Volume 5, Issue 5

Malaysia Journal of Invention and Innovation

https://journal.academicapress.org/aps/index.php/mjii

Research Article

RUBIQUANT: An Intelligent Computational Framework for Governed Rubric Assessment Using the Adaptive Weighted Penalty Model

Mohamad Irwan P. Harahap[1, *]

1 Pengajian Kejuruteraan Awam, UiTMPP, MALAYSIA 1; irwan1971@yahoo.com
* Correspondence: irwan1971@yahoo.com.

Keywords:
rubric assessment
decision support system
computational assessment
educational technology
formative evaluation

Abstract:
Rubric-assessment approach is commonly adopted in tertiary institutions, but the application of such assessment technique continues to be inconsistent due to manual computation of scores, assessor subjectivity in marking, and lack of transparency in the calculation of marks. This paper outlines the development and adoption of RUBIQUANT, an automatic rubric assessment approach, implemented in Microsoft Excel and Visual Basic for Applications, for governed and traceable mark computation. The system is based on the Adaptive Weighted Penalty Model and was iteratively designed from the experience obtained from real academic usage within four successive semesters.

1. INTRODUCTION

The use of rubrics-based assessment is common practice in higher education institutions since rubrics provide a way of structuring communication about what should be expected from the evaluated person or work and enhance transparency in the evaluation process (Brookhart, 2018; Reddy & Andrade, 2010). However, the availability of the rubric does not mean the scores will be calculated consistently. The similarity between qualitative judgment that evaluators can provide while evaluating someone's performance does not always imply consistency in applying arithmetic operations to calculate the total grade in case of partial completion, criteria multiplicity, and different severities of errors (Jonsson & Svingby, 2007; Panadero & Jonsson, 2020). Such inconsistency becomes critical when considering academic institutions that need to be moderated or accredited and require justification for their assessment processes and results.

While modern digital assessment tools and learning management technologies provide numerous possibilities for automation, many academic institutions are still relying on manually operated Excel sheets due to accessibility and familiarity issues (Bearman et al., 2020; Ifenthaler & Yau, 2020; Panko, 2008). Though these tools offer flexibility of working with rubrics, they fail to implement structured algorithms that would provide governed score calculation. To address this gap, this research proposes RUBIQUANT - an intelligent computing approach based on the application of the AWPM to structure score calculation within the accessible Excel sheet environment.

2. LITERATURE REVIEW

Assessment rubrics can be described as assessment tools which increase clarity, consistency, and transparency by providing explicit performance criteria and clear guidelines for evaluation in higher education contexts where projects, presentations, laboratory work, and competency assessment are involved (Brookhart, 2018; Reddy & Andrade, 2010). However, having rubrics does not preclude assessment inconsistencies because reliability issues stem from differences in assessment interpretations, scorer calibration, scorer fatigue, and judgement variations (Jonsson & Svingby, 2007; Panadero & Jonsson, 2020). These issues are accentuated when dealing with technical and performance assessments where partial completion, different severity grades for errors and complex evaluation criteria are taken into account (Bloxham et al., 2016; Suskie, 2018). Digital assessment technologies have increased efficiencies and automated some elements, but they are hampered by limited configurability, reliance on ecosystem-specific solutions, costs and discipline mismatch in terms of scoring algorithms (Bearman et al., 2020; Ifenthaler & Yau, 2020). In consequence, spreadsheet-based assessments are still common practice due to ease and familiarity, even though spreadsheets are known to be error-prone and lack operational controls (Panko, 2008).

Decision-support systems provide a conceptual base because they enhance consistency by embedding clear computational logic into decisions (Power, 2002; Shim et al., 2002). Modern educational innovations draw similar ideas as intelligent data-driven systems are developed to improve transparency, accountability, and standardization of assessments (Luckin et al., 2016; Zawacki-Richter et al., 2019). In competency-based or goal-oriented assessment environments, there is sometimes a need to consider both completion and quality of performance and create scoring logic that provides enough flexibility and consistency (Guskey, 2007; Harden, 2002). It was the rationale behind the creation of the Adaptive Weighted Penalty Model (AWPM) as a computational backbone of RUBIQUANT:

$$Marks = max\left\{0, C^{\beta}W - \lambda[1 + \gamma(1 - C)]\sum k_i E_i^{\alpha}\right\} + B$$

where:

