Computers and Education: Artificial Intelligence 5 (2023) 100157

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Are deeper reflectors better goal-setters? AI-empowered analytics of reflective writing in pharmaceutical education

Yuheng Li [a], Mladen Raković [a,*], Wei Dai [a], Jionghao Lin [a,b], Hassan Khosravi [c], Kirsten Galbraith [d], Kayley Lyons [e], Dragan Gašević [a], Guanliang Chen [a,*]

[a] Centre for Learning Analytics, Monash University, Clayton, 3168, Victoria, Australia
[b] Human-Computer Interaction Institute, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA
[c] Institute for Teaching and Learning Innovation, University of Queensland, Brisband, Queensland, 4072, Australia
[d] Experiential Development and Graduate Education, Faculty of Pharmacy and Pharmaceutical Sciences, Monash University, Parkville, Victoria, 3052, Australia
[e] Centre for Digital Transformation of Health, University of Melbourne, Parkville, Victoria, 3010, Australia

A R T I C L E   I N F O

Keywords:
Reflective writing
Goal-setting
Self-regulated learning
Machine learning
Deep learning
BERT

A B S T R A C T

Reflection and goal-setting are interrelated processes in well-established educational theories to promote in-depth self-reflection and self-regulated learning. Prior studies have considered reflection to be an important antecedent for meaningful goal-setting. Yet, there lacks empirical evidence to shed light on how students' abilities to reflect inform their abilities to set goals. Hence, in the present study, we aimed to quantify the connection between students' retrospective reflection and their subsequent goal-setting, and derive more in-depth insights to benefit educators in their teaching to promote deeper reflection, more specific goal-setting and better self-regulation. To this end, we utilised two fine-grained coding schemes, adapted from well-established reflection and goal-setting theories, respectively, as well as pertinent prior studies, to annotate the reflective and goal-setting elements within 600 student responses in pharmacy curricula. We visualised such elements as a network graph to study students' joint behavioural patterns in reflecting and setting goals. Then, we statistically analysed the correlation between students' reflective levels and the goal specificities using a Mann Whitney U test. We found that (1) descriptive reflection and goals that included content and actions with additional details more commonly presented jointly; (2) students who reflected deeply tended to set more specific goals. These findings are further summarised and discussed to guide educators to adopt reflective and goal-setting practices when designing teaching activities. Moreover, driven by these findings, we emphasised the significance of aiding instructors to provide timely assessment to students' written reflections so as to further ameliorate students' reflective abilities. Therefore, we attempted to automate such assessments using five traditional machine learning algorithms and one deep learning approach based on Bidirectional Encoder Representation of Transformers (BERT), and discovered that BERT gave the best performance in terms of identifying reflective sentences and differentiating various reflective elements.

1. Introduction

Self-Regulated Learning (SRL) is theorised as a cyclical and goal-oriented process in which learners set their learning goals and engage in different learning behaviours to accomplish those goals (Winne & Hadwin, 1998, Winne, 2018, Zimmerman, 2002). In this process, self-regulated learners often reflect on their learning progress relative to goals they have set and, if they deem it necessary, alter their learning behaviours or modify their goals to boost the effectiveness of their learning (Winne & Hadwin, 1998). Engagement in SRL is commonly considered to benefit students' academic success across different subjects and learning tasks (Cleary & Chen, 2009, Zimmerman, 2000), which, subsequently, may help students become more productive lifelong learners (Klug et al., 2011).

Goal-setting and reflection on prior studies are considered critical processes in most of the prominent SRL theoretical models (see (Panadero, 2017) for an overview). In particular, setting goals that are specific, measurable, attainable, realistic and time efficient (Doran et

* Corresponding authors.
E-mail addresses: mladen.rakovic@monash.edu (M. Raković), guanliang.chen@monash.edu (G. Chen).

https://doi.org/10.1016/j.caeai.2023.100157
Received 25 May 2023; Received in revised form 16 July 2023; Accepted 16 July 2023
Available online 31 July 2023
2666-920X/

al., 1981) has been documented to improve academic achievements, (e.g., (Mento et al., 1987, Kleingeld et al., 2011), (Locke et al., 1981), (Bryan & Locke, 1967), (Shell, 2020), (Acee et al., 2012), (Alessandri et al., 2020)), because such goals define explicit standards against which learners can compare their performance (Marzouk et al.). Even though researchers and educators have proposed several instructional approaches to support goal-setting, such as abiding well-established theoretical goal-setting guidelines (e.g., (McCardle et al., 2017, Morisano et al., 2010)), and these approaches have helped learners improve the quality of their goals (McCardle et al., 2017) and academic performance (Morisano et al., 2010), research to date appears to mainly focus on teaching learners how to set effective goals following the well-established set of rules. However, students commonly tend to set goals that lack specificity, even after the importance of setting specific goals was purposely emphasised during instruction (McCardle et al., 2017).

To improve students' abilities to set specific goals and promote their self-regulation, it is necessary to examine how students' goal-setting behaviours relate to other processes underlying SRL. It has been theorised (Winne & Hadwin, 1998) and empirically documented (Raković et al., 2022) that setting goals for future studying is associated to learners' reflection on prior studying, another critical process in SRL. By reflecting on prior studying, self-regulated learners recall their past learning experience (e.g., while they were studying before the exam) and feelings related to that experience, evaluate their prior studying behaviours against their goals and make decisions to adapt their future studying behaviours, often by setting new goals (Winne & Hadwin, 1998, Raković et al., 2022). Furthermore, goal-setting is commonly considered an important activity in other theoretical models that emphasise students' reflective practice in different educational contexts (Kolb, 1984, Gibbs, 1988, Killion & Todnem, 1991, Boud et al., 1985a, Driscoll, 2006). Even though students' reflective behaviours are theorised to be relevant to goal-setting and ensuring productive SRL, the relationship between students' abilities to reflect and set goals appeared to be insufficiently examined in prior research. For instance, more research is needed to identify how different, fine-grained reflective behaviours students engaged in may be related to goals that those students set after they engaged in reflections. This, in turn, may provide researchers with new opportunities to understand and support goal setting via learners' reflective practice, and ultimately enable productive SRL.

Educators oftentimes task students with reflective writing assignments to capture students' states of reflection (Charon & Hermann, 2012). Such states are represented in natural language which generally describes the consciously stated purpose and/or the exact context of reflection, with a specific outcome in regards to learning, action or clarification (Moon, 2007). Writing samples collected in this way may reveal more fine-grained information about students' reflective and goal-setting behaviours, contributing a new source of data for instructors to better understand students' learning process and potentially new knowledge to the learning sciences. Hence, the first aim of our current study was to determine the relationship between students' abilities to reflect and set goals. With this in mind, we collected a total of 14,908 reflective responses written by students in the setting of pharmaceutical education and manually annotated 600 of them. The responses included students' retrospective reflections, i.e., reflections on prior learning experience in an educational scenario such as a placement or an internship, and prospective goal-setting, i.e., goals learners set for future studying in the same scenario. To annotate these reflective responses, we utilised the fine-grained coding schemes rooted in well-established theories and prior studies of reflective practice and goal settings to enable the comprehensive evaluations of students' reflective and goal-setting behaviours. To accomplish the first aim of our study, we analysed the co-occurrence of reflective and goal-setting elements identified in students' reflective responses by visualising them in a network graph, and statistically examined the correlation between students' levels of reflections and qualities of their set goals using the Mann-Whitney U test. Our findings demonstrated that students more commonly reflected by describing their prior experience and feelings while setting goals that contained specific actions and contents (i.e., areas of improvement). More importantly, we observed that students who reflected more deeply on their prior studying tended to set more specific goals for future studying, whereas students whose reflective behaviours were not as deep tended to include fewer details in their set goals.

