Evaluating Machine Learning Approaches to Classify Pharmacy Students' Reflective Statements

Ming Liu[1] [0000-0003-4256-6531], Simon Buckingham Shum[1],
Efi Mantzourani[2] , Cherie Lucas[1]

[1] University of Technology Sydney, Sydney, 2008, Australia
`Ming.Liu@uts.edu.au`
[2] Cardiff University, Cardiff, CF10 3AT, United Kingdom

Abstract. Reflective writing is widely acknowledged to be one of the most effective learning activities for promoting students' self-reflection and critical thinking. However, manually assessing and giving feedback on reflective writing is time consuming, and known to be challenging for educators. There is little work investigating the potential of automated analysis of reflective writing, and even less on machine learning approaches which offer potential advantages over rule-based approaches. This study reports progress in developing a machine learning approach for the binary classification of pharmacy students' reflective statements about their work placements. Four common statistical classifiers were trained on a corpus of 301 statements, using emotional, cognitive and linguistic features from the Linguistic Inquiry and Word Count (LIWC) analysis, in combination with affective and rhetorical features from the Academic Writing Analytics (AWA) platform. The results showed that the Random-forest algorithm performed well (F-score=0.799) and that AWA features, such as emotional and reflective rhetorical moves, improved performance.

Keywords: Reflective writing, Automated feedback, Learning analytics.

1 Introduction

 "We do not learn from experience... we learn from reflecting on experience" is a well-known adage that summarizes Dewey's [1] foundational work on teaching and learning. Critical self-reflection has been increasingly recognized as central to the development of agentic, self-regulated learners. Moreover, finding ways to scaffold quality reflection becomes all the more important as we seek to provide learners with more authentic tasks and assessments, that are distinctive because of the rich, complex, social, psychological and embodied experiences they provide [2]. When students engage effectively in reflective learning activities, this can provide evidence of self-critical insight, identify challenging issues, connect academic with experiential knowledge, acknowledge emotions and feelings, and reflect on how they can apply such insights in the future [3, 4]. Reflective processes in learning have most impact when they are formative and future-oriented [5]. In addition, reflection is important for meta-cognitive adaptation, when students connect their thinking to the wider world [6].

Reflection can be a purely internal form of contemplation, but making it explicit can help clarify one's own thinking, benefit fellow learners, and help a student narrate

(e.g. in a job interview) succinct evidence of their personal and academic development. Moreover, if it is to serve as a form of evidence in a formal learning context, students must learn to express their insights. However, whatever medium is used (e.g. speech; video; writing), the student must be literate enough that they can effectively use that medium to do justice to themselves. Reflective writing is one of the most common approaches used, but teaching, learning, and grading reflective writing presents challenges, since it is often an unfamiliar genre for educators and students.

The evaluation of student reflective writing is traditionally accomplished by researchers using manual content-analysis methods to assess student reflective writings [7], and by educators by grading against a rubric. It is of course extremely labor-intensive to grade or otherwise code writing, and it is here that automated approaches have potential roles to play. Natural language processing could assist if student texts can be analyzed, automatically coded (classified) according to a scheme, and in a learning context, helpful feedback given. However, there is very little work in this field to date, with research, and products, in Automated Writing Evaluation (AWE) dominated by more common genres of writing such as persuasive essays, literature review or research proposals [8–11]. The work in reflective writing to date uses either a rule-based [12] or machine learning approach to classify reflective sentences [13], reflective passages[14], forum posts [15], with only one example of automated feedback deployed with students [3].

This present study contributes new empirical results, investigating the use of automated text analytics methods for evaluating the reflective statements written by pharmacy students on experiential work placements. Section 2 reviews the literature of reflective writing analytics, before Section 3 describes the evaluation method and Section 4 presents the results. Section 5 discusses the results and identifies directions for future work.

