Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: http://ees.elsevier.com

Effects of rubric quality on marker variation in higher education

Subrata Chakraborty PhD [a,*], Christopher Dann PhD [b], Anuradha Mandal MCA [a], Beverly Dann PhD [c], Manoranjan Paul PhD [d], Abdul Hafeez-Baig PhD [e]

[a] Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), School of Information, Systems & Modelling, Faculty of Engineering and IT, University of Technology Sydney, NSW, Australia
[b] School of Education, University of Southern Queensland, QLD, Australia
[c] School of Education, University of Sunshine Coast, QLD, Australia
[d] Machine Vision and Digital Health (MaViDH), School of Computing and Mathematics, Charles Sturt University, NSW, Australia
[e] School of Business, University of Southern Queensland, QLD, Australia

ARTICLE INFO

Keywords
Higher education
Assessment bias
Marker bias reduction
Assessment reliability
Assessment validity
Rubric analysis

ABSTRACT

Variation among markers has the potential to disadvantage students by contributing to a discrepancy in assessments in higher education settings. In this study, we extended a previous study that analyzed first-year students' results in a Business Faculty within an Australian university to understand the extent of variation within and between multiple markers and across multiple courses. The study investigated the potential influence of quality of rubrics and associated documentation provided as marker guidance. Results indicated that specific features of rubrics, such as the inclusion of clear indicators of quality, had an observable effect on marker variation.

1. Introduction

Higher education institutions have a duty to maximize the defensibility, quality and validity of their assessment processes (Fuller, Homer, Pell, & Hallam, 2017). This is complicated by the need for assessment to undertake a 'double duty' (Boud & Falchikov, 2006), with both a grading function and a learning support function, not necessarily prioritized in that order. Markers require an explicit and transparent framework (Jones, Allen, Dunn, & Brooker, 2017) to consistently identify academic performance which is the basis for criterion-referenced assessment (Biggs, 2011). This paper presents data on the grading of student work to illuminate the information that markers unwittingly provide through the overall grades presented for students as a summative indication of learning. In doing this, we highlight aspects that impact the marker and marking process. Technology is one such component that can add value to assessment practices (Devedzic & Devedzic, 2019; Grion, Serbati, Felisatti, & Li, 2019; JISC, 2005). This paper explores the opportunity to leverage the 'big data' available when technology such as a Learning Management System (LMS) is used to gather data on the markers' practices, experiences, influences and skills.

Marker variation sits among other factors impacting on student grades such as assessment task design (Biggs, 2011), direct instruction (Chase & Klahr, 2017) and practice (Saddler & Andrade, 2004; Sadler, 2010), feedback processes (Black & Wiliam, 1998; Boud & Molloy, 2013), and teacher content knowledge (Dann, Dann, & O'Neill, 2018). Marking variation impacts directly on the reliability and validity of the assessment (Brown & Knight, 2012; Panadero, Romero, & Strijbos, 2013). Thus, the marking depends on a number of factors including, experience, knowledge of the content, how closely markers work to a standardized task or criteria, the amount of freedom to make judgements and more (Bloxham, Boyd, & Orr, 2011; Sadler, 2007). In order to reduce the marking variation and increase reliability and validity, rubrics were introduced. The use of rubrics has increased because of their capacity to enhance student performance and promote student learning (Hanrahan & Isaacs, 2001), although it does not guarantee high reliability among markers (Albluwi, 2018; Stellmack, Konheim-Kalkstein, Manor, Massey, & Schmitz, 2009).

Various techniques have been adopted to reduce the subjective element in marking within a business program. They include multiple-choice questions (MCQ) (Brown, Irving, & Keegan, 2008), criterion-based assessment together with increased formative assessment techniques, and dynamic assessment scenarios that require the application of constructed knowledge rather than simple recall. A vertical marking approach has been suggested in which each marker marks a particular section of the assessment piece for all the students (Dong et al.,

* Corresponding author.
E-mail addresses: subrata.chakraborty@uts.edu.au (S. Chakraborty); Chris.Dann@usq.edu.au (C. Dann); Anuradha.Mandal@gmail.com (A. Mandal); bdann@usc.edu.au (B. Dann); mpaul@csu.edu.au (M. Paul); Abdul.Hafeez-Baig@usq.edu.au (A. Hafeez-Baig)

https://doi.org/10.1016/j.stueduc.2021.100997
Received 24 September 2020; Received in revised form 19 February 2021; Accepted 21 February 2021
Available online xxx
0191-491/

2014). This approach is appropriate if the assessment piece contains distinct and mutually exclusive parts. For instance, if an assessment task contains five separate questions in which to respond, five markers could each mark one specific question for all the participants (Chakraborty et al., 2016). This approach could improve the marking fairness, as each individual marker variation applies to all students equally but is not suitable if (a) the assessment piece cannot be segmented into mutually exclusive segments, (b) segments cannot be distributed equally among markers, or (c) markers are not able to be physically co-located. This paper adds to the earlier work of Chakraborty et al. (2016) by introducing rubric analysis to show alignment between the rubric characteristics and earlier results. In contrast to the Chakraborty, Zhou et al. (2016) course-based study, Zhao and Huang (2020) found that single task and single rater scenarios led to larger score variation in assessment marking of a high-stakes standardized English writing task in China.

Other approaches impacting the marker's role in assessment reliability and validity have included the creation of computer programs such as Developing Understanding of Assessment for Learning (DUAL) (Bird & Yucel, 2013) in an effort to improve the reliability and efficiency of the marking process. DUAL did improve efficiency (Bird & Yucel, 2013); however, it was developed specifically to improve the communication and knowledge exchange between staff and students in an environment where a community of practice was being built and is not universally applicable.

Despite such attempts to increase objectivity and eliminate marker variation, the problem still remains. This paper explores other influences that may impact on marker variation. These include the marker's approach to the marking process, the marker's workload, experience in the marking of the task, and documentation, such as assignment descriptions and rubrics, available to markers. Other factors influencing the marking process include, but are not limited to, time constraints imposed on the return of grades to students (many education providers have a two-week results release policy); the use of online systems such as Turnitin® impacting on a marker's ability to enhance the validity and reliability of assessment marking, and the complexity of assessment rubrics and associated documentation delivered to the markers which, in the context of this study, were often removed from the delivery of the course and were external to the university.

2. Literature review

Use of rubrics for assessment in higher education have been considered as beneficial and impact positively on student learning (Brookhart & Chen, 2015; Grainger & Weir, 2016) however, the criticisms are increasing. Panadero and Jonsson (2020) after reviewing the research on rubric concerns and limitations, have identified a lack of empirical evidence and advocate for a pragmatic approach to investigate benefits and limitations in a scientific manner. Marker reliability is one area that raises concerns, especially where large numbers of students may be impacted by inconsistent marking. Scoring student work is complex and requires deep knowledge of the content along with "application of both objective and subjective criteria" (Herbert, Joyce, & Hassall, 2014). Some researchers argue that the inconsistencies of the marking process and the variability of standards have increased the burden for markers, increased workload, created assessment limitations and lengthens the time for feedback to students (Bloxham, 2009; Sadler, 2009). There is currently limited research on marker reliability on course work in higher education. However, there is a range of research on the use of rubrics and their ability to promote objectivity and improved consistency (Boettger, 2010; Crusan, 2010, 2015; Dempsey, PytlikZillig, & Bruning, 2009).

Marker variation impacts on marker reliability. High variation in marking refers to the provision of substantially different scores for the same task by a group of markers. Low variation indicates that markers have very similar scores.

Current knowledge of the factors impacting on variations in marks for higher education students is broad and well established in the literature, particularly in the assessment of students in higher education generally (Biggs, 2011; Peeters, Schmude, & Steinmiller, 2014). For example, Peeters et al. (2014) have noted that there are different types of rubrics and that the use of rubrics in conjunction with training, improve marker consistency. In fact, holistic rubrics were found to be most successful for inter-rater agreement.

