SHORT COMMUNICATION
Volume 14 Issue 1 2022
DOI: 10.21315/eimj2022.14.1.8
ARTICLE INFO
Received: 10-05-2021
Accepted: 25-10-2021
Online: 30-03-2022

Inter-Rater Reliability of Reflective-Writing Assessment in an Undergraduate Professionalism Course in Medical Education

Diantha Soemantri[1,2], Rita Mustika[1,2], Nadia Greviana[1,2]

[1]*Department of Medical Education, Faculty of Medicine Universitas Indonesia, Jakarta, INDONESIA*

[2]*Medical Education Center, Indonesian Medical Education and Research Institute (IMERI), Faculty of Medicine Universitas Indonesia, Jakarta, INDONESIA*

To cite this article: Soemantri D, Mustika R, Greviana N. Inter-rater reliability of reflective-writing assessment in an undergraduate professionalism course in medical education. Education in Medicine Journal. 2022;14(1):87–97. https://doi.org/10.21315/eimj2022.14.1.8

To link to this article: https://doi.org/10.21315/eimj2022.14.1.8

ABSTRACT

Reflective writing is increasingly being used in the teaching of professionalism. Because assessment enhances the learning process, effective evaluation of students' reflective writing is needed. The aim of this study was to examine the inter-rater agreement between two different reflective writing assessment rubrics, which categorised reflective writings into four level of reflection, in an undergraduate medical professionalism course. The reflective writing assignments from 63 medical students enrolled in the 2017 medical professionalism course in the Faculty of Medicine Universitas Indonesia were randomly selected and independently assessed by two raters in September 2019. Intraclass correlation (ICC) analysis (two-way mixed effect, single measure) was carried out to determine the inter-rater agreement of the reflective writing assessment. The less detailed instrument showed a low ICC score of 0.43, which was classified into poor inter-rater agreement, whereas the more detailed rubric showed poor to moderate reliability, with ICC scores of 0.50, 0.50, and 0.36 for the score of each criterion, the total score of each assessed criterion, and the overall score of reflection, respectively. Utilising a more detailed (analytic) rubric to assess students' reflective writing produced a relatively higher score of inter-rater reliability, although the reliability achieved using this rubric was still categorised as moderate.

Keywords: *Reflective writing assessment, Self-reflection, Inter-rater reliability*

CORRESPONDING AUTHOR Nadia Greviana, Department of Medical Education, Faculty of Medicine Universitas Indonesia, Salemba Raya 6, Jakarta Pusat 10430, Jakarta, Indonesia

Email: nadiagreviana@ui.ac.id, nadia.greviana@gmail.com

INTRODUCTION

Reflection is an important feature of professional development (1–2), since reflection helps learners to integrate new knowledge with their prior knowledge and promotes lifelong learning. Studies have also shown that reflective writing contributes to the development of medical students' empathy, professionalism, respect for diversity and communication skills (3–4). Thus, it is important to provide students with appropriate pedagogical strategies, including teaching methods, assessment and a reflective learning environment (5). It has been argued that a well-designed assessment

system for reflective writing will increase students' engagement in the reflective task and improve their capacity to learn (6). Because reflection is contextual, assessment is needed to verify the authenticity of the significant events upon which the students have reflected (1, 5). However, Pee et al. (7) has suggested that the assessment of reflective writing has the potential to inhibit the development of students' reflective skills, since students may focus on obtaining high scores and neglect the reflective skills part of their training.

Koole et al. (8) highlighted two important aspects in the assessment of reflective writing. First, the content and process of reflection should be viewed as distinct entities. Focusing on the process will likely increase the objectivity of the assessment (6). Second, since reflection is closely related to the event that triggers it, a description of the triggering event should be included in the reflection to provide the assessors with a more objective frame of reference. Furthermore, both quantitative and qualitative approaches can be used to assess reflective writing. Based on a literature review by Plack et al. (9), the qualitative approach focuses only on the themes emerging from the writing, but does not show the level of reflection proficiency, whereas quantitative methods can be used to analyse the level of reflection (9). Two quantitative methods are commonly used, i.e., a questionnaire to assess reflective thinking ability, and a reflective rubric to assess the level of students' reflective writing (6).