2 Reflective Writing Analytics

With the advancement of text analytics, researchers are able to develop novel reflective writing analytics by using rule-base or machine learning approaches with rich textual features extracted from computer tools (e.g. LIWC [16] and Coh-Metrix [17]). Ullmann conducted some of the earliest work on reflective writing sentence classification (i.e. differentiating reflective and non-reflective sentences) by using both rule-based and machine learning approach [13, 18, 19]. Chen et al. [20] adapted topic modelling to analyze pre-service teachers' reflection journals, but topic models focus on the content rather than quality and depth of reflection. Extending the work of Buckingham Shum et al. [4], Gibson et al. [3] proposed a concept-matching rhetorical analysis framework [21] to automatically detect sentences performing three key reflective rhetorical functions summarized as Context, Challenge and Change. Kovanovic et al. [15] developed a random forest classifier using features extracted from the LIWC and Coh-Metrix for

arts students' reflective statements (observation, motive and goal). In only one case did the system generate actionable feedback to students [3].

Similar to previous studies [3, 13, 15] , this study focuses on binary classification of reflective statements. The contribution of this study includes the exploration of machine learning approach for classifying reflective statements written by pharmacy students, extensive evaluation of different machine learning algorithms with features extracted from theoretically-sound reflective rhetorical moves features and LIWCs, and the analysis of how these features affect the classification results.

3 Method

This section describes the dataset (see section 3.1) used for training and evaluating four different classifiers described in section 3.4 and the 101 features extracted by the LIWC and Academic Writing Analytic tool (see section 3.2). In addition, the imbalance class distribution problem is addressed in Section 3.3.

3.1 Dataset Description

The dataset comes from 43 pharmacy students enrolled in year 2 of the Master of Pharmacy degree at Cardiff University, United Kingdom. As part of a professional development module that needs to be passed to ensure progression, all students were required to complete two different experiential placements. The first placement comprised of a week in a community pharmacy, and the second involved a visit in a non-traditional setting such as an optician or a care home [22, 23]. All students completed a pre-placement workshop whereby emphasis was given on reflective skills and reflective writing, and a post-placement workshop where students exchanged experiences from their placements with the support of a pharmacy academic facilitator. Following this, students were asked to complete a reflective account based on prompts to facilitate reflection provided in template; all students were required to submit this reflective account in order to successfully pass the module. Examples of prompt template questions to facilitate student reflections included: "Thinking about your professional development, what went well during placements? What was the highlight? What have you learned? How was this different to what you thought/expected? Please tell us about something that happened in your placements that made you reflect on your role as a pharmacist in patient care and/or the role of other health and social care professionals?" The template had previously been developed by the authors after multiple cycles of action research involving placement supervisor and student input [24].

All student reflective accounts were assessed against a reflective rubric [25, 26], developed by integrating Mezirow's [27] and Gibbs' models of reflection, and related to different stages of reflection [24]: *1. attending to feelings; 2. relating new knowledge with previous knowledge; 3. integrating prior with new knowledge; 4. feelings or attitudes; 5. self-assessing beliefs, approaches and assumptions; 6. internalizing the knowledge or experience; 7. personally applied.*

Replicating a validated approach, four human experts assessed the same set of reflective accounts [25, 26]. Reflective accounts were assigned a score for each of these

stages of reflection: a score of 0 was assigned where the student had not demonstrated any reflective skills in the writing (*non-reflective*), a score of 0.5 when an attempt was made to relate experiences or feelings with prior knowledge and identify learning (*reflective*), and a score of 1 when clear links were made between experiences, feelings, learning and a change of behaviour was demonstrated (*critically reflective*).

Human experts reached moderate to substantial agreement (Intra-class correlation coefficient= 0.55-0.69, p<0.001) on rating these reflective elements. In our study, the human ratings are transformed into categorical values, reflective (average rating >=0.5) and non-reflective statement (average rating < 0.5). Collapsing the reflective and critically reflective categories to create this binary classification was a measure introduced in light of the relatively small dataset available for this preliminary experiment (301 statements), recognizing that future work will need to differentiate, to enable more effective feedback, which is our ultimate objective. Table 1 shows the description of the dataset used for training and evaluating statistical models. Some stages of reflection, such as stage 5 (Validation) and stage 6 (Appropriation), are harder than others because they require students to delve deeper and reflect on why their belief system is what it is (for example as a result of their upbringing and cultural background).

Table 1. *Reflective* and *Non-Reflective* categories in training and test datasets.

