ORIGINAL RESEARCH

Recognizing Reflection: Computer-Assisted Analysis of First Year Medical Students' Reflective Writing

Caitlin D. Hanlon[1] [ID] · Emily M. Frosch[2] · Robert B. Shochet[3] · Simon J. Buckingham Shum[4] · Andrew Gibson[5] · Harry R. Goldberg[6]

Abstract

Background Reflective writing is used throughout medical education to help students navigate their transformation into medical professionals. Assessment of reflective writing, however, is challenging; each available methodology of assessment has distinct advantages and disadvantages. We tested if combining two independent assessment mechanisms—a faculty-designed rubric and Academic Writing Analytics (AWA), an automated technique—could be used together to form a more robust form of evaluation.
Methods We obtained reflective essays written by first year medical students as part of a clinical skills course. Faculty scored essays using a rubric designed to evaluate Integration, Depth, and Writing. The same essays were subjected to AWA analysis, which counted the number of reflective phrases indicative of Context, Challenge, or Change.
Results Faculty scored the essays uniformly high, indicating that most students met the standard for reflection as described by the rubric. AWA identified over 1400 instances of reflective behavior within the essays, and there was significant variability in how often different types of reflective phrases were used by individual students.
Conclusions While data from faculty assessment or AWA alone is sufficient to evaluate reflective essays, combining these methods offer a richer and more valuable understanding of the student's reflection.

Keywords Reflective writing · Undergraduate medical education · Natural language processing

Introduction

Professional identity formation (PIF) is a critical component of medical education and a task that the LCME explicitly emphasizes as a domain of student learning and growth [1–3]. As students' professional identity forms, they gain perspective of what it means to become a physician and how their values are realized in practice [4]. How the trajectory of PIF is

✉ Caitlin D. Hanlon
  caitlin.hanlon@quinnipiac.edu

  Emily M. Frosch
  efrosch@jhmi.edu

  Robert B. Shochet
  rshochet@montefiore.org

  Simon J. Buckingham Shum
  Simon.BuckinghamShum@uts.edu.au

  Andrew Gibson
  andrew.gibson@qut.edu.au

  Harry R. Goldberg
  Goldberg@jhmi.edu

[1] Department of Biology, Quinnipiac University, Hamden, CT, USA

[2] Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD, USA

[3] Montefiore Health System, Yonkers, NY, USA

[4] Connected Intelligence Center, University of Technology Sydney, Sydney, Australia

[5] Information Science, Queensland University of Technology, Brisbane, Australia

[6] Department of Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, MD, USA

assessed, tracked, and otherwise defined, however, can be difficult. Consequently, reflective writing exercises are increasingly being used within medical curricula as a proxy for PIF as the act of reflecting helps students gain perspective on themselves and their experiences during their training [5, 6].

Reflective exercises such as narrative writing, appreciative inquiry, and clinical case discussions are designed to help students evaluate their lived experiences and revise their internal schemas [2]. By asking students to consider how incidents unfolded, to examine their emotional responses to situations, and to anticipate future events, elements of metacognitive and reflective behavior are nurtured and further refined. In addition to supporting the internal capacities of students, reflective exercises are a tangible, external output of a student's acculturation into a medical professional [7]. Reflective skills are correlated with higher academic performance [8, 9], increased professionalism [10], and better clinical decision-making [11, 12].

The benefits of reflective writing are not limited to students. Reading, writing, and assessing reflective writing positively affects faculty in many ways [13, 14]. Faculty report that the process of evaluating reflective writing made them feel personally fulfilled about their career choice and allowed them to know, understand, and appreciate their students more deeply. Evaluation of reflection, however, is a time-consuming process [15]. Faculty must devote a considerable amount of effort to training for how to evaluate reflective writing in addition to the reading and grading of the essays themselves. Other professional responsibilities often preclude faculty from giving extensive feedback on reflective exercises. Without this feedback, students may question the evaluation method as a whole [16] and the value of reflection [17, 18]. These challenges threaten to erode the benefits of reflection for both faculty and students.