Plack et al. (9) list several studies, each of which uses a particular instrument derived from a particular framework of reflection, including a study from Pee et al. (7) in which a structured worksheet, rather than a rubric, was used to assess the reflective writing of dental therapy students with good inter-rater agreement. A valid assessment instrument is important to minimise bias in assessing reflective practice, and should include a provision for feedback, which will ultimately promote better self-regulated learning (10–11). The inter-rater reliability assessment, which measures the correlation between two measurements, as well as the extent to which two or more raters agreed on certain measurements (12), is also essential, because the assessment of reflective writing with the aid of a rubric involves the expert judgement of the assessors.

Kember et al. (13) developed a reflective writing assessment rubric based on four levels of reflection, including no reflection, habitual action and critical reflection. In habitual action, students provide an answer to the prompt questions without any attempt to understand the underlying theory. Students might understand the theory, but it is not applied to any personal experiences or real-life applications. At the reflection level, students include their personal insights, and the theory is interpreted in the context of personal experiences. Critical reflection is the highest level of reflective practice, in which individuals obtain perspectives' on their transformation (14). By adapting the levels of reflection reported by Kember et al. (14), Wald et al. (15) developed a comprehensive analytic rubric called the Reflection Evaluation for Learners' Enhanced Competencies Tool (REFLECT). This instrument uses the same four levels of reflection developed by Kember et al. (14), but each level consists of more detailed criteria: writing spectrum, presence, description of conflict or disorienting dilemma, attending to emotions, analysis and meaning making, and attention to the assignment as the optional minor criteria (including how students address the assignment questions and provide rationale).

The details of the assessment criteria embodied within the two instruments described above differ in a number of ways. Wald et al.'s (15) instrument has more detailed descriptors for each scale compared to Kember et al.'s instrument, despite sharing a similar reflection framework.

Inter-rater reliability of a reflective writing assessment is an important factor in establishing the validity of the assessment process (14, 16). Hence, this study is aimed at examining the inter-rater reliability of the similar, but different, instruments developed by Kember et al. (14) and Wald et al. (15) in assessing reflective writing. Given the nature of reflection itself, assessing it may not be as straightforward as assessing skills in the cognitive or motor domains; however it is expected that our study will contribute to the discourse on the assessment of reflective ability in medical education.

METHODS

Context

This study was conducted in the Faculty of Medicine Universitas Indonesia. The medical school has a longitudinal course on empathy and professionalism for undergraduate medical students from their first to their third year of training. During the course, one of the students' tasks is reflective writing. The objective of reflective writing is for students to explore their lives or learning experiences in order to identify the attributes of empathy and professionalism they have acquired or need to attain. The reflective writing assignments assessed in this study were produced by second-year medical students in their empathy and professionalism course in response to the question, "Do I have empathy?". The students were asked to explore their experiences related to empathy and reflect on whether they have or do not have empathy. The 1,000-word-minimum reflective writing exercise was assessed for summative purposes, along with two other tasks, and this information was made available to the students through the course guidebook. Prior to taking part in the reflective writing exercise, general guidelines on what reflection was, how to develop reflective writing, and how it would be assessed were provided to the students in the form of lectures.

Design

This study was a cross-sectional, quantitative study to determine the degree of inter-rater agreement of reflective writing assessments. In order to determine the inter-rater agreement between two different instruments, this study used secondary data from reflective writing assignments obtained from the undergraduate medical professionalism course in our institution taught in 2017. All 175 submitted reflective assignments were eligible to be included in the analysis; however, after calculating the sample size, it was determined that only 63 anonymised reflective writing assignments were required, and these were randomly selected from the total number of reflective assignments. In 2019, two trained medical education experts familiar with the reflection concept, independently rated the anonymised reflective writing samples. Prior to the assessment, a discussion between the raters was held in order to agree on similar perceptions of the reflective writing level and the assessment instruments. The raters practiced beforehand by scoring together a sample of two reflective writing assignments. Any differences in scores were discussed until an agreement was reached. When assessing the reflective writing assignments using the REFLECT rubric, the raters followed the four steps as defined by Wald et al. (15) and also referred to the reflective narrative example as described in the published paper.