Rating type	N	Reflective	Non-reflective
All stages	301	243	58
Stage 1	43	34	9
Stage 2	43	42	1
Stage 3	43	42	1
Stage 4	43	42	1
Stage 5	43	32	11
Stage 6	43	10	33
Stage 7	43	41	2

3.2 Feature Extraction and Selection

In order to develop a statistical classifier for student reflections, we extracted several different types of features. The extracted features were inspired by existing work in reflective writing [3, 15]. *Linguistic Inquiry and Word Count (LIWC)* is a linguistic analysis product [16] which extracts approximately 90 linguistic measures indicative of a large set of psychological processes (e.g., affect, cognition, biological process, drives), personal concerns (e.g., work, home, leisure activities) and linguistic categories (e.g., nouns, verbs, adjectives). The psychological process categories consist of *social* (e.g. family, friend and humans), *affective* (e.g. positive and negative emotions, anxiety, anger and sadness), *cognitive* (e.g. insight, causation, discrepancy, tentative, certainty and inclusive), *perceptual* (e.g. heard, see, hear and feel), *biological* (e.g. body, blood and health) and *relativity* (e.g. before, space, motion and time) subcategories. Previous

work [16] indicated that LIWC measures, including perceptual words (punctuation, causal words, past-oriented words, passive voice, and connectives) were among the most important classification features. Thus, a total number of 94 features were extracted by using LIWC.

Academic Writing Analytics (AWA) is an open source software platform (heta.io) focusing on providing actionable feedback to support academic writing, such as analytical writing [11] and reflective writing [3]. Gibson et al. [3] used the concept-matching rhetorical analysis framework [21] to automatically detect sentences indicating three key reflective rhetorical moves: *Context* (initial thoughts and feelings about a significant experience), Challenge (the challenge of new surprising or unfamiliar ideas, problems or learning experience) and *Change* (potential solution and learning opportunities). Overlaid on these moves there may be further classification to indicate deeper reflection which references oneself, and three expression types, *Emotive* (expresses emotions and feelings), *Epistemic* (expresses beliefs, learning or knowledge) and *Critique* (expresses self-critique). As detailed in [3], these features were based on a model distilling a wide range of scholarship into reflective writing. *Emotive* expressions are detected based on lexical comparisons with a corpus based on a model of arousal, valence and dominance [28], while the *Critique* and *Epistemic* expressions are derived using techniques for identifying metacognition in reflective writing [6]. In addition, a metrics feature is used to indicate if sentences appear to be excessively long (more than 25 words). Therefore, a total number of eight reflective writing features is used based on Gibson et al.'s work [3].

To avoid over-fitting, feature selection is an important data pre-processing stage in machine learning since the dataset is relatively small (Table 3) and the feature size is moderate (N=102 features). In this study, correlation-based feature selection [29], one of the most popular feature selection methods, is used to rank the features so that top ranked features are highly correlated to the class label.

3.3 Addressing the Problem of Class Imbalance

As shown in Table 1, the dataset has the problem of class imbalance, containing more reflective statements (N=243) than non-reflective statements (N=58). Moreover, from the application perspective, high recall in the non-reflective category is very important in order to generate feedback for non-reflectors. Therefore, we used MetaCost [30], a popular method for addressing the class imbalance problem, which makes an arbitrary classifier cost-sensitive by wrapping a cost-minimizing procedure around it. In other words, the misclassification cost on non-reflective instances is higher than the cost on reflective instances. In this case, a misclassification cost ratio of 10 to 1 was chosen, based on Weiss et al's work [31], presented in Table 2. Research has shown this approach can improve the classification accuracy of the high cost category [32].

