and Computer Science*, *9*(2), 447–459.

Kushwah, J. S., Kumar, A., Patel, S., Soni, R., Gawande, A., & Gupta, S. (2021). Comparative study of regressor and classifier with decision tree using modern tools. *Materials Today: Proceedings*.

Marcoulides, G. A. (1993). Maximizing power in generalizability studies under budget constraints. *Journal of Educational Statistics*, *18*(2), 197–206.

McArdle, J. J., Paskus, T. S., & Boker, S. M. (2013). A multilevel multivariate analysis of academic performances in college based on NCAA student-athletes. *Multivariate Behavioral Research*, *48*(1), 57–95.

McFarland, D. A., Khanna, S., Domingue, B. W., & Pardos, Z. A. (2021). Education data science: Past, present, future. *AERA Open*, *7*, 23328584211052055.

Mueller, R. O., & Hancock, G. R. (2018). Structural equation modeling. In *The reviewer's guide to quantitative methods in the social sciences* (pp. 445–456). Routledge.

Office of Educational Technology. (2023). *The Future of Teaching and Learning with AI* [Accessed: July 16, 2024]. U.S. Department of Education. https://tech.ed.gov/ ai-future-of-teaching-and-learning/

Packer, M. J., & Maddox, C. (2016). Mapping the territory of the learning sciences. *Reflections on the learning sciences*, 126–154.

Piety, P. J., Hickey, D. T., & Bishop, M. (2014). Educational data sciences: Framing emergent practices for analytics of learning, organizations, and systems. *Proceedings of the fourth international conference on learning analytics and knowledge*, 193–202.

R Core Team. (2022). R: A language and environment for statistical computing. https://www.R-project.org/

Randall, J., Rios, J. A., & Jung, H. J. (2021). Graduate training in educational measurement and psychometrics: A curriculum review of graduate programs in the us. *Practical Assessment, Research, and Evaluation*, *26*(1), 2.

Rodrigues, M. W., Isotani, S., & Zarate, L. E. (2018). Educational data mining: A review of evaluation process in the e-learning. *Telematics and Informatics*, *35*(6), 1701–1717.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1355.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *40*(6), 601–618.

Romero, C., & Ventura, S. (2017). Educational data science in massive open online courses. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *7*(1), e1187.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN computer science*, *2*(3), 160.

Sawyer, R. K. (2005). *The Cambridge handbook of the learning sciences*. Cambridge University Press.

Shu, X., & Ye, Y. (2023). Knowledge discovery: Methods from data mining and machine learning. *Social Science Research*, *110*, 102817.

Sinharay, S. (2017). Detection of item preknowledge using likelihood ratio test and score test. *Journal of Educational and Behavioral Statistics*, *42*(1), 46–68.

Slater, S., & Baker, R. S. (2018). Degree of error in Bayesian knowledge tracing estimates from differences in sample sizes. *Behaviormetrika*, *45*(2), 475–493.

Slater, S., Joksimovic´, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2017). Tools for educational data mining: A review. *Journal of Educational and Behavioral Statistics*, *42*(1), 85–106.

Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press.

Suk, Y., & Han, K. T. (2022). A psychometric framework for evaluating fairness in algorithmic decision making: Differential algorithmic functioning. *Journal of Educational and Behavioral Statistics*, *49*(2), 151-172..

Thagard, P. (2013). Cognitive science. In *The Routledge companion to philosophy of science* (pp. 597–608). Routledge.

Yanagiura, T., Yano, S., Kihira, M., Okada, Y., et al. (2023). Examining algorithmic fairness for first-term college grade prediction models relying on pre-matriculation data. *Journal of Educational Data Mining*, *15*(3), 1–25.

Citation:
Collier, Z. K., Sukumar, J., & Barmaki, R. (2024). Discovering educational data mining: An introduction. *Practical Assessment, Research, & Evaluation*, 29(11). Available online: https://doi.org/10.7275/pare.2037

Corresponding Author:
Zachary K. Collier
University of Connecticut
249 Glenbrook Road, Unit 2064, Storrs, CT 06269
Email: zachary.collier@uconn.edu