Instrument

This study used two reflective writing assessment rubrics. One developed by Kember et al. (14) and the other by Wald et al. (15), which are based on the same reflection framework, but differ in terms of structure and detailed content. The quality of reflection that can be captured using these rubrics spans from merely describing the experiences to obtaining new perspectives based on the experiences. The rubric from Kember et al. (14) (Rubric A, Table 1) consists of four reflection

levels: (a) non-reflection, (b) evidence of understanding of a concept or topic, (c) reflection, and (d) critical reflection. The rubric from Wald et al. (15) (Rubric B, Table 2) uses the same four levels as the above (non-reflection to critical reflection), with additional axes for the highest level (critical reflection), including transformative reflection and confirmatory learning. The Wald et al. (15) instrument also includes specific criteria for each level of reflection that the assessment should be based on. These criteria are (a) writing spectrum, (b) presence, (c) description of conflict or disorienting dilemma, (d) attending to emotions, (e) analysis and meaning making, and (f) optional minor criterion: attention to assignment. Furthermore, there are two components in Rubric B that incorporate the score of each criterion and the global score of the reflective writing level. The score of each criterion in Rubric B was also added to produce a total score for the reflective writing exercise. Notwithstanding the recommendation from the original developer to use REFLECT for formative purposes, the psychometric properties of REFLECT (15) still support the use of this instrument for summative assessment.

Table 1: Rubric A – Reflective writing rubric from Kember et al. (14)

Level of reflection	Description
Level 1: Non-reflection (score 1)	The answer shows no evidence of the student attempting to reach an understanding of the concept or theory which underpins the topic. Material has been placed into an essay without the student thinking seriously about it, trying to interpret the material, or forming a view. Largely reproduction, with or without adaptation, of the work of others.
Level 2: Evidence of understanding of a concept or topic (score 2)	Evidence of understanding of a concept or topic. Material is confined to theory. Reliance upon what was in the textbook or the lecture notes. Theory is not related to personal experiences, real-life applications or practical situations.
Level 3: Reflection (score 3)	Theory is applied to practical situations. Situations encountered in practice will be considered and successfully discussed in relationship to what has been taught. There will be personal insights which go beyond book theory.
Level 4: Critical reflection (score 4)	Evidence of a change in perspective over a fundamental belief of the understanding of a key concept or phenomenon. Critical reflection is unlikely to occur frequently.

Table 2: Rubric B – Reflective writing rubric from Wald et al. (15)

Criterion	Level				Axis II for critical reflection	
	Level 1: Habitual action (non-reflective) (Score 1)	**Level 2: Thoughtful action (introspection) (Score 2)**	**Level 3: Reflection (Score 3)**	**Level 4: Critical reflection (Score 4)**	**Transformative reflection and learning**	**Confirmatory learning**
Writing spectrum	Superficial descriptive writing approach (fact reporting, vague impressions) without reflection or introspection.	Elaborated descriptive writing approach and impressions without reflection.	Movement beyond reporting or descriptive writing to reflecting (i.e., attempting to understand, question, or analyse the event).	Exploration and critique of assumptions, values, beliefs, and/or biases, and the consequences of action (present and future).	Frames of reference or meaning structures are transformed. Requires critical reflection. Integration of new learning into one's identity, informing future perceptions, emotions, attitudes, insights, meanings and actions. Conveys a clear sense of a breakthrough.	Frames of reference or meaning structures are confirmed. Requires critical reflection.
Presence	Sense of writer being partially present.	Sense of writer being partially present.	Sense of writer being largely or fully present.	Sense of writer being fully present.		
Description of conflict or disorienting dilemma	No description of the disorienting dilemma, conflict, challenge, or issue of concern.	Absent or weak description of the disorienting dilemma, conflict, challenge, or issue of concern.	Description of the disorienting dilemma, conflict, challenge, or issue of concern.	Full description of the disorienting dilemma, conflict, challenge, or issue of concern that includes multiple perspectives, exploring alternative explanations, and challenging assumptions.		
Attending to emotions	Little or no recognition or attention to emotions.	Recognition but no exploration or attention to emotions.	Recognition, exploration, and attention to emotions.	Recognition, exploration, attention to emotions, and gain of emotional insight.		
Analysis and meaning making	No analysis or meaning making.	Little or unclear analysis or meaning making.	Some analysis and meaning making.	Comprehensive analysis and meaning making.		
Optional minor criterion: attention to assignment (when relevant)	Poorly addresses the assignment question and does not provide a compelling rationale for choosing an alternative.	Partial or unclear addressing of assignment question; does not provide a compelling rationale for choosing an alternative.	Clearly answers the assignment questions, or if relevant, provides a compelling rationale for choosing an alternative.	Clearly answers the assignment questions, or if relevant, provides a compelling rationale for choosing an alternative.		