Table 2. Cost matrix (after Weiss et al. [31])

		Predict	
		non-reflection	reflection
Actual	non-reflection	0	10
	reflection	1	0

## 3.4   Model Selection and Evaluation

Following previous research in reflective text classification [15, 19], four common statistical classifiers were evaluated: Random Forests (high performance model), Support Vector Machine (high performance model), PART (rule-based model) and Naïve Bayes (high performance model). Random forests, one of the most popular ensemble classification techniques, combine a large number of decision-trees and bootstrap sampling to provide a low-bias, low-variance classification method [33]. PART generates accurate rule sets from partial decision trees [34]. Support Vector Machine aims at finding the optimal hyperplane to separate classes. By adjusting their kernel function, SVMs can be extended to classify patterns that are not linear [35]. Naïve Bayes is an algorithm based on Bayes' rule [36]. It is "naïve" because it assumes that all features are independent from each other. This has the benefit of rapid processing while producing good performance in many cases.

In this study, the implementation of these classifiers was performed in the Weka tool with the default settings [37]. Ten-fold cross validation method was used to evaluate the performance of each classifier.

4   Results and Discussion

4.1   Classification Results and Discussion

Figure 1 illustrates that the random forest classifier outperforms others, and the feature sets influence the performance of these classifiers, which have similarly sharp curves. The feature selection/ranking step (described in section 3.2) is performed before the 10-fold cross-validation. Initially, the performance of each classifier increases when the number of ranked features used in the classification model rises over the top 22 features, before plateauing. Then, the performance of PART and SVM becomes unstable as the number of features increases in their models. These results are consistent with Ullmann's findings that showed random forest outperforming other classifiers, such as PART and Naïve Bayes in reflective statement classification, because the random forest constructs multiple decision trees and aggregates them together to get a more accurate and stable prediction [13].
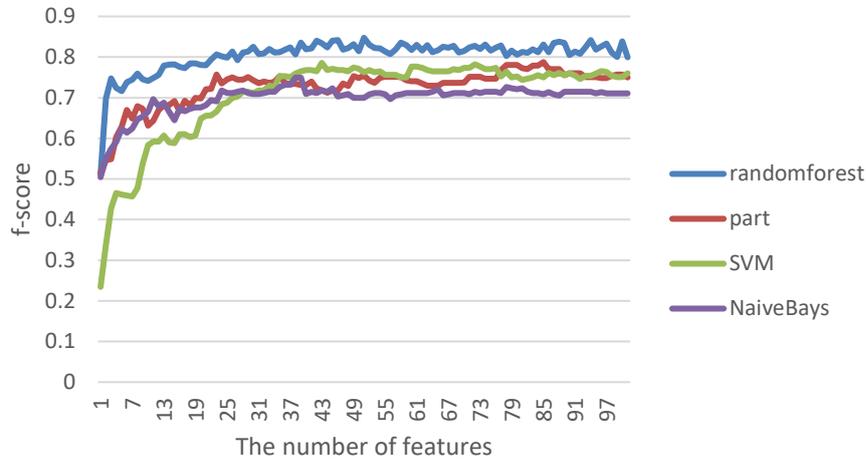
Fig. 1. The effects of ranked feature sets on the performance of
different reflective writing classifiers in the combined element dataset

Table 3 shows that the random forest classifier reached higher scores in all stages combined and stage 1 than others. Since other datasets (Stage 2, 3 and 4) had a serious class-imbalanced distribution problem and only contained a single instance of non-reflective writing in each stage, it was difficult to train the model to correctly classify non-reflective instances. The model failed to detect these instances causing the division by zero problem undefined issues (N/A) when calculating the precision. However, these results indicate that the classifier performed well in the combined 7 stages dataset, F-score=.799.

Table 3. Random forest classification results among different datasets.

Dataset	N	Precision	Recall	F-score
All stages	301	.831	.784	.799
Stage 1	43	.874	.860	.865
Stage 2	43	N/A	.977	N/A
Stage 3	43	N/A	.953	N/A
Stage 4	43	N/A	.977	N/A
Stage 5	43	.593	.395	.413
Stage 6	43	N/A	.767	N/A
Stage 7	43	N/A	.953	N/A

4.2 Feature Importance Analysis and Discussion

Next, we examine the correlations between features and human-graded reflective writing scores. Table 4 shows the Spearman correlation between the top 10 features and reflective writing scores.

Table 4. Top ten feature analysis ranked by their correlation (note both positive and negative) to human-graded reflective writing.