Furthermore, students may compare assessment results with each other, thus noticing inconsistencies in marking. This leads to student dissatisfaction that drives the improvement of assessment practices (Field & Kift, 2010). However, attempts to work towards standards-based assessment using rubrics (Allen & Tanner, 2006; Barney, Khurum, Petersen, Unterkalmsteiner, & Jabangwe, 2012) and criterion-referenced rubrics (Barrie, Brew, & McCulloch, 1999; Burton, 2015) still does not meet the needs of academics dealing with large student cohorts who are required to undertake complex written tasks as part of their professional assessment requirements. Student dissatisfaction may also appear within this context when there is a discrepancy between both students' and markers' interpretations of the marking criteria and standards (Dawson, 2017; Nicol, 2009; Nicol & Colin, 2006).

2.1. Developing rubrics

Rubrics are widely used in the higher education sector in Australia and worldwide for summative assessment. They have also been used in the formative space to support student achievement of summative tasks. The rubric presents a set of guidelines to students and markers, outlining the requirements and expectations of the assessment piece (Denner, Salzman, & Harris, 2002; Popp, Ryan, Thompson, & Behrens, 2003), which have the potential of promoting student learning and improving instruction. Effective rubrics provide clarity and explicit criteria and descriptions for assessment tasks that are viewed positively by students and teachers (Bissell & Lemons, 2006; Morrell & Ackley, 1999; Nkhoma, Nkhoma, Thomas, & Le, 2020; Schamber & Mahoney, 2006; Shaw, 2004). Both standards-based rubrics and criterion-based rubrics (Barrie et al., 1999; Burton, 2015; Buragga, Khan, & Zaman, 2013; Cain, 2013; Chow, Ko, Li, & Zhou, 2012; McKenzie & Wood-Bradley, 2014) require the attention of markers who in turn are required to interpret the intention of the rubric statements. In addition, the concrete nature of rubrics and their specific elements allow for student self-assessment and feedback on a task (Hanna & Smith, 1998; Humphry & Heldsinger, 2014; Schamber & Mahoney, 2006).

Developing an appropriate marking rubric is a challenge in itself (Czaplinski, Senadji, Adie, & Beutel, 2014) and alignment of the rubric to the learning outcomes of coursework and programs of instruction forms a separate and complex process often not completed by the markers involved. The complexities of rubrics may lead to a marker interpretation issue. If a marking rubric is too prescriptive, it may reduce misinterpretation but discourages students' creativity and freedom and is often viewed in terms of student improvement in marks from one cohort to another (Kite & Phongsavan, 2017). On the other hand, a rubric with generalized statements of achievement as the basis for guidelines tends to open up the possibility for different interpretations by different markers.

2.2. Marker consistency

A well-developed marking rubric can improve consistency between markers, but variation still exists as markers may have different toler-

ance levels to errors made by the students that sit outside the text of the rubric and assessment documentation. The original indicators of clarity in rubrics found that they have evaluative criteria, quality definitions for those criteria at particular levels, and a scoring strategy (Crusan, 2010, 2015; Popham, 1997) yet these terms are open to interpretation. Dawson (2017) presented a summary of fourteen design elements evident in the literature which could be used to guide analysis of the rubric characteristics as they impact on marker variation. Of these elements, specificity, score strategy, evaluative criteria, quality levels, judgement complexity, quality definition and explanation were identified and applied to the evaluation of rubrics in this study, where we extended the examination of data from a previous study (Chakraborty et al., 2016) to investigate the effects of rubric quality on variation among markers. A brief definition of the elements we selected can be found in Table 1. These elements were considered identifiable issues through an examination of the documentation. Dawson (2017) describes seven others that include secrecy, exemplars, users and uses, creators, quality process, accompanying feedback information and quality definition. These were identified as issues that could not be fully comprehended through document analysis and would require further investigation beyond the scope of this study.

2.3. Moderation of marking

Formalized moderation processes, of which there are numerous interpretations, involve the marking of a small number of sample assessments by all the markers, and subsequent discussion among the markers and the course leader. A moderation process is often adopted to reduce marking variability in higher education institutions including in the Australian higher education sector (Malouff, 2008; Orr, 2007). In a moderation process, the moderator may choose to randomly check the marked papers and advise adjustments or may conduct formal blind marking sessions followed by groups of markers discussing the resulting marks. This process is often done with a small random sample due to high costs associated with gathering markers for such a process and is focused on the interpretation of assessment documentation such as rubrics and assessment task information provided to students. This process is also impacted by the ability of the institution to provide for payment of markers for such moderation processes. Failure to implement moderation may adversely affect results for some students. Moderation and feedback processes are widely accepted as norms of practice in higher education. Such techniques and approaches have improved marking fairness, but the study of marker variation might assist with verifying the quality or operational reliability and validity of assessments.

Further, variables such as the time of day that marking is undertaken, marking load, time pressures and cohort size, can all impact the marking process while diluting the impact of student empowerment through enhanced formative feedback processes prior to the submission of the task for final marking. This has a major impact on the learning outcome of the student and plays a significant role in the student learning process (Boud, 2015). Much of this formative feedback is centered on the documentation provided to students and the rubric delivered as part of the assessment process.

This study is grounded in the literature related to summative assessment tasks based on the production of a technical written text that was marked with the support of rubrics and moderation processes (Andrade, 2005; Reddy & Andrade, 2010). The researchers accept a broad definition of rubrics to be described as a scoring tool for the qualitative rating of complex student work that includes criteria and standards of attainment for those criteria. These are important for both the student and the marker (Arter & McTighe, 2001; Perlman, 2013).

Other aspects of assessment, including assessment task design, assessment pedagogy, assessment alignment and assessment feedback for both formative and summative assessment processes, are recognized as valuable components of assessment. However, this study will remain focused on marker variation and the influence of rubrics on marker variation.

3. Methodology

In this study, we followed the generic data analysis stages used by Chakraborty et al. (2016) to analyze the variations among markers in terms of the marks they have given to different student cohorts in different courses. These analysis results were then compared with the quality of marking rubrics to understand how rubrics may have impacted the variation in marking processes.

As shown in Fig. 1, the data analysis method consists of three key stages. In Stage 1, the necessary data collection is completed. Firstly, the appropriate sources of data are identified, and necessary approval

Table 1
Definition of seven of Dawson (2017) rubric elements.

Dawson's elements for rubrics	Brief definition of elements
Specificity	The rubric is specific to the task rather than generic.
Score strategy	The rubric contains either holistic or analytic scoring strategies. Analytic scoring strategies require the user to make a series of judgements about individual criteria.
Evaluative criteria	The rubric contains the expected task criteria usually listed on the left side.
Quality levels	The rubric usually has the quality levels in the top row. These quality levels correspond to the grading scheme.
Judgement complexity	Distinguishing between levels requires making an expert qualitative judgement using the descriptive information provided. Analytic decisions are less complex.
Quality definition	The rubric usually has descriptors that describe the quality levels for each criterion.
Explanation	Additional information is provided to the users of the rubric. EG: discussions, exemplars, task specifics

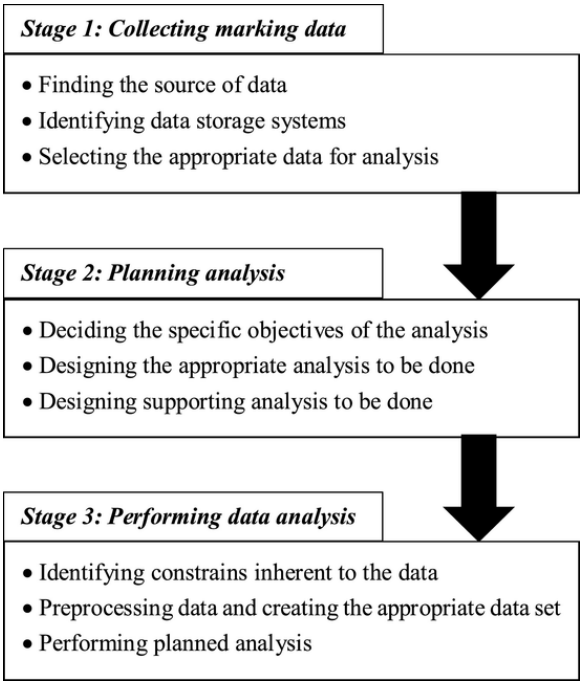


Fig. 1. Three stages of data analysis for detecting inconsistencies in marking (Chakraborty et al., 2016).

for data access is completed. The data storage mechanism and data collection support are then identified before actual data are collected. In Stage 2, the core purpose of data analysis is identified from the research hypothesis. Once the objectives are determined, appropriate analysis tools are selected. In Stage 3, various constraints in the data are identified and used to preprocess the data into a suitable set for analysis. Finally, data analysis is performed.

