Journal of Applied Testing Technology, Vol 23(Special Issue 1), 95-116, 2022

Artificial Intelligence: Transforming the Future of Feedback in Education

Tarid Wongvorachan[1]*, Ka Wing Lai[1], Okan Bulut[2], Yi-Shan Tsai[3] and Guanliang Chen[3]

[1]Department of Educational Psychology, University of Alberta, Canada; wongvora@ualberta.ca, kawing2@ualberta.ca

[2]Centre for Research in Applied Measurement and Evaluation, University of Alberta, Canada; bulut@ualberta.ca

[3]Department of Human Centred Computing, Monash University, Australia; yi-shan.tsai@monash.edu, guanliang.chen@monash.edu

Abstract

Feedback is a crucial component of student learning. As advancements in technology have enabled the adoption of digital learning environments with assessment capabilities, the frequency, delivery format, and timeliness of feedback derived from educational assessments have also changed progressively. Advanced technologies powered by Artificial Intelligence (AI) enable teachers to generate different types of feedback supporting student learning. Despite the rapid uptake of digital technologies in education, previous studies on educational feedback primarily focused on the theoretical underpinnings of feedback practices, which are limited in terms of their coverage of AI-based technologies. This paper aims to inform both researchers and practitioners about the present and future of AI applications in feedback practices, identify and organize potential areas for the use of AI for feedback purposes, and establish venues for AI research and practice in educational feedback. Furthermore, the role of the three branches of AI (i.e., natural language processing, educational data mining, and learning analytics) in feedback practices and potential areas for their future development are discussed.

Keywords: Artificial Intelligence, Educational Data Mining, Educational Feedback, Learning Analytics, Natural Language Processing

1. Introduction

Feedback—a process where learners make sense of the provided information to reduce the gap between their current and desired performance—is a crucial component of student learning (Carless & Boud, 2018; Watling & Ginsburg, 2019). Rather than being a piece of static information, this paper conceptualizes feedback as the instructional process that encompasses the component of information and communication strategy to enhance students' understanding of their learning (Gamlem & Smith, 2013; Hattie & Timperley, 2007). For example, feedback to improve students' math problems can be conveyed by verbally explaining, solution demonstrating, or both. As each student has their own condition and may be differently equipped to access, understand, and use their feedback, both feedback information and communication strategy need to be considered to maximize the benefit of feedback to students and provide a lasting change; hence, the importance of personalized feedback (Hattie & Timperley, 2007; Kochmar *et al.*, 2020).

Teachers have their role to effectively formulate and communicate their feedback while students use the information to update their knowledge and change the corresponding behavior (e.g., learning strategy, approach to the task, and the use of learning resources) to achieve the desired outcomes(s) (Boud & Molloy, 2013; Forsythe & Johnson, 2017). Feedback is used for both formative and summative purposes during the learning process.

**Author for correspondence*

Teachers use formative feedback to provide students with opportunities for continual improvements while using summative feedback to inform students about their performance in the course (Barana *et al.*, 2019; Marriott & Teoh, 2012).

As advancements in technology have enabled new ways of learning and changed the dynamics of education (e.g., the transition from traditional paper-and-pencil assessments to digital online assessments), the frequency, delivery format, and timeliness of feedback derived from educational assessments have also changed progressively to meet the needs of students (Jurs & Špehte, 2021). For example, digital score reporting has enabled students to receive immediate and personalized feedback from computerized assessments to best inform students in their learning (Bulut *et al.*, 2019; Zenisky & Hambleton, 2012). Similarly, intelligent tutoring systems (ITSs) can provide granular and specific feedback to students as they complete learning tasks personalized based on their unique interests and proficiency levels (Ai, 2017; Kulik & Fletcher, 2016). The examples (i.e., digital score reporting and ITS) can be further developed through AI.

Specifically, in the field of education, AI itself refers to the development of computer systems that can perform education-related tasks that require human intelligence, such as grading students' exams, personalizing learning materials, or providing recommendations for assignment tasks based on real-time data analysis (L. Chen *et al.*, 2020). In other words, the usage of AI includes all activities that involve applying computer systems to extract information, solve problems, and answer questions, from a simple rule-based decision to a complex process of image or voice recognition (L. Chen *et al.*, 2020; Loyola-Gonzalez, 2019). The definition also expands to the involvement of AI in human decisions, such as the usage of an unsupervised learning approach to assign students into groups (e.g., adaptive learners, deep learners) to inform teachers in their feedback communication strategy (Tempelaar, 2020). Therefore, any action that involves the assistance of computer systems to simulate human intelligence in information extraction and decision making, whether in part or in full, can constitute the usage of AI.

In the education field, AI also plays a role in supporting feedback practices by providing fully automated or semi-automated feedback in various forms such as written feedback (Zhang *et al.*, 2019), audio-based feedback (Rodway-Dyer *et al.*, 2011), or video-based feedback (Ketchum *et al.*, 2020). This article focuses on technology that enhances the provision of verbal feedback, both manually written by teachers as informed by results from AI (i.e., semi-automated process) or generated by computers (i.e., fully automated process) and non-verbal feedback such as computer-generated graphs. Some cutting-edge examples of AI applications for feedback include the usage of Machine Learning (ML) and Natural Language Processing (NLP) to evaluate student performance in real-time and produce personalized feedback for students who are at risk of low performance (Jimenez & Boser, 2021). Educators can also utilize learning analytics (LA), which involves using AI and relevant techniques to provide real-time personalized feedback to all students and thereby enhance their learning experience (de Laat *et al.*, 2020; Tsai *et al.*, 2021). Such technologies can empower students to use feedback through the speed and efficiency of AI applications (Zhang *et al.*, 2019).

In sum, using AI allows educators to provide feedback to a large number of students in a short time frame in contexts such as Massive Open Online Courses (MOOCs), minimal disruptions from time and space barriers, and an ability to process large-scale educational data such as Trends in International Mathematics and Science Study (TIMSS) (e.g., Bethany *et al.*, 2021) and Programme for International Student Assessment (PISA) (e.g., Organisation for Economic Co-Operation and Development [OECD], 2019) with educational data mining (EDM) and ML methods (Gardner *et al.*, 2021; Witten *et al.*, 2017).

Given the increasing use of advanced technologies and AI in education worldwide, it is important to inform both researchers and practitioners about the present and future of AI applications in feedback practice (United Nations Educational, Scientific and Cultural Organization [UNESCO], 2019). Previous studies on educational feedback primarily focused on the theoretical underpinnings of feedback practices, such as the development of feedback literacy (e.g., Carless & Boud, 2018; Carless & Winstone, 2020), learner-centered feedback (e.g., Molloy *et al.*, 2021), and the advocacy for formative feedback practices in education (e.g., Boud, 2020). However, these studies are generally limited in their coverage of AI-based technologies for generating and delivering student feedback. Additionally, most studies on the application of AI for feedback purposes only focus on a single application of each technology, such as the use of AI to develop ITS (Ubani & Nielsen, 2022) or personalized feedback (Chan & Zary, 2019). This situation warrants a systematic review of the application of AI for feedback provision to provide an introductory overview of the area for current and new researchers in the field.

This theoretical paper aims to inform researchers and practitioners about the current and future landscape of AI applications in feedback, identify and organize potential areas for the use of AI for feedback (e.g., AutoTutor or QuizBot chatbots with NLP, interactive feedback with online data visualizations, and intelligent recommender systems), and establish venues for AI research and practice in educational feedback (G. Chen *et al.*, 2020; Yildirim-Erbasli & Bulut, 2021). Educators could also use results from this paper to guide their implementation of AI in their feedback practice (e.g., LA-driven feedback or NLP for automatic feedback generation).

The paper begins by discussing the extent to which feedback can affect students with the four levels of feedback framework as one of our supporting frameworks in the paper. Then, we introduce the involvement of AI in educational feedback to lay out the groundwork for the contribution of AI to feedback practices. We discuss the application of AI technologies in the fields of NLP, EDM, and LA for feedback purposes by defining them and their scope and then introducing the feedback dimension framework and weaving it together with the four levels of feedback framework. We conclude the paper by discussing the directions where AI can be harnessed to transform the future of feedback practices. Our goal is to form an anchoring concept before revolving into the specific application of AI in each research field to show the current landscape of the area. Note that the organization of this paper takes on a specific pattern by discussing NLP, EDM, and LA in this order throughout the paper for ease of understanding.

2. The Four Levels of Feedback Framework

Feedback is an essential element in student learning as it helps facilitate student development by stimulating their learning process and optimizing their understanding of class materials for improved performance in the task (Hounsell, 2007). In a feedback spiral, students reflect on feedback from their instructors to update their task-related knowledge and behavior in response to the received feedback; for example, students who receive feedback from their mid-term exam can use it to adjust their learning strategy such as investing more time to study the course content they did not do well to prepare for the final exam (Carless, 2019). Characteristics of high-quality feedback include thorough coverage, appropriate tone, straightforward language, and transparency in its guidance (Hounsell, 2007). Feedback should also be timely and relevant to both the course itself and student circumstances to maximize its actionability, especially in the distance learning context where communication is

Table 1. The four levels of feedback

Feedback Level	The Effect on Students' Level of Change
Task Level	This level concerns how the tasks are performed (e.g., correctly vs. incorrectly).
Process Level	This level concerns the thought process needed to perform the task and its related variant.
Self-Regulated Learning Level	This level concerns how students monitor, direct, and regulate their actions toward learning goals.
Self Level	This level concerns personal aspects of the students themselves (e.g., well done).

impeded by the lack of physical proximity (Bulut *et al.*, 2020; Hounsell, 2007; Jurs & Špehte, 2021).