Top Ranked Features	Description	r
		* p < .05 **p <.001
1. AWA.Context	A personal response to a learning context	.193**
2. AWA.Emotive	Indicating affective elements and emotive expressions	.183**
3. LIWC.Quant	Quantifiers (e.g. few, many, much)	-.148*
4. LIWC.Adj	Common adjectives: free, happy, long	-.123*
5. LIWC.Compare	Comparative words. e.g. Greater, best, after	-.117*
6. AWA.Self-Critique	Learners criticize their pre-existing knowledge based on new information	-.095
7. LIWC.Differ	Differentiation (e.g. different, hasn't, but)	-.085
8. LIWC.WC	Word count	.066
9. LIWC.Authentic	Truthful words	.037
10. LIWC.Space	Space words (e.g. down, in)	.001

It can be seen that the top ten features include LIWC linguistic features (*LIWC.Quant/Compare/Adj*), emotional features, cognitive features (*AWA.Self-Critique and LIWC.Differ*), reflective rhetorical moves (*AWA.Context/Self-Critique*), authentic (*LIWC.Authentic*) and space features (*LIWC.Space*). These results confirm earlier reports that LIWC provides good classification indicators of reflective writing [15, 20]. Some of these features are significantly correlated to the level of reflection (the average rating scores of our human experts, as described in section 3.1). The expression of emotion and the presence of rhetorical moves (context) are positively significantly correlated to the level of reflection, while the number of quantifiers, comparative and adjective used are negatively significantly correlated to the level of reflection.

These correlation analysis results fit the reflective cycle model of Gibbs [38], who points to the importance of cognition-oriented elements (e.g. *AWA.Self-Critique*) of evaluation, analysis, conclusion and future plan in addition to the description of experience (*AWA.Context*) and emotions (*AWA.Emotive*). However, some study results have not reported a strong correlation between the emotional features provided by LIWC and the reflective writing grade [20].

There is no significant correlation between *LIWC.WordCount* and the writing grade, which indicates that the quality of reflection is more important than the length of writing in reflective writing. In addition, we find that non-reflectors tended to use many relative

(*LIWC.Space*), adjectival and differentiation words. We have not yet fully understood this, although a possible account for the latter is that non-reflective students made heavy use of rather general words (e.g. "*Different* pharmacists had *different* attitudes and *different* processes in place to achieve this." or The *whole* experience went well I really liked working there and definitely learnt a *lot* of new things.), rather than provide more specific details (e.g. about how exactly pharmacists differed from each other, and how these differences connected to their personal learning experiences).

5   Conclusion and Future Work

Manually analyzing reflective writing is time consuming, but a small number of researchers have proposed computational approaches to automatically detect the distinctive features of this genre [13, 15, 20]. This study presents a machine learning approach that is distinctive in two respects: (i) a training model based on a validated, graded corpus of reflective writing from authentic work-placements (specifically, pharmacy students), and (ii) using features extracted from the combination of a generic linguistic analysis (LIWC) and a rule-based parser developed specifically for academic reflective writing (AWA). The study results are promising (F-score=.799 using the Random-Forest classifier), and we are interested to note that of the ten most powerful features, while the top features and two others are AWA's rhetorical features, the remainder are LIWC features which we have discussed.

However, this initial study has some limitations. Firstly, using different parameters of the classifiers may influence the evaluation results. Kovanovic et al. [15] showed that by tuning the random forest parameters, such as the number of decision trees, the classifier performed better. Secondly, the dataset included only pharmacy students' reflective statements about their placement experiences, so more training and test sets are needed to evaluate the generalizability of the classifier. Despite these limitations, we have argued that our evaluation is sound and the dataset is authentic, annotated by four human experts.

