ORIGINAL RESEARCH

Construct Validity of an Instrument for Assessment of Reflective Writing-Based Portfolios of Medical Students

Salah Eldin Kassab [1]
Mubarak Bidmos [1]
Michail Nomikos [1]
Suhad Daher-Nashif [2]
Tanya Kane [2]
Srikant Sarangi [3]
Marwan Abu-Hijleh [1]

[1]Department of Basic Medical Sciences, College of Medicine, QU Health, Qatar University, Doha, Qatar; [2]Department of Population Medicine, College of Medicine, Qatar University, QU Health, Doha, Qatar; [3]Danish Institute of Humanities and Medicine (DIHM), Aalborg University, Aalborg, Denmark

Purpose: Assessment of reflective writing for medical students is challenging, and there is lack of an available instrument with good psychometric properties. The authors developed a new instrument for assessment of reflective writing-based portfolios and examined the construct validity of this instrument.

Methods: After an extensive literature review and pilot testing of the instrument, two raters assessed the reflective writing-based portfolios from years 2 and 3 medical students (n=135) on three occasions. The instrument consists of three criteria: organization, description of an experience and reflection on the experience. We calculated the reliability of scores using generalizability theory with a fully crossed design and two facets (raters and occasions). In addition, we measured criterion validity by testing correlations with students' scores using other assessment methods.

Results: The dependability (Φ) coefficient of the portfolio scores was 0.75 using two raters on three occasions. Students' portfolio scores represented 46.6% of the total variance across all score comparisons. The variance due to occasions was negligible, while the student–occasion interaction was small. The variance due to student–rater interaction represented 17.7%, and the remaining 27.7% of the variance was due to unexplained sources of error. The decision (D) study suggested that an acceptable dependability (Φ = 0.70 and 0.72) can be achieved by using two raters for one and two occasions, respectively. Finally, we found moderate to large effect-size correlations between students' scores in reflective writing-based portfolios and communication skills (r = 0.47) and PBL tutorials (r = 0.50).

Conclusion: We demonstrated the presence of different sources of evidence that support construct validity of the study instrument. Further studies are warranted before utilizing this instrument for summative assessment of students' reflective writing-based portfolios in other medical schools.

Keywords: G-theory, validity, reflective writing, portfolio, reliability, student assessment

Introduction

Since the evolution of competency-based education, there has been widespread use of reflective portfolios and reflective writing in medical curricula. Effective use of portfolios helps to enhance students' ability to reflect, provides evidence of their personal and professional development and promotes their critical thinking and communication skills.[1] A portfolio is a collection of material that is used as an evidence of achieving learning outcomes over a period of time. There are several types of portfolios, which vary according to the purpose and setting of use.[2]

Correspondence: Salah Eldin Kassab
Physiology & Medical Education, College of Medicine, QU Health, Qatar University, PO Box 2713, Doha, Qatar
Tel +974 – 4403 7843
Email skassab@qu.edu.qa

Usually, portfolios include the requirements for learners to write reflections on their learning experience using short reflective pieces. Furthermore, the quality of reflection appears to be the most significant contribution in explanation of the variance of regular portfolio ratings.[3] Therefore, the current study focuses on reflective writing-based portfolios, where reflection on a learning experience is the primary component of the students' portfolio along with providing the evidence to support their described experience.

Reflection is generally understood as a metacognitive process that aims to develop critical understanding of both the self and the situation, which can be transferred to inform situated encounters in the future.[4] Four necessary conditions have been identified for successful implementation of reflective portfolios: good coaching; structure and guidelines; adequate experiences and material for reflection; and summative assessment.[5]

In what follows, we offer a review of relevant literature organized into three sections: (i) theoretical underpinnings of reflection; (ii) studies describing reflection in medical education settings; and (iii) studies devoted to assessment of reflective writing.