Automated techniques for assessing the depth and quality of reflective writing are now emerging as a response to these obstacles. Academic Writing Analytics (AWA) is an initiative from the University of Technology Sydney (Australia) to improve academic writing by providing more timely, standardized, and personalized feedback on a student's written work. AWA employs natural language processing to identify phrases describing details (*Context*), explaining struggles (*Challenge*), and planning for subsequent events (*Change*; [19]). Context, challenge, and change are referred to as "reflective moves" as they indicate shifts to progressively more complex and deeper levels of reflection [20]. Pedagogically, AWA is used as a writing tool: students are able to submit drafts to a website hosting AWA and instantaneously view counts and highlighted phrases of reflective moves within their writing. Students can use this information to refine their writing before submitting another iteration. AWA has been

employed in several different disciplines, including pharmacy, law, and engineering [19, 21, 22], and the response has been overwhelmingly positive. Students repeatedly engage with AWA to improve their writing [19]. While AWA is able to flag phrases that indicate reflective behavior, it is unable to provide context or meaning to these phrases. Therefore, the feedback is limited to counts, which may not provide sufficient guidance to some students.

An ideal method of evaluation would be fast, reproducible, instructive, and empathetic to students' experiences. Since we have not found a single method to attain all of these goals, we hypothesized that using two independent methods would synergistically create a more meaningful evaluation. In this report, we describe a preliminary study to determine if AWA could complement the current assessment method for evaluating reflective writing in a medical school clinical skills course. The first step in this process is determining how AWA processes medical reflective writing. Our approach asked if AWA (1) can assess reflective writing in a medical domain, (2) can discern a dynamic range of reflective markers, and (3) can identify unique traits within reflective writing that may have been overlooked by faculty evaluation. Addressing these questions will allow us to capitalize on the benefits of each approach and develop a richer method of assessment and guidance for students.

Methods

Subjects and Setting

All medical students at The Johns Hopkins School of Medicine (JHSOM) take an introductory, 16-week clinical skills course in the first semester of their first year called *Clinical Foundations in Medicine* (CFM). This course includes training in communication skills, history-building, physical examination, and the practice of professionalism. CFM is taught weekly in 24 small groups integrated within JHSOM's learning community. Each group consists of five students and one faculty member who maintains a longitudinal connection to each student as their advisor. At the end of CFM, students write two 250–350 word reflective essays in response to prompts asking them to describe their perception of their learning in the course (see Appendix I). Essay 1 asks students to reflect upon changes that they noticed in themselves over the preceding semester, while essay 2 asks students to reflect on an experience as a member of their learning team. Each of the 24 faculty members score essays from students outside of their small group.

Faculty grade essays using a four-item Likert scale in three domains evaluating Integration (between experiences in the

Table 1 Evaluative methods for reflective essays. (a) CFM course holistic rubric domains (Integration, Depth, Writing) with score assignments (1–4) used by faculty. (b) AWA lexical categories (Context, Challenge, And Change), definitions used to identify each reflective move, and examples of each move type found within essays used in this study set

a

Categories	Unsatisfactory	Underdeveloped	Good	Excellent
Integration	Provides no connection(s) between their experience in CFM and personal growth.	2	3	Provides clear connection(s) between their experience in CFM and personal growth.
	1			4
Depth	Superficial appreciation of complexity of their own growth.	2	3	Considers complexity of their own growth with humility and openness.
	1			4
Writing	Poor organization, typos, grammatical errors, not at a graduate level	2	3	Makes no or very few errors. Organization, flow at graduate level.
	1			4

b

Categories	Description	Example
Context	Describing details of a situation; noticing what occurred and how it made the author feel	"I've had doctors with a wide range of clinical and interpersonal skills, so I was surprised at how much time was spent teaching relationship-building techniques."
Challenge	Explaining struggles that arose during a situation; identification of problems that occurred in pursuit of a goal	"I have trouble asking a patient to reveal their most personal stories and potential regrets without telling them anything about my personal life."
Change	Determining what can be learned from an experience; planning for future events	"In acquiring these skills, I have noticed how they relate to my future role as a physician."

course and personal growth), Depth (the complexity of personal growth), and Writing Quality (Table 1a). Students could receive a maximum score of 24 points: 12 points from each essay, 4 points from each domain. Prior to assessment, faculty members review the rubric as a group to discuss scoring strategies. Two members of our study team (EF, RS) supervised this assessment process.

Procurement of Essays

We obtained two-hundred forty essays written by the 120 students enrolled in the 2017–2018 CFM course. To maintain confidentiality, all essays were de-identified and randomly numbered before the AWA analysis began. Because the essays were in hard copy, we scanned each essay into a PDF format and used optical text recognition software to generate a plain text document.

Exclusion Criteria

Our study included only students who had both faculty rubric scores and AWA counts from essay 1 and essay 2. Optical processing failed for eight students' essay 2 and for six students for both essays 1 and 2. Therefore, these 14 students were excluded from the study. As a result, we evaluated 212 essays from 106 students.