Note on how to use the rubric: Assess the writing based on each criterion and provide the global score for the reflective writing level: (non-reflective (score 1), introspection (score 2), reflection (score 3), critical reflection (score 4); if it is a critical reflection, then decide on the second axis.

Analysis

To calculate the inter-rater reliability of assessing students' reflective writing with the aid of the two rubrics, a two-way mixed effect, single measure intra-class correlation (ICC) analysis was carried out, for which the model stipulates that the variations of measurement are random (17). The ICC was considered to be appropriate as a measure of reliability since it takes into account both correlation and agreement between measurements (12). According to Koo and Li (12), ICC values of less than 0.5 indicate poor reliability, values between 0.5 and 0.75 represent moderate reliability, values of 0.75–0.9 indicate good reliability. Values exceeding 0.9 indicate excellent reliability.

This study was exempt from ethical approval as it utilised anonymised secondary data and did not involve human subjects as participants. The assignments included in the analysis were taken from the course conducted in 2017 and the students have since completed their second-year professionalism course; thus, the results of the analysis will not impact their studies in professionalism.

RESULTS

Sixty-three reflective writing assignments were assessed using the two rubrics. The assessments made were purely based on the quality of the reflective writing.

The average reflective writing scores obtained from the two rubrics, and the ICC analysis are shown in Table 3. Reflective writing assessments using Rubric A showed the category of non-reflection (with an approximate mean score of 1) and when compared with the assessment results using the global score of Rubric B (with mean scores that were more evenly distributed).

Table 3: ICC coefficients of reflective writing assessment using two rubrics

Rubric	Mean reflective writing score		ICC coefficient (95% CI)
	Rater 1	Rater 2	
Rubric A: Kember et al. (14) (score between 1–4)	1.62	1.37	0.43 (0.21–0.61)*
Rubric B: Wald et al. (15)			
For the score of each criterion (score between 1–4)			0.50 (0.40–0.60)*
a. Writing spectrum	2.14	3.35	
b. Presence	1.41	2.49	
c. Description of conflict or disorienting dilemma	2.05	2.95	
d. Attending to emotions	2.19	2.67	
e. Analysis and meaning making	2.03	2.98	
f. Optional minor criterion: attention to assignment (when relevant)	2.06	2.84	
For the total score of all criteria (maximum score = 24)	11.89	17.29	0.50 (0.29–0.67)*
For the global score of the reflective writing level (score between 1–4)	1.81	2.48	0.36 (0.12–0.56)*

Note: *p-value < 0.05 which indicated significant differences with ICC = 0 (no agreement).

Based on the ICC analysis, the ICC coefficient was lower for Rubric A compared to that of Rubric B, when the analysis was conducted using the score of each criterion and the total score of all criteria, as seen in Table 3.

The above ICC coefficients, based on guidelines for their interpretation by Koo and Li (12), indicate poor reliability when the assessment was conducted using Rubric A. Whereas moderate reliability was shown in the assessments using Rubric B (for the score of each criterion and the total score of criteria). However, if the assessment was conducted using Rubric B, through selected of the most appropriate overall level (global score of the reflective writing level), the coefficient dropped to below 0.5, which indicates poor reliability.

