

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

# Currents in Pharmacy Teaching and Learning

journal homepage: [www.sciencedirect.com/journal/currents-in-pharmacy-teaching-and-learning](http://www.sciencedirect.com/journal/currents-in-pharmacy-teaching-and-learning)



## Research Note

# Mental effort in the assessment of critical reflection: Implications for assessment quality and scoring

Pilar Robinson Gonzalez<sup>a</sup>, Bridget Paravattil<sup>b</sup>, Kyle John Wilby<sup>c,\*</sup>

<sup>a</sup> School of Pharmacy, University of Otago, PO Box 56, Dunedin 9018, New Zealand

<sup>b</sup> College of Pharmacy, Qatar University, PO Box 2713, Doha, Qatar

<sup>c</sup> College of Pharmacy, Faculty of Health, Dalhousie University, 5968 College Street, Halifax, Nova Scotia, Canada



## ARTICLE INFO

### Keywords:

Assessment  
Cognitive load  
Mental effort  
Pharmacy  
Reflection

## ABSTRACT

**Introduction:** Critical reflection is a mainstay in the training of health professionals, yet assessment of reflection is commonly described as difficult, taxing, and resulting in inconsistent scoring across assessors. At the same time, there is evidence from experiential and simulation settings that assessors' mental effort may explain assessor variability, which could be a target for simplifications in assessment design. Assessors' mental effort for assessment of reflection is currently unknown. This study aimed to determine reliability of rubric scoring of critical reflection, variation in pass-fail rates, and the relationship between reflection scores and assessors' perceived mental effort.

**Methods:** Eleven assessors were recruited to assess six reflection assignments using a published rubric. Mental effort was measured using the Paas scale for each assignment assessed and was correlated with rubric scores for each assignment.

**Results:** Findings showed inconsistency in scoring between assessors, resulting in varying pass rates for each assignment (55–100%). All assignments demonstrated negative correlations between rubric scores and perceived mental effort ( $r = -0.115$  to  $-0.649$ ).

**Conclusions:** Findings support the notion that more work should be done to optimize assessment of critical reflection. Future studies should focus on disentangling the influence on mental effort of scoring tools, assignment structures, and writing quality.

## Introduction

Recognized as a necessary skill for future practice, critical reflection is becoming a mainstay in health professional training curricula worldwide.<sup>1</sup> Despite its importance, programs appear to be divided on how best to ensure students are graduating with the capacity to be reflective practitioners.<sup>2</sup> One of the key controversies is around assessment and whether critical reflection can be judged accurately and objectively by raters. It is also debated whether assessment of critical reflection should be scored as simply pass or fail or considered alongside other assessment data in the form of a portfolio.<sup>2,3</sup> Much of this controversy is driven by the ability of an assessor to accurately interpret and judge critical reflection in a consistent and objective manner.

The task of assessing reflection is not easy, as reflection can be deeply personal and must be deciphered, interpreted, and evaluated by the assessor.<sup>2</sup> Assessor cognition research points to many factors that may influence assessors' judgements, including expertise,

\* Corresponding author.

E-mail addresses: [bridgetj@qu.edu.qa](mailto:bridgetj@qu.edu.qa) (B. Paravattil), [kyle.wilby@dal.ca](mailto:kyle.wilby@dal.ca) (K.J. Wilby).

<https://doi.org/10.1016/j.cptl.2022.06.016>

perspectives, cultural orientations, attentional capacity, and others.<sup>4–7</sup> These factors have explained assessor variability in other contexts (e.g., experiential training, objective structured clinical examinations, simulation), but little data exists to explain assessor variability for assessment of critical reflection. Although assessor cognition research suggests that achieving ‘perfect reliability’ in rater-based assessment tasks should not likely be a goal, it is important to understand assessor variability across assessment contexts to improve assessment processes and practices, including decision-making on how to best judge student performance (e.g., scores, pass-fail, portfolios).<sup>5,8</sup>