AWA Analysis

Essays were analyzed using the AWA analytics engine (TAP and Athanor-server; [19]). TAP and Athanor NLP software is open source (https://github.com/heta-io). The simplest way to see how they perform is via the *AcaWriter Demonstrator*: http://acawriter-demo.utscic.edu.au (select the *reflective writing* genre). Algorithms within AWA identify lexical patterns that are indicative of reflective activity (Table 1b). AWA analyzed essays for "reflective moves": instances of (i) *Context* (describing initial thoughts about an event), (ii) *Challenge* (describing difficulties encountered during an event), and (iii) *Change* (indicating a future commitment to altering an approach or behavior when confronted with a similar event). AWA does not assign a summative grade to each essay, but instead highlights and counts instances of context, challenge, or change phrases.

A study member (CH) combined student identification number, essay identification number, AWA pattern counts for each essay, and the faculty score into a single spreadsheet. From these data, we performed statistical analysis using paired Student's t tests in Microsoft Excel. The datasets generated during and/or analyzed during the current study are available in the Open Science Framework repository at https://mfr.osf.io/render?url=https%3A%2F%2Fosf.io%2Fvhwuq%2Fdownload. Standard deviations for averages are reported in

text parenthetically next to the average value. This study was approved by the Johns Hopkins University Internal Review Board (IRB: 00145005).

Results

Faculty Evaluation of Essays

Twenty-four LC faculty scored 212 essays written by 106 students in December 2017 (end of 2017–2018 CFM course). The average score given by faculty was 23.1 out of 24 total possible points (average = 96.1%; Online Resource 2). The mean scores for Writing quality (97.5%) were significantly higher than the Depth scores (95%, $p < 0.01$; Fig. 1a), but there was no significant difference between Integration and Depth scores (96.3% vs 95%), or Integration and Writing scores (96.3% vs 97.5%; Fig. 1a). Additionally, there was no statistically significant difference between domain scores when comparing essay 1 to essay 2 (Online Resource 3).

AWA Assessment of Medical Reflective Writing

AWA analysis of 212 essays from 106 students identified 1415 discrete reflective moves (Context, Challenge, or Change; Online Resource 2). Students used an average of 4.9 Context phrases, 7.4 Challenge phrases, and 1.0 Change phrases across both essays (Fig. 1b). The difference between all of these values was statistically significant ($p < 0.001$). Comparing each essay, students used more Challenge phrases (4.0 phrases/student within Essay 1 (professional identity)) compared to Essay 2 (teamwork; 3.5 phrases/student; Online Resource 3), and more Context in Essay 2 (2.7 phrases/student) than in Essay 1 (2.3 phrases/student; Online Resource 3). Of note, AWA identified at least one Context, Challenge, or Change phrase for every student.

Distribution of Reflective Moves

To examine the distribution of reflective behavior within the class, histograms were created that plotted reflective behavior (either score from domains in the faculty rubric or AWA counts of moves) against the percent of the class that demonstrated said behavior. For faculty scoring, the distribution was skewed towards the right, as the majority of the class received a perfect score of 100%. Within the rubric domains, 83% of the class received a perfect score on Integration, 69% received a perfect score on Depth, and 86% received a perfect score on Writing (Fig. 1c). The histogram of AWA reflective counts was more broadly distributed. Challenge phrases appeared most frequently, ranging from 1 to 14 instances (Fig. 1d). Eighty-eight percent of the class ($n = 93/106$) had five or more instances of Challenge phrases. Context was the next most