W	=	criterion weightage
C	=	completion ratio
k_i	=	weighted penalty coefficients assigned to error category i
E_i	=	error frequency
α	=	error sensitivity exponent
β	=	completion sensitivity exponent
γ	=	adaptive penalty amplification factor
λ	=	overall penalty scaling factor
B	=	bonus adjustment

The AWPM follows a two-level architecture separating stable configuration parameters used for rubric governance from case-dependent performance characteristics. Configuration parameters include criterion weightage, cost coefficients for different types of errors, sensitivity factor, normalizer, scaler and, optionally, bonus coefficient. They are set only once for each assessment context and are shared across multiple candidates. Candidate-specific data includes completion ratio and frequency of error categories, which correspond to actual performance observations made during assessment. The separation is consistent with decision-support theory where logic governing the system operation is kept separate from specific input parameters (Power, 2002; Shim et al., 2002). In other words, the design allows for establishing common evaluative rules prior to applying assessment scores.

The mathematical form was selected specifically for consistency with assessment principles. Criteria weights reflect the fact that criteria rarely contribute to the overall score equally (Brookhart, 2018). Completion ratio indicates progress towards accomplishment similarly to how it is done in competency- and mastery-oriented assessments (Guskey, 2007; Harden, 2002). Nonlinear exponent for the completion term allows adjusting its significance. Penalty function takes into consideration that errors differ in terms of their severity and cannot be treated equally in many cases, especially competency- or technical-related assessments (Brookhart, 2018; Suskie, 2018). Sensitivity allows increasing the impact of repeated errors.

One of the unique aspects of the model is how penalties for incompleteness are calculated. Unlike typical approaches relying on fixed deduction values, AWPM treats incomplete performance adaptively based on the principle that incomplete performance can reflect general deficiency rather than occasional execution flaws, especially where achievement of necessary competencies is essential (Guskey, 2007; Harden, 2002). There are additional configuration controls allowing flexibility of use across various contexts with the same bounded output results mathematically. All the aspects mentioned above make RUBIQUANT an innovative spreadsheet-based assessment tool using principles of decision support.

3. METHODOLOGY

Design and Development Research (DDR) approach was employed in this study as a research approach that focuses on the systematic generation and evaluation of innovation in technology and education (Richey & Klein, 2007). This research approach was chosen because the main aim of this study was not hypothesis testing but rather the development of a decision-support computational model. The decision support model underwent development using a four-stage process of development based on the principles of prototyping as shown in Figure 1 below.

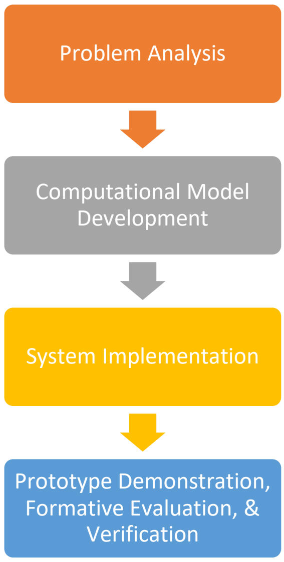


Figure 1. RUBIQUANT development methodology framework.

3.1 Problem Analysis

In the first phase, the objective was to analyse practical difficulties faced during the existing assessment processes, which would have led to the need for developing RUBIQUANT. The existing practices of using rubrics for assessment purposes, including manual use of spreadsheets and unstructured scoring sheets, were analysed for their inherent weaknesses, such as lack of uniformity in applying score computation rules, mathematical mistakes in score calculation, low transparency for score computation process, uncontrolled interactions in spreadsheets, and lack of customizable computing governance to handle different assessment needs. These difficulties were more prominent in academic settings where moderation is essential.

Another consideration in the analysis phase of the problem was the practicality of the solution to be provided, in order to make sure that any potential development was actually implementable within the institute. The development of any technological platform for the problem did not assume the availability of any proprietary enterprise-grade assessment tool and relied on technology easily accessible to an academic environment. Therefore, this phase helped establish the requirements and deployment setting for the innovation, and laid foundations for subsequent phases, including computational and implementation phases.

3.2 Computational Model Development

Phase two addressed the translation of assessment requirements from the previous problem analysis stage into a governed computational scoring system. Instead of resorting to typical methods such as the fixed point deduction scheme or spreadsheet scoring system, the development approach endeavoured to produce a computational scoring model that would facilitate weightings, differential errors, incomplete task achievement, and configurable scoring logic. The end result of this process was the design and formulation of the Adaptive Weighted Penalty Model (AWPM), forming the computational basis for RUBIQUANT, which was mathematically outlined and theoretically justified within the preceding literature review.