Despite the low ICC coefficients, it is important to note that the 95% confidence interval (CI) was quite wide for each of the ICC coefficients. Specifically, for Rubric B, although the reliability was moderate when using the ICC coefficient criteria as the basis of assessment, the width of the confidence intervals (95% CI) increased when the score of all criteria was added to produce a total score, indicating a low level of accuracy.

DISCUSSION

Studies have shown that professionalism can be nurtured with the support of elective learning (5, 18–20). Therefore, an assessment of a student's reflective ability is an important part of the educational process, and can be achieved with the use of a valid and reliable rubric.

The findings in the current study revealed that inter-rater reliability, as measured by ICC coefficients, was poor to moderate. The ICC coefficient for the more elaborate and detailed rubric (Rubric B) was higher than that of the more generic one (Rubric A). The REFLECT rubric developed by Wald et al. (15) was based on the four levels of reflection described by Kember et al. (14), with the addition of specific criteria for each level of reflection. The availability of appropriate criteria, along with their description, is likely to assist each rater in providing a more precise score. It is also likely that inter-rater reliability will increase when raters are provided with a comprehensive understanding of the reflection concept that is being assessed. Lucas et al. (21) developed a rubric based on a similar framework of reflection, and when they assessed the students' reflective writing with their own rubric, they found a very good ICC coefficient. However, the rubric developed by Lucas et al. (21), to the best of the authors' knowledge, has only been applied in the context of undergraduate pharmacy education.

The original developer of the REFLECT rubric reported an ICC coefficient of 0.632 (15), whereas a study by Moniz et al. (22) found an ICC of 0.457. Ottenberg et al. (23) slightly modified the REFLECT rubric and used it to assess students' reflective writing on professionalism in gross anatomy, resulting in an ICC of 0.68. The current study demonstrated an ICC of 0.50 when using the REFLECT rubric, which is comparable to that of other studies. The differences in ICC across a range of studies may be due to variability in the raters' ability, as well as the context in which the reflection took place. It is also important to note that, in the current study, when the raters were asked to provide a global rating score using the REFLECT rubric, the ICC decreased. This result further confirms the need for specific and detailed criteria for assessing the levels of reflection in order to produce reliable scores.

The problem of low reliability in the assessment of reflective writing may stem from the difficulty of operationalising the concept of reflection itself and, more fundamentally, the question of whether reflection is a necessary component of the education process and can be effectively assessed. The various definitions of reflection may be well known, but the most effective way to translate understanding

into a practical method of teaching and assessing reflection remains a challenge (24). Another challenge is the concept of what de la Croix and Veen (25) characterised as a "reflective zombie", that is, someone who is faking reflection. The writing sample produced by such an individual might fulfill all the criteria of critical reflection, but the student in question is not actually engaged in any reflecting. The authors argue that one of the causes of the reflective zombie is that reflective writing is assessed using a checklist containing a list of criteria of acceptable levels of reflection, which has encouraged a "socially desirable" form of reflective writing. Despite the challenges in assessing reflection, assessment might still be necessary to facilitate students' learning and practice.

Aside from the moderate inter-rater reliability found in our study, which was not included in its main objectives, the current study also found that more than 50% of the reflective writing was categorised as non-reflection. This finding is in line with that of Gadbury-Amyot et al. (26), who found that more than 50% of the reflective writing they assessed was at the low level of reflection. Based on these results, several assumptions can be made. The first assumption is that medical students, who were high-school leavers, have not yet acquired the ability to reflect. A low level of reflection skills was also identified in a study in Indonesian dental students with similar characteristics, as was found to be caused by multiple factors, one of which was the poor understanding of the concept of self-reflection (27). Moreover, this study was conducted among second year undergraduate students, thus different results may have occurred if the assessment had been carried out on the reflection of more senior students. In particular, those in the clinical years who have received more extensive training in reflective writing. In this case, such a finding emphasises the importance of continuously monitoring and assessing the students' ability to reflect, as assessment has been found to drive the learning process (28). Hence, the use of reliable rubrics may aid medical teachers in accessing their students' reflective ability. The second assumption is that these results may not reflect the true ability of the students. More writing samples may be necessary to adequately represent students' ability to reflect, as argued by Moniz et al. (22) who found that as many as 14 reflective writing samples from each student were required to obtain good reliability.