Feedback provided by teachers could affect students at four levels: 1) task level, 2) process level, 3) Self-Regulated Learning (SRL) level, and 4) self level, as suggested by Hattie and Timperley (2007)'s model of feedback level; these four levels are defined and compared in Table 1 on how they differently affect students' level of change in their task performance. An effective feedback process can influence students beyond task-level to process- or SRL-level *via* feedback-informed action; conversely, feedback disconnected from student context tends to be disregarded instead (Bulut *et al.*, 2020; Carless, 2019). An example of actionable feedback could be "*you could improve your performance in domain X by reviewing lecture Y on topic Z*"; this way, the feedback will be able to guide students in their actions. Instructors could also add statements that encourage students to reach out to them to maintain interaction between student and instructor for potential follow-up.

3. The Involvement of AI in Educational Feedback

The application of AI in feedback practices and education is increasing, and this trend is likely to continue as more than 50% of human proficiency levels such as literacy, numeracy, and problem-solving can be covered by AI. Specifically, the current capability of AI can fully cover level 2 proficiency that 53% of OECD adults can achieve and can increasingly cover level 3 proficiency that 36% of OECD adults can achieve (Elliott, 2017; Holmes *et al.*, 2019). The mentioned level 2 and level 3 proficiency include acting on less explicit mathematical information and ideas, comprehending lengthy and non-continuous texts, and solving problems requiring multiple steps and constant monitoring (National Center for Educational Statistics [NCES], 2022).

The early application of AI for feedback purposes is dated back to the 1950s, during which AI was used for adaptive learning (the self-adaptive keyboard instructor) or computerized assessment (Holmes *et al.*, 2019; Pask, 1982). AI will likely continue to play important roles in education in the future due to its benefits. The combination of AI technology and high-quality human instruction allows students to learn more efficiently while at the same time allowing instructors to address issues that can only be identified through results from a fine-grained level data analysis (Barana *et al.,* 2019; Jimenez & Boser, 2021). For example, AI can be used to grade a large number of exams and at the same time identify patterns of student performance with data mining to inform teachers in their feedback provision, such as providing more feedback detail in the content areas that the student cohort did not do well (L. Chen *et al.*, 2020; Jimenez & Boser, 2021).

Additionally, a number of testing organizations such as Educational Testing Service (ETS) and Pearson have implemented an automatic essay scoring system (AES) to assess written essays from test takers for a more efficient workflow; the system evaluates the essay based on elements such as grammatical error, writing style, and discourse structure not only to ease the scoring process but also to provide relevant feedback to the test takers for their improvements such as grammar usage, vocabulary diversity, or essay organization (Gardner *et al.*, 2021). Another instance of AI-related assessment is computerized adaptive testing (CAT), which is usually implemented in high-stakes testing such as the Graduate Management Admission Test (GMAT) or the Graduate Record Examination (GRE) *via* internet delivery (Gardner *et al.*, 2021). CAT automatically tailors item selection by matching the test taker's estimated ability to items to be administered with a rule-based computer system based on Item Response Theory, so that the test can maximize the gained information by delivering items with appropriate parameters (e.g., difficulty) to examinees (Magis *et al.*, 2017). Test providers can then provide personalized feedback from the information gained from the assistance of CAT as each examinee received a different set of test items (Economides, 2005). Aside from the educational assessment area, the intelligent tutoring system (ITS) has been used to provide corrective feedback and suggestions to student errors and human tutors as an enhanced teaching practice (Ai, 2017). A meta-analysis of 50 controlled evaluations of ITS found that students who receive assistance from ITS exhibited greater performance than students from conventional human-only classes (Kulik & Fletcher, 2016).

4. Primary Applications of AI Technologies in Educational Feedback

4.1 Definition and Scope of AI Technology in the Three Fields

AI is an umbrella term covering a wide area of machine capability, from basic problem solving such as rule-based message delivery to advanced decision-making such as multi-class machine learning-based classification. The most relevant AI technologies for feedback are situated in NLP, EDM, and LA (Gardner *et al.*, 2021; Lemay *et al.*, 2021; Zhang *et al.*, 2019). Table 2 presents the juxtaposition between the application of AI technologies in these three research fields in terms of their definition and capability. In terms of data requirement, the application of AI in the three research fields can process any kind of data as they work together; for example, ML techniques used in EDM can either process numerical data by themselves or process textual data with the help of NLP to support feedback practices; thus, it is impossible to attribute the application of AI in the three fields to any specific types of data as the three fields share overlapping space in the actual practice. Note that despite being applicable to the educational field, NLP's capability also spans to non-educational settings as well.

The application of AI in NLP focuses on manipulating unstructured textual data for understanding, interpreting, and potentially generating relevant textual output (Roberts, 2019). Textual data used in NLP can come in any shape ranging from simple words or sentences such as student-authored course reviews to complex essays with various structures and writing styles from GRE/GMAT examinees (Moreno & Redondo, 2016; Roberts, 2019). Algorithms used in NLP can automatically convert the data into understandable formats and extract non-trivial information from it by analyzing elements such as syntax, semantics, morphology, or even the basic frequency of words (Goddard, 2021; Moreno & Redondo, 2016). Some applications of NLP in education include text summarization to extract essential elements from unstructured documents (e.g., theses, essays, or reports), machine translation to bypass or mitigate language barriers, and sentiment analysis to gain insights into public opinion (Goddard, 2021).

The application of AI in EDM focuses on using ML techniques such as clustering or classification of educational databases for knowledge discovery (Guo *et al.*, 2015; Hussain *et al.*, 2018; Qazdar *et al.*, 2019). Data used in EDM can be both numerical and textual in nature and can take various forms, such as student performance as indicated by their GPA, their history of grade repetition as indicated by self-reported binary indicator (i.e., yes *vs.* no), or even the educational level of their parents; such data can come from large-scale sources such as student records or educational surveys (e.g., PISA, TIMSS) (Bethany *et al.*, 2021; Hussain *et al.*, 2018; OECD, 2019). Some applications of EDM include the prediction of student performance from school databases to identify

Table 2. Focus and capability of AI technologies in the three fields

Research Field	Focus of the Application of AI Technology	Capability
NLP	The understanding and manipulation of textual data	Convert textual data for translation or pattern extraction
EDM	Knowledge extraction from educational databases	Automatically extracting information using machine learning to discover insights
LA	The leveraging of student activity data for classroom optimization	Process learning activity data to support human judgment in classrooms

potential low achieving students or the development of adaptive learning systems with student log data to provide personalized lessons (Elatia *et al.*, 2016; Qazdar *et al.*, 2019). Such insights can support stakeholders in the educational context, such as principals, teachers, or even parents, to make informed decisions on matters such as curriculum design or school development (Elatia *et al.*, 2016).

The application of AI in LA focuses on collecting and leveraging accumulated data on student learning processes and activities for classroom optimization (Larusson & White, 2014; Sipes, 2017). Similar to EDM, the research focus of LA lies in the translation of data-driven insights into practical recommendations to guide the process of planning, decision making, and intervention (Larusson & White, 2014; Siemens, 2012). Like EDM, LA can also process numerical and textual data (Lemay *et al.*, 2021). However, the difference between the two fields is that LA operates from a holistic framework that considers student data as a whole while performing descriptive and diagnostic analyses with an emphasis on the process of teaching and learning. On the other hand, EDM focuses more on knowledge discovery from various analysis techniques (Lemay *et al.*, 2021).

EDM researchers focus more on the automated discovery aspect, such as predictive or descriptive analytics and machine learning models. In contrast, LA researchers focus on leveraging data to inform human judgment, such as investigating data patterns, resource allocation, and their effect on learning and teaching practice (G. Chen *et al.*, 2020; Siemens, 2012). In other words, we could say that LA is more end-users-oriented while EDM is automation-oriented. Some applications of LA is the leveraging of a data-rich environment by collecting digital footprints of learning activity from learning management systems (LMS) data such as records of course material access, student demographics, or course history to maximize teaching effectiveness with technology-enhanced learning environments or to develop an early warning system to assist with academic advising (Larusson & White, 2014; Sipes, 2017; Tempelaar, 2020). Rather than being mutually exclusive, the applications of AI in the three fields share overlapping space in their usage of educational data as they are often utilized together for innovative applications of AI in education. Figure 1 shows a visual representation of the intersection among NLP, EDM, and LA in solving educational problems.

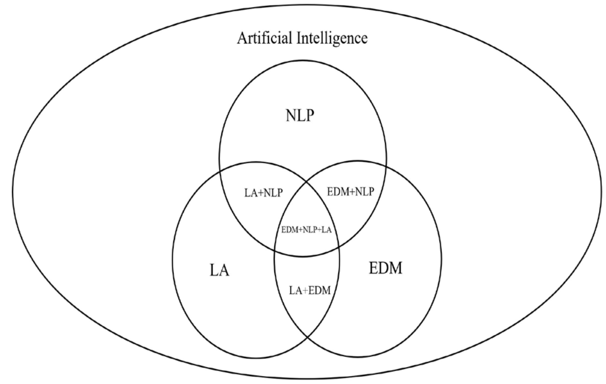

Figure 1. The scope of learning analytics, educational data mining, and natural language processing[1].