A critical design consideration was the isolation of governance elements in the form of constant rubric parameters separate from individualized performance data, ensuring that scoring rules were consistent within each assessment scenario while observed performance differed per candidate. This resulted in a repeatable process that minimized inconsistency in scoring and facilitated an adaptable assessment procedure without affecting scoring computation itself. AWPM scoring logic was extended to allow weighting for completion, differential penalty schemes, sensitivity tuning, bound protection, and potential positive adjustments to represent the nuance of each performance accurately in a reproducible and governed manner.

3.3 System Implementation

In terms of implementation, the computational model was converted to a working prototype based on Microsoft Excel coupled with Visual Basic for Application (VBA). This technology combination was chosen due to its ubiquity, familiarity, relatively low entry barrier, and flexibility for quick and easy customization. In contrast to being utilized for typical mathematical calculations on a spreadsheet, the implementation itself was developed as a governed assessment evaluation system with both protected formula functionality and controlled interaction from the user. The proposed system architecture involves three components: the input sheet for rubric governance, the automated output sheet for calculation, and assisted user interaction via the user interface. The input sheet (see Figure 2) is a spreadsheet where all the assessment parameters (common ones, shared between candidates, and individual performance data) were entered. The output sheet (see Figure 3) is an automated calculation mechanism where the scores are calculated automatically using the entered input values, thus minimizing arithmetic operations.

To make the operation easier and avoid potential issues, a separate VBA-assisted interaction interface was introduced using a modeless userform (see Figure 4). The user interface allows the user to work with the scoring system by providing guided interaction and preventing direct editing of worksheet cells, thus limiting any unintentional interference with the system. The interface offers guided input of parameters, dynamic input field populating, formula visualization, controlled navigation, validation control, and workflow restrictions that restrict any unintentional changes to the underlying formulas. Such an approach enabled the RUBIQUANT to be used not only as an interactive scoring worksheet but as a decision support tool for practical rubric assessment.