Future work will focus on more nuanced automatic detection of the *depth* of reflective states, based on the two dimensional depth/breadth reflection model proposed by Gibson et al. [3], in which the depth of reflection includes non-reflection, reflection and deep reflection, and the breadth of reflection contains initial thoughts, feelings (context), challenges, self-critique, potential solution and learning opportunities (changes). Following Milligan and Griffin's learner development progression model [39], the two dimensional reflection model will be used to locate the writing more precisely on a scale that reflects their progression towards critically reflective writing, in order to generate feedback based not only on the current reflective state, but on how they have *improved*, which is known to be a powerful motivator in feedback design. We also plan to draw on work that generates questions for reflection based on the depth of reflective state. For example, following the different questions identified in Gibson, et al.'s synthesis of the literature, if the system classifies the writing as non-reflective, containing only statements of knowledge and belief (and no affect), formative feedback could generate a question such as, *"Did this incident evoke any strong feelings?"* to move the

student forward in the depth/breadth reflection space. Or, in the absence of any reflection about the future, the student might be asked, *"Do you think you would handle this differently next time this arises?"* Thus, regardless of whether the writing progresses (or regresses) within the depth/breadth reflection space, questions could be generated to provoke deeper reflection, and help move the student the next step forward.

References

1. Dewey, J.: How we think. Prometheus Books (1933).
2. Herrington, T., Herrington, J.: Authentic learning environments in higher education. Information Science Publishing (2005).
3. Gibson, A., Aitken, A., Sándor, Á., Buckingham Shum, S., Tsingos-Lucas, C., Knight, S.: Reflective writing analytics for actionable feedback. Proc. Seventh Int. Learn. Anal. Knowl. Conf. - LAK '17. 153–162 (2017).
4. Buckingham Shum, S., Sándor, Á., Goldsmith, R., Bass, R., McWilliams, M.: Towards Reflective Writing Analytics: Rationale, Methodology and Preliminary Results. J. Learn. Anal. 4, 58–84 (2017).
5. Boud, D., Falchikov, N.: Aligning assessment with long-term learning. Assess. Eval. High. Educ. 31, 399–413 (2006).
6. Gibson, A., Kitto, K., Bruza, P.: Towards the Discovery of Learner Metacognition from Reflective Writing. J. Learn. Anal. 3, 22–36 (2016).
7. Krippendorff, K.: Content Analysis: An Introduction to Its Methodology. Sage Publications (2004).
8. Liu, M., Li, Y., Xu, W., Liu, L.: Automated essay feedback generation and its impact on revision. IEEE Trans. Learn. Technol. 10, (2017).
9. Wilson, J., Czik, A.: Automated essay evaluation software in English Language Arts classrooms: Effects on teacher feedback, student motivation, and writing quality. Comput. Educ. 100, 94–109 (2016).
10. McNamara, D.S., Raine, R.B., Roscoe, R., Crossley, S. a, Jackson, G.T., Dai, J., Cai, Z., Renner, A., Brandon, R., Weston, J., Dempsey, K., Carney, D., Sullivan, S., Kim, L., Rus, V., Floyd, R., McCarthy, P., Graesser, A.C.: The Writing-Pal: Natural Language Algorithms to Support Intelligent Tutoring on Writing Strategies. Appl. Nat. Lang. Process. Identification, Investig. Resolut. 2, 298–311 (2012).
11. Knight, S., Buckingham Shum, S., Ryan, P., Sándor, Á., Wang, X.: Designing Academic Writing Analytics for Civil Law Student Self-Assessment. Int. J. Artif. Intell. Educ. 28, 1–28 (2018).
12. Ullmann, T.D., Wild, F., Scott, P.: Comparing automatically detected reflective texts with human judgements. In: the Proceeding of 2nd Workshop on Awareness and Reflection in Technology-Enhanced Learning. pp. 101–116. , Saarbrucken, Germany (2012).
13. Ullmann, T.: Automated detection of reflection in texts - A machine learning based approach., PhD thesis The Open University (2015).
14. Cheng, G.: The impact of online automated feedback on students' reflective journal writing in an EFL course. Internet High. Educ. 34, 18–27 (2017).
15. Kovanović, V., Joksimović, S., Mirriahi, N., Blaine, E., Gašević, D., Siemens, G., Dawson, S.: Understand students' self-reflections through learning analytics. Proc. 8th Int. Conf. Learn. Anal. Knowl. - LAK '18. 389–398 (2018).

16. Tausczik, Y.R., Pennebaker, J.W.: The psychological meaning of words: LIWC and computerized text analysis methods, (2010).