In designing assessment procedures and tools, there are increasing calls to consider the theoretical perspective of assessors’ mental workload and mental effort.<sup>5</sup> This consideration may be important for assessment of critical reflection, as the deeply personal nature of the writing may require a high attentional capacity (i.e. mental effort). Within the realm of performance assessment, Eva<sup>5</sup> argues that a fundamental flaw with current assessment processes may be that the procedures and tools used do not account for limitations of human cognition (e.g., attentional capacity, working memory, idiosyncratic influences of previous experience). This argument is largely supported by work conducted by Tavaras and colleagues<sup>9,10</sup> that aimed to determine the influence of mental effort and workload (i.e. cognitive load) on assessment tasks. These researchers and others have found that the mental effort and workload associated with assessment tasks may influence assessment quality (measured typically as reliability) and that simplifying the assessment processes (or tools) may improve accuracy of results obtained.<sup>11</sup>

Given the increasing emphasis on critical reflection within health professional training programs and the potential for assessor variability to influence judgements and scores of student performance, it is important to better understand the drivers of variability to modify processes and tools to achieve a more accurate assessment. Assessors’ mental effort requires further investigation, due to the availability of new rubrics and the need for increased attentional capacity when assessing reflection. Therefore, the purpose of this study was to explore assessor variability arising from assessors’ mental effort in assessment of critical reflection. Specifically, this study aimed to:

- (1) determine interrater reliability of rater-based critical reflection scoring;
- (2) determine consistency of critical reflection pass-fail decisions scored using a rubric;
- (3) identify associations between assessors’ mental effort and rubric scores when assessing critical reflection.

**Table 1**

Adapted rubric for assessment of critical reflection.<sup>12,13</sup>

Elements of reflective practice	Non-reflector (0 marks)	Incomplete reflector (1 mark)	Developing reflector (2 marks)	Reflector (3 marks)	Critical reflector (4 marks)
<b>Description of experience</b> <i>What happened?</i>	No description of the experience	Incomplete description of experience	Description of experience is <b>clear</b>	Description of the experience is clear and <b>chronological</b>	Description of the experience is clear, chronological, and <b>free of judgements</b>
<b>Description of feelings</b> <i>How did/do I feel?</i>	No evidence of personal feelings or thoughts	Personal thoughts or feelings implied, but not expressed	Personal feelings or thoughts are expressed	Personal feelings and thoughts are expressed and <b>described</b>	Multiple personal feelings and thoughts are expressed and <b>evaluated</b> and/or <b>explained</b>
<b>Assessment of expectations</b> <i>How did this experience compare to expectations?</i>	No description of expectations	No link made between experience and expectations	Experience described as <b>consistent</b> with expectations	<b>Differences</b> between expectations and experience described	Expectations <b>explained</b> with previous experience/knowledge and differences between expectations and experience described
<b>Explanation of experience</b> <i>Why did this go the way it did?</i>	No connection between experience and prior knowledge, feelings, or attitudes	Connection between experience and prior knowledge, feelings or attitudes incomplete	<b>Connection</b> between experience and prior knowledge, feelings, or attitude made, but no suggestion of causative link	<b>Causative</b> connection <b>suggested</b> between experience and prior knowledge, feelings, or attitudes	<b>Clear</b> causative connection made between experience and prior knowledge, feelings, or attitudes
<b>Assessment of self</b> <i>What have I learned?</i>	No evidence of new knowledge, feelings, or attitudes	Evidence of change in knowledge, feelings, or attitudes implied, but not clearly	<b>Some</b> evidence of change in knowledge, feelings, or attitudes	<b>Clear</b> evidence of change in knowledge, feelings, or attitudes	Clear evidence of change in knowledge, feelings, or attitudes and <b>linked to future pharmacy practice</b>
<b>Learning goals</b> <i>What is there to learn now?</i>	No learning goals given	Learning goals implied, but not expressed	Learning goal/s given, but may be inadequate or irrelevant	<b>Relevant</b> and <b>adequate</b> learning goal/s given	Relevant and adequate learning goal/s given and <b>justified</b> in the context of future pharmacy practice
<b>Learning plans</b> <i>What am I going to do now?</i>	No plans for learning given	Plan for future learning is incomplete	Plan given for future learning, but it is under-detailed, lofty, unrealistic, or irrelevant to learning goals	<b>Reasonable</b> plan given for future learning goals	Good plan given for future learning given learning goals, including <b>clear “next step”</b>

## Methods

This was a rater-based assessment study of pharmacy students' reflective assignments using a published rubric known to assessors and the program.<sup>12,13</sup> The study was approved by the Human Ethics Committee at the University of Otago.