Driven by the findings described above, we posited that it was necessary for instructors to provide timely support and feedback to students, who were less capable of reflecting critically, not only to promote more in-depth reflection but also to facilitate more comprehensive goal-setting. In this respect, the instructors were expected to efficiently assess the reflective responses compiled by students. More importantly, educators need to instruct learners on how to identify, compare and contrast the features of descriptive reflections with those of evaluative or critical reflections and explicitly demonstrate such annotations on these features to promote students' attempts to write reflectively (Ryan, 2011). However, conducting assessments and annotations in a timely manner was impractical, especially under a high student-teacher ratio in higher education (Koc & Celik, 2015), as it heavily relied on laborious manual efforts. To address this challenge, researchers have developed various text analytical approaches based on Natural Language Processing Natural Language Processing (NLP) for automated analysis of reflective writing in different education scenarios (e.g., (Liu et al., 2017, Cheng, 2017, Kovanović et al., 2018, Ullmann, 2019, Wulff et al., 2022)). However, yet limited attempts have been made to enable in conjunction the assessment of students' overall reflective levels and the fine-grained analysis of the reflective elements embedded within students' reflective responses, especially in the context of pharmaceutical education. Additionally, none of the existing studies, to the best of our knowledge, have compared the performance of machine learning-based NLP approaches with that of the deep learning-based ones in the automatic identification of reflective elements.

To address the above research gap, the second aim of our study was to automate the evaluation of students' reflective writings in two steps: 1) to differentiate reflective sentences from non-reflective ones in students' reflective writings, and then 2) to identify the reflective elements each reflective sentence correspond to. To this end, we implemented models on the basis of the state-of-the-art pre-trained language model Bidirectional Encoder Representations of Transformers (BERT) (Devlin et al., 2019) which showed superior performance compared to other deep learning models in a prior study (Wulff et al., 2022). For comparison, we developed classifiers based on five popular machine learning algorithms adopted by pertinent prior studies. By doing so, we expected to find out the extent to which the deep learning and machine learning approaches were able to identify the reflective elements in our coding scheme that enabled overall assessment to students' levels of reflection. Our results indicated that the BERT-based models outperformed all machine learning-based ones, achieving F1 scores ranging from 0.82 to 0.92.

2. Background

Goal-setting is a central process to self-regulated learning (Panadero, 2017). Although the roles of goal-setting across SRL theoretical models may differ, researchers commonly agree that (1) goals provide the context to interpret tasks and (2) setting goals guides and motivates productive SRL behaviours. However, setting a general goal (e.g., "I will do my best") is typically insufficient to promote self-regulation (Schunk, 2001). Instead, high-quality goals, i.e., the goals that clearly specify time constrains, actions, standards and content for learning tasks, have been considered to benefit SRL (McCardle et al., 2017). To assist students set specific goals, researchers utilised different instructional frameworks, e.g., TASC (i.e., timeframe, action, standard, content) (McCardle et al., 2017) and SMART (i.e., specific, measurable, attainable, relevant, time-bounded) (Wollny et al., 2019). Researchers showed that setting specific goals can increase engagement in self-regulation (Schunk, 2001, Koch & Nafziger, 2011, Latham & Locke, 1991) and improve learning gains (Mento et al., 1987, Kleingeld et al., 2011, Locke et al., 1981, Bryan & Locke, 1967, Oistad, 2020). Likewise, being another central process to SRL, reflection on prior studying can also support self-regulation and help learners boost their abilities in e.g., critical thinking and problem solving (Dewey, 1933, Mann et al., 2009). Even though reflection was defined differently by several researchers (Boud et al., 1985b, Dewey, 1933, Mezirow, 1991, Moon, 1999), they generally agree that reflection should involve a purposeful and critical analysis of the past knowledge, feelings and experiences of oneself, which can lead to a better self-awareness and deeper comprehension of the knowledge. As a result of such reflection, students can increase their awareness about what did and what did not work well for them in prior studying to inform goal-setting and facilitate more effective future studying.

Educators have widely utilised reflective writing to capture students' states of reflection and measure the attainment of skills, while allowing students to share their thoughts, motives and feelings (Charon & Hermann, 2012). Ryan (2011) suggested that, to help students more effectively engage in reflection, educators should instruct them on how to identify, compare and contrast features of descriptive reflection and those of evaluative and critical reflection. To this end, educators need to comprehensively assess students' reflective writings to gain a deeper insight into their reflective behaviours, such as by identifying and annotating the presence of different reflective elements in students' reflective writings, e.g., learning strategies previously used, evaluation of effectiveness and anticipated usefulness of those strategies in similar learning tasks in the future. However, assessing students' written reflections to provide feedback in a timely manner is often hindered by the complexity and high rhetorical demand of the writing task itself, especially in pharmacy curricula (Tsingos et al., 2015), posing a need for automating such assessments not only to assist educators in timely intervention (Lucas et al., 2018) but also to help students improve their reflective ability (King, 2002).

Existing research on automatic reflective writing assessment mainly adopted three approaches: dictionary-based, rule-based and machine learning-based (Ullmann, 2019). While dictionary-based and rule-based approaches have been used to successfully identify the presence of reflective texts in student writing, they oftentimes relied on expert knowledge to derive words or patterns of expressions which may limit their reliability and generalisability (Ullmann, 2019). As an alternative, researchers have been increasingly using machine learning approaches to automatically detect linguistic reflective patterns.

For example, Liu et al. (2017) used Naive Bayes to classify teachers' reflective actions in online discussion posts, and achieved F1 scores between 0.79 and 0.85. Cheng (2017) used latent semantic analysis to classify students' reflective skills and provide a holistic reflection assessment, achieving accuracy scores between 0.72 and 0.82. Kovanović et al. (2018) classified students' reflective behaviours on a sentence level with a Random Forest classifier and attained an overall accuracy of 75%. Ullmann (2019) implemented several machine learning models to classify reflective elements of students' reflections and obtained accuracy scores between 0.71 and 0.96. Liu M. et al. (2019) extracted linguistic features from student reflective writing and trained multiple predictive models to differentiate reflective stages in corpora. The authors identified Random Forests as the best performing model with a F1 score of 0.79. Wulff et al. (2022) harnessed the power of deep learning models (i.e., BERT, Long Short-Term Memory, Convolution Neural Network and Feedforward Neural Network) in classifying reflective elements of reflective essays written by pre-service physics teachers', identifying BERT as the best-performing model with a weighted F1 score of 0.81.

2.1. Summary and RQs

Despite the strong connection between reflection and goal-setting in theories of SRL and reflective practice, limited attempts have been made to empirically examine the relationship between students' reflective and goal-setting behaviours, which contributes potential new insights about students' SRL for educators and researchers. To address this gap, we posed the following research question:

- **RQ1**: To what extent do students' reflective behaviours correlate to their goal-setting behaviours?