One example of the combination of EDM and LA (EDM+LA) is the usage of ML techniques such as cluster analysis to find potential groups of students (e.g., low achievement *vs.* high achievement) and predictive model to predict students' learning outcomes to extract insights from educational databases before displaying them along with other information such as learning process data (e.g., students' test-taking time) on a LA dashboard to inform students in their learning strategy (Larusson & White, 2014; Lemay *et al.*, 2021; Qazdar *et al.*, 2019); this combination leverages the capability

[1]Note. LA stands for learning analytics, EDM stands for educational data mining, and NLP stands for natural language processing.
The LA part is adapted from "Using learning analytics in SoTL," by S. Sipes, 2017, Center for Innovative Teaching and Learning @ IUB, (https://blogs.iu.edu/citl/2017/12/13/using-learning-analytics-in-sotl). Copyright 2021 by Indiana University Bloomington.
The EDM part is adapted from "A machine learning algorithm framework for predicting students performance: A case study of baccalaureate students in Morocco," by A. Qazdar, B. Er-Raha, C. Cherkaoui, and D. Driss, 2019, Education and Information Technologies, 24(6), p. 3579 (10.1007/s10639-019-09946-8). Copyright 2021 by Education and Information Technologies.
The NLP part is adapted from "5 Natural language processing examples: How NLP is used," by T. Roberts, 2019, Bloomreach, (https://www.bloomreach.com/en/blog/2019/09/natural-language-processing). Copyright 2021 by Bloomreach.

of EDM by applying ML models to educational data and leverages the power of LA with its dashboard for student improvements. Researchers can also combine the capacity of EDM and NLP (EDM+NLP) by extracting both textual data (e.g., writing quality of students' written assignment) with NLP and numerical data (e.g., course grade) with EDM. The combined information can predict students' class completion to aid teachers in their course planning, including feedback provision (Crossley *et al.*, 2015; Moreno & Redondo, 2016). For the combination of NLP and LA (NLP+LA), instructors could extract students' learning process data with LA and their textual data with NLP to automatically generate verbal feedback for students (Piotrkowicz *et al.*, 2017). The three fields (LA+NLP+EDM) can also work together by extracting students' learning process data from LA and combining it with results from machine learning analyses of students' data from EDM. This information can inform the generation of data-supported feedback in writing with NLP and non-verbal format (e.g., graphs, numbers) *via* a LA dashboard (Li & Xing, 2021; Pardo *et al.*, 2017). The mentioned combinations show that AI can support instructors in formulating feedback information (i.e., the what) and the communication of feedback (i.e., the how) in various ways to enhance students' learning experience and outcomes.

4.2 The Feedback Dimension Framework

Applying AI technology to feedback requires more specificity within the educational context to ensure that the result contributes to feedback delivery for teachers and feedback understanding for students. In this paper, we use Gamlem and Smith (2013)'s framework to organize aspects of feedback practice into two dimensions, namely, feedback strategy and feedback content. Each category consists of various sub-characteristics of feedback, such as focus, clarity, or honesty. We have summarized and defined the sub-characteristics that could be applied to the application of AI in the three fields in Table 3. The definition of feedback characteristics from Gamlem and Smith (2013)'s framework.

From the framework summarized in Table 3, the *feedback content* dimension covers the description of

Table 3. The definition of feedback characteristics from Gamlem and Smith (2013)'s framework

Dimension of Feedback	Sub-Characteristics of Feedback	Definition
Feedback Content	Focus	The targeted level of change (e.g., task level *vs.* thought process level).
	Comparison	The data to which students' performance was compared.
	Function	The purpose of the feedback itself (i.e., descriptive *vs.* judging).
	Valence	The tone of the feedback content (i.e., positive *vs.* negative)
Feedback Strategy	Mode	The format in which the feedback is delivered (i.e., verbal *vs.* non-verbal).
	Timing	The frequency and timeliness in which the feedback is delivered.
	Use	The utility of feedback related to the given timeframe for students to use the feedback.
	Management	The manageability of the feedback by students and teachers.

Table 4. The Summary of AI Applications in Feedback

Dimension of Feedback		**Natural Language Processing**	**Educational Data Mining**	**Learning Analytics**
Content	Focus	Provide task-level feedback for domain-specific tasks (e.g., academic writing, physics) Provide process-level feedback such as verbal hints through natural language generation Provide SRL-level feedback by generating context-relevant feedback to aid in self-reflection Assess manually written feedback in terms of feedback level (task-level *vs* process-level *vs* self-level)	Personalized process-level feedback as informed by student behavior and learning style from data mining techniques (e.g., clustering). Provide SRL-level feedback via findings from EDM models using student behavior and survey-based SRL-related construct as features.	Task-level feedback via score reporting and point-of-error identification. Personalized process-level feedback as informed by student behavior, profile, and other manually imported learning analytics data. Provide SRL-level feedback through the monitoring of student activity both individually and as a group.
	Comparison	Self-referenced feedback based on previous work. Norm-referenced feedback based on writing work in databases such as *Expertiza* or *SWoRD* Criterion-referenced feedback based on domain-specific standards such as grammatical rules or mathematics.	Primarily self-referenced from students' previous profiles.	Self-referenced from students' previous profiles. Criterion-referenced as indicated by the domain of the task. Norm-referenced as indicated by class performance.
	Function	Judging feedback based on aspects such as content type, coverage, and plagiarism. Outputs are in the form of scores/ grades. Descriptive feedback by flagging errors, their locations, and suggestions for improvements.	Data-driven feedback can be either descriptive or judgemental as intended by the instructor.	Primarily descriptive due to its formative nature.

	Valence	Implied neutral as generated by the system.	Instructor dependent.	Instructor dependent.
Strategy	Mode	Automatically generated verbal feedback on the quality of work and location of errors. Automatically generated verbal prompts on suggestions for revision. Non-verbal feedback in scores or metrics assessing the quality of elements as indicated by the rubric. Both feedback modes can be delivered *via* chatbots.	Verbal feedback *via* natural language generation or manual input from the instructor. Non-verbal feedback with data visualization.	Verbal feedback *via* manually input dialogues from the instructor. Non-verbal feedback with data visualization delivered via an interactive dashboard.
	Timing	Immediate feedback delivery via natural language generation.	Immediate output delivery with automatic feedback formulation.	Semi-automatic output by allowing instructors to co-author feedback with algorithms. Immediate output delivery with automatic feedback formulation. Longitudinal feedback (i.e., across time) is also available.
	Use	High usability as NLP-based feedback is usually formulated for formative purposes, implying opportunities for students to put feedback into action.	EDM-informed feedback is frequently used for formative purposes, which implies the opportunity to use the received information.	High usability due to its formative nature.
	Management	User-friendly interface with self-paced feedback navigation. Error localization makes the feedback more manageable.	Feedback personalization and user-friendly interface make the feedback easily manageable and relatable.	Feedback personalization at both individual- and group level makes the feedback easily manageable and relatable. An interactive dashboard for step-by-step guidance also increases manageability. Interface localization (e.g., Chinese version) makes the feedback more accessible.

feedback itself (i.e., the what) with four sub-characteristics that include: 1) the focus of feedback, 2) the comparison of feedback, 3) the function of feedback, and 4) the valence of feedback. The focus aspect of feedback synergizes well with Hattie and Timperley (2007)'s framework of feedback level (see Table 1) as they both concern levels of change in students from receiving feedback. Therefore, we describe the focus of feedback with the four levels of change (i.e., task-, process-, SRL-, and self-level) as suggested in Hattie and Timperley (2007)'s framework. The comparison of feedback concerns how students' performance was compared. For example, in criterion-referenced comparison, students' scores are compared with specific criteria (e.g., 95-100% = A, 90-95% = A-), whereas in norm-referenced comparison, students' scores are compared against each other (e.g., the top 10% of students receive an A, the next 30% gets a B). Also, it is possible to apply self-referenced comparisons where students' scores are compared against their previous scores (e.g., comparing scores from the first and second midterms). The function of feedback concerns the purpose of the feedback itself, such as a descriptive function that describes students' performance as is (e.g., *you got 80/100*) or a judging function that forms conclusions about students' performance (e.g., *you did well with 80/100*). The "*you did well*" part implies that students' performance is judged positively, whereas the descriptive feedback only provided students' scores. Lastly, the valence of feedback concerns the tone of the feedback itself (i.e., positive, negative, neutral).

The second dimension of feedback strategy covers how the feedback is delivered from instructors to students (i.e., the how) with four sub-characteristics that include: 1) the mode of feedback, 2) the timing of feedback, 3) the use of feedback, and 4) the management of feedback. The feedback mode concerns the format in which the feedback is delivered; for example, verbal feedback could be manually written feedback as informed by results from AI (i.e., semi-automatic process) or an automatically generated sentence (i.e., fully automated process). The timing of feedback concerns the frequency and timeliness in which the feedback is delivered in real time (i.e., immediately available) or not in real time (i.e., taking time to inspect the feedback before releasing it to the student). The use of feedback concerns the utility of feedback related to the given timeframe for students to use the feedback. For example, formative feedback can inform students to adjust their learning strategy before the final exam; therefore, it has high usability compared to summative feedback given after a course is over. Lastly, the management of feedback concerns the manageability of the feedback by students and teachers; for example, personalized feedback could be easier to grasp than generic feedback and, therefore, more manageable to students (Pardo *et al.*, 2019).