In this study, we identified the data source as the business faculty in an Australian university that uses the Moodle learning management system to manage student data, assessment submission, marking and results. Chakraborty et al. (2016) analyzed student results from the year 2015 for courses with two or more markers from the Management and Information Systems disciplines. The study used data from 71 undergraduate and postgraduate courses with student numbers in each course varying from 20 to 400. A total of 136 markers were involved in marking of these courses where some markers were involved in marking multiple courses. The earlier study showed marker variations of different degrees in different courses (Chakraborty et al., 2016). The data utilised in this study were collected retrospectively, which made it impossible to know what instructions were provided in each course other than the rubric. However, we found that the course coordinators in the selected university generally provide guidelines to markers on how to apply the rubrics and we assume that such instructions were provided in our selected courses.

In this study, we aimed to establish how the quality of marking rubrics related to marker variations. Before analyzing the documents and rubrics, the team identified Dawson's rubric design elements as the key framework for undertaking analysis. Dawson's summation of essential elements required to create a complete and successful rubric (Nkhoma et al., 2020) became the standard for analysis. The descriptions of each of the seven were taken from the original work of Dawson (2017) and were used as an on-balance interpretation of minimum requirements for an effective rubric. If on-balance, as interpreted by the researchers, the rubric content was as described by Dawson, then this determination was recorded. We selected a set of eight homogeneous courses from the larger set analyzed previously (Chakraborty et al., 2016) for the purpose of comparing their rubric quality with the level of variation among markers. All courses selected for this study were from postgraduate level with two major assignments as the assessment items. Courses were conducted in 2015 over at least two semesters and had at least two markers involved in the marking. All the courses had two essay type assignments. We selected four courses from both Management and Information System disciplines, to provide us with more understanding in two distinct areas. Of the four courses in each discipline, two were identified as having high variations among markers, and two had low variations in marks given. The data collected were marks given to students, the marker who marked the students, and the documents showing the marking rubrics with assignment specifications. A total of 43 markers were involved in marking these eight courses. Among the 43 selected markers, 7 were involved with two courses, while the rest only marked for one course only. All the markers involved were considered experienced as each had marked the same subject at least twice before this data collection period. All the data were anonymized using masking codes to de-identify student details, markers and courses.

Following Stage 2 of the methodology, we defined our objectives as i) identify and understand the marking variations in different courses, ii) understand the quality of assignment specification and marking rubrics of different courses, and iii) understand how rubrics and specifications could impact in marking variations. To achieve the objectives, we needed to analyze how the marking trends looked in different courses to identify variations. Self-consistency of individual markers needed to be analyzed over various marking tasks. To establish the impact that markers have on marking variation, we needed to analyze student performances in a course.

We followed the Stage 3 process to conduct the actual data analysis. To prepare the data set for analysis, we first defined data constraints and completed the preprocessing activities. The eight selected courses each had two assignments and were conducted in at least two semesters in 2015. The courses were coded as C1 through C8, and the markers were coded as M1 through M43. We defined a marking task as C*S*A* where C* is Course number, S* is the Semester number, and A* is the Assignment number. There was a total of 32 marking tasks defined in this study.

We then identified which of the markers marked each of the tasks. Two specific constraints were applied to the selected data set: i) Any marker marking fewer than ten students in a task was excluded from the final analysis as the small numbers would not provide meaningful trends for comparison, and ii) Students who did not submit a piece of assessment were excluded from the final analysis as those were not actually marked by markers and were skewing trends for markers with more of these students allocated to them.

After the preprocessed marking data set was generated, we conducted three distinct analyses on the marking data. We first conducted the course-specific trend analysis and analyzed the marking trends between markers in the selected eight courses to see how consistent each course was in terms of the marks given to students. Although the results would be influenced by the student cohort each marker marked, with random student allocation, the effect can be assumed to be evenly distributed. With large student cohorts, the trends tend to indicate the general leniency/toughness pattern of any individual marker. We then conducted the marker specific analysis. Marker specific data were analyzed across different marking tasks completed by each marker to understand marking consistency of individuals. Finally, students' performance data were analyzed to compare the performance trend of a student cohort in two different assignment pieces in a course.

Rubric analysis was undertaken using seven of Dawson (2017) fourteen design elements evident in the literature which could be used to guide analysis of the rubric characteristics as they impact on marker variation. These elements included specificity of the information to the course; evidence of a scoring strategy; the presence of evaluative criteria; the presence of quality levels; reliance on judgement complexity based on professional knowledge; presentation of the rubric information and explanation of task requirements.

4. Results

4.1. Results for variations between markers in courses

We calculated the general statistics of the datasets to understand the extent of the variations among markers in a single assessment piece and for a marker marking across multiple marking pieces. The extent of variations observed for the 43 markers across 32 assessment pieces is summarized in Table 2.

The variation in range was high but can be attributed to outliers where students did not perform well, or students dropped out. The standard deviation variation indicated there was high variability among markers in some courses. The high variation in IQR indicated that some markers marked in a flatter manner than others so that students received average marks within a narrow band. Although this could be highly dependent on the student cohort, our observation and analysis indicated that some markers indeed had a tendency towards narrow band marking, highlighting the need for further analysis. The high variation for the Coefficient of variation (Cv) indicated the high variability between some markers in several courses. Our analysis indicated that in all courses there were variations between markers highlighting the existence of bias among markers. Analysis results in the

Table 2
Overall variations in general statistics.

General statistics on variability	Variation among markers observed over the 32 assessments
Range	Up to 70 %
Quartile 1	Up to 15 %
Quartile 3	Up to 20 %
Median	Up to 20 %
Inter quartile range (IQR)	Up to 50 %
Mean	Up to 15 %
Standard deviation	Up to 60 %
Coefficient of variation (Cv)	Up to 50 %

next sections provide a better understanding of the bias among markers.

We completed the trend analysis for all the 32 marking tasks across the eight selected courses. The trendlines were drawn by applying the following process:

(i) Identify and separate students marked by each marker for a marking task in a course.
(ii) Sort in ascending order each set of student marks obtained in the earlier step.
(iii) Draw line graphs for the marks (Y-axis) against number of students (X-axis) for all markers in a task within a course.
(iv) Repeat the process for each task in each course.

Fig. 2 shows sample results from courses C1, C2, C3, and C4, which were found to exhibit low trend variations over their marking tasks. In these courses, we observed that marking trends between different markers involved were very similar. On the contrary, we observed higher variations in the marking trends among markers for courses C5, C6, C7, and C8, as shown in Fig. 3. We observed that marker variations in these courses were highly prominent. For example, for Course six in Fig. 3(b), marker M12 exhibits a very different trend profile than other markers, while M13 appears to be much more lenient than the remaining two while exhibiting a very narrow marking range.

Markers M10 and M11 appeared to have similar patterns, although it can be said overall M11 appeared to be lenient in his/her marking compared to M10. These results warrant further investigation into the consistency levels of individual markers as well as performance consistency among students.

4.2. Results for variations within markers across tasks

In this analysis, we saw the marking trends of individual markers while marking multiple tasks which may represent different courses, semesters and assignments. From the selected data set we identified that out of the 43 markers: six marked 1 task, 25 marked two tasks, four marked three tasks, three marked four tasks, and five marked six tasks. We analyzed the trends for each marker who marked two or more tasks and found that all markers were generally self-consistent in marking different tasks; hence markers were deemed consistent in marking.