For the purposes of this study, six reflective assignments that were previously graded as a summative assessment in a third professional year (of four total professional years) course were chosen (with student permission). The assignment required students to critically reflect on an experience working with a team. Students were asked to identify the experience, critically reflect on the experience based on previous reflection training, and relate the experience to future pharmacy practice. Students were able to choose any experience, including course work (e.g., group project, interprofessional education experience, professional skills training), employment, or extra-curricular activities. Students had previously written a critical reflection and received feedback according to the same rubric the previous year. Reflections were capped at 880 words. The six assignments chosen for this study were selected based on grading performance: two high performers, two average performers, and two low performers. All reflections were de-identified and assigned an identification number ranging from one to six.

The rubric used for this study was the same rubric used for grading reflections within the program (Table 1). It had been adapted from a previously published rubric to meet the program specifications.<sup>12,13</sup> The rubric utilizes five points ranging from zero to four and scores critical reflection over seven domains. The highest possible total score on the rubric is 28 points.

### Participants

A total of 11 assessors who had not originally graded the reflections included in this study as part of course requirements were recruited for this study. Assessors were eligible if they had previously assessed students' critical reflection (for a different assignment or year group) according to the rubric as part of course requirements and were University of Otago staff members. As previous assessors, all participants had received comprehensive training and feedback on the rubric from an instructional designer and a faculty member with an educational background. Of 12 assessors that met these criteria, 11 agreed to participate in the study. Assessors were recruited by email and provided written informed consent prior to initiation of study procedures.

### Procedures

Once consent was obtained, assessors were oriented to the procedures and provided with a 20-min individual training session to introduce the assignment expectations and refresh their memories on the rubric and criteria for assessment. Assessors were able to ask questions to the senior investigator at that time. Once training was complete, assessors were provided with the six critical reflective essays and six scoring rubrics. The rubrics also included a nine-point rating scale (Table 2) for assessors to rate their perceived mental effort in assessing each reflection (Paas Scale).<sup>14</sup> This scale was chosen based on known validity, as well as ease of use for rating mental effort for multiple assignments. Assessors were given two weeks to complete all assessments.

### Data analysis

Total rubric scores were calculated for each returned assessment. Passing scores were determined to be 14 out of 28 points, as per usual grading policies. Interrater reliability was calculated for each of the six assignments using a two-way random intraclass correlation coefficient. To determine if perceived mental effort was associated with rubric scores, Spearman's Rank Correlation Coefficient was used to identify correlations between mental effort ratings and rubric scores. All statistical analyses were conducted using SPSS, version 25 (IBM SPSS Statistics).

## Results

All 11 assessors completed all study procedures. Assessors were all academic staff members at the University of Otago. The median rubric score for each assignment, percentage of assessments that met passing criteria, median mental effort per assignment, and the correlation coefficient between mental effort and rubric score are provided in Table 3. Pass rates varied from 55% to 100% across

**Table 2**  
The Paas scale for measuring mental effort.<sup>14</sup>

Score	Mental effort
1	Very, very low mental effort
2	Very low mental effort
3	Low mental effort
4	Rather low mental effort
5	Neither low nor high mental effort
6	Rather high mental effort
7	High mental effort
8	Very high mental effort
9	Very, very high mental effort

**Table 3**

Summary of rubric scores, pass-rates, assessors' mental effort, and correlations between rubric scores and mental effort.