Both feedback aspects are equally important to convey information about students' performance for their improvement. The content part ensures that the message is meaningful and actionable for the students, while the strategy part ensures that the feedback is appropriately delivered to students at the right time, with the right amount, and on the right channel (Brookhart, 2008). We applied the literature on the application of AI to the three research fields to Gamlem and Smith (2013)'s framework as the anchoring point and discussed how AI could benefit the formulation and delivery of feedback at a fine-grained level. See Table 4 for the summary of AI applications in feedback. This Table could serve as the current landscape of what AI can do to benefit feedback practices Table 4. The summary of AI applications in feedback

4.3 The Application of NLP, EDM, and LA to Educational Feedback

As summarized in Table 4, the primary benefit of NLP in feedback practices is the additional capability to generate feedback on students' written performance based on linguistic components such as writing quality, syntactic complexity, and grammatical errors before providing verbal feedback or numerical scores to students. The benefit of EDM to feedback practices relies on the results of machine learning techniques to establish and deliver data-supported feedback *via* data visualization such as normative curves or bar charts. EDM can also be used to provide verbal feedback by using NLP-based systems or relying on manual input from the instructors as informed by results from the algorithm. The benefit of LA for feedback practices relies on monitoring students' activity data such as time use or interaction history with course material to provide personalized feedback *via* an interactive dashboard. Like EDM, LA feedback can be either semi-automatic or fully automatic, depending on the implemented system. The application of AI can provide immediate feedback with high usability and

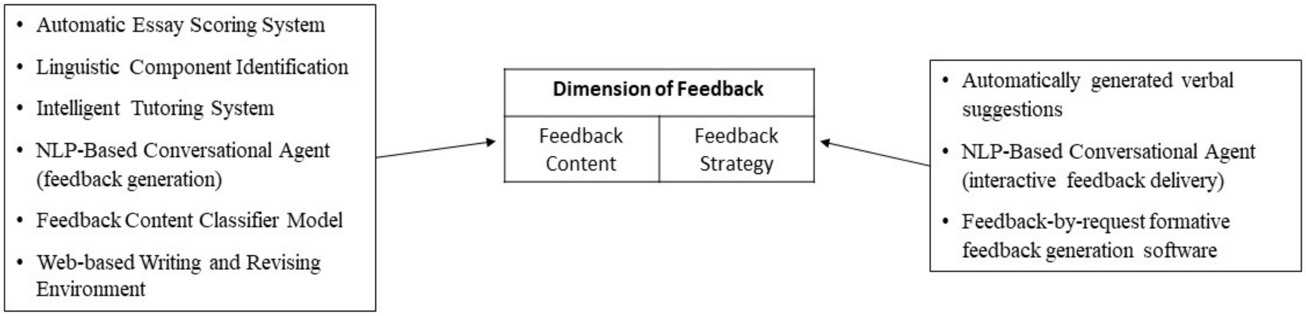

Figure 2. Feedback practice in NLP.

manageability due to its formative nature and user-friendly interface.

NLP. The application of NLP to feedback primarily focuses on text processing for constructed response tasks such as essay writing through the usage of NLP-based models for text-based feature extraction, language recognition for feedback selection, or natural language generation for automatic feedback generation (Li & Xing, 2021; Zhang *et al.*, 2019). Figure 2 visualizes how NLP technologies can enhance feedback practices in both feedback content and feedback strategy dimensions.

For its application to feedback content, task-level feedback is generated to assess the quality of domain-specific tasks such as essay structure *via* the automatic essay scoring system or constructed response tasks in Physics (Dzikovska *et al.*, 2014; Zhang *et al.*, 2019). NLP-based feedback software programs can also flag the location of the error, assess the clarity of the content, and provide feedback to assist students in their improvement (Lan *et al.*, 2015; Xiong *et al.*, 2012). For example, *AcaWriter* is a web-based writing assistance tool that assesses students' analytical and reflective writing and provides real-time feedback on academic writing characteristics such as clarity, conciseness, and rhetorical connotation (Knight *et al.*, 2020). Process-level feedback is usually provided *via* Intelligent Tutoring System (ITS) by generating step-by-step hints on working toward the correct answer. The ITS can also use students' meta-cognitive data from the behavioral log to enhance its procedural feedback (Kochmar *et al.*, 2020; Perikos *et al.*, 2017). In addition, NLP-based conversational agents can also generate context-relevant feedback by comparing students' level of knowledge to the course material and provide feedback that could stimulate self-reflection as well as encourage learners with motivational prompts (e.g., "*you are on the right track*", "*keep going*") at the same time to support student self-regulated learning with SRL-level feedback (Desai & Chin, 2020; Pengel *et al.*, 2021). In the case of manually written feedback, NLP can be used to develop a content classifier model to assess whether the written feedback falls into task-level, process-level, SRL-level, or self-level feedback to allow instructors to provide personalized feedback to students and target the intended level of learning (Cavalcanti *et al.*, 2020).

In terms of the comparison aspect, feedback can be provided based on the previous work of the students or a normative database *via* a web-based writing and revising platform such as the *eRevise*, the *Expertiza*, and the *Scaffolded Writing and Rewriting in the Discipline* (SWoRD) project where students' work is compared with instances of other students (Ramachandran *et al.,* 2017; Zhang *et al.*, 2019). The system can also provide feedback based on grading standards (or criteria) of domain-specific tasks such as mathematics, grammatical rules, or physics formulas (Dzikovska *et al.*, 2014; Kochmar *et al.*, 2020; Lan *et al.*, 2015; Perikos *et al.*, 2017).

NLP-based feedback systems can provide both judging and descriptive feedback. The system can evaluate students' writing to provide judging feedback such as scores/grades on content type, coverage, tone, volume, and plagiarism or even categories of the performance itself (high *vs.* low quality) (Ötleş *et al.*, 2021; Ramachandran *et al.*, 2017; Zhang *et al.*, 2019). The system can also provide descriptive feedback to flag and localize the detected elements and assess their clarity with keyword detection (Dzikovska *et al.*, 2014; Lan *et al.*, 2015; Perikos *et al.*, 2017; Xiong *et al.*, 2012). Lastly, the valence of the feedback provided by NLP is implied to be neutral as generated by the system (Kochmar *et al.*, 2020).

By applying NLP to the feedback strategy, the system can provide verbal prompts and suggestions for students to improve their work through a user interface (Xiong *et al.*, 2012; Zhang *et al.*, 2019). Descriptors of work quality (e.g., effective-mediocre-ineffective-other) can be delivered *via* a user interface in natural language sentences (Ötleş *et al.*, 2021; Solano *et al.*, 2021). The developers can combine the mentioned features with a pop-up chatbot to automate the feedback process and increase interactions with the students (Kochmar *et al.*, 2020). Non-verbal feedback is also available on the same interface as scores or metrics of each element (e.g., relevance or plagiarism) or as locations of the error, such as "*paragraph X*" or "*page Y*" (Lan *et al.*, 2015).

NLP-based feedback software programs can provide immediate feedback with its automation in terms of feedback timing. Feedback provided by NLP-based software programs through their user interface is usually used for formative purposes; that is, the feedback is supposed to be used to improve students' work before their next submission or post-test in subjects such as mathematics or academic writing (Kochmar *et al.*, 2020; Lan *et al.*, 2015; Zhang *et al.*, 2019). For some software programs, such as Natural Language to First-order Logic (NLtoFOL), students can also request formative feedback before their submission (Perikos *et al.*, 2017; Xiong *et al.*, 2012; Zhang *et al.*, 2019). For the feedback management aspect, feedback is delivered *via* a user-friendly interface that students can navigate at their own pace (Perikos *et al.*, 2017; Zhang *et al.*, 2019). Additional prompts (e.g., "*provide more detail from…*" or "*use more evidence from*"), error localization, and clarity assessment are also available on the feedback provision platform such as LMS for students to better understand their feedback as well (Dzikovska *et al.*, 2014; Xiong *et al.*, 2012; Zhang *et al.*, 2019).

EDM. EDM primarily contributes to feedback *via* insights extracted from data mining techniques such as association rule mining, clustering, and classification, most of which fall under machine learning (Pechenizkiy *et al.*, 2008; Ray & Saeed, 2018). Additionally, EDM can also process both numerical data and textual data with the help of NLP, thus, enabling it to leverage data of many forms to support feedback in education. Figure 3 visualizes how EDM technologies can enhance feedback practice in both feedback content and feedback strategy dimensions.

