

Development of an analytical rubric and estimation of its validity and inter-rater reliability for assessing reflective narrations

VINAYAGAMOORTHY VENUGOPAL, AMOL DONGRE,
RAJENDRAKUMAR NIVARATIRAO KAGNE

ABSTRACT

Background. Reflective practice is an integral component of continuing professional development. However, assessing the written narration is complex and difficult. Rubric is a potential tool that can overcome this difficulty. We aimed to develop, validate and estimate inter-rater reliability of an analytical rubric used for assessing reflective narration.

Methods. A triangulation type of mixed-methods design (Qual-Nominal group Technique, Quan-Analytical follow-up design and Qual-Open-ended response) was adopted to achieve the study objectives. Faculties involved in the active surveillance of Covid-19 participated in the process of development of assessment rubrics. The reflective narrations of medical interns were assessed by postgraduates with and without the rubric. Steps recommended by the assessment committee of the University of Hawaii were followed to develop rubrics. Content validity index and inter-rater reliability measures were estimated.

Results. An analytical rubric with eight criteria and four mastery levels yielding a maximum score of 40 was developed. There was a significant difference in the mean score obtained by interns when rated without and with the developed rubrics. Kendall's coefficient of concordance, which is a measure of concordance of scorers among more than two scorers, was higher after using rubrics.

Conclusion. Our attempt to develop an analytical rubric for assessing reflective narration was successful in terms of the high content validity index and better inter-rater concordance. The same process can be replicated to develop any such analytical rubric in the future.

Natl Med J India 2023;36:323–6

INTRODUCTION

Reflection is a metacognitive process that creates a greater

All India Institute of Medical Sciences, Deoghar 814152, Jharkhand, India
VINAYAGAMOORTHY VENUGOPAL Department of Community and Family Medicine

Pramukh Swami Medical College (PSMC), Karamsad, Gujarat
388325, India

AMOL DONGRE Department of Community Medicine

Sri Manakula Vinayagar Medical College and Hospital, Puducherry,
India

RAJENDRAKUMAR NIVARATIRAO KAGNE Department of
Forensic Medicine

Correspondence to AMOL DONGRE; amolrdongre@gmail.com

[To cite: Venugopal V, Dongre A, Kagne RN. Development of an analytical rubric and estimation of its validity and inter-rater reliability for assessing reflective narrations. *Natl Med J India* 2023;36:323–6. DOI: 10.25259/NMJ1_732_21]

© The National Medical Journal of India 2023

understanding of both the self and the situation so that future actions can be informed by this understanding.¹ Reflection is essential for self-directed and lifelong learning. Reflective narrations of any sort of learning allows the learner to identify lacunae and thereby improve their performance for similar future events.^{2,3} Assessing such narrations gives an opportunity for providing structured and constructive feedback to students and educators.⁴

Scoring the written narrative experience is a complex and difficult process, which is the single most important obstacle in practising reflective narration for learning and research.⁵ The major issues of assessing such narrative essays are maintaining inter-rater reliability of scoring and ensuring the transparency of scoring between students and faculty. The assessment rubric is the tool that takes care of these two important issues. Rubrics list the criteria established for a particular task and the level of achievement associated with each criterion.⁶

We aimed to develop an analytical rubric for assessing the reflective narrations of medical interns on active surveillance of Covid-19, to estimate the content validity and inter-rater reliability of the rubrics, and to explore the rater perception on the usage of rubrics.

METHODS

Setting and design

We did this study at the department of Community Medicine in a tertiary care teaching hospital. The department was involved in active surveillance of Covid-19 during the first wave of the pandemic. We adopted a triangulation type of mixed-methods design.⁷ Nominal Group Technique (NGT, Qual) was used to develop rubrics.⁸ The validity and inter-rater reliability of the rubrics was estimated using analytical follow-up design (Quan). Finally the rater experience of using a rubric was captured using open-ended responses (Qual).

Study participants

Faculties involved in active surveillance ($n=5$) participated in NGT and all of them were trained in Qualitative Research Methods (QRM). Medical interns ($n=30$) were involved in field surveillance. Postgraduates ($n=4$) who supervised them in the field were the raters. The entire team was guided by a consultant in QRM and Health Professions Education.

Development of rubrics

The analytical rubric has four major elements.⁹ They are task description, characteristics to be rated (rows), levels of mastery (columns), and description of each characteristic at each level of mastery (cells).

Recommendations of the Assessment and Curriculum Support Center, University of Hawaii were followed for developing analytical rubrics.¹⁰ First, the type of rubrics to be developed was determined. Holistic scoring has no details of description of mastery thus lacks reliability.¹¹ Hence it was

decided to develop an analytical rubric. Second, the task to be assessed was identified by literature review and brainstorming of faculties. This resulted in development of four open-ended questions. Third, the criteria against which student's reflection will be compared and rated were framed. Fourth, the levels of mastery were determined. Fifth, the description of each level of mastery for each characteristic was done. Sixth and final, the rubric developed was reviewed and revised based on the feedback.