common reflective move. Instances of Context varied from one to eleven; 89% of the class ($n = 94/106$) had between two and eight instances of Context. Only 7.5% of the class ($n = 8/106$ students) had more than two Change phrases; 41.5% ($n = 44/106$ students) of the class did not have any Change phrases.
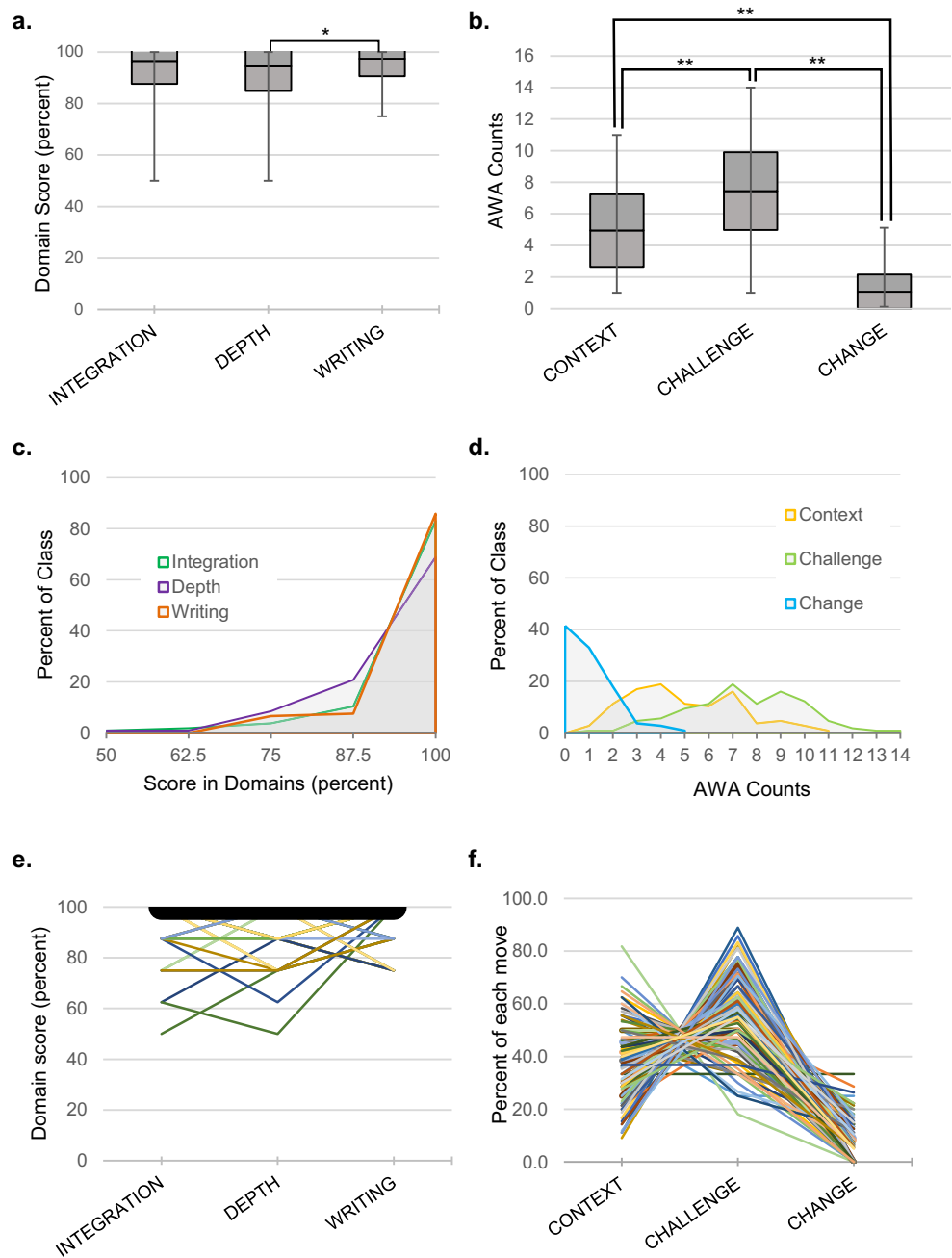
Patterns of Reflective Moves

To visualize the data in a more granular and individualized way, each student's reflective behavior was plotted. Although there was some variance in faculty scoring, the distribution of rubric scores was most towards the top of the graph (Fig. 1e). The majority of students (63/106 = 59.4%) received a perfect score in all three domains. Twenty-two students (22/106 = 20.8%) received perfect scores in two domains, while the 19 students received a perfect score in one domain. Altogether, only two students did not receive a perfect score in at least one domain. For AWA, the data was normalized to show each student's reflective moves in each domain relative to the student's total number of moves. This normalization allowed us to visualize the different patterns between students (Fig. 1f). Over 75% of students (75/106 = 75.8%) had a unique pattern of how many Context, Challenge, and Change phrases used within their essays. The most commonly occurring pattern was 50% Context:50% Challenge:0% Change, which occurred for five individuals. The other recurring patterns occurred between only two or three individuals.

Discussion

In this study, we sought to determine if AWA software could provide a rapid, quantitative evaluation of medical student reflective writing that would supplement our current evaluative methods. Our first step was to determine if AWA could process first year medical students' reflective essays. Indeed, AWA was able to identify over 1400 instances of reflective behavior and at least one count of Context, Challenge, or Change for all students. We next asked if AWA could capture the range of reflective behavior within the class. While the average student used 13.3 reflective moves in two essays, the standard deviation was 3.2, indicating significant differences in reflective behavior within the class. Finally, we wondered if AWA could identify unique reflective traits for individual students. We found that the majority of students used distinct lexical patterns of Context, Challenge, and Change within their work, suggesting that AWA highlights the extraordinary individuality of the students. Therefore, we believe that AWA is a useful tool that complements our current method of assessment. To our knowledge, this is the first use of a natural language processing tool to evaluate reflective writing in a medical curriculum.