The current study also highlights the need to select appropriate rubrics for assessing reflective writing. The rubrics can be based on different frameworks of reflection, as long as the framework is comprised of an analytic or detailed description for each component of the reflection being assessed. It has been shown that such an approach is supported by other studies, that rubrics based on an analytic framework can improve inter-rater reliability.

This study is not without limitations. First, the focus of the study was only on quantitative scoring with the use of rubrics. Further studies could expand the scope of the research to include analysis on narrative feedback and the inclusion of other institutions. Second, the study was conducted at a single institution; thus, generalising our findings may have limited validity. The ICC model used in this study also limits the generalisability of the results given that the raters were not randomly selected. However, the findings in this study confirm those of other studies and have further elucidated the issues involved in the assessment of reflective writing. The moderate reliability of reflective writing assessments calls for further consideration for the most appropriate way to use reflective writing as an assessment tool, including examining the educational impact of assessing reflective writing with specific checklists. It is likely that the assessment of reflection practices by a rubric alone is not sufficient, and needs to be complemented with constructive feedback.

Observing students' reflective practices in the workplace, while creating a supportive and nurturing environment, are perhaps more reliable strategies for equipping students with positive reflective skills (25). Furthermore, prior to establishing the use of rubrics in assessing reflective writing, it will be important to equip raters with in-depth knowledge about the rubric employed and proper training to assess reflective writing based on the criteria in the rubric, in order to achieve high inter-rater reliability (6).

CONCLUSION

This study has shown the potential of using a more detailed (analytic) rubric to assess medical students' reflective writing in order to produce a relatively higher inter-rater reliability, although the reliability of the rubric was still categorised as moderate. Good reliability is one of the criteria of a valid assessment. Several approaches should be explored to formulate the best possible method to assess the reflective ability of medical students, including the ability to provide valid and reliable results, with the expected impact on their education.

ACKNOWLEDGEMENTS

This study was presented in the 12th Jakarta Meeting on Medical Education, 18th to 20th October 2019.

REFERENCES

1. Sandars J. The use of reflection in medical education: AMEE guide no. 44. Med Teach. 2009;31:685–95. https://doi.org/10.1080/01421590903050374

2. Bernard AW, Gorgas D, Greenberger S, Jacques A, Khandelwal S. The use of reflection in emergency medicine education. Acad Emerg Med. 2012;19(8):978–82. https://doi.org/10.1111/j.1553-2712.2012.01407.x

3. Chen I, Forbes C. Reflective writing and its impact on empathy in medical education: systematic review. J Educ Eval Health Prof. 2014;11:20. https://doi.org/10.3352/jeehp.2014.11.20

4. Arntfield SL, Slesar K, Dickson J, Charon R. Narrative medicine as a means of training medical students toward residency competencies. Patient Educ Couns. 2013;91(3):280–6. https://doi.org/10.1016/j.pec.2013.01.014

5. Aronson L. Twelve tips for teaching reflection at all levels of medical education. Med Teach. 2011;33:200–5. https://doi.org/10.3109/0142159X.2010.507714

6. Tsingos C, Bosnic-Anticevich S, Lonie JM, Smith L. A model for assessing reflective practices in pharmacy education. Am J Pharm Educ. 2015;79(8): Article 124. https://doi.org/10.5688/ajpe798124

7. Pee B, Woodman T, Fry H, Davenport ES. Appraising and assessing reflection in students' writing on a structured worksheet. Med Educ. 2002;36:575–85. https://doi.org/10.1046/j.1365-2923.2002.01227.x

8. 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 Med Educ. 2011;11:104. https://doi.org/10.1186/1472-6920-11-104

9. Plack MM, Driscoll M, Blisset S, McKenna R, Plack TP. A method for assessing reflective journal writing. J Allied Health. 2005;34:199–208.