The analysis of reflective writing is empirically challenging and the advancement of NLP approaches made the automation of these processes more probable. Researchers to date have automatically assessed reflective writing either on a sentence level (e.g., (Kovanović et al., 2018, Ullmann, 2019, Wulff et al., 2022)) or on a corpus level (e.g., (Liu et al., 2017, Cheng, 2017, Liu M. et al., 2019)). While assessing students' reflective writings on a corpus level identifies students' overall capability in reflection, analysing and classifying reflective sentences into single categories can reveal more fine-grained characteristics of reflective writing to delineate student's "reflection profile". However, few studies have been conducted to date to automatically analyse both students' overall reflective capability and the fine-grained characteristics of reflective writing, including reflective writing in pharmaceutical education. Additionally, even though machine learning and deep learning-based NLP methods have demonstrated certain potential in the classification of reflective texts, some reflective elements were still challenging to be identified and all studies, to our knowledge, assumed that one unit of analysis (e.g., sentence) can be classified into one, rather than multiple reflective categories. There is also a lack of empirical study to compare the performance of machine learning and deep learning-based methods in addressing this challenging classification task. To address this gap, we posed the following research question:

- **RQ2**: To what extent can machine learning and deep learning-based NLP techniques accurately identify fine-grained rhetorical elements in students' written reflection?

3. Methods

3.1. Task and dataset

The reflective responses used in this study were collected from 1,321 pharmacy students at an Australian University. Ethics approval has been obtained from Anonymous University for the collection of students' reflective writing. The reflections were written between 2017 and 2019, as a part of a coaching program aimed at helping students improve their professional skills in pharmacy domain, e.g., communication, teamwork, empathy, integrity, inquiry, and problem-solving (Malone et al., 2021). Before each coaching session, students were required to write a reflection on different topics, including exams, internships and placements.

The students were instructed to develop reflections following the framework *"What"*, *"So What"* and *"Now What"* (Driscoll, 2006). Specifically, in the *"What"* section, the students were prompted to describe their prior learning experience in an educational scenario (e.g., rehearsing a skill for a practical exam), reason about/evaluate their feelings related to that experience, identify any changes in their thoughts/beliefs as a result of that experience, and offer new insights into their learning. In the *"So What"* section, the students were prompted to reflect on how their learning experiences, and reactions or feelings related to those experiences may have impacted their life at the moment, how those may impact their future development and/or how their previous studying approaches may be improved. In the *"Now What"* section, the students were prompted to provide concrete and practical plans about how they would improve their skills, based on what they reflected upon in the previous two sections. The students were asked to set their goals following the SMART framework (i.e., Specific, Measurable, Attainable, Relevant and Time-bounded) (Rubin, 2002).

We collected 14,908 reflections. The average word count of students' responses for the *"What"*, *"So What"* and *"Now What"* sections was 108.14 ($SD = 78.18$), 121.03 ($SD = 75.84$) and 106.36 ($SD = 66.45$) words, respectively. The total number of reflections was 2,649 created in 2017, 5,230 in 2018, and 7,029 in 2019.

3.2. Coding schemes

As students' reflective responses followed the *"What, So-What, Now-What"* framework, these responses may reveal students' retrospective reflections (i.e., reflections on past knowledge, feelings and experiences) and students' prospective learning goals, based on retrospective reflections. We noted that the reflective writing framework (Driscoll, 2006) utilised in our study was inspired by the theories of reflection and experiential learning by Boud et al. (1985a), Dennison and Kirk (1990) and Kolb (1984). The reflective prompts provided to students aligned with the well-regarded theoretical framework proposed in Boud et al. (1985a). The framework posits that reflection involves three main stages:

1. *Returning to Experience*: learners chronologically review their past experience;
2. *Attending to Feelings*: learners reflect on their feelings related to that experience;
3. *Re-Evaluating Experience* includes five fine-grained and interrelated processes:
   (a) *Association*: learners associate their feelings and knowledge from past experience and the present moment, and identify changes, e.g., inconsistent assumptions and improved comprehension;
   (b) *Integration*: learners integrate the associations of the past and the present into new feelings and perspectives;
   (c) *Validation*: learners validate their newly integrated perspectives and feelings against the existing ones to find (in)consistency;
   (d) *Approximation*: learners attach their newly integrated ideas and feelings to their current personal life or to their future development;
   (e) *Outcomes and Action*: learners list possible actions to be implemented as a result of the reflection.

An alignment between the expectations for the *"What"* section and the descriptions to *Returning to Experience*, *Attending to Feelings*, *Association* and *Integration* could be observed, and so did the expectations for the *"So What"* section and the descriptions to *Validation* and *Approximation*. Similarly, the *"Now What"* section was supposed to elicit *Outcomes and Action* which can be used to understand students' ability to set goals. Additionally, as mentioned in Boud et al. (1985a) that ones' reflective behaviours commonly proceeded in linear stages within their framework, it enabled us to quantify students' levels of reflection.

Rubric for Evaluating Retrospective Reflection: We adopted and modified the rubric proposed by Tsingos et al. (2015) to evaluate students' retrospective reflections on a sentence level. This rubric was used in a context similar to our study and it built-upon the theoretical framework proposed by Boud et al. (1985a), i.e., the framework that informed the reflective prompts in this study. We excluded *Outcomes and Action* (goal-setting related categories) from (Tsingos et al., 2015), because students' set goals were assessed by the rubric described in Section 3.2 to more comprehensively reveal the goal specificities. The rubric by Tsingos et al. (2015) also included another reflection model by Mezirow (1991) to differentiate non-reflective behaviours from reflective and critically reflective ones. Similar to Liu M. et al. (2019), to better fit the context of our study, we further adopted their rubric (Tsingos et al., 2015) by combining the specifications for reflectors and critical reflectors as such a combination was on a higher level of Mezirow's hierarchical model (Mezirow, 1991). To this end, phrases like "show some evidence" and "clearly provide edivence" were grouped together, simplifying the identification of reflective behaviours and avoiding ambiguities. Furthermore, we adjusted the wording of the specifications (e.g., change "little or no evidence" to "no evidence") to ensure a precise characterising of non-reflective behaviours. Accordingly, our rubric included six reflective processes, each with two levels (i.e., reflective and non-reflective) enabling the assessment of students' reflective writing from both the breadth and depth dimensions (for rubric details, see Table 1).

Rubric for Evaluating Prospective Goal-setting:For the in-depth evaluation of students' goal-setting, we composed a rubric based on the most common SMART interpretation (Rubin, 2002) which delineated a high-quality goal as being specific, measurable, attainable, relevant and time-bounded (for rubric details, see Table 2). It was expected that students set goals as specific as possible by including 1) the specific future actions or routines to be followed instead of vague intentions (Locke et al., 1981, Schut & Stam, 1994) (i.e., *action specific*), 2) the specific skills to be improved and/or the outcomes to be achieved given the planned actions (i.e., *content specific*), 3) the specific metrics to enable the measurement of improvements and/or achievements instead of the subjective estimates of task or goal difficulty (Locke et al., 1981) (i.e., *standard specific*), 4) the specific time to measure such improvements and/or achievements, or the specific frequency for the planned actions (i.e., time-frame specific), and 5) as many details as possible (e.g., locations, people involved, possible challenges, possible benefits) in regards to the planned actions to ensure they were relevant and attainable (i.e., *details specific*).

We formulated the rubric in a way that all criteria involved could be assessed objectively to ensure an accurate and consistent evaluation to students' prospective goal-setting on a sentence level.