17. Graesser, A.C., McNamara, D.S., Louwerse, M.M., Cai, Z.: Coh-metrix: analysis of text on cohesion and language. Behav. Res. Methods. Instrum. Comput. 36, 193–202 (2004).

18. Ullmann, T.D.: An architecture for the automated detection of textual indicators of reflection. In: the proceeding of 1st European Workshop on Awareness and Reflection in Learning Networks. , Palermo, Italy (2011).

19. Ullmann, T.D.: Automated Analysis of Reflection in Writing: Validating Machine Learning Approaches. Int. J. Artif. Intell. Educ. 1–41 (2019).

20. Chen, Y., Yu, B., Zhang, X., Yu, Y.: Topic modeling for evaluating students' reflective writing. In: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16. pp. 1–5. ACM Press, Edinburgh, UK (2016).

21. Sandor, A.: Modeling metadiscourse conveying the author's rhetorical strategy in biomedical research abstracts. Rev. Fr. Linguist. Appl. (2007).

22. Mantzourani, E., Hughes, M.L.: Role-emerging placements in pharmacy undergraduate education: Perceptions of students. Pharm. Educ. 16, 88–91 (2016).

23. Mantzourani, E. (Efthymia), Deslandes, R., Ellis, L., Williams, G.: Exposing Pharmacy Students to Challenges Surrounding Care of Young Children via a Novel Role-Emerging Placement. J. Curric. Teach. 5, (2016).

24. Deslandes, R., Lucas, C., Louise, M., EfiMantzourania, H.: Development of a template to facilitate reflection among student pharmacists. Res. Soc. Adm. Pharm. 14, 1058–1063 (2018).

25. Tsingos, C., Bosnic-Anticevich, S., Lonie, J.M., Smith, L.: A model for assessing reflective practices in pharmacy education, (2015).

26. Lucas, C., Bosnic-Anticevich, S., Schneider, C.R., Bartimote-Aufflick, K., McEntee, M., Smith, L.: Inter-rater reliability of a reflective rubric to assess pharmacy students' reflective thinking. Curr. Pharm. Teach. Learn. 9, 989–995 (2017).

27. Mezirow, J.: How Critical Reflection Triggers Transformative Learning. Foster. Crit. Reflect. Adulthood. (1990).

28. Warriner, A.B., Kuperman, V., Brysbaert, M.: Norms of valence, arousal, and dominance for 13,915 English lemmas. Behav. Res. Methods. 45, 1191–1207 (2013).

29. Hall, M.A.: Correlation-based Feature Selection for Machine Learning, (1998).

30. Domingos, P.: MetaCost : A General Method for Making Classifiers. In: Proceedings of the 5th International Conference on Knowledge Discovery and Data Mining. pp. 155–164. ACM Press, San Diego, CA (1999).

31. Weiss, G.M., Mccarthy, K., Zabar, B.: Cost-Sensitive Learning vs . Sampling : Which is Best for Handling Unbalanced Classes with Unequal Error Costs ? In: The Proceeding of 2007 International Conference on Data Mining. pp. 35–41. ACM Press, Las Vegas (2007).

32. Kim, J., Choi, K., Kim, G., Suh, Y.: Classification cost: An empirical comparison among traditional classifier, Cost-Sensitive Classifier, and MetaCost. Expert Syst. Appl. 39, 4013–4019 (2012).

33. Breiman, L.: Random forests. Mach. Learn. 45, 5–32 (2001).

34. Frank, E., Witten, I.H.: Generating accurate rule sets without global optimization. In: Proceedings of the Fifteenth International Conference on Machine Learning. pp. 144–151. Morgan Kaufmann Publishers Inc., Madison, Wisconsin (1998).

35. Christianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press (2000).

36. Murphy, K.P.: Naive Bayes classifiers Generative classifiers. Bernoulli. (2006).
37. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The WEKA data mining software: an update. ACM SIGKDD Explor. 11, 10–18 (2009).
38. Gibbs, G.: Learning by Doing: A guide to teaching and learning methods. FEU (1988).
39. Milligan, S., Griffin, P.: Understanding Learning and Learning Design in MOOCs: A Measurement-Based Interpretation. J. Learn. Anal. 3, 88–115 (2016).