Fig. 1 Numerical analysis of reflective essays. **a** Average score (in percent) per student (*n* = 106 students, 212 total essays) in each domain (Integration, Depth, and Writing) in the faculty rubric. The average score between Depth (95%) and Writing (97.5%) was statistically significantly different (**p* < 0.05). No statistically significant difference was observed between Integration (average score = 96.3%) and Depth or Writing. **b** Average number of AWA counts (*n* = 106 students, 212 total essays) for each type of move. Students had 4.9 *Context* moves/student, 7.4 *Challenge* moves/student, and 1.0 *Change* moves/student. The difference between these averages was statistically significant (***p* < 0.001). The top and bottom of the box are the standard deviation of the score. The caps on the vertical line are the minimum and maximum score. **p* < 0.05; ***p* < 0.001. **c** Histogram of the percent of the class versus the average rubric score for each domain (Integration, Writing, and Depth). **d** Histogram of the percent of the class versus the number AWA counts for each domain (*Context*, *Challenge*, and *Change*). One hundred six students were included in the study. **e** Scores within each rubric domain for each individual student. **f** AWA counts of each type of reflective move for each individual student

One of the more interesting patterns that emerged from the AWA analysis was that *Challenge* was the reflective move used most often (Fig. 1b). While one explanation could be the inherent stress, anxiety, and tension of medical school (a theme that emerges in several thematic analyses of reflective writing; [23–26]), an alternative explanation could be that students were prompted to write about challenges. Indeed, essay 1 asked students about a personal "dissonant" moment, and essay 2 asked students about a difficult moment they experienced while working in a team. In fact, essay 2 explicitly asked students, "What challenged you[?]". The high level of *Challenge* phrases used by students seems to be driven by

the directives within the prompts. This hypothesis also explains the lack of *Change* phrases within the essays: neither prompt asked students how they changed throughout their training. Although AWA was not tailored to our prompts, its output was sensitive enough to reveal that the prompts were promoting different types of reflective behavior. This conclusion aligns with previous studies showing that students who were provided with structured guidelines scored better on reflective writing exercises than students who were only given a prompt [27–29]. One benefit of using AWA to evaluate how prompts or other interventions are affecting student behavior is its speed: in the Miller-Kuhlmann study, 2 to 4 h of training

were needed to achieve sufficient levels of inter-rater reliability [27], while a similar analysis conducted using AWA would take a significantly shorter amount of time.

Another surprising finding was that AWA identified a high degree of individuality within the essays. AWA analysis revealed significant variations in phrasing patterns among students (Fig. 1f), indicating that students are completing the act of reflection in many diverse ways. Much like a fingerprint, the lexical patterns revealed by AWA were largely unique to each student. The assertion that there are many different pathways to reflection would be quite apparent to any reader of a reflection [14, 30, 31]. This information, however, was not evident in the data that resulted from faculty-based analysis. While the faculty-based evaluation showed that the majority of students are meeting or exceeding the bar for reflection (Fig. 1e), it was unable to represent the individuality within the reflections. The data from AWA easily showed that students take unique approaches within their reflections, but because AWA does not assign meaning or value to these reflective counts, the information lacks context. This finding supports our belief that a single method of evaluation cannot capture the full extent of a student's reflective activity. We believe that together these two approaches (software-based AWA accounting and rubric-based faculty assessment) enhance our understanding of a student's reflective capacity.

One data set that was deliberately not presented in this study was a direct comparison between the faculty rubric score and AWA counts. While we did complete this analysis (and found no correlation between faculty scoring and AWA counts), we later realized that this result should be expected because the faculty rubric and AWA are couched in two different conceptual frameworks of reflection. The faculty rubric aligns most closely with Moon's input-outcome model, which posits that reflection itself is purposeful and the outcomes (such as a self-development, a critical review of events, or resolution of an uncertainty) are beneficial but not an all-encompassing goal [32]. AWA is similar to Boud's reflection-in-learning model [33], Schön's reflective practitioner concept [34], Korthagen's ALACT model [35], and Kolb's experiential learning cycle [36]. These models incorporate the concept of planning for future actions based on lessons learned while reflecting, which is similar to the *Change* metric in AWA. Each of these models value different aspects of reflective behavior. Consequently, our lack of correlation between the faculty rubric score and AWA counts supports the idea that these two methods of evaluation are quite different from one another and can therefore be complementary to one another.