3.3. Data annotation

In our initial analysis of students' written reflections, we found that some students did not comply with the given instructions (e.g., not following the sequential order of the reflective writing prompts and not clearly identifying sections within the response). Such responses from students with formatting issues posed challenges to our annotation and were thus excluded, resulting in a total number of 9,828 students' reflective responses for analysis. Subsequently, given that the manual assessment of reflective writing is time-consuming, we randomly selected 600 reflective responses to be annotated utilising the rubrics presented in Section 3.2. It was worth noting that in our annotation, we allowed one sentence to be assigned to more than one process/criterion.

Two coders with relevant expertise were involved in the annotation process. The two coders were trained by annotating 10 reflective responses together based on the rubric to ensure consistency. Then, they independently annotated a subset of 20 reflective responses to measure the inter-coder agreement. Due to the large number of categories and differences in student writing, it took the coders five rounds of training, independent annotations and conflict resolutions to achieve the inter-rater agreement of 80%, a commonly accepted reliability threshold in discourse analysis (Artstein & Poesio, 2008), and each round with a newly sampled set of responses. Some annotation rules were refined during this stage (e.g., though we expected students to use "I" as pronoun in *Returning to Experience*, we accepted the use of "we" as pronoun; though we expected students to use future tense in *action specific*, we accepted the use of past future tense). As a result, the two coders achieved 81.2% agreement in the training stage to proceed to the annotation of the 600 selected reflective responses. The conflicts of the annotation stage were addressed in a case-by-case manner where the two coders discussed and resolved the disagreements. We noticed that a portion of

Table 1
Rubric for Evaluating Retrospective Reflection.

Stage and Process		Non-reflective	Reflective
Returning to Experience		Experience was not clearly described	Experience was clearly described (the description may include chronological information or personal judgements)
Attending to Feelings		Personal feelings were not described or were described without evaluation	Personal feelings were described with judgements/reasons provided
Re-evaluating Experience	Association	No links between prior knowledge, feelings or attitudes, and newly acquired knowledge, feelings or attitudes	Links between prior knowledge, feelings or attitudes, and newly acquired knowledge, feelings or attitudes
	Integration	Association between prior and new knowledge, feelings or attitude provided, but new insights not provided	Association between prior and new knowledge, feelings or attitude provided, and new insights provided
	Validation	No self-assessment of the new insights or no reference to prior experience provided	Self-assessment of the new insights provided, with reference to prior experience
	Approximation	New insights not related to current life and/or future development	New insights related to current life and/or future development (may specifically infer the impact of the connections)

Table 2
Rubric for Evaluating Prospective Goal-setting.

Criteria	Not specific	Specific
Action specific	Goal only vaguely stated a plan without in depth descriptions	Goal specifically described the actions or routines to be used to achieve the goal
Content specific	Goal did not mention specific aspects to be improved	Goal specified what exactly would be achieved or improved
Standard specific	Goal was not measurable to mark progress	Goal clearly described metrics to measure progress of achievement
Time frame specific	Goal did not include an exact time or frequency for the action or a deadline for the goal	Goal supplied an exact time for the actions to be done and/or a frequency of the actions and/or a deadline for the goal
Details specific	Goal did not include other details about the actions	Goal included additional details about the actions (any of locations, people involved, evaluations of the action, possible challenges, possible solutions to challenges, possible benefits)

students' reflective responses did not present reflective processes and/or goal-setting criteria at all. Such responses were excluded in our later analyses, retaining a total number of 493 student responses.

3.4. Data analysis

To identify the relationship between students' reflective and goal-setting behaviours (RQ1), we visualised student reflective processes and goal-setting criteria as a network graph. Each node in the network graph represented either a reflective process or a goal characteristic. The nodes were presented in different sizes to indicate frequency of corresponding categories in the 493 annotated samples. The nodes were connected by edges that had various thicknesses depicting frequencies of co-occurrence between each pair of nodes. By delineating the elements and their connections as a network graph, we were able to identify the elements more commonly expressed by the students both individually and jointly and thus, obtain a better understanding to students' behavioural patterns in reflection and goal-setting.

To reveal the correlation between students' abilities to reflect and set goals, we further grouped students into *shallow reflector* and *deep reflector* to statistically analyse the specificities of their set goals. Similar to Sen (2010), we characterised a student as either a *shallow reflector*, i.e., students with only reflections of *Returning to Experience* and/or *Attending to Feelings*, which were delineated to be descriptive in Boud et al.'s model (Boud et al., 1985a); or a *deep reflector*, i.e., students with reflections reaching any of the *Association*, *Integration*, *Validation* and *Approximation*, where they started to re-assess their prior experience in these processes to establish connections between the past and the present (Boud et al., 1985a).

Such a grouping followed the sequential reflective stages Boud et al. established in their model (Boud et al., 1985a) that reflection started from *Returning to Experience* to *Attending to Feelings* and eventually to *Re-evaluating Experience*. We then calculated a specificity score for each prospective goal-setting based on the number of criteria met, and investigate whether the specificity scores of *shallow reflector* differed from those of *deep reflector* significantly by adopting the Mann-Whitney U test (Mann & Whitney, 1947) provided in the toolkit *SciPy*[1] as it had no requirements in regards to distribution or sample size equality.

3.5. Predictive models

To address RQ2, we implemented several predictive models 1) first to differentiate reflective sentences from non-reflective ones, and then 2) to classify the reflective sentences into their corresponding reflective stages specified in Table 1. By doing so, our predictive results could enable the categorisation of a retrospective reflection as either a *shallow reflector* or a *deep reflector* in alignment with the grouping in Section 3.4.

We utilised machine learning algorithms including Naive Bayes (NB), Support Vector Machine (SVM) and Random Forest (RF) as they had evinced potentialities in similar prior studies (Liu et al., 2017, Ullmann, 2019, Kovanović et al., 2018). Additionally, we included Logistic Regression (LR) and XGBoost as they were widely demonstrated to be effective in prior studies classifying educational texts (Sha et al., 2021, Li et al., 2022). We also harnessed the power of the state-of-the-art pre-trained language model BERT (Devlin et al., 2019) to implement deep learning models inspired by their superb performance (Sha et al., 2021, Li et al., 2022, Wulff et al., 2022).

As pointed out in Ullmann (2019) that not all sentences within students' reflective writing delivered reflective elements, our first aim was to examine the performance of the predictive models in distinguishing between reflective and non-reflective sentences within the reflections. To this end, sentences annotated with corresponding reflective processes were considered reflective while those without were marked as non-reflective, leading to 1,858 reflective sentences and 2,679 non-reflective ones. To address the class imbalance issue, we utilised the random under-sampling approach, as prior study has suggested that it outperformed other sampling techniques in binary classification tasks when data were not highly unbalanced (Chen et al., 2011). Hence, we implemented five traditional machine learning classifiers and a BERT-based classifier for the classification of reflective and non-reflective sentences. Subsequently, we examined the performance of the same

[1] https://scipy.org/.

approaches in distinguishing among the three reflective stages (i.e., *Returning to Experience*, *Attending to Feelings* and *Re-Evaluating Experience*) for the reflective sentences. We implemented the models in two manners. The first was to implement binary classifiers to distinguish whether a reflective sentence corresponds to a specific reflective stage. The same under-sampling class balancing technique was used in these binary classifications. The second was to implement multi-class classifiers to identify the reflective stages for the sentences.