Theoretical Underpinnings of Reflection
Several scholars have conceptualized the theoretical underpinnings of reflection by proposing different explanatory models of learning.[6–10] Dewey[6] has demonstrated that reflection is an active and deliberate learning process that makes sense of situations or events that are difficult to explain. He argues that reflective thinking transforms a situation from an experience of perplexity and ambiguity into a balanced state of clarity, coherence, settlement and harmony.

Schön[7,8] is the first to link reflection to professional development and practice. For Schön,[7,8] reflection is a process that makes the hidden theoretical knowledge more explicit and transforms it into practical knowledge, ie, reflection enables professionals to improve their practice and become progressively more experts in their areas. According to him, there are two types of reflections, which are triggered during professional practice: "reflection in action" and "reflection on action". While "reflection in action" involves the awareness of the situation and the use of professional knowledge on the spot to plan for contingent situational changes, "reflection on action" involves retrospectively visiting the experience and building on it.[7,8]

Boud et al[9] emphasized the importance of emotions in reflective thinking, which influence the ways in which individuals recall events. For these authors, reflection is an iterative process of effective learning that begins with personal experiences. Learners are encouraged to go back, revisit their personal experiences, and evaluate the values and beliefs underlying specific actions and decisions. In the last step (outcomes), new perspectives of the experience are generated that lead to commitment to action and change in behavior.[9] Among others, Moon[10] describes reflection as a stimulus for transforming the superficial knowledge into deep knowledge. He defines reflection as

> a form of mental processing with a purpose and/or anticipated outcome that is applied to relatively complex or unstructured ideas for which there is not an obvious solution.[10]

Studies Describing Reflection in Medical Education Settings
In her empirically grounded monograph, Locher[11] systematically compares reflective writings from experts and medical students and identifies the recurrence of a typical "description-reflection-conclusion/aims" format, complemented by a number of textual features. The use of reflections for personal development involves students examining their own values, beliefs and assumptions.[12] The understanding of the person's values and beliefs is essential for developing a therapeutic relationship with the patients, which is essential for empathy and caring of these patients.[4] Several authors have demonstrated that developing reflective skills in medical education improves diagnostic reasoning,[13–15] communication skills, collaboration and empathy,[16,17] professional identity,[18] and development of expertise.[4,14,15] The role of mentors, whether a faculty member or a peer, is essential for scaffolding the reflection by medical students. A mentor will provide the supportive environment for reflection by facilitating the awareness and making sense of an experience.[4]

Studies Devoted to Assessment of Reflective Writing
Despite the potential usefulness of reflection in medical practice, there are contradicting findings in the current literature regarding the psychometric properties of instruments for measuring this construct.[19,20] Wald et al[19] developed an analytic rubric for scoring reflective writing and called it the Reflection Evaluation for Learners' Enhanced

398 submit your manuscript | www.dovepress.com
DovePress
Advances in Medical Education and Practice 2020:11

Competencies Tool (REFLECT). The rubric consisted of four reflective capacity levels: habitual action, thoughtful action, reflection, and critical reflection. They demonstrated adequate interrater reliability, face validity, feasibility, and acceptability of the rubric.[19] However, the rubric was recommended for formative evaluation of students and as a guide for faculty feedback to students. In contrast, another study casted doubt on the reliability of reflective writing scores of undergraduate medical students using the same rubric.[20] These authors demonstrated that at least 14 reflective essays assessed by four or five raters were required to achieve acceptable reliability. They also demonstrated a non-significant correlation between reflective writing scores and scores using other assessment measures such as multiple choice questions and Objective Structured Clinical examinations (OSCEs).[20]

Against this backdrop, we designed the current study to develop an instrument for evaluation of portfolios with a focus on reflective writing. We also aimed to assess the different sources of evidence that support the construct validity of the study instrument.[21] The sources of validity evidence include the content-related evidence, internal structure by measuring the generalizability of scores, and relations to other variables by testing the correlations of reflective writing-based portfolios with written examination scores (divergent validity) and communication skills scores (convergent validity).[21] Primarily, the study aims to answer the following research questions:

1. To what extent can we generalize medical students' scores in reflective writing-based portfolio scores across raters and occasions?
2. What is the relationship between students' scores in reflective writing-based portfolios and their scores in written examinations and communication skills?