Feedback content-wise, EDM can inform teachers to provide feedback at either task-, process-, or SRL-level. Personalized process-level feedback can be formulated based on data on student behavior, learning style, and web usage behavior as logged by the LMS (Anjewierden *et al.*, 2007; Romero & Ventura, 2010). Instructors can also adjust their feedback to task level to fit their needs (Merceron & Yacef, 2005). Feedback from EDM is primarily self-referenced as students can reflect on data of their previous lessons, but instructors can provide both norm- and criterion-referenced feedback by tailoring their feedback to fit the needs of the class (Merceron & Yacef, 2005; Romero & Ventura, 2010). For SRL-level feedback, predictive models (e.g., decision tree) or clustering models (e.g., k-means clustering) from EDM can identify and operationalize student behavior (e.g., system gaming, careless response) to provide feedback

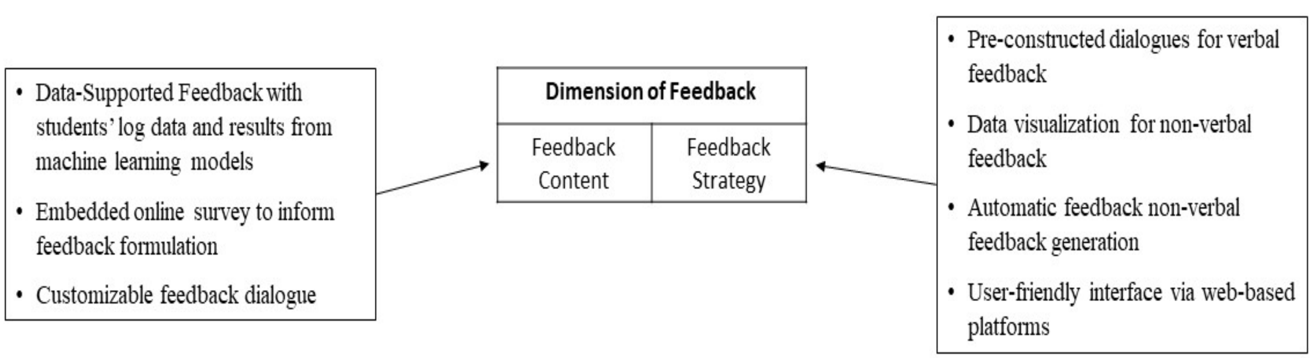

Figure 3. Feedback practice in EDM.

that is relevant to their motivation to inform students in developing their self-regulation (Winne & Baker, 2013). In addition, instructors can include surveys that measure SRL-related constructs such as the Online Self-Regulated Learning Questionnaire (OSLQ) and the Motivated Strategies for Learning Questionnaire (MSLQ) in the LMS to extract features that can inform EDM models in providing SRL-level feedback as well (Araka *et al.*, 2019). EDM-driven feedback can also be either judging or descriptive. Descriptive feedback is provided in the form of recommendations such as "*from your performance, you may want to check out course material X*" (Pechenizkiy *et al.*, 2008; Romero & Ventura, 2010). However, instructors may also set up the dialogue to be judging, such as "*you have a satisfactory performance*" (Merceron & Yacef, 2005; Romero & Ventura, 2010). In terms of feedback valence, feedback provided by EDM can be either positive, negative, or neutral, as instructors who give the feedback can adjust the tone to align with their teaching style.

For EDM application to feedback strategy, EDM-informed feedback can provide both verbal and non-verbal responses. Predetermined dialogues and natural language generation can provide written answers from teachers, while the system can give graphical information with data visualization (Anjewierden *et al.*, 2007; Ray & Saeed, 2018; Romero & Ventura, 2010). Depending on the system, EDM-informed feedback can take time if the feedback is human-processed (Merceron & Yacef, 2005). Some systems allow automatic feedback formulation to provide immediate non-verbal feedback to student users *via* an online learning platform or LMS (Pechenizkiy *et al.*, 2008; Ray & Saeed, 2018; Romero & Ventura, 2010). The feedback can be both formative and summative, so students may use feedback and the provided resource to inform their subsequent activities, such as second submission or post-test (Merceron & Yacef, 2005; Ray & Saeed, 2018). Lastly, EDM-informed feedback is usually provided *via* an online learning environment, which can be designed in a user-friendly format for better manageability (Merceron & Yacef, 2005; Romero & Ventura, 2010). Feedback personalization based on student profiles also makes the message more relatable to students (Pechenizkiy *et al.*, 2008; Ray & Saeed, 2018).

LA. Like EDM, LA contributes to feedback by providing insights extracted from student profiles with differences in its primary focus to inform human decisions in learning and teaching instead of model generalization (Gardner *et al.*, 2021; Lemay *et al.*, 2021). Despite being similar to EDM-driven feedback, LA-driven feedback sets itself apart based on its capability to monitor student activity and provide feedback in both cross-sectional formats (e.g., one time) or longitudinal format (e.g., across the semester). Also, LA-driven feedback is delivered *via* an interactive dashboard as opposed to a static format, which is the characteristic that emphasizes the practical aspect of the data more when compared to EDM-driven feedback (G. Chen *et al.*, 2020). Figure 4 visualizes how LA technologies can enhance feedback practice in both feedback content and feedback strategy dimensions.

Content-wise, LA can provide task-level feedback by displaying students' performance in each task in the format of scores or error localization (i.e., where they

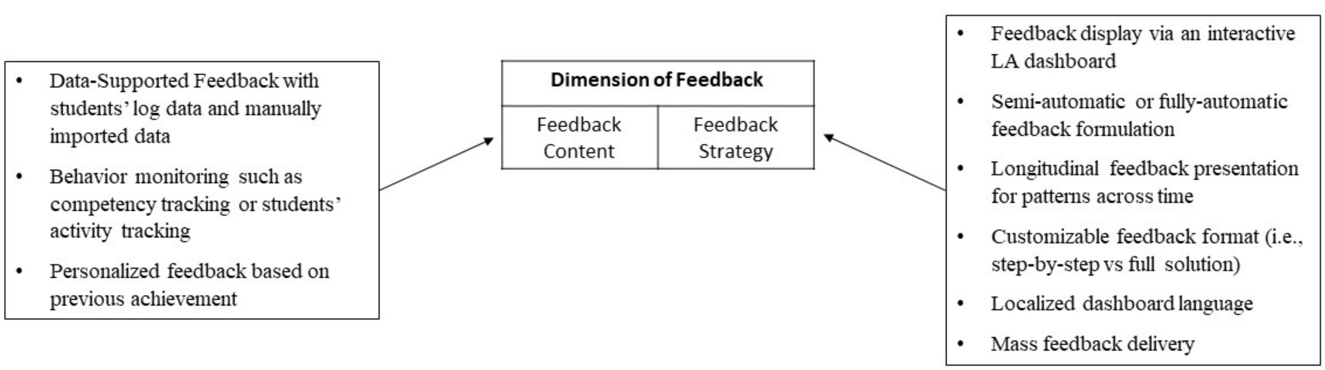

Figure 4. Feedback practice in LA.

did wrong) (Sedrakyan *et al.*, 2020). LA can also provide process-level feedback by using additional information from learners' behavior data such as their approach to the task, course history, academic proficiency profile, or even surface-level data such as clickstream in addition to the task performance data (Gardner *et al.*, 2021; Tempelaar *et al.*, 2015). Instructors can also manually import their data into the system to inform their feedback formulation (Tsai *et al.*, 2021). For SRL-level feedback, LA can track SRL-related features to inform its feedback formulation, such as a competency tracking dashboard that monitors students' domain knowledge and motivational factor that could affect them in their self-regulation, learning design tracking that monitors students' activity to inform instructors of course design, and even teamwork tracking that monitor students' activity in relation to their roles in the group through social network analysis; information can then be processed and displayed in a dashboard as SRL-level feedback to inform students of their progress and create awareness in their learning strategies (Matcha *et al.*, 2020; Viberg *et al.*, 2020).

LA-based feedback can be either self-referenced as informed by previous profiles of the students for personalized suggestions, criterion-referenced as informed by domain-specific criteria (e.g., mathematical solutions), or even norm-referenced should the instructors wish to give feedback in the form of peer-comparison (e.g., class average) (Barana *et al.*, 2019; Pardo *et al.*, 2019; Wang & Han, 2021). For the function aspect, LA-based feedback is usually descriptive to fit its formative purposes (Knight *et al.*, 2020; Sedrakyan *et al.*, 2020). Content of the feedback could be suggestions for improvements, descriptions of the current performance, and prescriptive feedback for correction (Pardo *et al.*, 2019; Wang & Han, 2021). Lastly, in terms of feedback valence, instructors can frame their feedback as either positive, negative, or neutral to fit their teaching style.

For the application of LA to feedback strategy, LA-based feedback can be both verbal and non-verbal. Verbal feedback such as suggestions is usually prepared by the instructors as informed by results from LA, while non-verbal feedback takes the form of a graphical display of numerical information (e.g., line chart) *via* interactive dashboards (Barana *et al.*, 2019; Gardner *et al.*, 2021; Wang & Han, 2021). LA-based feedback software programs can be semi-automatic in the case of *OnTask* where the program processes student data to inform the instructors in their feedback write-up. Also, it can be fully automatic in the case of *AcaWriter* where feedback is automatically and immediately generated based on the document that students provide to the software (Knight *et al.*, 2020; Tsai *et al.*, 2021). Feedback can also be presented across time with a longitudinal design to inform students of their improvements (Sedrakyan *et al.*, 2020). For the feedback use aspect, LA-based feedback is intended for formative purposes, so students usually have time to process and apply it to their learning strategy (Sedrakyan *et al.*, 2020). LA-based feedback has high manageability as it is personalized to match the profile of each student, as well as presented *via* an interactive dashboard with step-by-step guidance to inform students about their thinking (Barana *et al.*, 2019; Sedrakyan *et al.*, 2020; Tsai *et al.*, 2021). With customization, students can choose whether they want to view step-by-step feedback hints, consider an example of the task, or send a message to the instructor (Tempelaar *et al.*, 2015). Instructors can also localize the output dashboard with the local language (e.g., Chinese) to make the feedback and its software even more accessible for non-English speakers (Wang & Han, 2021). Lastly, instructors can choose to deliver feedback to groups of students or tailor it to one student, making the feedback process more manageable for both students and instructors (Pardo *et al.*, 2019).

5. Transforming the Future of Feedback in Education

The previous section discussed the current state of the application of AI to the fields of NLP, EDM, and LA for educational feedback. However, as technology evolves, the AI applications could be improved to maximize the benefit that students receive from their feedback as technology continues to advance. This section discusses future directions where AI can be enhanced to transform the future of feedback practice.