The NGT was used to finalize the third to sixth step. The purpose was briefed to the faculties involved in surveillance. They were asked to answer four open-ended questions. Those questions captured their experiences of surveillance, people's reactions, challenges faced and the lessons learnt. Manual content analysis was done. The results were shown to the faculties. Different perceptions were explored and clarified. At the point of saturation, the final criteria of assessment were established. By serial discussion of ideas and the voting system four levels of mastery for each criterion and the explanation of all identified criteria were arrived. Scores were then assigned from 0 to 5.

Finally they were instructed to independently rate the appropriateness of each criterion of the rubric and its scoring system with 'Yes' or 'No'. The content validity index was then calculated.

Assessment of reflective narration

Medical interns were hesitant to participate in active surveillance during the first wave of the pandemic where vaccine development was in the pipeline. Hence they were briefed, counselled and motivated. To reiterate the importance and consolidate their experience, they were asked to reflect on their learning in the field. The purpose and process of reflective narration was explained. Following which they were asked to type the reflective narrations for the four open-ended questions in a word document and submit it to the department e-mail. All those narrations that fulfilled the minimum word count of 1000 were coded. Raters were instructed to use their self criteria for scoring.

After a week's interval they were demonstrated the use of the developed rubrics for scoring by the investigator. The same answer papers were provided again to all raters. No time limitation was imposed to complete the assessment. Upon completion of the assessment, they were invited to answer two open-ended questions based on their experience of using the rubric. As this research was planned on already collected data as a part of teaching and training programme for medical interns, Institutional Ethics Committee clearance was obtained subsequently (IEC/99/2021 dated 12-08-2021).

Statistical analysis

Comparison of mean scores of all interns by different raters without and with scoring rubrics was done using paired *t*-test. Rubric-CVI/UA (Universal Agreement method) was calculated by summing the number of criterion that had 100% agreement and dividing that by the total number of criterion.¹² Inter-rater reliability analysis of the scores given with and without rubrics was carried out using (i) Cronbach alpha that measured internal consistency; (ii) intraclass correlation co-efficient; and (iii) Kendall coefficient of concordance (W).¹³ SPSS version 24.0 was used for analysis.

Manual content analysis was carried out on the experiences shared by faculty about surveillance and the rater experience of

using rubrics. We read through the entire responses to familiarize ourselves with the data. Initial codes were then manually generated by highlighting relevant aspects of the responses. Any discrepancy arrived was sorted out and final consensus was arrived to minimize bias. Similar codes were then collated under various categories.¹⁴

RESULTS

Of the 30 narrative essays, 23 (76.6%) were eligible for assessment. Of the 23 eligible essays 13 (56.5%) were written by women interns. Among the four scorers, three were women. Of the five faculties who shared their experience, three were women.

Manual content analysis of reflective narrations of faculty resulted in the emergence of seven categories and 44 text codes. From these seven categories eight criteria (rows) were generated for the assessment rubrics. They were observation of field activities, learning about the topic, learning around the topic, leadership skills, interpersonal communication skills, professional values and behaviours, writing skills and challenges faced. The Rubric-CVI/UA of 1.0 was obtained after presenting all criteria to all four faculties. The detail of the assessment rubrics are given in Annexure I.

The mean (SD) scores obtained by all interns by all four raters without rubrics was 27.8 (3.5) and with rubrics was 21.6 (1.5). There was a statistically significant difference in their scores awarded without rubrics and with the developed rubrics. Individual variations in score between raters are provided in Table I. Cronbach alpha, a measure of internal consistency between the scores given by different raters, improved from 0.69 to 0.87 after using rubrics for assessment. The intraclass correlation coefficient also improved from 0.61 to 0.71 between scorers. Kendall coefficient of concordance that is a measure of concordance of scorers among more than two scorers improved from 0.52 to 0.73 after using rubrics (Table II).

The advantages of rubrics as felt by scorers were: comprehensive nature (more criteria to check), widened their learning, feedback for improvement, and guided them to plan future field activities. The disadvantages mentioned were: predefined criteria hampers their thinking process, the flow of scoring was uneasy, and difficult to search for the criteria and rate.

DISCUSSION

Assessment rubrics with eight criteria and four mastery levels that were scored between 0 and 5 against each criterion yielding a maximum score of 40 were developed to assess the reflective narrations of interns on their experience of active Covid-19 surveillance. The average scores given by all four scorers were significantly different after using assessment rubrics. All measures of inter-rater reliability improved statistically after using rubrics for assessing narrative reflections.