Fig. 4 shows some representative results. Marker M4 in Fig. 4(a) marked tasks only from courses with low variations, marker M13 in Fig. 4(b) marked tasks only from high variation courses, while marker M27 marked tasks from both low and high variation courses. Despite marking different courses with different variations, it was evident that markers were generally self-consistent. Further investigation into the experience levels of the markers uncovered that all the markers had marked the same course at least once before. Thus, we can safely assume that markers were experienced enough to understand the requirements of the assessments. The results suggest that with marker specific variation being negligible in this study, we need to put more focus on the variations among markers and what guides their judgments. In the following section, we discuss the results for student cohort variation to understand the impact of student performances on marker variation.

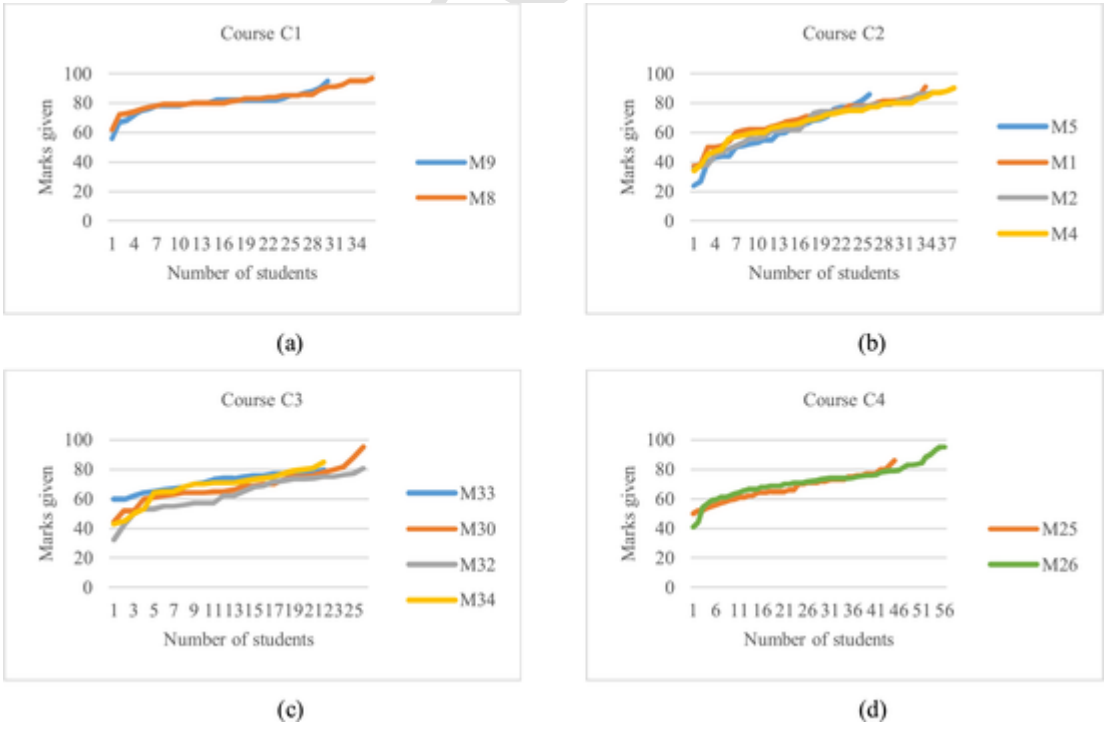


Fig. 2. Results for course trends for low variation courses: (a) In course C1 for task C1S3A2, (b) In course C2 for task C2S1A2, (c) In course C3 for task C3S3A1, (d) In course C4 for task C4S3A1.

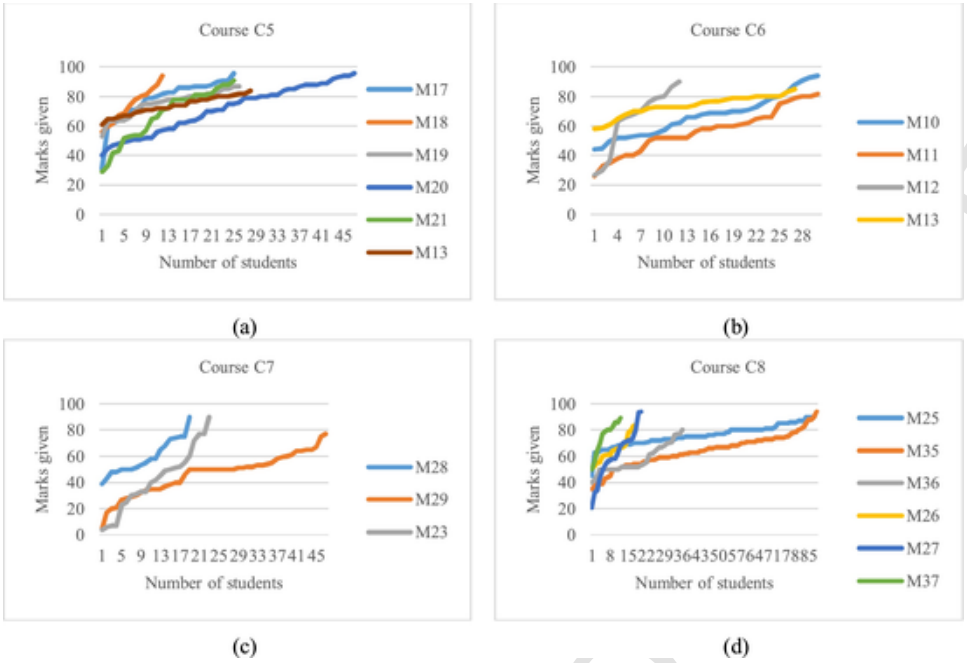


Fig. 3. Results for course trends for high variation courses: (a) In course C5 for task C5S3A1, (b) In course C6 for task C6S3A1, (c) In course C7 for task C7S1A2, (d) In course C8 for task C8S1A1.

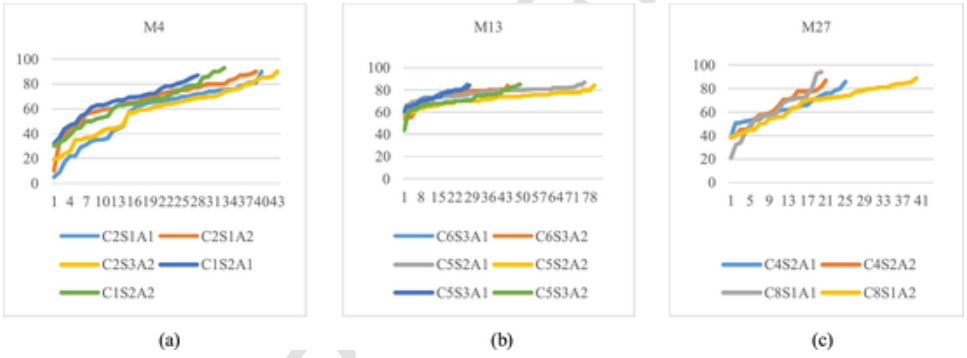


Fig. 4. Results for trend variations for individual markers: (a) Trends for Marker M4, (b) Trend for Marker M13, (c) Trends for Marker M27.

4.3. Results from document analysis

In this analysis, the task sheets/specifications and task rubrics were compared according to the variation between grades and categorized into four groups. These were low variation between grades with rubrics demonstrating the seven rubric qualities identified earlier (Dawson, 2017). The second phase was to examine the course rubric in order to identify commonalities and differences between the two groups. A poor rubric did not indicate higher qualities within the seven elements and did not provide an explicit and transparent framework for markers (Biggs, 2011) or used language that was difficult to interpret. The groups were characterized by

- Low variation with sound assessment rubrics –two courses
- Low variation with poor rubrics –two courses
- High variation with poor rubrics –four courses
- High variation with sound assessment rubric – None identified

Table 3 shows the eight courses with an indication of the presence of the seven elements selected from the fourteen identified by Dawson (2017). The greyed 'Yes' indicates that the element was evidenced in the rubric while the 'No' indicates where a rubric did not demonstrate the element as described in Dawson (2017). The low variation in marks found in the data have positive indicators for the elements, while those with high variations hold fewer positive indications. It is noted in the data that courses three, four and five only had two, and two and one positive indicators respectively yet still had low variation. This is critical and indicates further examination is required to understand how the variation emerged. The final row of the table shows the length of assessment associated documentation presented to support the rubric and the length of the rubric. It is interesting that for courses three, four and five, the support information was lengthy and wordy, particularly given that these courses had no quality definitions on the rubrics.