Table 5 summarizes the actions, characteristics, and functionalities we can acquire to improve the application of AI technologies in the three research fields to further the capability of educational feedback based on the reviewed literature. The suggestion for improvement is based on Gamlem and Smith (2013)'s Framework of

Table 5. The future direction of AI applications to improve feedback in education

Dimension of Feedback	Natural Language Processing	Educational Data Mining	Learning Analytics
Content	More data for coverage in terms of corpus structure, topics, and examples. Increase accessibility with localization and machine translation Contextualization with domains-related factors for accurate content interpretation.	Contextualization with more variable features for deeper personalization. Utilize appropriate measures to extract domain knowledge for enhanced contextualization.	More fine-grained profiling for deeper personalization (e.g., student preference, interest, bias, and biofeedback) Shared open-source databases across institutions for more domain content coverage. Consider informal social learning and personal learning environments for more profile coverage.
Strategy	Interoperability with AI in other research fields (e.g., LA) for innovative modes of delivery (e.g., dashboard). Eliminate redundant feedback cycle for efficiency. Contextualization with different domains and student profiles for appropriate feedback delivery strategies.	Fully integrated EDM with LMS for enhanced automation capability.	Utilizing longitudinal records across the course and school year for future-oriented feedback information. Multidirectional communication between learners, the system, and instructors for appropriate feedback strategy.

feedback dimension. To further improve the application of AI to benefit feedback content, more data is needed for AI algorithms to expand the scope of vocabulary and contextual-relevant patterns for the feedback system to recognize and inform its results (Bengfort *et al.*, 2018; Elatia *et al.*, 2016). To improve the application of AI in feedback practices, researchers could consider applying AI technologies across the three fields together. Also, they can utilize the capability of different technologies for enhanced effectiveness, such as using machine translation with an LA dashboard to expand the accessibility of a learning platform to international students or integrating EDM into a learning environment at a deeper level (i.e., using the computer assistance more) for enhanced automation in the feedback process.

Future studies applying NLP techniques to enhance feedback practices could consider expanding the coverage of the system with more data on domain-related topics to improve relevance and practicality (Ötleş *et al.*, 2021; Solano *et al.*, 2021; Zhang *et al.*, 2019). Output relevance in both feedback content and feedback delivery could also be improved by considering contextual factors such as student profile, knowledge gap, domain-specific convention (e.g., mathematical expression), writing complexity, differences between first and second language learners, and the interaction of item difficulty and feedback effectiveness to make the feedback more context-specific; some of these data can be obtained from LA (Goddard, 2021; Kochmar *et al.*, 2020; Lan *et al.*, 2015; Perikos *et al.*, 2017; Xiong *et al.*, 2012). Further, improving error localization and clarity assessment of students' work with higher precision by expanding the keyword database and optimized algorithm could reduce redundancy in the feedback cycle, resulting in a quick turnaround and more efficiency in the feedback process (Dzikovska *et al.*, 2014; Ramachandran *et al.*, 2017). Researchers could also increase the accessibility of the output with localization and machine translation, which could be combined with an interactive dashboard for enhanced feedback understanding (Barana *et al.*, 2019; Moreno & Redondo, 2016; Wang & Han, 2021).

Future research in the application of EDM to feedback practice could consider extracting domain-specific knowledge and contextual factors (e.g., students' performance in related courses) with appropriate procedures to add more variable features into the predictive model for deeper feedback personalization (Lan *et al.*, 2015; Merceron & Yacef, 2005; Pechenizkiy *et al.*, 2008). Researchers could also consider developing a fully-integrated LMS with a built-in EDM component for a higher automation capability that can handle a large amount of data with minimal human processing (i.e., manually data importing) (Ray & Saeed, 2018; Romero & Ventura, 2010).

For the application of LA to feedback, researchers could collect student profiles at a deeper level, such as learner differences in preference, interest, and bias *via* survey, more performance data *via* stealth assessment, and biological data *via* wearables to expand coverage across student characteristics, disciplines, and grades (i.e., K-12 *vs.* higher education); however, ethical issues on privacy should be considered as well (Gardner *et al.,* 2021; Karaoglan Yilmaz & Yilmaz, 2021; Pardo *et al.*, 2019; Sedrakyan *et al.*, 2020; Tsai *et al.*, 2021). Researchers could also expand data coverage by sharing the database for writing patterns across institutions, or they could collect data on informal social learning and personal learning environments (e.g., forum discussion, homework, in-class activities) into account as they are primary predictors of academic performance (Knight *et al.*, 2020; Sedrakyan *et al.*, 2020; Tempelaar *et al.*, 2015). Future research could also explore the use of future-oriented feedback by utilizing a longitudinal record to identify the impact of feedback across the course duration. This would allow instructors to link feedback information to its relevant learning outcome in future subjects to examine how past feedback informs future performance (e.g., introduction to calculus to intermediate calculus) (Ryan *et al.*, 2019). The multidirectional channel between students, instructors, and the system could also be established to enable the instructors to reflect on their instructional design *via* learners' results to formulate appropriate feedback strategies (Sedrakyan *et al.*, 2020).

6. Discussion

This theoretical paper aims to identify the current applications of AI technology to feedback in education and to identify venues for future research in technology-driven feedback practice. The application of AI in each of the three fields (i.e., NLP, EDM, and LA) has its unique contribution to the feedback of different natures, some examples are the usage of AES from NLP, process mining from EDM, and an analytics dashboard from LA. With adequate expertise, researchers could develop a feedback system that is situated across the three fields such as a predictive model that can process textual data by combining the capability of NLP with EDM (Lan *et al.*, 2015).

An AI-driven feedback system could leverage the capability of online educational platforms by increasing the quality of provided feedback and enhancing the efficiency of the feedback process. A real-time conversation-based system embedded in online assessments could stimulate information exchange between students and the digital tutor and thus improve their learning engagement and motivation (Yildirim-Erbasli & Bulut, 2021). Further, displaying verbal and non-verbal feedback together *via* a LA dashboard could synergize the capability of both feedback formats, help students understand patterns of their learning progress with data visualization, and receive personalized messages as informed by predictive models (Qazdar *et al.*, 2019; Wang & Han, 2021). Thus, using AI technologies in feedback could constitute the best practice of effective feedback communication in terms of design, contents, and ancillary materials (Zenisky & Hambleton, 2012).

Automated or semi-automated feedback systems could also enhance the efficiency of the feedback process by reducing the time and resources required from the instructors to provide formative feedback (Knight *et al.*, 2020). Specifically, automated feedback systems enable students to learn and receive feedback asynchronously, allowing the instructors to focus on addressing concerns or providing a one-on-one feedback session to students who need extra guidance (Perikos *et al.*, 2017). The use of EDM also allows instructors to discover patterns of student performance in large-scale learning environments (e.g., MOOCs) with relative ease to provide feedback to students as a group (Romero & Ventura, 2010). Finally, by integrating a venue for students to respond to instructors about the quality of feedback, instructors can use such data to inform their development of feedback models (Flodén, 2017; Zenisky & Hambleton, 2012).

As technology rapidly evolves, this paper could serve as a starting point in the field of AI in educational feedback by introducing the current landscape and future possibilities of the research area. However, there are limitations to what AI can currently do. First, predictive results about students' learning outcomes can be computed based on immutable variables such as gender, age, or socio-economic status that could hardly be used to form actionable recommendations (Ramaswami & Bhaskaran, 2010). Instead of immutable predictors, instructors are encouraged to harness malleable learning predictors based on learning theory, such as their formative assessment activities (Ramaswami & Bhaskaran, 2010; Tempelaar, 2020). Second, the automatically-generated feedback could overlook minor but meaningful background information such as students' relevant medical conditions (e.g., learning disorder history). If possible, instructors should constantly check the record of automatically generated feedback to ensure its relevance and improve the system whenever it is appropriate.

Future studies are needed to further explore and investigate the growing capabilities of technologically-enhanced feedback practices for the betterment of education. On the system side, instructors could develop EDM predictive models or NLP models for feedback generation with features relevant to their fields, such as students' programming background in the STEM field (Mbunge *et al.*, 2021). On the student side, future research could explore topics about students' data literacy, which influences their interpretation of data, and their experience with non-verbal feedback displayed *via* a LA dashboard (e.g., data literacy skills for students to meaningfully interpret graphs provided in a LA dashboard) to help them make the most of their feedback (Wasson *et al.*, 2016).

7. References

Ai, H. (2017). Providing graduated corrective feedback in an intelligent computer-assisted language learning environment. *ReCALL*, 29(3), 313–334. https://doi.org/10.1017/S095834401700012X

Anjewierden, A., Kolloffel, B., & Hulshof, C. (2007). Towards educational data mining: Using data mining methods for automated chat analysis to understand and support inquiry learning processes. *International Workshop on Applying Data Mining in E-Learning (ADML 2007)*, 27–36.

Araka, E., Maina, E., Gitonga, R., & Oboko, R. (2019). A conceptual model for measuring and supporting self-regulated learning using educational data mining on learning management systems. *2019 IST-Africa Week Conference (IST-Africa)*, 1–11. https://doi.org/10.23919/ISTAFRICA.2019.8764852

Barana, A., Conte, A., Fissore, C., Marchisio, M., & Rabellino, S. (2019). Learning analytics to improve formative assessment strategies. *Journal of E-Learning and Knowledge Society*, 15(3), 75–88. https://doi.org/10.20368/1971-8829/1135057

Bengfort, B., Bilbro, R., & Ojeda, T. (2018). Applied text analysis with Python: Enabling language-aware data products with machine learning (First edition). O'Reilly Media, Inc.