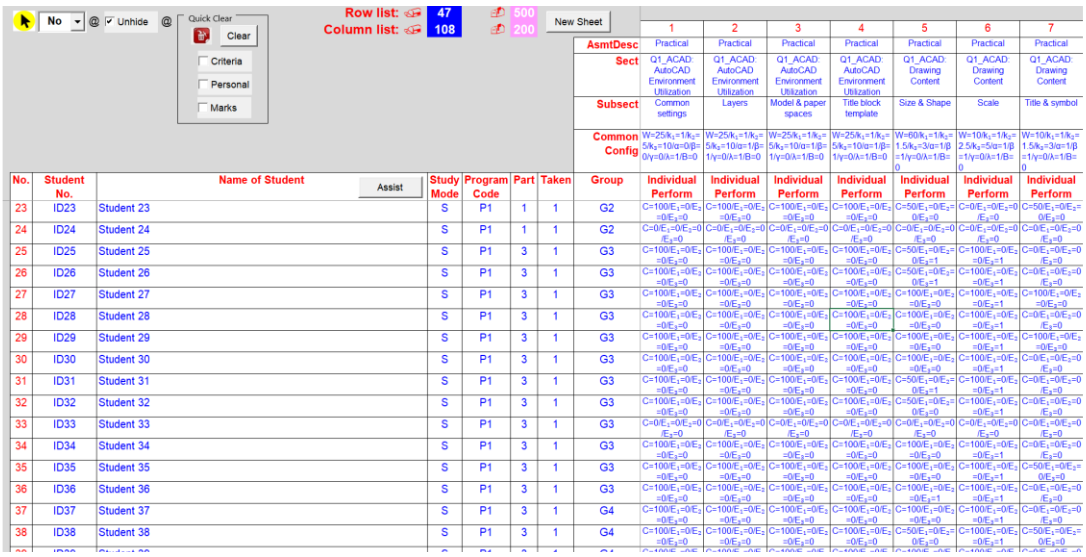
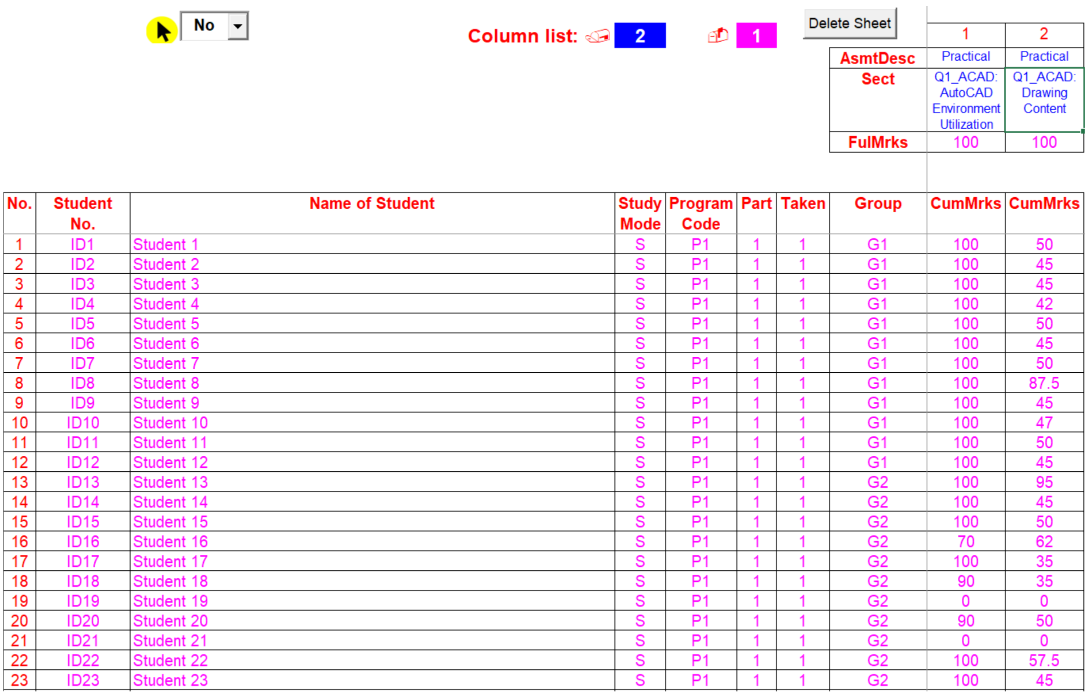


Figure 2. RUBIQUANT input worksheet interface.



									1	2
								AsmtDesc	Practical	Practical
								Sect	Q1_ACAD: AutoCAD Environment Utilization	Q1_ACAD: Drawing Content
								FulMrks	100	100
No.	Student No.	Name of Student	Study Mode	Program Code	Part	Taken	Group	CumMrks	CumMrks	
1	ID1	Student 1	S	P1	1	1	G1	100	50	
2	ID2	Student 2	S	P1	1	1	G1	100	45	
3	ID3	Student 3	S	P1	1	1	G1	100	45	
4	ID4	Student 4	S	P1	1	1	G1	100	42	
5	ID5	Student 5	S	P1	1	1	G1	100	50	
6	ID6	Student 6	S	P1	1	1	G1	100	45	
7	ID7	Student 7	S	P1	1	1	G1	100	50	
8	ID8	Student 8	S	P1	1	1	G1	100	87.5	
9	ID9	Student 9	S	P1	1	1	G1	100	45	
10	ID10	Student 10	S	P1	1	1	G1	100	47	
11	ID11	Student 11	S	P1	1	1	G1	100	50	
12	ID12	Student 12	S	P1	1	1	G1	100	45	
13	ID13	Student 13	S	P1	1	1	G2	100	95	
14	ID14	Student 14	S	P1	1	1	G2	100	45	
15	ID15	Student 15	S	P1	1	1	G2	100	50	
16	ID16	Student 16	S	P1	1	1	G2	70	62	
17	ID17	Student 17	S	P1	1	1	G2	100	35	
18	ID18	Student 18	S	P1	1	1	G2	90	35	
19	ID19	Student 19	S	P1	1	1	G2	0	0	
20	ID20	Student 20	S	P1	1	1	G2	90	50	
21	ID21	Student 21	S	P1	1	1	G2	0	0	
22	ID22	Student 22	S	P1	1	1	G2	100	57.5	
23	ID23	Student 23	S	P1	1	1	G2	100	45	

Figure 3. RUBIQUANT output worksheet interface.