4.3.1. Rubric examples

Examples of a rubric from Course C1 and Course C8 are provided in Appendices A and B, respectively. The rubric for Course C1 (Appendix A) demonstrates a somewhat effective rubric with all seven elements provided. The researchers note that while this is one of the better rubrics found in the business courses that participated in the study, there is still room for improvement to provide even further clarity for

Table 3
Course element alignment.

	Course C1	Course C2	Course C3	Course C4	Course C5	Course C6	Course C7	Course C8
	Management	Management	Information System	Information System	Management	Management	Information System	Information System
	Low Variation	Low Variation	Low Variation	Low Variation	High Variation	High Variation	High Variation	High Variation
Specificity	Yes	Yes	No	No	No	No	No	No
Scoring strategy	Yes	Yes	Yes	Yes	Yes	Yes	Limited	Yes
Evaluative Criteria	Yes	Yes	Yes	Yes	No	No	No	No
Quality Levels	Yes	Yes	No	No	No	No	No	No
Judgement complexity	Yes	Yes	Yes	Yes	No	No	No	No
Quality definitions	Yes	Yes	Limited	Limited	No	No	No	No
Explanation	2 tasks 3000 words 2-3 pages of detailed support each	2 tasks 3000 words 6 pages of detailed support each	3 tasks 2500 words 4 pages of detailed support each	3 tasks 3000 words 5 pages of detailed support each	2 tasks 2000 words 4 pages of detailed support each	2 tasks 2500 words 2-4 pages of technical support each	2 tasks 2000 words 3-4 pages of functional support each	2 tasks 3000 words 2-3 pages of functional support each
	4-5 pages criteria each	2 pages criteria each	2 pages criteria each	2 pages criteria each	2 pages criteria each	2 pages criteria each	1 pages criteria each	1 pages criteria each

markers and student understanding with the quality definitions. In Table 4, the quality definitions include 'no introduction provided'; 'poor introduction provided, or relevant information' and so forth. Table 4 shows an excerpt from the rubric for Course C1:

However, when compared with Course C8 (Appendix B), it was clear that many of the elements were missing in the Course C8 rubric and therefore could lead to greater disparity in marking and student understanding of the task as shown in Table 5. While both contained 'evaluative criteria' and a 'scoring strategy', one of the most powerful elements to assisting with marker consistency were the 'quality definitions'. The C1 rubric had rudimentary definitions which allowed for some disparity in marking, however, C8 allowed for greater disparity in judgement. To improve C1, more detail with the 'quality definitions' is needed. For example, a description or list of what is required to have a good clear introduction, an adequate introduction and so forth will improve marking decisions and lead to greater marker consistency (Dawson, 2017)

Experienced educators who are familiar with rubrics, especially effective rubrics, may likely view Table 4 and think that it is not a rubric. These authors would agree with that evaluation and argue that rubrics, like those in Course C8, are more similar to task score sheets.

5. Discussion

5.1. Marker consistency

This research is unique in that it has examined marker variance both within and between courses. Our results reveal that regardless of the rubric quality, markers demonstrated strong self-consistency across the written essay type texts they marked within courses. Markers involvement with moderation processes and their previous experience with the courses have likely improved consistency. This ability to maintain consistency with each marker is important for overall marking consistency. The clarity of the criteria, scoring levels, scoring strategy, judgement complexity and quality definitions of the rubric contributes to the markers' ability to interpret the assessment intentions which leads to consistent marker judgement across all markers. We contend that the clarity of these elements influence marker consistency both within and across courses and also increases reliability (Brookhart & Chen, 2015).

Research focused on inter-rater reliability between groups has identified that there are multiple ways to identify consistency and reliability with rubrics and marking (Graham, Milanowski, & Miller, 2012; Iacobucci & Duhachek, 2003). Our results identified that there is consistency and inconsistency demonstrated between courses. Although markers were self-consistent, other variables such as cohorts, types of

Table 4
Example from Course C1 rubric – added Dawson's (2017) elements found in the rubric.

	(QUALITY LEVELS) Non-compliant	Poor/inadequate	Basic	Adequate	Good	Excellent	Mark
Assignment requirements out of 60	(SCORING STRATEGY)0	0-29	30–38	39–44	45–50	51–60	/60
(EVALUATIVE CRITERIA) 1a. Introduction-provides an explanation of the reason for preparing the document and a brief summary of the contents.	(QUALITY DEFINITION) No introduction provided.	Poor introduction provided, or irrelevant information.	Basic introduction provided.	Adequate introduction provided.	Good clear introduction provided.	Clear, concise and comprehensive introduction providing essential details.	

Table 5
Example from Course C8 rubric- added Dawson (2017) elements found in the rubric.

Item	Max marks	Marks obtained
(EVALUATIVE CRITERIA)	(SCORING STRATEGY)	
In-Depth discussion on implementation strategy The quality and depth, relevance and applicability to the content chosen and solution conceived.	10	

tasks, varying amounts of experience with each course may impact on the inconsistency across courses. According to Wolfe, Song, and Jiao (2016), other factors related to writing tasks that influence marker variance include long essays, essays with less lexical difficulty and high-quality essays. The written tasks for this study were between 2000 words and 3000 words. We consider a 2000-word assignment to be relatively average assignment length and a 3000-word assignment to be somewhat high, we noticed in this study that there was low marker variance with courses that had sound rubrics for 3000-word assignment. In contrast, the high marker variation included assignments from 2000 to 3000 words but included poor rubrics. This contradicts the findings by Wolfe et al. (2016). Further, we found that 3000-word assignments were part of both low and high marker variance and consequently believe it is the rubric that influenced the difference in marker variation in this study.

5.2. Alignment to Dawson's rubric elements

Variation in the range of marks offered is not unusual, but the results showed high levels of variation in the marks given by different markers. To explore possible causes of this variation, we overlaid rubric and assessment document analysis which indicated implications for rubric and assessment design. The following discusses the impact of the seven elements identified by Dawson (2017) on marker variation to illuminate how sound assessment rubrics can result in lower marker variation.

The researchers looked for seven of Dawson (2017) fourteen rubric design elements because we felt those were the basic elements or the starting point for effective rubrics. We therefore sought to determine if any of those elements were included in assessment rubrics from eight different business courses and then compared for marker variance for written essay type assignments. We expected that courses with low marker variance would meet more of Dawson's seven key rubric elements, and those with high marker variance would meet fewer of the elements. What we found was that two of the four low marker variance courses aligned with all seven of the elements; however, the other two only met about half of the elements. In contrast, the four high marker variance courses met one or none of Dawson's elements. This raises questions about the elements and their importance for marker variance. These authors argue that the more of Dawson's components that are effectively created within a rubric, the less marker variance there will be, as discussed below.

Two of Dawson's elements were included consistently in the four low marker variance courses. 'Judgement complexity' was identified in all four courses. This refers to the amount of difficulty and complexity to make a judgement which is based on the markers' ability to make judgements for each of the criteria in a consistent manner. Often this judgement complexity is increased when there is detailed information provided within the rubric to assist with the decision making. We found that the markers were able to make sound judgements for each criterion, increasing reliability between markers. 'Quality definitions' which refers to the clarity of definitions required for each level of scoring and may be referred to as a benchmark to attain a particular score, was the second element met in all four low marker variance courses. Benchmarks may increase agreement, but they should be carefully written to maintain consistency in scoring (Denner et al., 2002; Popp et al., 2003). A rubric that is clear, descriptive and has a scoring system that describes the required expectations is likely to be more objective and less subjective. Therefore, we can conclude that rubric effectiveness will impact on consistency and reliability of marking within and across courses.