Methods
Design and Study Setting
We conducted this study at the College of Medicine, Qatar University (CMED-QU). The undergraduate program is of six years duration, divided into three phases: 1) Phase I (one year) is traditional, course-based, 2) Phase 2 (two and half years) is integrated, problem-based, and Phase 3 (two and half years) is hospital-based clinical rotations. During all phases of the program, reflective portfolio is a core-learning tool for medical students. This specific study involved year 2 (n=67) and year 3 (n = 68) medical

students during their study in Phase II for the academic years 2018/2019 and 2019/2020. The study has received Research Ethics approval No. QU-IRB 697-E/16 issued by Institutional Review Board, Office of Academic Research, Qatar University.

Students submit their reflective portfolios at the end of each semester. In each portfolio, students are expected to submit three writing entries to demonstrate their ability to describe and reflect on their learning experience related to three of the six curriculum competency domains. The competency domains include 1) patient care and clinical skills, 2) population health, 3) knowledge for practice, 4) interpersonal communication and collaboration, 5) personal development & professionalism, and 6) research. By the end of the second semester, students are required to reflect on the experiences related to the other three competency domains of the program. In addition to their reflection entries, students provided an evidence to support their described experience. Students were not provided with prompts for their reflections, but were rather given the liberty to reflect on their personal experiences. Students received two hours of training about the portfolios during their study of a "Health Professions Education" course. In addition, they attended a supplementary workshop in year 2 on how to use the portfolio as a learning tool and how they are assessed at the CMED program. Before submitting their portfolio for summative assessment, students were strongly encouraged to review their portfolios with their mentors, who provide formative assessment for them. Students were also provided with a supplementary guide describing the purpose and benefits of the portfolio, expected learning outcomes, how to describe and reflect on experiences related to different competency domains of the curriculum, and the assessment instrument.

Development of the Study Instrument
Based on a systematic review of relevant literature, the authors developed the reflective portfolio-scoring rubric. Two rather different scholarly traditions guided the development of the study instrument: 1) the concept of reflection and its role in professional education/development; and 2) the linguistic/rhetorical manifestations of reflection in writing. Schön's[7] characterization of "knowing-in-action" comes closer to our conceptualization, especially his distinction between "reflection in practice" and "reflection on practice". The reflective writing portfolio in the context of the medical curriculum falls within the latter but

with a future orientation – although there are bound to be elements of the former during the writing process. The study instrument is based on three main premises: 1) reflection is an iterative process used for learning from revisiting and analyzing a previous experience;[22] 2) reflection is triggered by the presence of a complex, unrecognized problem;[7,14] and 3) recognition of the boundaries between description of an experience and deeper levels of reflection.[19,23] Although the primary aim in developing the instrument was to produce a streamlined template that would assist faculty raters in the assessment process, it was important to bear in mind its comprehensibility and usefulness as far as medical students were concerned in terms of actionable feedback.

The final instrument (Appendix 1) was refined after pilot testing in a training workshop with faculty members at CMED-QU. During the workshop session, participants were introduced to the process of portfolio assessment and reflective writing. Participants (n=18) were then divided into two groups (A and B), and each member of a group was requested to grade two samples of students' portfolio (one each from Years 2 and 3) using the study instrument. Furthermore, faculty were provided with a questionnaire to indicate their degree of agreement, on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree), with five items related to the instrument. The evaluation items were: 1) the instrument relates to the learning outcomes of the portfolio, 2) assessment criteria are clearly defined and accurate, 3) the descriptors accurately describe each level of performance, 4) the instrument will be useful for providing feedback to students, and 5) language is clear with no ambiguity. Results of the questionnaire indicated 78.8% agreement (strongly agree or agree), 5.9% disagreement (strongly disagree or disagree) and 15.3% were neutral. In addition, faculty members provided qualitative comments for improvement of the instrument. The results of the questionnaire were discussed in the workshop and final refinement of the instrument was done accordingly.