Preliminary data pre-processing was uniformly done for all analyses including lowercasing and removing contents between brackets because, during annotation, we found that such contents (e.g., "the evidence can be found in appendix") tended to be meaningless for the context of the sentences. Afterwards, we extracted textual features demonstrated to be promising in a similar prior study (Kovanović et al., 2018) and other educational text classification studies (e.g., educational forum posts (Sha et al., 2021), learning outcomes (Li et al., 2022)) for our machine learning models. Specifically, the features we extracted were: 1,000 most frequent uni-grams, 1,000 most frequent bi-grams, and 1,000 most frequent terms using TF-IDF all exclusive of stopwords; the automated readability index (Senter & Smith, 1967); and a total number of 93 LIWC features (Pennebaker et al., 2015). In the meantime, word tokens were generated as inputs to the BERT-based models with a *HuggingFace* (Wolf et al., 2019) pre-trained tokenizer `BERT-base-uncased`.

The NB, SVM, LR, RF and XGBoost models were implemented with the aid of *Scikit-Learn* (Pedregosa et al., 2011) and *XGBoost* (Chen & Guestrin, 2016) while the `BERT-base-uncased` by *HuggingFace* (Wolf et al., 2019) was fine-tuned in our study. The dataset was split in an 80:20 ratio – 80% of the data were used for model training while the remaining 20% served the purpose of evaluating the performance of the classifiers. As machine learning and deep learning models employed different strategies for optimisation, for the machine learning models, we conducted hyperparameter tuning using grid search with 3-fold cross-validation given the 80% training data to reliably select the optimal sets of parameters for each model, with F1 score as the evaluation metric. For deep learning models, we further split the 80% training data, pertaining 80% for model fine-tuning and 20% for validation. The BERT-based models had a total of 12 hidden layers, each accommodating 768 neurons. For the binary classifications, an additional linear layer, with two output neurons and *softmax* as the activation function, was integrated to `BERT-base-uncased`. For the multi-class classification, an additional linear layer, with three output neurons and *softmax* as the activation function, was appended to `BERT-base-uncased`. To predominantly avoid overfitting in the deep learning models, we introduced roll-back mechanism to return the models to their best-performing state based on the F1 scores on the evaluation set. Thereupon, we measured the performance of the models on the testing set using several metrics including accuracy, Cohen's κ, Area Under ROC Curve (ROC-AUC) and F1 score for their reliability in measuring model performance (Sha et al., 2021, Li et al., 2022).

4. Results

In 493 reflective responses we analysed, 1,165 sentences were labelled as *Returning to Experience*, 432 as *Attending to Feelings*, 168 as *Association*, 67 as *Integration*, 11 as *Validation*, and 15 as *Approximation*. Students mainly described their prior experience. Many reflective responses also contain students' personal feelings, evaluations and/or inference related to those feelings. Further, approximately one third of the students re-evaluated their prior experience to different extents. At the same time, 738 sentences of goal setting responses were labelled as *action specific*, 376 as *content specific*, 298 as *details specific*, 19 sentences as *standard specific*, and 167 sentences as *time frame specific*. Students commonly included planned actions in their set goals. Most of these actions target specific areas of improvements (i.e. content) and include additional details. Roughly one third of these goals have incorporated specified frequencies or time-frames for the planned actions. However, students rarely defined standards to measure the achievement of their goals.

Table 3
Classification Performance for differentiating reflective vs. non-reflective sentences by the six approaches (i.e., NB, SVM, LR, RF, XGBoost, and BERT). The best performance for each evaluation metric is highlighted in bold.

Methods	Accuracy	Cohen's κ	AUC	F1
NB	0.5745	0.1735	0.5905	0.6040
SVM	0.6596	0.3125	0.6568	0.6220
LR	0.6684	0.3252	0.6619	0.6176
RF	0.6649	0.3084	0.6514	0.5846
XGBoost	0.6844	0.3553	0.6762	0.6292
BERT	**0.7730**	**0.5326**	**0.7619**	**0.7217**

4.1. Relation between students' abilities to reflect and set goals

The connections among the reflective processes and goal-setting criteria visualised as a network graph (i.e., Fig. 1) illustrates students' behavioural patterns in reflection and goal-setting. We can observe that *Returning to Experience*, *Attending to Feelings*, *action specific*, and *content specific* were most regularly delivered both individually and in conjunction followed by *Details specific* in students' reflective writing. We found that *Association*, *Integration* and *time frame specific* co-occurred less frequently with others in the network, and *Validation*, *Approximation* and *standard specific* were rarely delivered jointly with other elements in students' reflective responses. Interestingly, we discovered that *Validation*, *Approximation* and *standard specific* were not connected to each other in the network. These results imply that:

1. students who returned to their experience when reflecting were able to plan future actions in their set goals; for these students:
   (a) most of them tended to attend to their feelings and specify areas of improvements; for these students:
      i. a good number of them included additional details in their planned actions;
      ii. a moderate number of them managed to associate prior experience with present one to re-evaluate their experiences;
2. higher-level reflective processes (i.e., *Validation* and *Approximation*) and more concrete levels of details in goal-setting (i.e., *standard specific*) were harder to achieve both respectively and in conjunction, and hence, educators need to contribute more efforts to promote deeper reflection and more comprehensive goal-setting.

Our grouping regarding the different reflective levels led to 352 *shallow reflectors* and 141 *deep reflectors*. The respective sets of specificity scores for their set goals were compared. The *deep reflector* ($Mdn = 4$; $Q1 = 3$; $Q3 = 4$) seemed to set more specific goals than the *shallow reflector* ($Mdn = 2$; $Q1 = 2$; $Q3 = 3$) and our one-sided Mann-Whitney U test (i.e., the distribution of the specificity scores for *deep reflectors* is stochastically greater than the distribution underlying the specificity scores for *shallow reflectors*) confirmed that this difference was statistically significant, $U(N_{shallow\,reflector} = 352, N_{deep\,reflector} = 141) = 7278.5, z = 2.294, p = .009$ with an effect size of $r = 0.216$. Therefore, we can conclude from this result that for students who reflect deeply, their set goals tend to be more specific than those from students who reflect superficially.

4.2. Predictive model performance

We can observe that XGBoost is the best-performing model compared to other traditional machine learning models, whereas the BERT-based deep learning model remarkably outperformed all machine learning models, achieving an accuracy of 77.3%, a Cohen's κ of 0.53, an AUC score of 0.76 and a F1 score of 0.72 (Table 3). In the cases of binary classifications, RF tended to be the most robust classifier, except in the