Reflection	Median score (range)	Number of passing scores (%)	Median mental effort (range)	Correlation coefficient between score and mental effort
1 (Low performer)	14 (8–26)	6 (55)	6 (2–8)	–0.160
2 (Low performer)	23 (14–27)	11 (100)	6 (2–9)	–0.319
3 (Middle performer)	23 (13–28)	10 (91)	6 (2–7)	–0.446
4 (Middle performer)	21 (11–27)	9 (82)	5 (2–7)	–0.573
5 (High performer)	26 (13–28)	10 (91)	5 (2–8)	–0.115
6 (High performer)	23 (18–28)	11 (100)	5 (2–7)	–.649 <sup>a</sup>

<sup>a</sup> This had a  $P < .05$ .

reflections, with two reflections achieving a passing score consistent across all assessors. Assessors used a range of 2 (very low mental effort) to 9 (very, very high mental effort) to describe their mental effort. Median mental effort scores ranged from 5 (neither high or low mental effort) to 6 (rather high mental effort). The correlation between rubric scores and mental effort for reflection 6 was statistically significant ( $r = -0.649$ ). All correlation coefficients were negative, demonstrating an inverse pattern between perceived mental effort and rubric scores. In other words, as rubric scores increased, perceived mental effort decreased. Single measures interrater reliability was low (Intraclass correlation coefficient [ICC] = 0.274, 95% CI 0.084 to 0.726). Average measures interrater reliability was high (ICC = 0.806, 95% CI 0.501 to 0.967).

## Discussion

The purpose of this study was to explore the associations between assessors' perceived mental effort with scoring of critical reflective essays. The study also aimed to determine reliability of rubric scoring for critical reflection and the influence of grading inconsistencies on pass-fail decision-making. Overall, findings showed that rubric-scoring of critical reflection was prone to variation between assessors, and this may have implications for pass-fail decision-making.

The inconsistencies in grading between assessors must be addressed. Despite achieving an ICC > 0.8, this was based on average measures and can be interpreted that reliability is excellent when 11 assessors grade each assignment. This is of course not practical and raises questions about the best way to grade reflection assignments. Pass-fail decisions also differed between assessors. Reflection numbers 1 and 4 appeared to pass or fail the assignment depending on which assessor scored them. This finding does not align with previous studies using this rubric that demonstrated acceptable interrater reliability.<sup>13,15</sup> While it is possible that the assessors in our study were not as experienced (or as well trained) as those in the previous studies, the assessors in this study did have previous training and experience (with feedback). Other reasons for discrepancy may be the small assessment sample size compared to this study or the inclusion of more assessors within our study. In any case, the assessors that participated in our study represent a typical cohort with typical training for a program required to distribute grading across assessors due to large student numbers.

Another key finding of this study was the inverse association between assessors' perceived mental effort and rubric score. Assessors likely found it more mentally taxing to determine rubric scores across the domains for poorer performers. This may have been due to the rubric itself, assessors' ability to interpret the reflective writing if not presented cohesively, or the writing style (spelling, grammar, etc.) of the student. With respect to the rubric, studies should determine how wording of descriptors may influence mental effort and any resulting grading decision. With respect to students' work, assessors may need to be provided with greater guidance when interpreting reflection from poor performers. While very relevant to reflection, this consideration aligns with findings across other high stakes assessments in health professionals' education.<sup>16</sup>

This study has limitations that must be addressed. Although the number of assessors was low ( $N = 11$ ), it was almost the maximum that could be recruited at our program and resulted in 66 data points of assessment. Future studies could recruit across settings, but assessment procedures and assessor experience and training should be similar. This study did not investigate reasons to why assessors rated their mental effort to be high or low for each assignment. If the study is repeated, this would be worthwhile to capture to determine if higher mental effort could be attributed to a particular finding (e.g., rubric, writing). Finally, the assessors recruited for this study were not all experts in critical reflection, and this may have led to inconsistencies in scoring. As discussed, however, they represent a typical cohort from an institution that would likely be responsible for grading such as assignment in real practice.

## Conclusions

This study found that reliability of assessors' critical reflection rubric scoring was suboptimal, and this resulted in differing pass-fail decisions across student work. The study also found that assessors perceive greater mental effort in assessment when assessing poor

performers. Findings support the notion that more work should be done to optimize assessment of critical reflection. Future studies should focus on disentangling the influence on mental effort of scoring tools, assignment structures, and writing quality to better identify and understand modifiable factors that may negatively impact assessment decisions.

### Disclosure(s)

None.

### Author statement

Pilar Gonzalez contributed to study design, collected the data, assisted with data analysis, and drafted parts of the manuscript. She edited the final manuscript and approved its submission.