Faculty raters assessed each reflective portfolio along three main criteria: organization and quality of presentation, description of a personal experience and reflection on that experience. Subsequently, assessment of reflection included three main criteria: 1) critical awareness of self and others, 2) experiential knowledge, and 3) ability to identify and manage uncertainty and ambivalence. The assessment of students' writing was based on the above-mentioned rubric with a three-point scale (0, 1 and 2). The total score (out of 10) was distributed mainly on description and reflection of personal experience (8 points) and only allocated a maximum of 2 points on organization and quality of presentation. Faculty raters received training workshops on the assessment of student portfolios with a focus on using the study instrument. The final score for students in the portfolio was the mean of the two raters' scores. The final score of the portfolio represented 10% of the summative scores of the study units.

Data Analysis
Generalizability of Reflective Writing-Based Portfolio Scores

We measured the reliability of the reflective writing-based portfolio scores using generalizability theory (G-theory) analysis, which includes both a generalizability study (G-study) and a decision study (D-study). The G-study calculated the variance attributed to the facets of the study (raters and occasions), while the D-study predicts the most favorable mix of raters and occasions that are needed to attain an acceptable reliability. The details of the methods used and the equations have been reported in previous studies.[24–27]

In the G-theory analysis, we selected a fully crossed design because the same raters assessed all students on the three study occasions. In addition, the study facets were considered random because we were interested in generalizing the study findings beyond the present study settings. The G-theory analysis calculates the variances attributed to differences between students, differences between ratings of portfolio assessors (two raters), and differences across occasions (three occasions). Furthermore, it allows the calculation of the variance due to interactions between students' scores, occasions and raters. For the D-study, we calculated the dependability (Φ) coefficient because we were interested in absolute performance of students without comparison to other students' scores in reflective portfolios (criterion-referenced assessment).

We have also analyzed the standard error of measurement (SEM), which is a measure of the spread of the scores for a single student if he/she is tested multiple times. Therefore, it is helpful in determining the degree of precision with which student measurements are made using the instrument in a certain way (ie two raters and three occasions).[24]

Relations Between Reflective Portfolio Scores and Scores in Other Areas of Competence (Criterion-Related Evidence of Validity)

The relationships between the students' scores in reflective writing-based portfolios and their scores measured by

other assessment tools such as MCQs, communication skills in OSCE, and evaluations of students by PBL tutors were ascertained using the Pearson's Product-moment correlation coefficient. Based on the previous research findings of the link between the reflective writing of medical students and development of communication skills, collaboration and empathy,[16,17] we hypothesized that scores in reflective writing-based portfolios correlate with students' scores in communication skills and PBL tutorials (convergent validity), but not with their scores in knowledge-based examinations (divergent validity). Assessment of knowledge is based on MCQs (from 60 to 120 items depending on the unit) of the A-type, mostly with context-rich scenarios.

The faculty facilitators at the end of each system-based unit assessed students' performance in PBL tutorials in phase II of the program after longitudinal exposure with students within a range from 8 to 20 sessions, depending on the length of the unit. Evaluation criteria include items related to professionalism such as accountability (eg being punctual, exhibiting leadership, trying one's best to complete assigned objectives), displaying respect & integrity (ie respecting group members, admitting mistakes, providing and accepting constructive feedback, establishing rapport with the group), and communication (eg expresses opinions well, does not interrupt group discussion, uses proper body language). In addition, students are evaluated on their participation in group dynamics and generation of learning objectives. Each item is assessed based on a scale of 1 (Very poor) to 10 (Excellent) and the total PBL scores represent 10% of the summative end-unit evaluation.