10. Koole S, Vanobbergen J, De Visschere L, Aper L, Doman T, Derese A. The influence of reflection on portfolio learning in undergraduate dental education. Eur J Dent Educ. 2012;1–7. https://doi.org/10.1111/**j**.1600-0579.2012.00766.x

11. Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. Stud High Educ. 2006;31(2):199–218. https://doi.org/10.1080/03075070600572090

12. Koo TK, Li MY. A guideline of selecting and reporting intra-class correlation coefficients for reliability research. J Chiropr Med. 2016;15:155–63. https://doi.org/10.1016/**j**.jcm.2016.02.012

13. Kember D, Leung DYP, Jones A, Loke AY, McKay J, Sinclair K, et al. Development of a questionnaire to measure the level of reflective thinking. Assessment & Evaluation in Higher Education. 2000;25(4):381–95. https://doi.org/10.1080/713611442

14. Kember D, McKay J, Sinclair K, Wong FKY. A four-category scheme for coding and assessing the level of reflection in written work. Assess Eval High Educ. 2008;33(4):369–79. https://doi.org/10.1080/02602930701293355

15. 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. https://doi.org/10.1097/ACM.0b013e31823b55fa

16. Downing SM. Validity: on the meaningful interpretation of assessment data. Med Educ. 2003;37:830–7. https://doi.org/10.1046/j.1365-2923.2003.01594.x

17. Field A. Discovering statistics using SPSS. 2nd ed. London: Sage Publications; 2009.

18. Pololi L, Frankel RM, Clay M, Jobe AC. One year's experience with a program to facilitate personal and professional development in medical students using reflection groups. Educ Health. 2001;14(1):36–49. https://doi.org/10.1080/13576280010015074

19. Langendyk V, Mason G, Wang S. How do medical educators design a curriculum that facilitates student learning about professionalism? Int J Med Educ. 2016;7:32–43. https://doi.org/10.5116/ijme.5683.c2e0

20. Smith L, Adam L, Moffat S, Meldrum A, Ahmadi R. How do educators in one New Zealand undergraduate Bachelor of Oral Health course teach and nurture professionalism? Eur J Dent Educ. 2017;22:e212–20. https://doi.org/10.1111/eje.12274

21. Lucas C, Bosnic-Anticevich S, Schneider CR, Bartimote-Aufflick K, McEntee M, Smith L. Inter-rater reliability of a reflective rubric to assess pharmacy students' reflective thinking. Currents in Pharmacy Teaching and Learning. 2017;9(6):989–95. https://doi.org/10.1016/j.cptl.2017.07.025

22. Moniz T, Arntfield S, Miller K, Lingard L, Watling C, Regehr C. Considerations in the use of reflective writing for student assessment: issues of reliability and validity. Med Educ. 2015;49:901–8. https://doi/org/10.1111/medu.12771

23. Ottenberg AL, Pasalic D, Bui GT, Pawlina W. An analysis of reflective writing early in the medical curriculum: the relationship between reflective capacity and academic achievement. Med Teach. 2016;38:724–9. https://doi.org/10.3109/0142159X.2015.1112890

24. Kalk K, Luik P, Taimalu M, That K. Validity and reliability of two instruments to measure reflection: a confirmatory study. TRAMES. 2014;18(2):121–34. https://doi.org/10.3176/tr.2014.2.02

25. de la Croix A, Veen M. The reflective zombie: problematizing the conceptual framework of reflection in medical education. Perspect Med Educ. 2018;7:394–400. https://doi.org/10.1007/s40037-018-0479-9

26. Gadbury-Amyot CC, Godley LW, Nelson Jr. JW. Measuring the level of reflective ability of predoctoral dental students: early outcomes in an e-portfolio reflection. Eur J Dent Educ. 2019;83(3):275–80. https://doi.org/10.21815/JDE.019.025

27. Greviana N, Mustika R, Soemantri D. Development of e-portfolio in undergraduate clinical dentistry: how trainees select and reflect on evidence. Eur J Dent Educ. 2020;24:320–7. https://doi.org/10.1111/eje.12502

28. Van der Vleuten CPM, Schuwirth LWT, Driessen EW, Dijkstra J, Tigelaar D, Baartman LKJ, et al. A model for programmatic assessment fit for purpose. Med Teach. 2012; 34:205–14. https://doi.org/10.3109/0142159X.2012.652239