Rubrics are an expectation at this university, and course coordinators are expected to seek required training and execute rubrics in their courses. There is a discrepancy between courses in the knowledge of rubric elements and the quality of rubrics within the business faculty. Unfortunately, the lack of clarity for markers in how to judge for points can lead to a lack of consistent marking within a course, as was demonstrated in this study. A clear indicator of high variation was the prevalence in rubrics with terms such as 'Excellent, Very Good, Good, Average, Poor and Very Poor' and no other definition to assist with judging marking. Rubrics can be viewed as a device to regulate scoring of assessment and may be more reliable than not using one. Rubrics are also used as support for learners, support of learning and teaching approaches, and to improve reliability and validity (Nkhoma et al., 2020). This is important to students as they link this to fairness.

5.3. Support documentation and rubric characteristics

An interesting aspect to the business courses' approach to preparing students for assessment is the overwhelming amount of support documentation provided to students. These documents were additional information regarding task-specific information which in many cases provided the deeper level of information that was not included in the rubrics. While this information is useful for students, it does not assist the use of the rubric for marking. However, the researchers believe it may contribute to marker self-consistency and ease student concerns related to task uncertainty.

In summary, the key differences between the courses in this study were found to lie in the presence of quality definitions within the rubrics. When quality definitions were provided to a marker, they reduced the level of variation in marks and although outside of our scope, would appear to increase the transparency of the assessment process for students (Hendry, Bromberger, & Armstrong, 2011). The quality definitions act as a standard for each of the quality levels and when written with clear descriptions of the task expectations for each level of scoring, will reduce marker variation. In contrast, the generic nature of the criteria and/or quality definitions of the rubrics, for example, "Based on the information provided, summarize the industry and market background…" was unlikely to provide clarity for a marker and consequently not lead to consistency between markers in a course.

The use of task explanation documents did not indicate a distinct difference between course documentation as all courses in this study had extensive prescriptive information about the expectations of the task. It was thought that the explanation documents would decrease variation in student marks and may have with each markers' ability to be self-consistent while marking.

5.4. Implications for practice

The findings support the use of effective rubrics to reduce marker variation and contribute to increased quality in marker judgement. Our results indicated that some university staff are unclear about the components of effective rubrics and therefore we suggest there is a need for training of rubric components and rubric creation, especially around developing an understanding of quality definitions. It is the clarity of these definitions that aid in marker consistency and are, therefore, a vital part of effective rubrics. The findings also indicate that markers who have experience with courses and undergo moderation processes are able to have strong self-consistency while marking. Therefore, we support the repeated use of markers and moderation processes that encourage high marker consistency.

5.5. Limitations

The authors recognize the limitations of this study to include the omission of seven of the Dawson (2017) elements and that these elements are also worthy of further research in relation to rubric design and marker variation. This study is based on retrospective data that limits some detail around markers and their use of the rubrics to assist in marking assignments. There are limitations to this study which include the narrow nature of the assessment tasks used and the focus on the traditional summative assessment process. There is further need to study individual student performance fluctuations using long term performance, such as interim GPA to quantify the impact of student performance on marks variation in courses demonstrating clear quality elements. The individual marker consistency requires further study using long term trends of individual markers over multiple courses. A further limitation was the detail of course information available since these courses and rubrics were from previous course iterations and held in university storage systems. Future research using a single course and the rubrics from that course over an extended period of time to investigate marker consistency, marker qualifications and marker training would be beneficial. Another area of research to investigate is rubric interpretation by students. Future research should investigate the effect of rubric quality and consequent interpretation by students on the learning outcomes as assessed by markers.

6. Conclusions

The results indicate that failure to employ high-quality rubrics does have an impact on the resultant reliability and validity of the assessment marks where the task is text-based and requires disparate markers of large cohorts. The seven rubric qualities used here have been shown to have a collective impact with elements such as quality definitions being the most impactful on marker variation. The authors recognize that the omission of seven of the elements identified by Dawson (2017) is a limitation of this study and offers future work in the area.

This study extended on previous research that conducted an analysis of variation between and within markers across 71 courses in a higher education program (Chakraborty et al., 2016). Close analysis of a subset of those courses, including the rubrics and other documentation used to guide markers, found that consistency of an individual marker is influenced by the construction and content of the rubric and assessment documentation. Marker variation turns out to be a key factor impacting on the reliability of an assessment task that is text-based and highly technical in nature and can be influenced by the presence of quality definitions within the rubric. The study establishes that there is significant impact of marker variation on marks obtained by students and that sound rubric elements can reduce the variation in marker assessments.

Declaration of Competing Interest

The authors report no declarations of interest.

Appendix A.

RUBRIC FOR COURSE 1

ASSIGNMENT 2 MARK SHEET - PROJECT-BAS

Note to markers - indicate level of achievement for each row by highlighting relevant text

	Non-compliant	Poor/inadequate	Basic	Ad
ASSIGNMENT REQUIREMENTS (out of 60)	0	0 to 29	30 to 38	39
Executive summary (ES) to indicate the nature of the assignment, background details, research findings and recommendations where appropriate, and to function as a stand-alone document	No ES provided.	Poor quality ES that does not provide the required information	Basic ES that provides limited information, or does not provide information in a clear and concise manner	Ad pro inf
1. Introduction – why your report has been prepared, what project is being analysed, and how your assignment is structured	No introduction provided	Poor introduction provided, or irrelevant information	Basic introduction provided	Ad pro
2. Background and environment of project – enough detail for the reader to understand the nature and objectives of the project and its sponsor	No background provided	Inadequate information provided	Basic or limited background information provided	Ap de
3. Audience and structure of the PMP				
• **Audience of PMP** – who is the intended audience for this PMP? Why? What do they need? In what way do they need it? Etc.	No discussion on the topic	Little or poor discussion on the topic	Basic discussion on the topic, with limited use of theory to support the discussion	Ad top the
• **Structure of the PMP** - Discuss how the PMP has been structured, and why the respective sections have been included	No discussion on the topic	Little or poor discussion on the topic	Basic discussion on the topic, with limited use of theory to support the discussion	Ad top the
4. Contents of the PMP sections				
• Discuss each of the major sections of the PMP and provide some theoretical framework for what is included	No discussion on the topic	Little or poor discussion on the topic	Basic discussion on the topic, with limited use of theory to support the discussion	Ad top the
4. Conclusions – Summarise this discussion of the PMP in relation to your project. What did you find out about the PMP? Does it conform to the recognised theory? What aspects are important and are of significance to the management of this, and similar, projects?	No conclusions provided	Poor or inadequate conclusions provided	Basic conclusions provided	Ad de pr
7. Recommendations – What are your recommendations flowing from this analysis? What changes could or should be made for management of this, or similar, projects.	No recommendations provided	Poor or inadequate recommendations provided	Basic recommendations provided	Ad re pr
RESEARCH & ACADEMIC THEORY (out of 30)	0	1 to 14	15 to 19	2

Appendix B.

RUBRIC FOR COURSE 8

Marking criteria - **Assignment B**
Marking Sheet for
Student name: ---------- Student # ----------

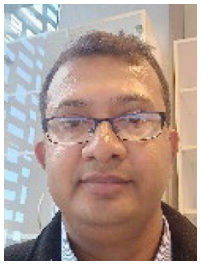
Item	Ma Ma
Cover letter including fee schedule, time and resource schedule, executive summary, other document presentation aspects	10
How feedback provided in the first report (Assignment A) can be addressed	5
In-Depth discussion on implementation strategy The quality and depth, relevance and applicability to the context chosen and solution conceived	10
Discussion on Information Technology Infrastructure Critical analysis and implications of the above to the chosen context including why such an infrastructure was chosen, justification, relevance to the solution, various considerations identified for the pilot project and its duration, long term future of the infrastructure	15
Discussion on strategies for operational excellence through pilot project Various KPI – relevant to the solution suggested, how to measure these KPIs, how did you validate the KPI measurement, what operational aspects have been considered for measurement, articulation of these leading to relevance, appropriateness and feasibility	15
Discussion on enhancement of decision making What decision making aspects have been considered, how did you integrate these with strategies, how did you demonstrate 'excellence', the impact of cost resources	15
Management of Global Market How did you identify various management aspects (within the context), the quality of these, justification, assumptions in identifying these management aspects, global constraints and variables	15
Overall quality of presentation, depth of arguments, professional presentation, grammar, punctuation, adhering to word count, cohesion of various document structures	15
Late submission deduction	