We followed the principles and process recommend by the Assessment and Curriculum Support Center, University of Hawaii and Center of Teaching Excellence, University of Florida.^{9,10} The steps mentioned by these centres were a conglomeration of various works already done in the field of rubrics development. Apart from the published literature, field experts were also consulted while developing the rubrics. All these contributed to better validity of the tool developed.

Our study found that the mean score of students reduced when assessed using the rubrics. Though there was a reduction

Annexure I. Rubrics developed for scoring reflective narration

Component and its description with suitable examples	Scoring system			
	0	1–2	3–4	5
		Student has simply observed	Student has analysed and interpreted his observation	Based on analysis student can infer and conclude
<p><i>Observation of field activities</i> Ability of the student to observe, understand and realize the importance of planning field activities <i>Examples:</i> Vehicle arrangement, drinking water, refreshments, infection prevention measures, mapping of area, household details, coordination with primary health centre (PHC) team members, handling locked house, missing households</p>	Not attempted	Social workers have used village map for survey	Village map helped us to identify the houses easily and helped to identify missed households	Village mapping activity is an important prerequisite to carry out field activity. It helps in easy identification of houses in any unfamiliar area
		PHC staff such as health inspector and auxiliary nurse midwives (ANM) were with us for the survey	They helped us to solve sensitive queries raised by the villagers	Frontline health workers of the locality helped us to solve field-level problems. Hence, their involvement/collaboration is important
<p><i>Learning about the topic (Covid-19)</i> Ability of the student to gain clinical knowledge about the topic of survey <i>Examples:</i> Epidemiological triad, modes of transmission, clinical features, management, prevention and control measures</p>	Not attempted	Mentioned the learning details of Covid-19	Ability to apply the learning details of Covid-19. (Suspect identification, risky behaviours)	Able to suggest recommendations based on his observation and application of learning. (BCC for preventive and control measures)
<p><i>Learning around the topic/awareness about the local community</i> Ability to realize the customs, beliefs and attitudes, socioeconomic, cultural, environmental factors related to Covid-19 and their impact on its management <i>Examples:</i> Awareness, attitude, practice of the community, loss of employment, traditional treatment, religious belief, perceived seriousness, etc.</p>	Not attempted	Failed to link medical knowledge with the social and environment factors	Student was able to relate social, cultural, and environmental factors around management of Covid-19	Apart from clinical details and complications of Covid-19, there are social, economic, cultural and environmental factors around management of Covid-19
<p><i>Leadership skills</i> Ability to report the importance of leadership qualities in handling an inter-professional team <i>Examples:</i> Team spirit, conflict management, decision-making, delegation of roles and responsibilities, leading by example, etc.</p>	Not attempted	Inadequate and irrelevant content	Some relevant content but evident weakness in meeting the criteria	Discussed impact of leadership skills comprehensively and critically
<p><i>Interpersonal communication skills</i> Ability to realize and report the importance of interpersonal communication <i>Examples:</i> Rapport building, dealing with sensitive questions, showing empathy, etc.</p>	Not attempted	Minimal mention	Some mention about impact of communication	Discussed impact of communication skills comprehensively and critically
<p><i>Professional values and behaviours.</i> Conduct, behaviour, attitude, respect, loyalty towards the community, commitment towards work. <i>Examples:</i> Greeting, using culturally sensitive words, committed towards work, self-satisfaction</p>	Not attempted	Minimal mention	Some mention about professional values and behaviours	Discussed importance and impact of professional values and behaviours with examples
<p><i>Writing skills</i> Ability to write their reflection using proper English (complete sentence, grammar and spelling, use of appropriate words, logical flow)</p>	Extensive errors	Such errors in more than half of the writeup	Such errors in less than one-third of the writeup	Meets standards of academic writing
<p><i>Challenges faced</i> Ability to reflect and report the challenges faced by the student, team and community and suggests appropriate recommendations to overcome. <i>Examples:</i> Hot weather, difficulty in physical distancing, use of face mask, handle locked house</p>	Not attempted	The climate is sunny	Hot weather affected the optimal functioning of team members	Weather changes are unpredictable and unavoidable so it is better to be prepared for them. (Use of umbrella, drinking water/ORS solution)

BCC behaviour change communication

TABLE I. Comparison of marks obtained by interns on their reflective narration, given by different scorers with and without scoring rubrics ($n=23$)

Scorer (maximum score 40)	Without rubrics	With rubrics	p value [#]
1	29.4 (4.2)	21.8 (1.4)	<0.001*
2	25.2 (3.4)	20.6 (1.6)	<0.001*
3	27.7 (4.3)	23.3 (1.4)	<0.001*
4	28.9 (2.1)	20.7 (1.7)	<0.001*
All scorers	27.8 (3.5)	21.6 (1.5)	<0.001*

All values are mean (SD) * p value based on paired sample *t*-test

TABLE II. Inter-rater reliability analysis of the scores given for the reflective narration of interns ($n=23$), by different scorers ($n=4$) with and without scoring rubrics

