



The Use of Binary Scales in Rubrics to Evaluate Computer Science Assessments

Jacqui Muller^(✉)  and Japie J. Greeff 

School of Computer Science and Information Systems, Unit for Data Science and Computing,
North-West University, Vanderbijlpark, South Africa
jacqui.muller@outlook.com, japie.greeff@nwu.ac.za

Abstract. Creating a rubric for project based assessment in ICT modules is a complex process and can be quite subjective if the wrong scales are used for rubric elements. The aim of this paper is to present a framework for developing rubrics that use binary scales for the rubric elements which will both lessen the risk of subjectivity (with focus on objective criteria) entering into the marking process and also create an opportunity for automation to be used for the marking of parts of the assessment. The framework and process used to develop a rubric are explained and an example is presented of what the final result would look like if created in this way. The framework was developed using a Design Science Research approach which is also the suggested approach to be used in the creation of a rubric. It is believed that this approach, once automated, could lead to a more dynamic approach to assessment where students can submit their work partially and get feedback immediately and lecturers can track in real time the progress of the class.

Keywords: Rubric development · Binary scale · Project-based assessment

1 Introduction

A large part of education systems is the ability to measure the growth in knowledge around the concepts that have been taught within the applicable system or environment. There are multiple different ways to measure knowledge development, some of which include evaluations and assessments (Novita 2017). Novita (2017) suggests that an assessment is used to evaluate the attainment of learning outcomes which can be conducted as formal or informal assessments. The learning outcomes that an assessment aims to assess can be measured through the use of a rubric (Zano 2015).

Alves et al. (2020) defines a rubric as a tool used as a measurement in performance-based assessments. Rubrics can be used to measure individual portions (or items) of an assessment, referred to as rubric elements. Rubrics are good at assessing algorithms and programming concepts in the context of Information Systems, Computer Science or Information Technology project-based assignments and tests. A rubric is made up of elements that have a weighting of marks associated to each element being assessed or measured. In this paper a proposed framework is presented for creating rubrics for project-based assessments in computer science or information systems modules using rubric elements that are comprised of binary scales.

2 Literature Review

According to Janse van Rensburg and Goede (2019) the kind of work graduates will do when employed in the Information Technology space will most likely be more akin to project-based assessment than to traditional assessments found in tertiary education. In project-based assessments there is a recognition that subjectivity can enter into the evaluation of work even with a rubric as can be seen by the work of Mustapha et al. (2016) where, even with a standard rubric that is employed for all assessments, there is still some risk of subjectivity. A well-designed rubric however can guard against this risk, provided that the criteria is defined in enough granular detail to reduce subjective interpretation as far as possible. Dawson (2017) puts forward 14 rubric design elements obtained from the literature to assist in the creation of well-designed rubrics. The problem with creating a single rubric though that applies to all assessments is that it does need to be very general, removing the ability to test specific skills in different assessments.

Jones and Tadros (2010) put forward that the weighting of marks in a rubric can be assessed using different scales, some of which could include categorical scaling, sliding scaling and binary scaling. These scales are considered in turn to indicate their characteristics.

2.1 Categorical Scale

Jones and Tadros (2010) position categorical scaling as the mark allocation of each rubric element being split up into categories with numerical values associated to each category. An example of this would be measuring each element in terms of the following categories:

- Not acceptable (attached to the numerical value of 1): The students' approach to addressing the rubric element in question is way below the standard or expectation. This could be because the element is not addressed at all within the assessment or that the students' approach displays a complete misunderstanding of the element.
- Below expectation (attached to the numerical value of 2): The students' approach to addressing the rubric element does not meet the standard results that were expected as an outcome of the assessment.
- Meets expectation (attached to the numerical value of 3): The students' approach to addressing the rubric element meets the standard results expected as an outcome of the assessment.
- Exemplary (attached to the numerical value of 4): The students' approach to addressing the rubric element exceeds the standard results expected as an outcome of the assessment.

The South African basic education system comprising of the foundation phase (grades R to 3), intermediate phase (grades 4 to 6), senior phase (grades 7 to 9) and FET phase (grades 10 to 12) curriculums governed by the Curriculum and Assessment Policy Statement (CAPS), divides the term results of students into categories based on the marks they received for the term (Education 2021). According to Morolong (2009), these categories are:

- 7 (80%–100%): Outstanding achievement
- 6 (70%–79%): Meritorious achievement
- 5 (60%–69%): Substantial achievement
- 4 (50%–59%): Adequate achievement
- 3 (40%–49%): Moderate achievement
- 2 (30%–39%): Elementary achievement
- 1 (0%–29%): Not achieved.

2.2 Sliding Scale

Park and Yan (2019) present the concept of a sliding scale as the consideration of different factors when deriving an overall score. These factors may vary between rubric elements and may not even be formally declared before the rubric is used to evaluate an assessment. Often these factors that need to be considered are determined by the assessor as they evaluate the assessment (Imbault et al. 2018). The objective of a sliding scale is to allow room for interpreting different influencing factors as part of the mark awarded for each rubric element. This is supported by Imbault et al. (2018) who suggest that data obtained through the use of sliding scale evaluations are, by their nature, interval.