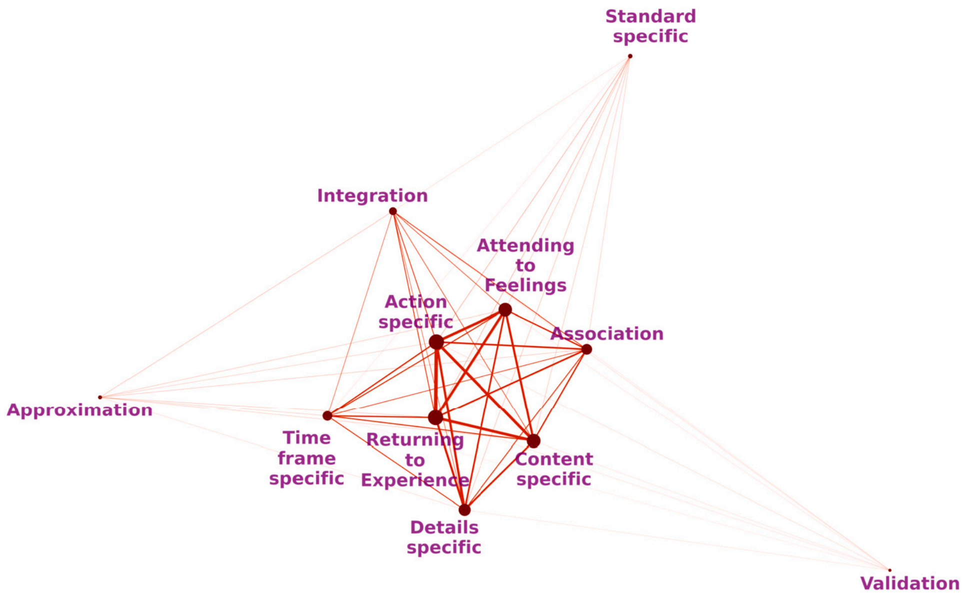

Fig. 1. Network Graph for Reflective Processes and Goal-specific Criteria (Node Size: Occurrence Frequency of Each Node; Edge Thickness: Co-occurrence Frequency of Two Nodes).

Table 4
Classification Performance for differentiating among reflective stages by the six approaches (i.e., NB, SVM, LR, RF, XGBoost, and BERT) in two manners (i.e. binary and multi-class). The best performance for each evaluation metric is highlighted in bold for each task.

Methods	Binary Classification												Multi-Class Reflective Stage Classification			
	Returning to Experience				Attending to Feelings				Re-Evaluating Experience							
	Acc.	Cohen's κ	AUC	F1	Acc.	Cohen's κ	AUC	F1	Acc.	Cohen's κ	AUC	F1	Acc.	Cohen's κ	AUC	F1
NB	0.6643	0.3281	0.6639	0.6245	0.7090	0.4179	0.7090	0.7273	0.5638	0.1277	0.5638	0.5773	0.5366	0.2512	0.6379	0.5039
SVM	0.7385	0.4770	0.7385	0.7376	0.7015	0.4030	0.7015	0.7143	0.5426	0.0851	0.5426	0.5905	0.6992	0.4624	0.8388	0.6291
LR	0.7244	0.4487	0.7244	0.7214	0.7239	0.4478	0.7239	0.7176	0.6170	0.2340	0.6170	0.6400	0.6992	0.4691	0.8342	0.6368
RF	0.7314	0.4628	0.7314	0.7246	0.8134	0.6269	0.8134	0.8000	0.6702	0.3404	0.6702	0.7103	0.7114	0.4682	0.8625	0.6193
XGBoost	0.7032	0.4063	0.7031	0.6978	0.7910	0.5821	0.7910	0.7879	0.6383	0.2766	0.6383	0.6600	0.6992	0.4681	0.8456	0.6466
BERT	**0.8304**	**0.6609**	**0.8305**	**0.8356**	**0.9179**	**0.8358**	**0.9179**	**0.9160**	**0.8191**	**0.6383**	**0.8191**	**0.8211**	**0.8577**	**0.7567**	**0.9506**	**0.8391**

classification of whether a reflective sentence corresponded to *Returning to Experience* where SVM slightly outperformed RF (Table 4). However, the BERT-based deep learning models still achieved considerably better performance than all machine learning models. The reflective stage *Attending to Feelings* seemed to be the easiest one to differentiate followed by *Returning to Experience* and then *Re-Evaluating Experience*. For the multi-class classification of reflective stages, SVM, LR, RF and XGBoost achieved similar performance. The BERT-based deep learning model significantly outperformed the machine learning models in all evaluation metrics, achieving accuracy scores and F1 scores greater than 0.82 and Cohen's κ greater than 0.64 which indicated substantial agreement between the models and the coders.

By comparison, we found that:

- for the classification of reflective stages:
  - with machine learning approaches, binary classifications achieve overall better performance than multi-class classification in most cases;
  - with BERT-based deep learning approach, binary classifications achieve similar overall performance compared to multi-class classification;
- the overall performance for differentiating reflective vs. non-reflective sentences is worse than that of distinguishing reflective stages.

5. Discussion

Prior research has provided theoretical foundations (Panadero, 2017, Kolb, 1984, Gibbs, 1988, Killion & Todnem, 1991, Boud et al., 1985a, Driscoll, 2006) and empirical evidence that reflection and goal-setting are interrelated processes within SRL (Raković et al., 2022). In the present study, we investigated the relationship between students' reflections on prior learning and goals students set for future learning. We did so by examining students' reflective and goal-setting behaviours recorded in the reflective writing task. Our results align with prior studies showing that students who appeared to engage in lower-level reflective processes (e.g., *Returning to Experience* and *Attending to Feelings*), as defined by Boud et al. (1985a), formulated their reflective responses in a rather descriptive way (Sen, 2010) with limited evaluative/critical reflection about their prior studying (Abdul Rabu & Badlishah, 2020). On the other hand, students who appeared to more deeply reflect on their prior studying included more detailed evaluations in their reflective responses, e.g., how their new insights differ from their prior believes and how they connect to personal development. We also found that many students in the dataset observed clearly described their planned actions for future studying, including areas of improvement (e.g., communication skills) and additional details to the actions (e.g., anticipated challenges and potential solutions). We speculate this finding may be due to the detailed instructions on goal-settings that learning coaches provided to pharmacy students. For instance, these instructions closely followed the SMART framework for goal settings, e.g., guide students to set specific goals that are relevant to them to meet course learning objectives. This further confirms prior findings that using theory-aligned frameworks to guide students' goal setting may improve the specificity of the goals students set (Shell, 2020, Acee et al., 2012, Alessandri et al., 2020, Kleingeld et al., 2011). Our findings also conform to prior research showing that students' reflective behaviours may correlate with their goal-setting behaviours (Raković et al., 2022). To this end, educators may opt in to administer reflective and goal-setting activities concurrently to more effectively facilitate deeper evaluation of prior studying from students which, in turn, leads to the development of more specific, measurable and attainable goals. Possible instructional designs include but are not limited to reflective writing tasks, think-aloud sessions (Whitehead et al., 2016), mindfulness-based practices (Nugent et al., 2011) and goal-setting worksheets (Nordengren, 2019). The integration of such instructional designs has the potential to improve students' metacognitive and goal-setting skills, which may further benefit the development of students' self-regulated learning in the long run.

We also demonstrated that the BERT-based deep learning models may outperform traditional machine learning models in identifying reflective sentences and classifying those sentences relative to the assessment rubric. Apart from these, the BERT-based automatic evaluation can facilitate classification of students' reflective responses. Our classification models can differentiate among students' reflective behaviours relative to breadth and depth dimensions (Tsingos et al., 2015), thus providing a potential computational means for educators to not only identify different reflective elements within an individual reflection, but also to perform at-scale overall assessment of students' reflective levels. Moreover, high classification performance documented for the BERT-based predictive models in the present study aligns with prior research (Wulff et al., 2022) that showed BERT-based models were able to accomplish considerable performance in classifying student reflective writing with limited data. Accordingly, BERT-based deep learning modelling may be considered a viable and more promising approach to automatically evaluate reflective writing in educational settings given that the acquisition of larger scales of reflective essays has traditionally been considered challenging. In addition, as our BERT-based models could be reliably employed for such automatic evaluation it is probable to implement a tool with an interactive interface (e.g., a web application) that colorcodes various reflective elements in students' written reflections relative to the coding schema articulated in this study and provides formative feedback to students on their reflective writings. In this way, the tool may indicate to students the existing elements in their reflective responses and identify specific reflective elements that are still missing in the responses. Such a tool can not only facilitate efficient evaluation so as to enable timely feedback from educators, but also benefit students to conduct self-assessment, obtaining an idea of their limitations while identifying and contrasting reflective elements to potentially promote their reflective ability (Ryan, 2011).