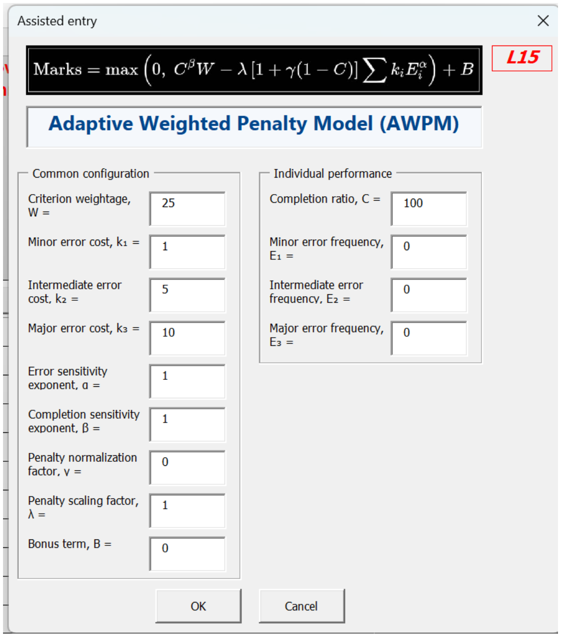


Figure 4. RUBIQUANT assisted entry user interface.

3.4 Prototype Demonstration, Formative Evaluation, and Verification

Upon its implementation, RUBIQUANT became available for the targeted group of academic users in the form of a prototype demonstration and subsequent practical application within authentic assessment situations. This method aligns with design and development research, wherein innovations are tested and refined through iterations of their application and the user input they generate, rather than being developed and delivered in a finished product format (Richey & Klein, 2007). The prototype's implementation involved four academic lecturers regularly applying rubric-based assessment practices in teaching.

The prototype application occurred in real-world academic assessment exercises carried out in four consecutive semesters with an improvement process gradually developing throughout this timeframe. User feedback generated in each iteration of use led to successive modifications aimed at improving navigation, assisted entry, validation operations, workflow management, and overall usability of the prototype. Several changes were introduced as a result of the experiences of earlier implementations, especially when avoidable issues arose due to a certain lack of usability or effective workflow management. Therefore, the prototype used in this study can be regarded as an improved implementation that has undergone several cycles of authentic use and refinement, rather than a single attempt at implementing an innovation.

Evaluation of the prototype was performed by observations of implementation processes complemented by direct feedback from users and focused on assessing practical aspects related to usability, workflow functionality, computational reliability, and applicability in academic setting rather than inferential data analysis. In order to allow for more objective interpretation of the results achieved by different users, the criteria of success employed were formulated according to a five-level linguistic scale based on usability assessment methodology of formative evaluation (Nielsen, 1994): Low, Moderate, Good, Very Good, and Excellent. After the conversion of individual feedback into numerical data, aggregation and retranslation into the corresponding linguistic terms was performed for each semester of the prototype's use. Verification criteria comprised user operational confidence, reliance on assisted entry features, efficiency and reliability of workflow management, effectiveness of worksheet control, accuracy and consistency of formulas, results' reproducibility, and scoring consistency.

4. FINDINGS

In relation to RUBIQUANT, its implementation over four consecutive academic semesters resulted in significant insights regarding its usability, as well as its development into a more usable prototype through the refinement process. The software was constantly developed and refined based on lecturer feedback prior to each semester, which means that the observed findings are influenced by improvements made to the prototype, rather than being solely based on growing user proficiency. A number of usability concerns became evident in the first round of implementation. Some lecturers would interact with parts of the program outside the intended input zones, assistance with the editing process was not always sufficient, and some workflow processes did not have sufficient guidance to avoid interaction with the structure of the worksheets themselves. This feedback led to improvements such as better worksheet navigation control, enhanced assisted-entry interaction, improved validation mechanisms and more structured data-entry workflows. As a result, later implementation cycles were significantly easier to manage. Proficiency findings are described in Table 1.

Table 1. Aggregated lecturer proficiency outcomes across iterative implementation.