Assessment of communication skills during the OSCE is based on direct observation of performance while communicating with a standardized patient in addition to a Clinical Multimedia-based Exam for Diagnosis and Decision-making (CMEDD). The CMEDD is a computer-based test which includes series of video-recorded encounters in clinical settings where students are asked about an aspect that relates to the clinical encounter.

All the study data were analyzed using IBM SPSS Statistics for Windows Version 26.0 (IBM Corp., Armonk, NY, USA). The G-theory analyses were conducted using the G1.sps program as previously described.[28] A p-value of <0.05 was considered statistically significant.

Results
Generalizability of Reflective Writing-Based Portfolio Scores

The G-study indicated an acceptable level of reliability (Φ = 0.75) of reflective portfolio scores using 2 raters across 3 occasions (Table 1). The percentage of variance imputed to the object of measurement (students) is 46.6% of total variance. Since this is the highest variance, it indicates that raters were able to discriminate to a large extent between students' quality levels of their reflective portfolios. On the other hand, the estimated variance component for raters accounted for a negligible percentage (0.8%) suggesting that the raters' scores did not vary across students and occasions. However, the interaction between students and raters accounted for 17.7% of the total variance, indicating that the raters' assessments of certain students vary to a significant extent. The facet of occasion contributed a -negligible percent of variance (1.9%) to the model, suggesting very low fluctuations in the overall ratings of students from one occasion to the next. Furthermore, the small percentage of variance (5.2%) attributed to the interaction between students and occasions suggests that the scores of students did not change significantly across occasions and there were small changes in rating behavior across occasions. Finally, the interaction among students, raters and occasions represented 27.7% of total variance. This large component represents both the variance imputed to the three-way interaction and the residual variance imputed to

Table 1 Generalizability Theory Study (G-Study) Results for the Scores of Medical Students (n=129) in Reflective Portfolios Using Two Raters and Three Measurement Occasions

Facets	df	SS	MS	Variance Component	Percent Variance
Students (s)	128.00	501.24	3.92	0.49	46.60%
Raters (r)	1.00	4. 17	4.17	0.01	0.80%
Occasions (t)	2.00	11.27	5.63	0.02	1.90%
s× r	128.00	109.25	0.85	0.19	17.70%
s× t	256.00	102.99	0.40	0.06	5.20%
r× t	2.00	0.44	0.22	0.00	0.00%
s× r × t	256.00	74.96	0.29	0.29	27.70%
Φ = 0.75					
Absolute SEM: 0.41					

Note: The proportion of observed variance explained by each facet is calculated by dividing the individual variance component by the total observed variance.
Abbreviations: SS, sum of squares; MS, mean of squares; *df*, degree of freedom; SEM, standard error of measurement.

facets which are not included in the current study. Finally, the SEM for the study model using two raters and 3 occasions was 0.41 resulting in a confidence interval of ± 0.80.

Results of the D-Study

Figure 1 illustrates the results of the decision (D) study, which predicted reliability of the instrument by using different combinations of raters and occasions. If we use one rater, even increasing the number of occasions to five achieves a dependability coefficient (Φ) of only 0.66, which is below the acceptable level. Using two raters, however, leads to increasing levels of dependability to an acceptable level (Φ = 0.72) across two occasions and to Φ = 0.75 across three occasions. To achieve a good level of reliability (Φ = 0.80), results demonstrate that we need three raters on three occasions. As illustrated in Figure 1, the effect of increasing the number of raters on reliability is considerably higher compared with increasing the number of occasions. In fact, adding a fourth, or fifth occasion led to little improvements in the overall dependability.

Relations Between Reflective Portfolio Scores and Scores in Other Areas of Competence (Criterion-Related Evidence of Validity)

In order to evaluate the criterion-related validity evidence of the instrument, we tested the relationship between the scores of students in reflective portfolios with their scores in written (MCQs) examinations, communication skills in objective structured clinical examinations (OSCE) and PBL tutorials. There were moderate to large positive effect size correlations between the communication skills and

Figure 1 Decision study (D-study) for the medical students (n=129) scores in reflective portfolio. The dependability coefficients are the estimated Φ-coefficients of combining different numbers of raters and occasions.