Reflective writing creates opportunities for medical students to consider, document, and share their progress in learning about themselves, their roles, and their challenges [2]. This intentional examination of their metacognitive processes allows students and faculty advisors to track the student's developing professional identity [7]. As faculty assess reflections in a formative or summative context, they should consider the relative value of both qualitative and quantitative methods in the measurement of metacognition. Although current approaches to assessing students' reflective writing are primarily through a manual, qualitative approach using a rubric [37], linguistic analyses [19, 38] and narrative methodologies [39] offer instructors a unique level of insight and into students' metacognitive processes. Combining both approaches allows faculty to recognize elements of reflection that may not have been previously noticed.

There were several limitations to this study. First, this was a one institution study involving one class of first year students, so we cannot yet determine the generalizability of results. Moreover, while we did confirm that there was no difference in reflective behavior between male and female students, we did not thoroughly examine rubric scores and AWA counts for other demographic sets. However, as this study's purpose was to assess the feasibility of using AWA for student reflections in a medical school course, we believe its scope was appropriate. Secondly, inter-rater reliability for rubric scoring by faculty was not rigorously addressed, and the rubric used was not developed as a validated scale. Although this may have resulted in inconsistencies in scoring, it represented the faculty's best efforts in the context of an actual student course and, in that way, is a realistic estimation of current assessments for reflections. Lastly, the significance of the number and types of reflective moves identified by AWA cannot be definitively linked to students' professional development at this time, although this is a central objective of future studies.

Future Directions

The results of this study open up several exciting avenues for future work. Before we fully adopt AWA and our current methods in a complementary fashion, we need to develop a complete understanding of faculty impressions of AWA and how instructors might use and interpret AWA counts. AWA can only be a complementary tool insofar as those performing the assessment view it as useful, interesting, and reliable [19]. Likewise, statistical validation of AWA scores is needed to demonstrate to what degree AWA counts are congruent with faculty evaluation. This type of assessment could lead to the AWA algorithm being refined specifically for discipline-specific terms and phrases. Altogether, completing these studies will be the first step in determining the meaning of *Context*, *Challenge*, and *Change* as applied to medical reflective writing.

Because we view AWA as a complement to our current method of assessment, we pondered how this partnership would be operationalized. AWA's strength is its ability to rapidly provide reproducible counts of reflective phrases, while the advantage of the faculty-based approach is assigning

meaning to these phrases and the interpersonal relationships that develop during the reading, writing, and sharing of the reflections. One potential use of AWA would be a tool to facilitate mentorship conversations. For example, if a faculty member notices that a student uses a high number of *Challenge* phrases but few *Change* phrases, they could use this information to help the student resolve the challenges and plan for future events. In this way, AWA would provide additional information that faculty members could use to better understand their students and help guide them on their academic journey. Pairing writing with reading and further discussion synergistically enhances the reflective process and can lead to a deeper perception of self [39, 40]. Using reflective writing and AWA during the advising process may also help students value reflection as more than an activity to simply be completed [2, 18, 20]. AWA could also be used to longitudinally track a student's reflective behavior over time [26, 41, 42], which would parallel their professional identity formation over the same period [10]. Of note, we do not envision AWA as a method to assign grades to reflective writing exercises. Rather, AWA should be used as a tool to help mentors further support the reflective process [43, 44].

AWA was developed to be used by students as a drafting tool to improve their writing [19]. While we did not use AWA in this manner, we are not opposed to this possibility. However, we are also interested in ways that AWA could be used at a programmatic level. Following refinement and validation, one potential use of AWA is that it could serve as an independent, standardized arbiter for prompt refinement. For example, if a program values the narrative or descriptive elements of reflection, they could use AWA to determine if students are predominantly using *Context* phrases. If the phrase distribution does not match their expectations, they could refine the prompt and determine if AWA counts are affected. Another option would be to determine if specific interventions (like giving students detailed guidelines [27, 29] or increased feedback [45, 46]) or specific clinical experiences [26] would change the distribution of AWA counts. Because AWA is an algorithm, it is able to provide standardized counts that can be used to answer these questions, and it is not subject to personal biases or semantic differences that lead to discrepancies between assessors [47]. Removing these barriers within assessment allows for a meta-analysis of the evaluative process within a program. Altogether, these approaches would make the reflective process more valuable to a program and more meaningful to students and faculty.

Conclusion

This study is the first to apply automated writing analysis to medical reflective writing and demonstrates the value of using multiple methods of assessment to paint a richer picture of students' reflective journeys.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Ethics Approval Johns Hopkins University Internal Review Board (IRB:00145005).