Evaluation Indicator	Semester 1	Semester 2	Semester 3	Semester 4
Independent system operation	Moderate	Good	Very Good	Very Good
Confidence in system use	Moderate	Good	Good	Very Good
Dependence on assisted-entry guidance	Very High	High	Moderate	Moderate

Input handling accuracy	Moderate	Good	Very Good	Excellent
Workflow efficiency	Moderate	Good	Good	Very Good

It appears that the level of lecturer's operational proficiency improved over time due to increased familiarity with the system and its functions, while also reflecting improvements that were made to the software as a result of feedback. Early implementation rounds necessitated a more hands-on approach to guidance, while later semesters had more proficient and independent lecturers who completed their tasks more efficiently. Verification findings are described in Table 2.

Table 2. Prototype verification outcomes across iterative deployment.

Verification Criterion	Semester 1	Semester 2	Semester 3	Semester 4
Computational execution stability	Good	Very Good	Excellent	Excellent
Formula integrity preservation	Moderate	Very Good	Excellent	Excellent
Output reproducibility	Excellent	Excellent	Excellent	Excellent
Workflow reliability	Moderate	Good	Very Good	Excellent
Worksheet governance effectiveness	Moderate	Good	Very Good	Excellent
Assisted-entry operational usability	Moderate	Good	Very Good	Very Good

The findings indicate that the prototype has proven stable in its performance, having significantly improved workflow and worksheet governing capabilities thanks to refinements made. Output reproducibility has been consistent throughout, with scoring outputs remaining stable in cases of repeated identical input conditions. Implementation findings concerning practical impact on academic assessment are described in Table 3.

Table 3. Aggregated implementation outcomes in academic assessment practice.

Evaluation Indicator	Semester 1	Semester 2	Semester 3	Semester 4
Reduction in manual scoring errors	Good	Very Good	Excellent	Excellent
Perceived scoring consistency	Moderate	Good	Good	Very Good
Transparency of score derivation	Good	Very Good	Excellent	Excellent
Assessment workflow efficiency	Moderate	Good	Very Good	Very Good
Practical suitability for continued deployment	Good	Good	Very Good	Excellent

As can be seen from the findings above, RUBIQUANT appeared to be a successful tool to implement in an academic assessment workflow due to its ability to assist in computation as well as in providing traceable scoring output. Score consistency also appears to have improved somewhat as a result of adoption of RUBIQUANT.

5. DISCUSSION

It can thus be hypothesised that the four semesters of implementation were necessary not only to gain user experience with RUBIQUANT but also to iteratively refine the technology according to the actual experiences gained in an academic context. Unforeseen difficulties encountered at early stages such as unintended user navigation out of designated input fields, workflow ambiguity, and inadequate safety controls led to further iterations of improvements to worksheet control, validation logic, and assisted input features, which resulted in a relatively stable and functional solution. This hypothesis is consistent with the paradigm of design and development research in education technology that emphasises the iterative maturation of artefacts through progressively deploying solutions in actual settings and eliciting end-user feedback rather than relying on a snapshot post-release evaluation (Richey & Klein, 2007).

One of the main contributions of RUBIQUANT to existing practice is its introduction of computational governance to a process that is known to be consistently inconsistent despite the common use of rubric-based assessments. While rubrics provide clarity in assessment criteria and increase transparency of expectations, previous studies have demonstrated that assessor inconsistency in interpreting criteria, performing arithmetic operations, and implementing the rubric still significantly impacts test reliability (Jonsson & Svingby, 2007; Panadero & Jonsson, 2020). Observations reported here indicate that RUBIQUANT partially alleviates this issue through moving score calculation into the domain of computation governed by formulae rather than by the human mind. The evaluator's discretion in assessing quality remains unchanged while their inconsistency in applying scoring rules is mitigated. The reported improvement in score traceability also carries practical significance in terms of moderation and quality assurance. Finally, the relatively consistent behaviour of the implemented computation across all phases of implementation seems to support the assumption that the employed safeguards are sufficient. Guided user interface and assisted data entry tools, in turn, helped make the computational framework more user-friendly and allowed lecturers to become increasingly confident in using the system. From this perspective, the implementation appears consistent with the established usability framework (Nielsen, 1994).

Despite the lack of an elaborate framework and infrastructure behind the technology in question, the implementation indicates that useful innovations do not necessarily entail proprietary enterprise software. Through combining two powerful technologies of Microsoft Excel and Visual Basic for Applications, the researcher managed to create a tool that could be used as a building block of a more robust framework for conducting assessments digitally. It is, however, important to point out that the present study had certain limitations. The evaluation process was formative and limited to a small sample of intended academic users, with user observation and feedback being its basis rather than formalised measurements based on reliable and valid instruments. Hence, the findings are better perceived as an implementation report and not as psychometric evaluation. It would be necessary to validate the reported results in future research involving broader institutional deployment and usability testing.