PBL scores and reflective portfolio scores (r = 0.47 and 0.50, respectively, P < 0.01), but with small effect size correlation (r = 0.28) with written MCQ examination.

Discussion

This study demonstrates different sources of evidence that support the construct validity of the study instrument. Content-related evidence is supported by the theory-informed construction of the study instrument, training of faculty raters and pilot testing of the instrument with faculty. The evidence for an acceptable internal structure of the instrument is demonstrated by the G-theory analysis. The study demonstrated that measuring students' portfolio scores when using two raters can achieve acceptable levels of reliability (Φ = 0.72 and 0.75) on two and three occasions, respectively. Finally, the large effect-size correlations between scores in reflective portfolios and scores in both communication skills and PBL tutorials support the evidence of convergent validity (relations to other variables). These findings suggest that the study instrument exhibits acceptable psychometric properties to be used for summative assessment of medical students in reflective portfolios. Other studies, which used G-theory analysis, proved much lower reliability of reflective portfolio scores of medical students.[3,19,20,29] The difference in findings could be related to the content of the study instrument, level of training of raters and sample size used. Further studies in other medical programs are warranted to examine the reliability and validity of the study instrument beyond the current study setting.

The 46.6% variance for the subject of measurement indicates that, averaging over raters and occasions, medical students differed systematically in their reflective portfolio scores. This finding suggests an acceptable degree of variability in faculty ratings of student reflective portfolios due to unsystematic sources of error. The large variance attributable to differences among students than other facets of the study could be explained by the training of faculty assessors and their familiarity in using the study instrument. Another study reported a much lower percent of variance (25%) due to year 1 student's performance using an analysis of two portfolios.[29] They explained the low variance, and the overall low G-coefficient, due to lack of training of students and assessors.

The study findings indicate that the facet of occasion contributed to only 1.9% of the variance, suggesting very low temporal fluctuations in the raters' scores. In addition,
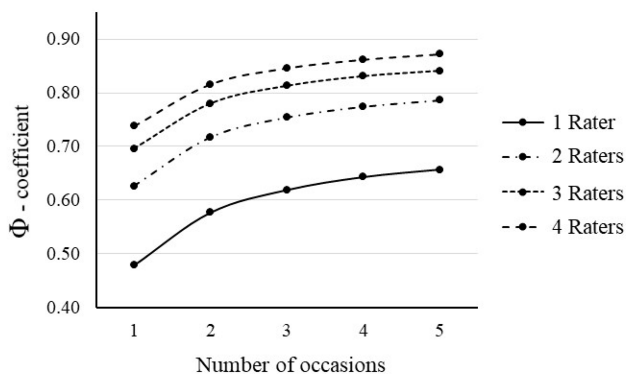
the interaction between students and occasions contributed to 5.2% of variance, suggesting that the students' scores in reflective portfolios did not change considerably across occasions and there were little changes in rating behavior across occasions. Rees et al[29] reported a much higher percent of variance (69.2%) due to student–occasion interactions. Finally, the large source of variance (27.7%) reflected by the interactions between students, raters and occasions, suggests that a significant proportion of the variability is caused by facets not included in the study or by a random error. This unexplained error could be due to variance related to the evaluations items or the study setting.

The current findings of the D-study demonstrated that increasing the number of raters from one to two over two occasions resulted in increasing levels of reliability ranging from G=0.58 to 0.72, respectively. Even increasing the number of raters to three can achieve an acceptable level of reliability ($\Phi = 0.70$) on one occasion. This clearly demonstrates that increasing the number of raters has more impact on reliability of reflective portfolio scores than increasing the number of occasions. Because we assess the students' reflective portfolios at the end of the semester (one occasion), using three reflective entries, it will be recommended to use three raters in order to achieve an acceptable reliability in our study context. The requirement for more than two raters to achieve higher reliability may, however, pose practical constraints in terms of human resource utilization.