Consent to Participate Not applicable.

Consent for Publication All authors read and approved the final manuscript.

Code Availability https://github.com/heta-io

References

1. Liaison Committee on Medical Education. LCME accreditation standards. 2008. http://www.lcme.org/functions2008jun.pdf.
2. Sandars J. The use of reflection in medical education: AMEE guide no. 44. Med Teach. 2009;31(8):685–95.
3. Cruess RL, Cruess SR, Boudreau JD, Snell L, Steinert Y. A schematic representation of the professional identity formation and socialization of medical students and residents: a guide for medical educators. Acad Med. 2015;90(6):718–25.
4. Weaver R, Peters K, Koch J, Wilson I. 'Part of the team': Professional identity and social exclusivity in medical students. Med Educ. 2011;45(12):1220–9.
5. Branch WT Jr, George M. Reflection-based learning for professional ethical formation. AMA J Ethics. 2017;19(4):349–56.
6. Wald HS, Anthony D, Hutchinson TA, Liben S, Smilovitch M, Donato AA. Professional identity formation in medical education for humanistic, resilient physicians: pedagogic strategies for bridging theory to practice. Acad Med. 2015;90(6):753–60.
7. Quirk M. Intuition and metacognition in medical education: keys to developing expertise: Springer publishing company; 2006.
8. Tsingos-Lucas C, Bosnic-Anticevich S, Schneider CR, Smith L. Using reflective writing as a predictor of academic success in different assessment formats. Am J Pharm Educ. 2017;81(1).
9. Di Stefano G, Gino F, Pisano G, Staats B. An appraisal of medical students' reflection-in-learning. Med Educ. 2000;34(3):182–7.

10. Hoffman LA, Shew RL, Vu TR, Brokaw JJ, Frankel RM. Is reflective ability associated with professionalism lapses during medical school? Acad Med. 2016;91(6):853–7.

11. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. Med Educ. 2008;42(5):468–75.

12. Croke E. The use of structured reflective journal questions to promote fundamental development of clinical decision-making abilities of the first-semester nursing student. Contemp Nurse. 2004;17(1–2):125–36.

13. Sukhato K, Sumrithe S, Wongrathanandha C, Hathirat S, Leelapattana W, Dellow A. To be or not to be a facilitator of reflective learning for medical students? A case study of medical teachers' perceptions of introducing a reflective writing exercise to an undergraduate curriculum. BMC Emerg Med. 2016;16(1):102.

14. Branch WT Jr, Frankel R. Not all stories of professional identity formation are equal: an analysis of formation narratives of highly humanistic physicians. Patient Educ Couns. 2016;99(8):1394–9.

15. Koole S, Dornan T, Aper L, Scherpbier A, Valcke M, Cohen-Schotanus J, et al. Factors confounding the assessment of reflection: a critical review. BMC Emerg Med. 2011;11(1):104.

16. Aitchison C, Catterall J, Ross P, Burgin S. 'Tough love and tears': learning doctoral writing in the sciences. High Educ Res Dev. 2012;31(4):435–47. https://doi.org/10.1080/07294360.2011.559195.

17. McCarthy J. Reflective writing, higher education and professional practice. J Educ Built Environ. 2011;6(1):29–43.

18. Grant A, Kinnersley P, Metcalf E, Pill R, Houston H. Students' views of reflective learning techniques: an efficacy study at a UK medical school. Med Educ. 2006;40(4):379–88.

19. Gibson A, Aitken A, Sandor A, Buckingham-Shum S, Tsingos-Lucas C, Knight S. Reflective writing analytics for actionable feedback. In: Proc. 7th Int. Conf. Learning Analytics & Knowledge; 2017. p. 13–7. https://doi.org/10.1145/3027385.3027436.

20. Buckingham Shum S, Sándor Á, Goldsmith R, Bass R, McWilliams M. Towards reflective writing analytics: rationale, methodology and preliminary results. J Learn Anal. 2017;4(1):58–84. https://doi.org/10.18608/jla.2017.41.5.

21. Lucas C, Gibson A, Buckingham SS. Utilization of a novel online reflective learning tool for immediate formative feedback to assist pharmacy students' reflective writing skills. Am J Pharm Educ. 2018;83:6800. https://doi.org/10.5688/ajpe6800.

22. Knight S, Buckingham Shum S, Ryan P, Sandor A, Wang X. Designing academic writing analytics for civil law student self assessment. Int J Artif Intell Educ. 2018;28(1):1–28. https://doi.org/10.1007/s40593-016-0121-0.