References

I., Albluwi (2018). A closer look at the differences between graders in introductory computer science exams. IEEE Transactions on Education, 61(3), 253–260.
D., Allen, & K., Tanner (2006). Rubrics: Tools for making learning goals and evaluation criteria explicit for both teachers and learners. CBE—Life Sciences Education, 5(3), 197–203.
H.G., Andrade (2005). Teaching with rubrics: The good, the bad, and the ugly. College Teaching, 53(1), 27–31.
J., Arter, & J., McTighe (2001). Scoring rubrics in the classroom: Using performance criteria for assessing and improving student performance. Corwin Press.
S., Barney, M., Khurum, K., Petersen, M., Unterkalmsteiner, & R., Jabangwe (2012). Improving students with rubric-based self-assessment and oral feedback. IEEE Transactions on Education, 55(3), 319–325.
S., Barrie, A., Brew, & M., McCulloch (1999). Qualitatively different conceptions of criteria used to assess student learning. Proceedings of the AARE-NZARE Conference.
J.B., Biggs (2011). Teaching for quality learning at university: What the student does. McGraw-hill education (UK).
F.L., Bird, & R., Yucel (2013). Improving marking reliability of scientific writing with the Developing Understanding of Assessment for Learning programme. Assessment & Evaluation in Higher Education, 38(5), 536–553.
A.N., Bissell, & P.P., Lemons (2006). A new method for assessing critical thinking in the classroom. BioScience, 56(1), 66–72.
P., Black, & D., Wiliam (1998). Assessment and classroom learning. Assessment in Education: Principles, Policy & Practice, 5(1), 7–74.
S., Bloxham (2009). Marking and moderation in the UK: False assumptions and wasted resources. Assessment & Evaluation in Higher Education, 34(2), 209–220.
S., Bloxham, P., Boyd, & S., Orr (2011). Mark my words: The role of assessment criteria in UK higher education grading practices. Studies in Higher Education, 36(6), 655–670.
R.K., Boettger (2010). Rubric use in technical communication: Exploring the process of creating valid and reliable assessment tools. IEEE Transactions on Professional Communication, 53(1), 4–17.
D., Boud (2015). Feedback: Ensuring that it leads to enhanced learning. Clinical Teacher, 12(1), 3–7.
D., Boud, & N., Falchikov (2006). Aligning assessment with long-term learning. Assessment and Evaluation in Higher Education, 31(4), 399–413.
D., Boud, & E., Molloy (2013). Rethinking models of feedback for learning: The challenge of design. Assessment and Evaluation in Higher Education, 38(6), 698–712.
S.M., Brookhart, & F., Chen (2015). The quality and effectiveness of descriptive rubrics. Educational Review, 67(3), 343–368.
S., Brown, & P., Knight (2012). Assessing learners in higher education. Routledge.
G., Brown, S., Irving, & P., Keegan (2008). An introduction to educational assessment, measurement, and evaluation: Improving the quality of teacher-based assessment. New Zealand: Pearson Education.
K.A., Buragga, A.R., Khan, & N., Zaman (2013). Rubric based assessment plan implementation for Computer Science program: A practical approach. Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE).
K., Burton (2015). Continuing my journey on designing and refining criterion-referenced assessment rubrics. Journal of Learning Design, 8(3), 1–13.
A., Cain (2013). Developing assessment criteria for portfolio assessed introductory programming. Proceedings of 2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE).
S., Chakraborty, X., Zhou, A., Hafeez-Baig, R., Gururajan, M., Paul, A., Mandal, … P.D., Barua (2016). Objective analysis of marker bias in higher education. 2016 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE).
C.C., Chase, & D., Klahr (2017). Invention versus direct instruction: For some content, it'sa tie. Journal of Science Education and Technology, 26(6), 582–596.
T., Chow, E., Ko, C., Li, & C., Zhou (2012). The systematic development of rubrics in assessing engineering learning outcomes. Proceedings of IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE) 2012.
D., Crusan (2010). Assessment in the second language writing classroom. University of Michigan Press.
D.J., Crusan (2015). Dance, ten; looks, three: Why rubrics matter. Assessing Writing, 26, 1.
I., Czaplinski, B., Senadji, L., Adie, & D., Beutel (2014). Analysis of moderation practices in a large STEM-focused faculty. 2014 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE).
C., Dann, B., Dann, & S., O'Neill (2018). Formative assessment via video feedback on practicum: Implications for higher education and professional teacher accreditation bodies. Formative Assessment Practices for Pre-Service Teacher Practicum Feedback: Emerging Research and Opportunities (pp. 158–183).
P., Dawson (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. Assessment & Evaluation in Higher Education, 42(3), 347–360.
M.S., Dempsey, L.M., PytlikZillig, & R.H., Bruning (2009). Helping preservice teachers learn to assess writing: Practice and feedback in a Web-based environment. Assessing Writing, 14(1), 38–61.
P.R., Denner, S.A., Salzman, & L.B., Harris (2002). Teacher work sample assessment: An accountability method that moves beyond teacher testing to the impact of teacher performance on student learning.
V., Devedzic, & M., Devedzic (2019). Technology-Enhanced Assessment at universities and in schools: An initiative. International Journal of Learning and Teaching, 11(3), 89–98.
Z., Dong, H., Fan, M., Zhang, L., Ye, J., Song, & Q., Wang, et al. (2014). Individual Evaluation for Freshman in Small Size Group (pp. 453–456).
R., Field, & S., Kift (2010). Addressing the high levels of psychological distress in law students through intentional assessment and feedback design in the first year law curriculum. Student Success, 1(1), 65.
R., Fuller, M., Homer, G., Pell, & J., Hallam (2017). Managing extremes of assessor judgment within the OSCE. Medical Teacher, 39(1), 58–66.
M., Graham, A., Milanowski, & J., Miller (2012). Measuring and promoting inter-rater agreement of teacher and principal performance Ratings Online Submission.
P., Grainger, & K., Weir (2016). An alternative grading tool for enhancing assessment practice and quality assurance in higher education. Innovations in Education and Teaching International, 53(1), 73–83.
V., Grion, A., Serbati, E., Felisatti, & L., Li (2019). Peer feedback and technology-enhanced assessment as critical issues to foster student learning. Italian Journal of Educational Research, 9–14.
M.A., Hanna, & J., Smith (1998). Using rubrics for documentation of clinical work supervision. Counselor Education and Supervision, 37(4), 269–278.
S.J., Hanrahan, & G., Isaacs (2001). Assessing self-and peer-assessment: The students' views. Higher Education Research & Development, 20(1), 53–70.
G.D., Hendry, N., Bromberger, & S., Armstrong (2011). Constructive guidance and feedback for learning: The usefulness of exemplars, marking sheets and different types of feedback in a first year law subject. Assessment & Evaluation in Higher Education, 36(1), 1–11.
I.P., Herbert, J., Joyce, & T., Hassall (2014). Assessment in higher education: The potential for a community of practice to improve inter-marker reliability. Accounting Education, 23(6), 542–561.
S.M., Humphry, & S.A., Heldsinger (2014). Common structural design features of rubrics may represent a threat to validity. Educational Researcher, 43(5), 253–263.
D., Iacobucci, & A., Duhachek (2003). Advancing alpha: Measuring reliability with confidence. Journal of Consumer Psychology, 13(4), 478–487.
JISC (2005). Effective practice with e-assessment: An overview of technologies, policies and practice in further and higher education. Joint Information Systems Committee.
L., Jones, B., Allen, P., Dunn, & L., Brooker (2017). Demystifying the rubric: A five-step pedagogy to improve student understanding and utilisation of marking criteria. Higher Education Research & Development, 36(1), 129–142.
J., Kite, & P., Phongsavan (2017). Evaluating standards-based assessment rubrics in a postgraduate public health subject. Assessment & Evaluation in Higher Education, 42(6), 837–849.
J., Malouff (2008). Bias in grading. College Teaching, 56(3), 191–192.