6. Conclusion

In conclusion, we comprehensively assessed students' reflective writings, which contained their retrospective reflection and prospective goal-setting, to evaluate their reflective and goal-setting behaviours and discovered that the majority of students managed to jointly include descriptive reflection, and specify planned actions, targeted areas of improvement and additional details to the planned actions in their set goals. However, the higher-level evaluative reflection and more concrete goal-setting seemed harder to achieve by students. To the best of our knowledge, our study is the first to look into students' behavioural patterns of reflecting and setting goals jointly. More importantly, we revealed that students who reflected deeply tended to set more specific goals. Though researchers have suggested the potential benefits of incorporating reflection to inform students' goal-setting (Raković et al., 2022), to our knowledge, our study is the first to provide empirical evidence on such a correlation. Our findings contribute prospectively new knowledge to learning sciences. Additionally, we identified BERT-based approach to be the most promising one in the automatic identification of reflective sentences and their corresponding reflective stages within students' written reflection in pharmacy curricula. Such automation can clearly and holistically reveal students' levels of reflection with an assessment rubric containing both breadth and depth dimensions. Thereupon, we posit that 1) educators, course designers and educational institutions should adopt jointly reflective and goal-setting activities in their instructional design to more effectively boost students' metacognitive skills and self-regulation; 2) educational researchers aiming to automate the effective evaluation of reflective writings and/or implementing educational technologies to assist the evaluation of such written responses should harness the power of state-of-the-art pre-trained language models.

7. Limitations and future work

In the current study, the evaluation frameworks for retrospective reflection and prospective goal-setting were developed based on the theory by Boud et al. (1985a), Mezirow (1991) and Rubin (2002). However, there are many other reflection models (e.g., (Moon, 1999, Dewey, 1933)) and goal-setting guidelines (e.g., (McCardle et al., 2017)). Even for the same guideline, there could be more than one interpretation (e.g., SMART (Rubin, 2002)). Future studies may be conducted to understand whether students show similar reflective and goal-setting behaviours assessed by other frameworks. In our current study, we identified BERT-based deep learning approach as the most effective one. However, the model for identifying reflective sentences didn't perform as satisfactorily as the ones differentiating reflective elements. Future work may experiment more advanced pre-trained language models (e.g., GPT-3 (Brown et al., 2020), RoBERTa (Liu Y. et al., 2019)) to potentially achieve better performance. Additionally, as deep learning models lacked interpretability, further studies could be conducted utilising Explainable Artificial Intelligence approaches to incorporate interpretability to the automatic evaluation of students' reflective writings. Due to the subjective nature of assessing students' set goals (e.g., whether goals are attainable depends on instructors' subjective evaluation of individual student's capabilities), we excluded the automatic evaluation of students' goal-settings from the scope of the current study. However, future studies may investigate the possibility of incorporating students' characteristics (e.g. past performance) to enable such automation.

8. Statements on open data and ethics

Ethics approval has been obtained from Anonymous University for the collection of students' reflective writing. The authors declare that this paper is in alignment with the publishing ethics guidelines.[2] The data are not publicly available as we haven't obtained approval to share the data and the participants didn't provide written consent to allow the public sharing of data.

SRL Self-Regulated Learning
BERT Bidirectional Encoder Representations of Transformers
NLP Natural Language Processing
NB Naive Bayes
SVM Support Vector Machine
RF Random Forest
LR Logistic Regression
ROC-AUC Area Under ROC Curve

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was in part supported by funding from Deutsche Forschungsgemeinschaft (BA 2044/10-1), Nederlandse Organisatie voor

[2] https://www.elsevier.com/about/policies/publishing-ethics.

Wetenschappelijk Onderzoek (GA 2739/1-1), Economic and Social Research Council of the United Kingdom (MO 2698/1-1) through Open Research Area (Call 5), the Australian Research Council (DP220101209), and Jacobs Foundation (CELLA 2 CERES).

References

Abdul Rabu, S. N., & Badlishah, N. S. (2020). Levels of students' reflective thinking skills in a collaborative learning environment using Google docs. *TechTrends*, *64*, 533–541.

Acee, T. W., Cho, Y., Kim, J. I., & Weinstein, C. E. (2012). Relationships among properties of college students' self-set academic goals and academic achievement. *Educational Psychology*, *32*, 681–698.

Alessandri, G., Borgogni, L., Latham, G. P., Cepale, G., Theodorou, A., & De Longis, E. (2020). Self-set goals improve academic performance through nonlinear effects on daily study performance. *Learning and Individual Differences*, *77*, Article 101784.

Artstein, R., & Poesio, M. (2008). Inter-coder agreement for computational linguistics. *Computational Linguistics*, *34*, 555–596.

Boud, D., Keogh, R., & Walker, D. (1985a). Promoting reflection in learning: A model. *Boundaries of Adult Learning*, *1*, 32–56.

Boud, D., Keogh, R., & Walker, D. (1985b). Reflection: Turning experience into learning. Routledge.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Bryan, J. F., & Locke, E. A. (1967). Goal setting as a means of increasing motivation. *Journal of Applied Psychology*, *51*, 274.

Charon, R., & Hermann, M. N. (2012). A sense of story, or why teach reflective writing? Academic medicine. *Journal of the Association of American Medical Colleges*, *87*, 5.

Chen, E., Lin, Y., Xiong, H., Luo, Q., & Ma, H. (2011). Exploiting probabilistic topic models to improve text categorization under class imbalance. *Information Processing & Management*, *47*, 202–214.

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). New York, NY, USA: ACM.

Cheng, G. (2017). Towards an automatic classification system for supporting the development of critical reflective skills in l2 learning. *Australasian Journal of Educational Technology*, *33*.

Cleary, T. J., & Chen, P. P. (2009). Self-regulation, motivation, and math achievement in middle school: Variations across grade level and math context. *Journal of School Psychology*, *47*, 291–314.

Dennison, W. F., & Kirk, R. (1990). Do, review, learn, apply: A simple guide to experiential learning. Blackwell Education.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Association for computational linguistics*. *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 4171–4186).

Dewey, J. (1933). How we think: A restatement of the relation of reflective thinking to the educative process. *The American Journal of Psychology*, *46*, 528.

Doran, G. T., et al. (1981). There'sa smart way to write management's goals and objectives. *Management Review*, *70*, 35–36.

Driscoll, J. (2006). *Practising clinical supervision: A reflective approach for healthcare professionals*. Elsevier Health Sciences.

Gibbs, G. (1988). Learning by doing: A guide to teaching and learning methods. *Further Education Unit*.

Killion, J. P., & Todnem, G. R. (1991). A process for personal theory building. *Educational Leadership*, *48*, 14–16.

King, T. (2002). Development of student skills in reflective writing. In *Proceedings of the 4th world conference of the international consortium for educational development*. *Spheres of influence: Ventures and visions in educational development*.