In the current study, students reflected on the six medical curriculum competency domains. The advantage of this model is addressing one of the main problems previously reported in the literature,[30] by offering an educational structure of integration of the portfolios within the curriculum. This model also provides a broader scope of reflection on essential competency domains required for any medical graduate. It provides different and meaningful experiences for reflection by students, which have been reported as key factors for success of the portfolio.[3] Furthermore, the students were given the freedom to reflect on personal experience rather than providing them with reflective prompts, which have been previously shown to restrict the ability of the students to engage in reflective writing.[31]

Study Limitations and Future Directions

This study has some limitations that warrant reporting. The design of the study was restricted to year 2 and 3 medical students in a problem-based curriculum. Therefore, future studies are needed to test the replication of the study findings in different years of study, in other educational settings such as the clinical environment, and in other cultures. Although the study instrument has proved acceptable psychometric properties, further refinement of the reflection construct is required in future studies. The study instrument focuses on measuring the outcome of reflection, which may not capture important dimensions of this rich construct. Further studies should focus on developing instruments for measuring both the "process" and the "outcome" of reflection. Finally, the effectiveness of mentors on the quality of students' reflective-writing based portfolios requires further investigation.

Conclusions

This study provides an evidence of acceptable reliability and validity of an instrument to be used for summative assessment of students' reflective portfolios in undergraduate medical programs. An acceptable Φ-coefficient value (≥ 0.7) could be achieved by having two raters scoring students over two occasions or three raters on one occasion. Further studies are warranted to reproduce these findings before utilizing it for summative assessment of students in other medical schools.

Acknowledgment

Disclosure

The authors report no conflicts of interest in this work.

References

1. Driessen EW, Van Tartwijk J, Van Der Vleuten C, Wass V. Portfolios in medical education: why do they meet with mixed success? A systematic review. *Med Educ*. 2007;41:1224–1233. doi:10.1111/j.1365-2923.2007.02944.x
2. Smith K, Tillema H. Clarifying different types of portfolio use. *Ass Eval Higher Educ*. 2003;28(6):625–648. doi:10.1080/0260293032000130252
3. Driessen EW, Overeem K, van Tartwijk J, van der Vleuten CP, Muijtjens AM. Validity of portfolio assessment: which qualities determine ratings? *Med Educ*. 2006;40(9):862–866. doi:10.1111/j.1365-2929.2006.02550.x
4. Sandars J. The use of reflection in medical education: AMEE guide no. 44. *Med Teach*. 2009;31:685–695. doi:10.1080/01421590903050374
5. Driessen EW, van Tartwijk J, Overeem K, Vermunt JD, van der Vleuten CPM. Conditions for successful reflective use of portfolios in undergraduate medical education. *Med Educ*. 2005;39:1230–1235. doi:10.1111/j.1365-2929.2005.02337.x

6. Dewey J. *Experience and Education*. New York, NY: Kappa Delta Phi, Touchstone; 1938.

7. Schon DA. *The Reflective Practitioner: How Professionals Think in Action*. New York, NY: Basic Books; 1983.

8. Schön DA. *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions*. San Francisco, CA: Jossey-Bass; 1987.

9. Boud D, Keogh R, Walker D. Promoting reflection in learning: a model. In: Boud D, Keogh R, Walker D, editors. *Reflection: Turning Experience into Learning*. London: Kogan Page; 1985:18–40.