S., McKenzie, & G., Wood-Bradley (2014). Using rubrics in IT: Experiences of assessment and feedback at Deakin University. 2014 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE).
P.D., Morrell, & B.C., Ackley (1999). Practicing what we teach: Assessing pre-service teachers' performance using scoring guides.
D., Nicol (2009). Assessment for learner self-regulation: Enhancing achievement in the first year using learning technologies. Assessment & Evaluation in Higher Education: Emerging Frameworks for the Use of E-Assessment in Higher Education, 34(3), 335–352.
D., Nicol, & M., Colin (2006). Rethinking technology-supported assessment practices in relation to the seven principles of good feedback practice. Innovative assessment in higher education. (pp. 84–98). Routledge.
C.A., Nkhoma, M., Nkhoma, S., Thomas, & N.Q., Le (2020). The role of rubrics in learning and implementation of authentic assessment: A literature review. SITE 2020: Informing Science + IT Education Conferences. Online.
S., Orr (2007). Assessment moderation: Constructing the marks and constructing the students. Assessment & Evaluation in Higher Education, 32(6), 645–656.
E., Panadero, & A., Jonsson (2020). A critical review of the arguments against the use of rubrics. Educational Research Review, 30, 100329.
E., Panadero, M., Romero, & J.-W., Strijbos (2013). The impact of a rubric and friendship on peer assessment: Effects on construct validity, performance, and perceptions of fairness and comfort. Studies in Educational Evaluation, 39(4), 195–203.
M.J., Peeters, K.A., Schmude, & C.L., Steinmiller (2014). Inter-rater reliability and false confidence in precision: Using standard error of measurement within PharmD admissions essay rubric development. Currents in Pharmacy Teaching and Learning, 6(2), 298–303.
D.J., Perlman (2013). Effective teaching and motivation: Application of self-determination theory.
W., Popham (1997). What's wrong–and what's right–with rubrics. Educational Leadership, 55(2), 72–75.
S.E.O., Popp, J.M., Ryan, M.S., Thompson, & J.T., Behrens (2003). Operationalizing the rubric: The effect of benchmark selection on the assessed quality of writing.
Y.M., Reddy, & H., Andrade (2010). A review of rubric use in higher education. Assessment and Evaluation in Higher Education, 35(4), 435–448.
B., Saddler, & H., Andrade (2004). The writing rubric. Educational Leadership, 62(2), 48–52.
D.R., Sadler (2007). Perils in the meticulous specification of goals and assessment criteria. Assessment in Education, 14(3), 387–392.
D.R., Sadler (2009). Indeterminacy in the use of preset criteria for assessment and grading. Assessment & Evaluation in Higher Education, 34(2), 159–179.
D.R., Sadler (2010). Beyond feedback: Developing student capability in complex appraisal. Assessment & Evaluation in Higher Education: Approaches to Assessment that Enhance Learning, 35(5), 535–550.
J.F., Schamber, & S.L., Mahoney (2006). Assessing and improving the quality of group critical thinking exhibited in the final projects of collaborative learning groups. The Journal of General Education, 103–137.
K.M., Shaw (2004). Using feminist critical policy analysis in the realm of higher education: The case of welfare reform as gendered educational policy. The Journal of Higher Education, 75(1), 56–79.
M.A., Stellmack, Y.L., Konheim-Kalkstein, J.E., Manor, A.R., Massey, & J.A.P., Schmitz (2009). An assessment of reliability and validity of a rubric for grading APA-style introductions. Teaching of Psychology, 36(2), 102–107.
E.W., Wolfe, T., Song, & H., Jiao (2016). Features of difficult-to-score essays. Assessing Writing, 27, 1–10.
C., Zhao, & J., Huang (2020). The impact of the scoring system of a large-scale standardized EFL writing assessment on its score variability and reliability: Implications for assessment policy makers. Studies in Educational Evaluation, 67, 100911.

Biography



Subrata Chakraborty (M'06 SM'19) is a Senior Lecturer in the School of Information, Systems & Modelling within the Faculty of Engineering & IT, University of Technology Sydney (UTS), Australia. He is also a core member of the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS) at UTS. He received his PhD in Decision Support Systems from Monash University, Australia. Previously Dr Chakraborty worked as an academic with University of Southern Queensland, Charles Sturt University, Queensland University of Technology, and Monash University. Dr Chakraborty's current research interests include Optimisation Models, Data Analytics, Machine Learning, and Image Processing with decision support applications in diverse domains including Business, Agriculture, Transport, Health, and Education. Dr Chakraborty is a Senior Member of IEEE and a Certified Professional Senior Member of ACS.

Christopher Dann has 34 years of experience in all areas of Education, moving into Initial Teacher Education in 2005. His Master of Research into the Internationalization of teacher education was followed by a PhD form Edith Cowan University in the use of mobile technologies in the formative assessment of preservice teachers while on practicum. His experience across educational sectors and between the education industry and the technology industry give him a unique view of the forces behind commercial influences, institutional influences and school-based essentials. Teachers and leadership teams are being presented with emerging possibilities to enhance their programs at a time when student assessment processes are becoming key components for graduating preservice teachers. Chris' personal goal is to improve the feedback and feed forward process using technology for learners.

Anuradha Mandal is a researcher with the School of Information, Systems & Modelling within the Faculty of Engineering & IT, University of Technology Sydney (UTS), Australia. She obtained her Master of Computer Application degree from University of Pune, India. Anuradha has over 15 years of industry and 5 years of academic experiences. Her research focus has been in the areas of education data analytics.

Beverly Dann has worked in primary schools as a teacher, deputy, and curriculum leader for 26 years in the USA and Australia. She is involved in teacher education with the Bachelor of Primary and the Bachelor of Early Years programs. She is currently part of a team writing the new Master of Teaching (Primary) program that will begin offerings in 2018. Beverly has a passion for science that is contagious and clearly evident in her teaching. Her contemporary approach to teaching engages learners of all ages through an inquiry and hands-on approach to teaching and learning. She often incorporates digital technologies in her courses that are transferable to the classroom setting. Beverly has recently completed her PhD in Primary Science Education.

Manoranjan Paul (M'03 SM'13) received PhD degree from Monash University, Australia in 2005. He was a Post-Doctoral Research Fellow in the University of New South Wales, Monash University, and Nanyang Technological University. Currently he is a full Professor, Director of Computer Vision Lab, and Leader of Machine Vision & Digital Health (MaViDH) Research Group at Charles Sturt University, Australia. His major research interests are in the field of Video Coding, Image Processing, Digital Health, Wine Technology, Machine Learning, EEG Signal Processing, Eye Tracking, and Computer Vision. He has published around 200 peer reviewed publications including 72 journals. He was an invited keynote speaker in IEEE DICTA-17 & 13, CWCN-17, WoWMoM-14, and ICCIT-10. Currently, he is an Associate Editor of three top ranked journals: IEEE Transactions on Multimedia, IEEE Transactions on Circuits and Systems for Video Technology, and EURASIP Journal in Advances on Signal Processing. He was a General Chair of PSIVT-19 and Program Chair of PSIVT-17 and DICTA-18. He was awarded the ICT Researcher of the Year 2017 by Australian Computer Society. He obtained more than $3.6 million competitive external grant including Australian Research Council (ARC) Discovery grants, Australia-China grant. He has supervised 15 PhD students to completion.

Abdul Hafeez Baig holds a PhD in the domain of healthcare and information systems from USQ, Australia and master's degrees in MIS and MBA from Griffith University, Brisbane, Australia. Since joining USQ in Jan 2004, he has published more than 100 referred publications and research grants of 3 million in the domains of health and education regarding technology management. He is very conversant with wireless technology as well as emerging technologies and learning management systems. He teaches information systems concepts to both undergraduate and postgraduate students, including MBA students. He has also supervised numerous post graduate research students. He also has numerous publications in academic and scholarly journals, and has a vast array of scholarly conference papers, all of which have focused on the domain of information systems. He has extensive experience in the area of information systems, especially relates to the healthcare sector. He is very interested in wireless healthcare applications, systems analysis and design, adoption, the infusion and diffusion of information technology, knowledge management, technology management, digital innovation, M-learning and E-Commerce, outsourcing, networking, healthcare and information technology, and the re-engineering of business processes.