Scoring	Cronbach alpha	Intraclass correlation coefficient (ICC)			Kendall coefficient of concordance	
		ICC	95% CI	p value	W	p value
Without rubrics	0.69	0.61	0.29–0.81	<0.001	0.52	0.002
With rubrics	0.87	0.71	0.30–0.88	<0.001	0.73	<0.001

ICC using two-way random effect model for absolute agreement

in the mean scores, structuring the assessment scheme minimizes subjectivity and improves accuracy. The reliability of scoring written narrations is comparatively lower than the other methods of written assessment, namely multiple choice and short answer questions.¹⁵ We developed an assessment rubric that has explicit criteria in separate components for scoring which to a greater extent overcome this reliability issue.¹⁶ Moreover, we developed an analytical rubric whose reliability is better than a holistic rubric.¹⁷ The improvement in reliability measures without rubric (holistic assessment) and with rubric (analytical assessment) in our study also confirms the same.

The following could be the reasons for the observed better inter-rater concordance in the present study: (i) the structure and the criteria mentioned in the rubric were simple to understand and score; (ii) the demonstration provided to use rubrics has the potential to minimize the difference in level of knowledge and experience of the rating among rater; (iii) sufficient time was provided to go through the content legibly and optimally; (iv) the answer sheets were coded and thus raters were blinded to the personal information; and (v) as the contents were typed by the interns the influence of handwriting was avoided. The study had few limitations. First, the lesser number of a raters might have positively influenced the reliability level. Second, individual level variation of a rater such as interest and cognitive levels were difficult to eliminate.

Conclusion and recommendation

The rubric was developed following the standard guidelines. Our attempt to develop an analytical rubric for assessing reflective narration was successful in terms of the high content validity index and better inter-rater concordance. Training a rater to use rubrics avoided misinterpretation of the rubrics and helped to improve the inter-rater concordance. This article aids health profession educators to develop analytical rubrics and enhances their proper usage in assessing reflective narrations, which are usually not done or done subjectively.

REFERENCES

- Sandars J. The use of reflection in medical education: AMEE Guide No. 44. *Med Teach* 2009;**31**:685–95.
- Smyth J. Teachers' work and the politics of reflection. *Am Educ Res J* 1992;**29**: 267–300.
- Loughran JJ. Effective reflective practice: In search of meaning in learning about teaching. *J Teach Educ* 2002;**53**:33–43.
- Beyreli L, Arý G. The use of analytic rubric in the assessment of writing performance -inter-rater concordance study. *Kuram Ve Uygulamada Egitim Bilim* 2009;**9**: 105–25.
- Cole J, Haley K, Muenz T. Written expression reviewed. *Research in the schools*. 1997;**4**:17–34.
- Stevens DD, Levi AJ. *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback and promote student learning*. 2nd ed. Sterling, Va: Stylus Publishing; 2012.
- Creswell J, Clark V. *Designing and conducting mixed methods research*. 3rd ed. Washington DC: Sage; 2011.
- Abramson JH. *Survey methods in community medicine: Epidemiological research, programme evaluation, clinical trials*. 5th ed. Edinburgh, New York: Churchill Livingstone; 1999.
- Brophy TS. Writing effective rubrics. Available at <https://assessment.aa.ufl.edu/faculty-resources/> (accessed on 10 Oct 2021)
- University of Hawaii. Creating and using rubrics. Assess. Curric. Support Cent. Available at <https://manoa.hawaii.edu/assessment/resources/creating-and-using-rubrics/> (accessed on 15 Oct 2021)
- Espin CA, Weissenburger JW, Benson BJ. Assessing the writing performance of students in special education. *Exceptionality* 2004;**12**:55–66.
- Polit DF, Beck CT. The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Res Nurs Health* 2006;**29**: 489–97.
- Legendre P. Species associations the Kendall coefficient of concordance revisited. *JABES* 2005;**10**:226–45.
- UCLA Center for Health Policy Research: Section 4: Key informant interviews. Available at https://healthpolicy.ucla.edu/programs/health-data/trainings/Documents/tw_cba23.pdf (accessed on 15 Oct 2021).
- Schuwirth LWT, van der Vleuten CPM. Different written assessment methods: What can be said about their strengths and weaknesses? *Med Educ* 2004;**38**: 974–9.
- Dunsmuir S, Kyriacou M, Batuwitige S, Hinson E, Ingram V, O'Sullivan S. An evaluation of the writing assessment measure (WAM) for children's narrative writing. *Assess Writ* 2015;**23**:1–18.
- Hayes JR, Hatch JA, Silk CM. Does holistic assessment predict writing performance?: Estimating the consistency of student performance on holistically scored writing assignments. *Writ Commun* 2000;**17**:3–26.