Bridget Paravattil contributed to study design, assisted with data analysis, and drafted parts of the manuscript. She edited the final manuscript and approved its submission.

Kyle John Wilby, conceived the idea for this manuscript, assisted with data collection, led data analysis, drafted parts of the initial manuscript, and edited and approved the final version.

### Declaration of Competing Interest

None.

### References

- 1 Delaney C, Watkin D. A study of critical reflection in health professional education: learning where others are coming from. *Adv Health Sci Educ Theory Pract.* 2009;14(3):411–429. <https://doi.org/10.1007/s10459-008-9128-0>.
- 2 Sandars J. The use of reflection in medical education: AMEE guide no. 44. *Med Teach.* 2009;31(8):685–695. <https://doi.org/10.1080/01421590903050374>.
- 3 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(12):1230–1235. <https://doi.org/10.1111/j.1365-2929.2005.02337.x>.
- 4 Berendonk C, Stalmeijer RE, Schuwirth LWT. Expertise in performance assessment: assessors' perspectives. *Adv Health Sci Educ Theory Pract.* 2013;18(4):559–571. <https://doi.org/10.1007/s10459-012-9392-x>.
- 5 Eva KW. Cognitive influences on complex performance assessment: lessons from the interplay between medicine and psychology. *J Appl Res Mem Cogn.* 2018;7(2):177–188. <https://doi.org/10.1016/j.jarmac.2018.03.008>.
- 6 Oudkerk Pool A, Govaerts MJB, Jaarsma DADC, Driessen EW. From aggregation to interpretation: how assessors judge complex data in a competency-based portfolio. *Adv Health Sci Educ Theory Pract.* 2018;23(2):275–287. <https://doi.org/10.1007/s10459-017-9793-y>.
- 7 Wilby KJ, Govaerts MJB, Austin Z, Dolmans DHJM. Exploring the influence of cultural orientations on assessment of communication behaviours during patient-practitioner interactions. *BMC Med Educ.* 2017;17:61. <https://doi.org/10.1186/s12909-017-0899-y>.
- 8 Gingerich A, Kogan J, Yeates P, Govaerts M, Holmboe E. Seeing the 'black box' differently: assessor cognition from three research perspectives. *Med Educ.* 2014;48(11):1055–1068. <https://doi.org/10.1111/medu.12546>.
- 9 Tavares W, Eva KW. Exploring the impact of mental workload on rater-based assessments. *Adv Health Sci Educ Theory Pract.* 2013;18(2):291–303. <https://doi.org/10.1007/s10459-012-9370-3>.
- 10 Tavares W, Ginsburg S, Eva KW. Selecting and simplifying: rater performance and behavior when considering multiple competencies. *Teach Learn Med.* 2016;28(1):41–51. <https://doi.org/10.1080/10401334.2015.1107489>.
- 11 Paravattil B, Wilby KJ. Optimizing assessors' mental workload in rater-based assessment: a critical narrative review. *Perspect Med Educ.* 2019;8(6):339–345. <https://doi.org/10.1007/s40037-019-00535-6>.
- 12 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):124. <https://doi.org/10.5688/ajpe798124>.
- 13 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. *Curr Pharm Teach Learn.* 2017;9(6):989–995. <https://doi.org/10.1016/j.cptl.2017.07.025>.
- 14 Paas F. Training strategies for attaining transfer of problem-solving skills in statistics: a cognitive load approach. *J Educ Psychol.* 1992;84(4):429–434. <https://doi.org/10.1037/0022-0663.84.4.429>.
- 15 Lucas C, Smith L, Lonie JM, Hough M, Rogers K, Mantzourani E. Can a reflective rubric be applied consistently with raters globally? A study across three countries. *Curr Pharm Teach Learn.* 2019;11(10):987–994. <https://doi.org/10.1016/j.cptl.2019.06.004>.
- 16 De Champlain AF. Chapter 24: Standard setting methods in medical education: High-stakes assessment. In: Swanwick T, Forrest K, O'Brien BC, eds. *Understanding Medical Education: Evidence Theory and Practice*. 3rd ed. Wiley-Blackwell; 2019:347–359.