Kleingeld, A., van Mierlo, H., & Arends, L. (2011). The effect of goal setting on group performance: A meta-analysis. *Journal of Applied Psychology*, *96*, 1289.

Klug, J., Ogrin, S., Keller, S., Ihringer, A., & Schmitz, B. (2011). A plea for self-regulated learning as a process: Modelling, measuring and intervening.

Koc, N., & Celik, B. (2015). The impact of number of students per teacher on student achievement. *Procedia – Social and Behavioral Sciences*, *177*, 65–70.

Koch, A. K., & Nafziger, J. (2011). Self-regulation through goal setting. *Scandinavian Journal of Economics*, *113*, 212–227.

Kolb, D. (1984). Experiential learning: Experience as the source of learning and development*Vol. 1*.

Kovanović, V., Joksimović, S., Mirriahi, N., Blaine, E., Gašević, D., Siemens, G., & Dawson, S. (2018). Understand students' self-reflections through learning analytics. In *LAK18* (pp. 389–398).

Latham, G. P., & Locke, E. A. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes*, *50*, 212–247.

Li, Y., Rakovic, M., Poh, B. X., Gasevic, D., & Chen, G. (2022). Automatic classification of learning objectives based on bloom's taxonomy. In *EDM, international educational data mining society* (pp. 530–537).

Liu, M., Shum, S. B., Mantzourani, E., & Lucas, C. (2019). Evaluating machine learning approaches to classify pharmacy students' reflective statements. In *Proceedings, part I: Vol. 20. Artificial intelligence in education: 20th international conference* (pp. 220–230). Springer.

Liu, Q., Zhang, S., Wang, Q., & Chen, W. (2017). Mining online discussion data for understanding teachers reflective thinking. *IEEE Transactions on Learning Technologies*, *11*, 243–254.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin*, *90*, 125.

Lucas, C., Gibson, A., & Buckingham Shum, S. (2018). Utilization of a novel online reflective learning tool for immediate formative feedback to assist pharmacy students' reflective writing skills. *American Journal of Pharmaceutical Education*.

Malone, D., Galbraith, K., White, P. J., Exintaris, B., Nicolazzo, J. A., Brock, T., Bruno-Tomé, A., Short, J. L., & Larson, I. (2021). Development of a vertically integrated pharmacy degree. *Pharmacy*, *9*. https://doi.org/10.3390/pharmacy9040156. https://www.mdpi.com/2226-4787/9/4/156.

Mann, H. B., & Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 50–60.

Mann, K., Gordon, J., & MacLeod, A. (2009). Reflection and reflective practice in health professions education: A systematic review. *Advances in Health Sciences Education*, *14*, 595–621.

Marzouk, Z., Patzak, A., Rahimi, D., Raković, M., Pakdaman-Savoji, A., Vytasek, J., & Winne, P.H. Guiding goal setting through learning analytics feedback loops: A conceptual framework. Manuscript in preparation.

McCardle, L., Webster, E. A., Haffey, A., & Hadwin, A. F. (2017). Examining students' self-set goals for self-regulated learning: Goal properties and patterns. *Studies in Higher Education*, *42*, 2153–2169.

Mento, A. J., Steel, R. P., & Karren, R. J. (1987). A meta-analytic study of the effects of goal setting on task performance: 1966–1984. *Organizational Behavior and Human Decision Processes*, *39*, 52–83.

Mezirow, J. (1991). *Transformative dimensions of adult learning*. London, England: ERIC.

Moon, J. (2007). Getting the measure of reflection: Considering matters of definition and depth. *Journal of Radiotherapy in Practice*, *6*, 191–200. https://doi.org/10.1017/S1460396907006188.

Moon, J. A. (1999). *A handbook of reflective and experiential learning: Theory and practice*. Routledge.

Morisano, D., Hirsh, J. B., Peterson, J. B., Pihl, R. O., & Shore, B. M. (2010). Setting, elaborating, and reflecting on personal goals improves academic performance. *Journal of Applied Psychology*, *95*, 255.

Nordengren, C. (2019). Goal-setting practices that support a learning culture. *Phi Delta Kappan*, *101*, 18–23.

Nugent, P., Moss, D., Barnes, R., & Wilks, J. (2011). Clear (ing) space: Mindfulness-based reflective practice. *Reflective Practice*, *12*, 1–13.

Oistad, M. (2020). Using smart goals in the 8th grade health classroom to increase student success in school.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. Technical Report.

Raković, M., Bernacki, M. L., Greene, J. A., Plumley, R. D., Hogan, K. A., Gates, K. M., & Panter, A. T. (2022). Examining the critical role of evaluation and adaptation in self-regulated learning. *Contemporary Educational Psychology*, *68*, Article 102027.

Rubin, R. S. (2002). Will the real smart goals please stand up. *The Industrial-Organizational Psychologist*, *39*, 26–27.

Ryan, M. (2011). Improving reflective writing in higher education: A social semiotic perspective. *Teaching in Higher Education*, *16*, 99–111.

Schunk, D. H. (2001). Self-regulation through goal setting.

Schut, H., & Stam, H. (1994). Goals in rehabilitation teamwork. *Disability and Rehabilitation*, *16*, 223–226.

Sen, B. A. (2010). Reflective writing: A management skill. *Library Management*.

Senter, R., & Smith, E. A. (1967). Automated readability index. Technical Report, Cincinnati Univ. OH.

Sha, L., Rakovic, M., Li, Y., Whitelock-Wainwright, A., Carroll, D., Gaševic, D., & Chen, G. (2021). Which hammer should I use? A systematic evaluation of approaches for classifying educational forum posts. EDM.

Shell, K. (2020). The effect of goal setting, goal tracking, and feedback on reading fluency.

Tsingos, C., Bosnic-Anticevich, S., Lonie, J. M., & Smith, L. (2015). A model for assessing reflective practices in pharmacy education. *American Journal of Pharmaceutical Education*, *79*, Article 124, 26690718[pmid].

Ullmann, T. D. (2019). Automated analysis of reflection in writing: Validating machine learning approaches. *International Journal of Artificial Intelligence in Education*, *29*, 217–257.

Whitehead, A. E., Cropley, B., Huntley, T., Miles, A., Quayle, L., & Knowles, Z. (2016). 'Think aloud': Toward a framework to facilitate reflective practice amongst rugby league coaches. *International Sport Coaching Journal*, *3*, 269–286.

Winne, P. H. (2018). Theorizing and researching levels of processing in self-regulated learning. *British Journal of Educational Psychology*, *88*, 9–20.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated engagement in learning. In metacognition in educational theory and practice. *Metacognition in Educational Theory and Practice*, 277–304.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., & Brew, J. (2019). Huggingface's transformers: State-of-the-art natural language processing. CoRR, arXiv:1910.03771.

Wollny, S., Schneider, J., Rittberger, M., & Drachsler, H. A. (2019). I planning smart?–analyzing student goals. In *LAK19* (pp. 261–266).

Wulff, P., Mientus, L., Nowak, A., & Borowski, A. (2022). Utilizing a pretrained language model (bert) to classify preservice physics teachers' written reflections. *International Journal of Artificial Intelligence in Education*, 1–28.

Zimmerman, B. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice*, *41*, 64–70.

Zimmerman, B. J. (2000). Chapter 2 - attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). San Diego: Academic Press.