23. Nevalainen MK, Mantyranta T, Pitkala KH. Facing uncertainty as a medical student—a qualitative study of their reflective learning diaries and writings on specific themes during the first clinical year. Patient Educ Couns. 2010;78(2):218–23.

24. Cunningham H, Taylor D, Desai UA, Quiah SC, Kaplan B, Fei L, Catallozzi M, Richards B, Balmer DF, Charon R. Looking back to move forward: first-year medical students' meta-reflections on their narrative portfolio writings. Acad Med. 2018;93(6):888–94.

25. Gaufberg EH, Batalden M, Sands R, Bell SK. The hidden curriculum: what can we learn from third-year medical student narrative reflections? Acad Med. 2010;85(11):1709–16.

26. Kristiansson MH, Troein M, Brorsson A. We lived and breathed medicine-then life catches up: medical students' reflections. BMC Emerg Med. 2014;14(1):66.

27. Miller-Kuhlmann R, O'Sullivan PS, Aronson L. Essential steps in developing best practices to assess reflective skill: a comparison of two rubrics. Med Teach. 2016;38(1):75–81.

28. Aronson L, Niehaus B, Lindow J, Robertson PA, O'sullivan PS. Development and pilot testing of a reflective learning guide for medical education. Med Teach. 2011;33(10):e515–21.

29. Aronson L, Niehaus B, Hill-Sakurai L, Lai C, O'Sullivan PS. A comparison of two methods of teaching reflective ability in year 3 medical students. Med Educ. 2012;46:807–14.

30. Sharpless J, Baldwin N, Cook R, Kofman A, Morley-Fletcher A, Slotkin R, et al. The becoming: students' reflections on the process of professional identity formation in medical education. Acad Med. 2015;90(6):713–7.

31. Bernard AW, Malone M, Kman NE, Caterino JM, Khandelwal S. Medical student professionalism narratives: a thematic analysis and interdisciplinary comparative investigation. BMC Emerg Med. 2011;11(1):11.

32. Moon JA. Reflection in learning and professional development: theory and practice. London: Kogan Page; 1999.

33. Boud D, Keogh R, Walker D. Reflection: turning experience into learning. London: Kogan Page; 1985.

34. Schön DA. The reflective practitioner: how professionals think in action. New York: Basic Books; 1983.

35. Korthagen FAJ. Reflective teaching and preservice teacher education in the Netherlands. J Teach Educ. 1985;9(3):317–26.

36. Kolb DA. Experiential learning: experience as a source for learning and development. Upper Saddle River: Prentice Hall; 1984.

37. Wald HS, Borkan JM, Taylor JS, Anthony D, Reis SP. Fostering and evaluating reflective capacity in medical education: developing the REFLECT rubric for assessing reflective writing. Acad Med. 2012;87(1):41–50.

38. Shibani A, Knight S, Buckingham Shum S. Contextualizable learning analytics design: a generic model, and writing analytics evaluations. In: Proc. 9th Int. Conf. Learning Analytics & Knowledge; 2019. p. 210–9. https://doi.org/10.1145/3303772.3303785.

39. DasGupta S, Charon R. Personal illness narratives: using reflective writing to teach empathy. Acad Med. 2004;79(4):351–6.

40. Charon R, Hermann MN. A sense of story, or why teach reflective writing? Acad Med. 2012;87(1):5–7 (Jan).

41. Cayley JW, Schilling R, Suechting R. Changes in themes over time from medical student journaling. WMJ. 2007;106(8):486–9 (Dec).

42. Sobral DT. An appraisal of medical students' reflection-in-learning. Med Educ. 2000;34(3):182–7.

43. Heeneman S, de Grave W. Tensions in mentoring medical students toward self-directed and reflective learning in a longitudinal portfolio-based mentoring system–an activity theory analysis. Med Teach. 2017;39(4):368–76.

44. Kalén S, Ponzer S, Seeberger A, Kiessling A, Silén C. Continuous mentoring of medical students provides space for reflection and awareness of their own development. Int J Med Educ. 2012;3:236–44.

45. Sargeant JM, Mann KV, Van der Vleuten CP, Metsemakers JF. Reflection: a link between receiving and using assessment feedback. Adv Health Sci Educ. 2009;14(3):399–410.

46. Boehler ML, Rogers DA, Schwind CJ, Mayforth R, Quin J, Williams RG, et al. An investigation of medical student reactions to feedback: a randomised controlled trial. Med Educ. 2006;40(8):746–9.

47. Eckes T. Rater types in writing performance assessments: a classification approach to rater variability. Lang Test. 2008;25(2):155–85.