Bethany, F., Foy, P., & Yin, L. (2021). TIMSS 2019 user guide for the international database (2nd ed.). TIMSS & PIRLS International Study Center.

Boud, D. (2020). Challenges for reforming assessment: The next decade. International Virtual Meeting: Teaching, Learning & Assessment in Higher Education, Universidad del Desarrollo. https://educacionsuperior2020.udd.cl/

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698–712. https://doi.org/10.1080/02602938.2012.691462

Brookhart, S. M. (2008). How to give effective feedback to your students (1st ed.). Association for Supervision and Curriculum Development.

Bulut, O., Cutumisu, M., Aquilina, A. M., & Singh, D. (2019). Effects of digital score reporting and feedback on students' learning in higher education. *Frontiers in Education*, 4(65). https://doi.org/10.3389/feduc.2019.00065

Bulut, O., Cutumisu, M., Singh, D., & Aquilina, A. M. (2020). Guidelines for generating effective feedback from e-assessments. *Hacettepe University Journal of Education*, 1–13. https://doi.org/10.16986/HUJE.2020063705

Carless, D. (2019). Feedback loops and the longer-term: Towards feedback spirals. *Assessment & Evaluation in Higher Education*, 44(5), 705–714. https://doi.org/10.1080/02602938.2018.1531108

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Carless, D., & Winstone, N. (2020). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, 1–14. https://doi.org/10.1080/13562517.2020.1782372

Cavalcanti, A. P., Diego, A., Mello, R. F., Mangaroska, K., Nascimento, A., Freitas, F., & Gašević, D. (2020). How good is my feedback?: A content analysis of written feedback. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 428–437. https://doi.org/10.1145/3375462.3375477

Chan, K. S., & Zary, N. (2019). Applications and challenges of implementing artificial intelligence in medical education: Integrative review. *JMIR Medical Education*, 5(1), e13930. https://doi.org/10.2196/13930

Chen, G., Rolim, V., Mello, R. F., & Gašević, D. (2020). Let's shine together!: A comparative study between learning analytics and educational data mining. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 544–553. https://doi.org/10.1145/3375462.3375500

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Crossley, S., McNamara, D. S., Baker, R., Wang, Y., Paquette, L., Barnes, T., & Bergner, Y. (2015, June 26). Language to completion: Success in an educational data mining massive open online class. *International Conference on Educational Data Mining (EDM)*. International Conference on Educational Data Mining (EDM), Spain.

de Laat, M., Joksimovic, S., & Ifenthaler, D. (2020). Artificial intelligence, real-time feedback and workplace learning analytics to support in situ complex problem-solving: A commentary. *The International Journal of Information and Learning Technology*, 37(5), 267–277. https://doi.org/10.1108/IJILT-03-2020-0026

Desai, S., & Chin, J. (2020). An explorative analysis of the feasibility of implementing metacognitive strategies in self-regulated learning with the conversational agents. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 64(1), 495–499. https://doi.org/10.1177/1071181320641112

Dzikovska, M., Steinhauser, N., Farrow, E., Moore, J., & Campbell, G. (2014). BEETLE II: Deep natural language understanding and automatic feedback generation for intelligent tutoring in basic electricity and electronics. *International Journal of Artificial Intelligence in Education*, 24(3), 284–332. https://doi.org/10.1007/s40593-014-0017-9

Economides, A. A. (2005). Personalized feedback in CAT. *WSEAS Transactions on Advances in Engineering Education*, 2(3), 174–181.

Elatia, S., Ippercieli, D., & Zaiane, O. (Eds.). (2016). Data mining and learning analytics: Applications in educational research. John Wiley & Sons.

Elliott, S. W. (2017). Computers and the future of skill demand. OECD. https://doi.org/10.1787/9789264284395-en

Flodén, J. (2017). The impact of student feedback on teaching in higher education. *Assessment & Evaluation in Higher Education*, 42(7), 1054–1068. https://doi.org/10.1080/02602938.2016.1224997

Forsythe, A., & Johnson, S. (2017). Thanks, but no-thanks for the feedback. *Assessment & Evaluation in Higher Education*, 42(6), 850–859. https://doi.org/10.1080/02602938.2016.1202190

Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice*, 20(2), 150–169. https://doi.org/10.1080/0969594X.2012.749212

Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment: "Breakthrough? or buncombe and ballyhoo?" *Journal of Computer Assisted*

Learning, 37(5), 1207–1216. https://doi.org/10.1111/jcal.12577

Goddard, W. (2021, September 7). Natural language processing in education. IT Chronicles. https://itchronicles.com/natural-language-processing-nlp/natural-language-processing-in-education/

Guo, B., Zhang, R., Xu, G., Shi, C., & Yang, L. (2015). Predicting students performance in educational data mining. *2015 International Symposium on Educational Technology (ISET)*, 125–128. https://doi.org/10.1109/ISET.2015.33

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. https://doi.org/10.3102/003465430298487

Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education: Promises and implications for teaching and learning. The Center for Curriculum Redesign.

Hounsell, D. (2007). Chapter 8: Towards more sustainable feedback to students. In D. Boud & N. Falchikov (Eds.), Rethinking assessment in higher education: Learning for the longer term (pp. 101–113). Routledge.

Hussain, S., Atallah, R., Kamsin, A., & Hazarika, J. (2018). Classification, clustering and association rule mining in educational datasets using data mining tools: A case study. In R. Silhavy (Ed.), Cybernetics and algorithms in intelligent systems, (Vol. 765, pp. 196–211). Springer International Publishing. https://doi.org/10.1007/978-3-319-91192-2_21

Jimenez, L., & Boser, U. (2021, September). Future of testing in education: Artificial intelligence. Center for American Progress. https://www.americanprogress.org/article/future-testing-education-artificial-intelligence

Jurs, P., & Špehte, E. (2021). The role of feedback in the distance learning process. *Journal of Teacher Education for Sustainability*, 23(2), 91–105. https://doi.org/10.2478/jtes-2021-0019

Karaoglan Yilmaz, F. G., & Yilmaz, R. (2021). Learning analytics as a metacognitive tool to influence learner transactional distance and motivation in online learning environments. *Innovations in Education and Teaching International*, 58(5), 575–585. https://doi.org/10.1080/14703297.2020.1794928

Ketchum, C., LaFave, D. S., Yeats, C., Phompheng, E., & Hardy, J. H. (2020). Video-based feedback on student work: An investigation into the instructor experience, workload, and student evaluations. *Online Learning*, 24(3). https://doi.org/10.24059/olj.v24i3.2194

Knight, S., Shibani, A., Abel, S., Gibson, A., Ryan, P., Sutton, N., Wight, R., Lucas, C., Sándor, Á., Kitto, K., Liu, M., Mogarkar, R. V., & Buckingham Shum, S. (2020). AcaWriter: A learning analytics tool for formative feedback on academic writing. *Journal of Writing Research*, 12(1), 141–186. https://doi.org/10.17239/jowr-2020.12.01.06

Kochmar, E., Vu, D. D., Belfer, R., Gupta, V., Serban, I. V., & Pineau, J. (2020). Automated personalized feedback improves learning gains in an intelligent tutoring system. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education* (Vol. 12164, pp. 140–146). Springer International Publishing. https://doi.org/10.1007/978-3-030-52240-7_26

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42–78. https://doi.org/10.3102/0034654315581420

Lan, A. S., Vats, D., Waters, A. E., & Baraniuk, R. G. (2015). Mathematical language processing: Automatic grading and feedback for open response mathematical questions. *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, 167–176. https://doi.org/10.1145/2724660.2724664

Larusson, J. A., & White, B. (Eds.). (2014). Chapter 1: Introduction. In *Learning analytics from research to practice* (pp. 1–12). Springer.

Lemay, D. J., Baek, C., & Doleck, T. (2021). Comparison of learning analytics and educational data mining: A topic modeling approach. *Computers and Education: Artificial Intelligence*, 2, 100016. https://doi.org/10.1016/j.caeai.2021.100016

Li, C., & Xing, W. (2021). Natural language generation using deep learning to support MOOC learners. *International Journal of Artificial Intelligence in Education*, 31(2), 186–214. https://doi.org/10.1007/s40593-020-00235-x

Loyola-Gonzalez, O. (2019). Black-box *vs.* white-box: Understanding their advantages and weaknesses from a practical point of view. *IEEE Access*, 7, 154096–154113. https://doi.org/10.1109/ACCESS.2019.2949286

Magis, D., Yan, D., & von Davier, A. A. (2017). Chapter 1: Overview of adaptive testing. In R. Gentleman, K. Hornik, & G. Parmigiani (Eds.), *Computerized adaptive and multistage testing with R* (pp. 1–5). Springer International Publishing.

Marriott, P., & Teoh, L. K. (2012). Using screencasts to enhance assessment feedback: Students' perceptions and preferences. *Accounting Education*, 21(6), 583–598. https://doi.org/10.1080/09639284.2012.725637

Matcha, W., Uzir, N. A., Gasevic, D., & Pardo, A. (2020). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies*, 13(2), 226–245. https://doi.org/10.1109/TLT.2019.2916802

Mbunge, E., Fashoto, S. G., & Olaomi, J. (2021). COVID-19 and online learning: Factors influencing students' academic performance in first-year computer programming courses in higher education. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3757988

Merceron, A., & Yacef, K. (2005). Educational data mining: A case study. In *Artificial intelligence in education: Supporting learning through intelligent and socially informed technology* (Vol. 125, pp. 467–474). IOS Press.

Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education*, 45(4), 527–540. https://doi.org/10.1080/02602938.2019.1667955

Moreno, A., & Redondo, T. (2016). Text analytics: The convergence of big data and artificial intelligence. *International Journal of Interactive Multimedia and Artificial Intelligence*, 3(6), 57. https://doi.org/10.9781/ijimai.2016.369

National Center for Educational Statistics [NCES]. (2022). Program for the international assessment of adult competencies (PIAAC). Organisation for Economic Co-Operation and Development. https://nces.ed.gov/surveys/piaac/about.asp

Organisation for Economic Co-Operation and Development [OECD]. (2019). PISA 2018 results (volume I): What students know and can do. OECD Publishing.

Ötleş, E., Kendrick, D. E., Solano, Q. P., Schuller, M., Ahle, S. L., Eskender, M. H., Carnes, E., & George, B. C. (2021). Using natural language processing to automatically assess feedback quality: Findings from 3 surgical residencies. *Academic Medicine*, 96(10), 1457–1460. https://doi.org/10.1097/ACM.0000000000004153

Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). Using learning analytics to scale the provision of personalised feedback: Learning analytics to scale personalised feedback. *British Journal of Educational Technology*, 50(1), 128–138. https://doi.org/10.1111/bjet.12592

Pardo, A., Poquet, O., Martinez-Maldonado, R., & Dawson, S. (2017). Provision of data-driven student feedback in la & edm. In C. Lang, G. Siemens, A. Wise, & D. Gasevic (Eds.), Handbook of learning analytics (First, pp. 163–174). Society for Learning Analytics Research (SoLAR). https://doi.org/10.18608/hla17.014

Pask, G. (1982). SAKI: Twenty-five years of adaptive training into the microprocessor era. *International Journal of Man-Machine Studies*, 17(1), 69–74. https://doi.org/10.1016/S0020-7373(82)80009-6

Pechenizkiy, M., Calders, T., Vasilyeva, E., & De Bra, P. (2008). Mining the student assessment data: Lessons drawn from a small scale case study. *Educational Data Mining 2008*.

Pengel, N., Martin, A., Meissner, R., Arndt, T., Neumann, A. T., de Lange, P., & Wollersheim, H.-W. (2021). TecCoBot: Technology-aided support for self-regulated learning. *ArXiv Preprint ArXiv:2111.11881*. http://arxiv.org/abs/2111.11881

Perikos, I., Grivokostopoulou, F., & Hatzilygeroudis, I. (2017). Assistance and feedback mechanism in an intelligent tutoring system for teaching conversion of natural language into logic. *International Journal of Artificial Intelligence in Education*, 27(3), 475–514. https://doi.org/10.1007/s40593-017-0139-y

Piotrkowicz, A., Dimitrova, V., Treasure-Jones, T., Smithies, A., Harkin, P., Kirby, J., & Roberts, T. (2017). Quantified self analytics tools for self-regulated learning with myPAL. *Proceedings of the 7th Workshop on Awareness and Reflection in Technology Enhanced Learning Co-Located with the 12th European Conference on Technology Enhanced Learning (EC-TEL 2017)*, 1997, 16.

Qazdar, A., Er-Raha, B., Cherkaoui, C., & Mammass, D. (2019). A machine learning algorithm framework for predicting students performance: A case study of baccalaureate students in Morocco. *Education and Information Technologies*, *24*(6), 3577–3589. https://doi.org/10.1007/s10639-019-09946-8

Ramachandran, L., Gehringer, E. F., & Yadav, R. K. (2017). Automated assessment of the quality of peer reviews using natural language processing techniques. *International Journal of Artificial Intelligence in Education*, 27(3), 534–581. https://doi.org/10.1007/s40593-016-0132-x

Ramaswami, M., & Bhaskaran, R. (2010). A CHAID based performance prediction model in educational data mining. *IJCSI International Journal of Computer Science*. http://arxiv.org/abs/1002.1144

Ray, S., & Saeed, M. (2018). Applications of educational data mining and learning analytics tools in handling big data in higher education. In M. M. Alani, H. Tawfik, M. Saeed,

& O. Anya (Eds.), *Applications of Big Data Analytics* (pp. 135–160). Springer International Publishing. https://doi.org/10.1007/978-3-319-76472-6_7

Roberts, T. (2019). 5 Natural language processing examples: How NLP is used. *Bloomreach*. https://www.bloomreach.com/en/blog/2019/09/natural-language-processing.html

Rodway-Dyer, S., Knight, J., & Dunne, E. (2011). A case study on audio feedback with geography undergraduates. *Journal of Geography in Higher Education*, 35(2), 217–231. https://doi.org/10.1080/03098265.2010.524197

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. https://doi.org/10.1109/TSMCC.2010.2053532

Ryan, T., Gašević, D., & Henderson, M. (2019). Identifying the impact of feedback over time and at scale: Opportunities for learning analytics. In M. Henderson, R. Ajjawi, D. Boud, & E. Molloy (Eds.), The Impact of Feedback in Higher Education (pp. 207–223). Springer International Publishing. https://doi.org/10.1007/978-3-030-25112-3_12

Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. A. (2020). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, 107, 105512. https://doi.org/10.1016/j.chb.2018.05.004

Siemens, G. (2012). Learning analytics and educational data mining: Towards communication and collaboration. 3.

Sipes, S. (2017). Using learning analytics in SoTL. Center for Innovative Teaching and Learning @IUB. https://blogs.iu.edu/citl/2017/12/13/using-learning-analytics-in-sotl

Solano, Q. P., Hayward, L., Chopra, Z., Quanstrom, K., Kendrick, D., Abbott, K. L., Kunzmann, M., Ahle, S., Schuller, M., Ötleş, E., & George, B. C. (2021). Natural language processing and assessment of resident feedback quality. *Journal of Surgical Education*, 78(6), e72–e77. https://doi.org/10.1016/j.jsurg.2021.05.012

Tempelaar, D. (2020). Supporting the less-adaptive student: The role of learning analytics, formative assessment and blended learning. *Assessment & Evaluation in Higher Education*, 45(4), 579–593. https://doi.org/10.1080/02602938.2019.1677855

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, 47, 157–167. https://doi.org/10.1016/j.chb.2014.05.038

Tsai, Y. S., Mello, R. F., Jovanović, J., & Gašević, D. (2021). Student appreciation of data-driven feedback: A pilot study on Ontask. *LAK21: 11th International Learning Analytics and Knowledge Conference*, 511–517. https://doi.org/10.1145/3448139.3448212

Ubani, S., & Nielsen, R. (2022). Review of collaborative intelligent tutoring systems (CITS) 2009-2021. *2022 11th International Conference on Educational and Information Technology (ICEIT)*, 67–75. https://doi.org/10.1109/ICEIT54416.2022.9690733

UNESCO. (2019). Section 2: Preparing learners to thrive in the future with AI. In Artificial intelligence in education: Challenges and opportunities for sustainable development (pp. 17–24). the United Nations Educational, Scientific and Cultural Organization,. https://unesdoc.unesco.org/ark:/48223/pf0000366994/PDF/366994eng.pdf.multi

Viberg, O., Khalil, M., & Baars, M. (2020). Self-regulated learning and learning analytics in online learning environments: A review of empirical research. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 524–533. https://doi.org/10.1145/3375462.3375483

Wang, D., & Han, H. (2021). Applying learning analytics dashboards based on process-oriented feedback to improve students' learning effectiveness. *Journal of Computer Assisted Learning*, 37(2), 487–499. https://doi.org/10.1111/jcal.12502

Wasson, B., Hansen, C., & Netteland, G. (2016). Data literacy and use for learning when using learning analytics for learners. *CEUR Workshop Proceedings*. http://hdl.handle.net/11250/2429460

Watling, C. J., & Ginsburg, S. (2019). Assessment, feedback and the alchemy of learning. *Medical Education*, 53, 76–85. https://doi.org/10.1111/medu.13645

Winne, P. H., & Baker, R. S. J. d. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *Journal of Educational Data Mining*, 5(1), 8.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). Data mining: Practical machine learning tools and techniques (4th ed.). Elsevier.

Xiong, W., Litman, D., & Schunn, C. (2012). Natural language processing techniques for researching and improving peer feedback. *Journal of Writing Research*, 4(2), 155–176. https://doi.org/10.17239/jowr-2012.04.02.3

Yildirim-Erbasli, S. N., & Bulut, O. (2021). Conversation-based assessments: Real-time assessment and feedback. *ELearn*, *2021*(12). https://doi.org/10.1145/3508017.3495533

Zenisky, A. L., & Hambleton, R. K. (2012). Developing test score reports that work: The process and best practices for effective communication. *Educational Measurement: Issues and Practice*, 31(2), 21–26. https://doi.org/10.1111/j.1745-3992.2012.00231.x

Zhang, H., Magooda, A., Litman, D., Correnti, R., Wang, E., Matsmura, L. C., Howe, E., & Quintana, R. (2019). eRevise: Using natural language processing to provide formative feedback on text evidence usage in student writing. *Proceedings of the AAAI Conference on Artificial Intelligence*, 9619–9625. https://doi.org/10.1609/aaai.v33i01.33019619