10. Moon J. *A Handbook of Reflective and Experiential Learning*. London, UK: Routledge; 1999.

11. Locher MA. *Reflective Writing in Medical Practice: A Linguistic Perspective*. Bristol: Multilingual Matters; 2017.

12. Chaffey LJ, de Leeuw EJ, Finnigan GA. Facilitating students' reflective practice in a medical course: literature review. *Educ Health (Abingdon)*. 2012;25(3):198–203. doi:10.4103/1357-6283.109787

13. Sobral DT. An appraisal of medical students' reflection-in-learning. *Med Educ*. 2000;34:182–187. doi:10.1046/j.1365-2923.2000.00473.x

14. Mamede S, Schmidt HG, Penaforte JC. Effects of reflective practice on the accuracy of medical diagnoses. *Med Educ*. 2008;42:468–475. doi:10.1111/j.1365-2923.2008.03030.x

15. Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review. *Adv Health Sci Educ Theory Pract*. 2007;14(4):595–621. doi:10.1007/s10459-007-9090-2

16. Charon R. Narrative medicine: a model for empathy, reflection, profession and trust. *JAMA*. 2001;286(15):1897–1902. doi:10.1001/jama.286.15.1897

17. Johna S, Woodward B, Patel S. What can we learn from narratives in medical education? *Perm J*. 2014;18(2):92–94. doi:10.7812/TPP/13-166

18. Niemi PM. Medical students' professional identity: self-reflection during the pre-clinical years. *Med Educ*. 1997;31:408–415. doi:10.1046/j.1365-2923.1997.00697.x

19. Wald H, Borkan J, Taylor J, Anthony D, Reis S. Fostering and evaluating reflective capacity in medical education: developing the REFLECT rubric for assessing reflective writing. *Acad Med*. 2012;87(1):41–50. doi:10.1097/ACM.0b013e31823b55fa

20. Moniz T, Arntfield S, Miller K, Lingard L, Watling C, Regehr G. Considerations in the use of reflective writing for student assessment: issues of reliability and validity. *Med Educ*. 2015;49(9):901–908. doi:10.1111/medu.12771

21. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119(2):166.e7–16. doi:10.1016/j.amjmed.2005.10.036

22. Swanwick T. *Understanding Medical Education: Evidence, Theory, and Practice*. 2nd ed. Hoboken, NJ: Wiley-Blackwell; 2014.

23. Hatton N, Smith D. Reflection in teacher education: towards definition and implementation. *Teach Teach Educ*. 1995;11:33–49. doi:10.1016/0742-051X(94)00012-U

24. Briesch AM, Swaminathan H, Welsh M, Chafouleas SM. Generalizability theory: a practical guide to study design, implementation, and interpretation. *J Sch Psychol*. 2014;52(1):13–35. doi:10.1016/j.jsp.2013.11.008

25. Bloch R, Norman G. Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. *Med Teach*. 2012;34(11):960–992. doi:10.3109/0142159X.2012.703791

26. Kassab S, Du X, Toft E, et al. Measuring medical students' essential professional competencies in a problem-based curriculum: a reliability study. *BMC Med Educ*. 2019;19:155. doi:10.1186/s12909-019-1594-y

27. Vispoel WP, Morris CA, Kilinc M. Applications of generalizability theory and their relations to classical test theory and structural equation modeling. *Psychol Methods*. 2018;23(1):1–26. doi:10.1037/met0000107

28. Mushquash C, O'Connor BP. SPSS and SAS programs for generalizability theory analyses. *Behav Res Methods*. 2006;38(3):542–547. doi:10.3758/BF03192810

29. Rees CE, Shepherd M, Chamberlain S. The utility of reflective portfolios as a method of assessing first year medical students' personal and professional development. *Reflective Pract*. 2005;6(1):3–14. doi:10.1080/1462394042000326770

30. Ahmed MH. Reflection for the undergraduate on writing in the portfolio: where are we now and where are we going? *J Adv Med Educ Prof*. 2018;6(3):97–101.

31. Arntfield S, Parlett B, Meston CN, Apramian T, Lingard L. A model of engagement in reflective writing-based portfolios: interactions between points of vulnerability and acts of adaptability. *Med Teach*. 2016;38(2):196–205. doi:10.3109/0142159X.2015.1009426