6. CONCLUSION

Thus, RUBIQUANT was designed as a pragmatic framework that would enhance consistency, clarity, and governance in terms of rubric-guided assessment processes. Though rubrics offer clear guidance in terms of evaluating academic performance, pragmatic grading is prone to inconsistencies due to the subjective handling of arithmetic and translation of judgment into scores. By adopting the AWPM in an accessible Excel-VBA environment, this project demonstrates that governed computation can enhance consistency and repeatability without interfering with academic judgment. As the result of the project's four-semester-long operation, it became possible for the system to mature by undergoing real-world implementation process.

Though this particular research can be described as being formative rather than confirmatory – as a matter of fact, the current project relied on a small number of participants and implementation-based observation only – the findings show the viability of applying a governed computational approach to improve assessment practices guided by rubrics. This project also proves that meaningful innovations in educational assessment could be developed with the help of common software solutions provided by institutions of learning rather than expensive specialized applications. In addition to its current implementation, the project should be extended to include deployment throughout the institution, inter-rater reliability assessment, comparative benchmarking against traditional assessments, and usability testing.

Acknowledgments: Faculty of Civil Engineering, Universiti Teknologi MARA.

References

Bearman, M., Dawson, P., Ajjawi, R., Tai, J., & Boud, D. (2020). Re-imagining University Assessment in a Digital World. *Assessment & Evaluation in Higher Education*, *45*(7), 1021–1033. https://doi.org/10.1080/02602938.2020.1776897

Bloxham, S., Den-Outer, B., Hudson, J., & Price, M. (2016). Let's Stop the Pretence of Consistent Marking: Exploring the Multiple Limitations of Assessment Criteria. *Assessment & Evaluation in Higher Education*, *41*(3), 466–481. https://doi.org/10.1080/02602938.2015.1024607

Brookhart, S. M. (2018). *How to Create and Use Rubrics for Formative Assessment and Grading*. ASCD.

Guskey, T. R. (2007). Closing Achievement Gaps: Revisiting Benjamin S. Bloom's "Learning for Mastery." *Journal of Advanced Academics*, *19*(1), 8–31.

Harden, R. M. (2002). Developments in Outcome-based Education. *Medical Teacher*, *24*(2), 117–120. https://doi.org/10.1080/01421590220120669

Ifenthaler, D., & Yau, J. Y.-K. (2020). Utilising Learning Analytics for Study Success: Reflections on Current Empirical Findings. *Research and Practice in Technology Enhanced Learning*, *15*, 1–17. https://doi.org/10.1186/s41039-020-00142-z

Jonsson, A., & Svingby, G. (2007). The Use of Scoring Rubrics: Reliability, Validity and Educational Consequences. *Educational Research Review*, *2*(2), 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. Pearson.

Nielsen, J. (1994). *Usability Engineering*. Academic Press.

Panadero, E., & Jonsson, A. (2020). A Critical Review of the Arguments against the Use of Rubrics. *Educational Research Review*, *30*, 100329. https://doi.org/10.1016/j.edurev.2020.100329

Panko, R. R. (2008). What We Know about Spreadsheet Errors. *Journal of End User Computing*, *10*(2), 15–21.

Power, D. J. (2002). *Decision Support Systems: Concepts and Resources for Managers*. Quorum Books.

Reddy, Y. M., & Andrade, H. (2010). A Review of Rubric Use in Higher Education. *Assessment & Evaluation in Higher Education*, *35*(4), 435–448. https://doi.org/10.1080/02602930902862859

Richey, R. C., & Klein, J. D. (2007). *Design and Development Research: Methods, Strategies, and Issues*. Lawrence Erlbaum Associates.

Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002). Past, Present, and Future of Decision Support Technology. *Decision Support Systems*, *33*(2), 111–126. https://doi.org/10.1016/S0167-9236(01)00139-7

Suskie, L. (2018). *Assessing Student Learning: A Common Sense Guide* (3rd ed.). Jossey-Bass.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic Review of Research on Artificial Intelligence Applications in Higher Education. *International Journal of Educational Technology in Higher Education*, *16*, 39. https://doi.org/10.1186/s41239-019-0171-0