2.3 Binary Scale

Jones and Tadros (2010) position the binary scale as the use of 1 and 0 to determine whether an element has been addressed adequately within an assessment or not. Where the sliding scale and categorical scale allows for some level of subjectivity and interpretation of elements between students, the binary scale aims to remove the level of subjectivity and only focus on whether the element is present/addressed or not. Haghdoost (2012) shows how the binary scale can be used to distinguish between two levels of evaluation. In the case of their research, Haghdoost (2012) used 0 to denote illiteracy and 1 to denote literacy. Park and Yan (2019) support the use of binary scale to segregate categories through 0 denoting ‘no’ and 1 denoting ‘yes’ as answers to the questions posed as rubric elements. Dimitrov (2016) presents an argument for the benefits of binary scale scoring with large-scale assessments. In this paper the binary scale is used to create rubric elements so as to remove the subjectivity (as far as possible) inherent in the other scales as well as to lay the groundwork for a higher level of automation in the assessing of project-based work.

3 Research Design

In this paper a framework is presented that can be used to develop a rubric with binary scaled elements for project-based assessments. This framework was developed using the Design Science Research (DSR) approach which is not only applicable to the creation of the framework but can also be used to iteratively develop rubrics for project-based assessments. It is likely that, once defined, a rubric can be further refined over a number of years each time the assessment is presented in a module.

Simon (1996) and Hevner (2007) propose Design Science Research as the introduction of new and innovative artefacts that motivate the desire to improve an environment. Lukka (2003) expands on this by stating that DSR can be used to solve real-life problems by making a contribution to the applied theory.

Hevner (2007) also supports using DSR to introduce new knowledge, executed in three cycles, namely relevance, design and rigour. The development of theories and artefacts will accept environmental needs (of people, processes and technologies) as inputs, utilising any applicable knowledge from the available knowledge base. The relevance cycle will represent the flow of all environmental needs and environmental application between the environment and the information system (IS) research. The rigour cycle will represent the flow of knowledge between the knowledge base and the IS research. Although used to great effect in IS research, it is also shown by Plomp and Nieveen (2013) that educational design research is a valid approach to developing curricula and learning artefacts. The framework of DSR used is shown in Fig. 1.

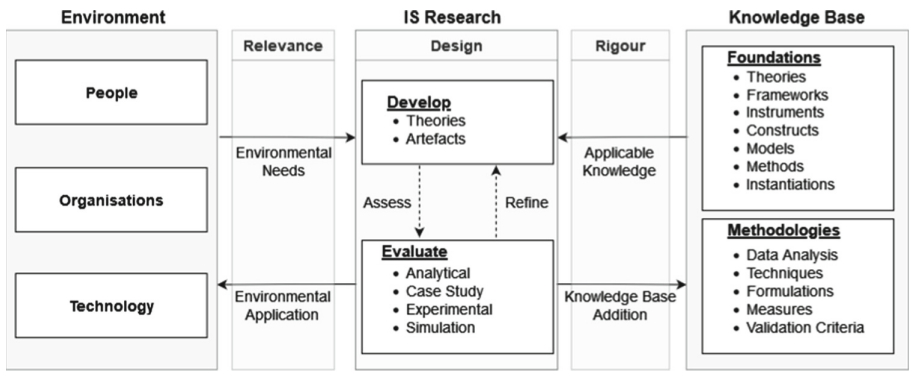


Fig. 1. Design science research cycles adapted from Hevner et al. (2004)

4 Background

The initial implementation of the proposed approach was trialed in a hackathon environment that was subject to a more open-ended project brief. A rubric was developed for evaluating submissions to the hackathon. The goal given to participants in the hackathon was to create innovative cloud solutions that address the world’s toughest problems as represented by the United Nations Sustainable Development Goals. 33 teams with between one and eight members were given 48 h to develop either a solution or an idea. These were then evaluated by 15 judges of which nine were technical experts and six were not. The rubric was developed using a looser implementation of a similar technique presented in this paper – with less focus on the objectivity of the criteria. The rubric with its categories, subcategories and weight are presented in Fig. 2.

Category	Sub-Category	Weight
Global Relevancy and Impact	Global aspects consideration	3.75%
	Global impact design	3.75%
	Alignment with global initiatives	3.75%
	Multiple continents	3.75%
Technical Execution	Mockups provided	2.14%
	Working prototype provided	2.14%
	Complex or sophisticated solution	2.14%
	Emerging or advanced technology	2.14%
	Solution scalability	2.14%
	Indexed on source control	2.14%
	Stands out	2.14%
Alignment to UN Sustainable Development Goals	One or more of the development goals addressed	4.00%
	At least one development goal addressed	4.00%
	At least one sub goal addressed	4.00%
	Two or more goals addressed	4.00%
	Two or more subgoals addressed	4.00%
Cloud Technology	Cloud technology used	4.00%
	Cloud software used	4.00%
	Cloud API integration	4.00%
	More than one cloud technology used	4.00%
	Cloud used novelly	4.00%
Creativity and Innovation	Non-traditional creative concept	3.75%
	Design thinking applied	3.75%
	Innovative	3.75%
	Unexpected and interesting	3.75%
Market and Business Feasibility	Addressed business feasibility	3.75%
	Practicality	3.75%
	Appealing marketing	3.75%
	Funding model provided	3.75%

Fig. 2. Weighted Hackathon rubric (including high-level categories and criteria)

Due to the nature of a hackathon, the evaluation of submissions was a lot more open-ended and even with binary scales there was a large amount of variation in results indicated by judges due to the subjective interpretation of some of the criteria. For example, one of the subcategories was whether the submission was innovative, a characteristic that would be very desirable from a hackathon entry. This is something that would inevitably be subjective between different judges. As such, the results obtained between different judges seemed to be consistent in terms of how they marked various entries but didn't correlate well between the marks allocated by different judges for the same submission. Figure 3 presents the box whisker diagram indicating the marks scored by different judges for the different submissions to the hackathon, with specific focus on the marks awarded by Judge 14, denoted by the dots.

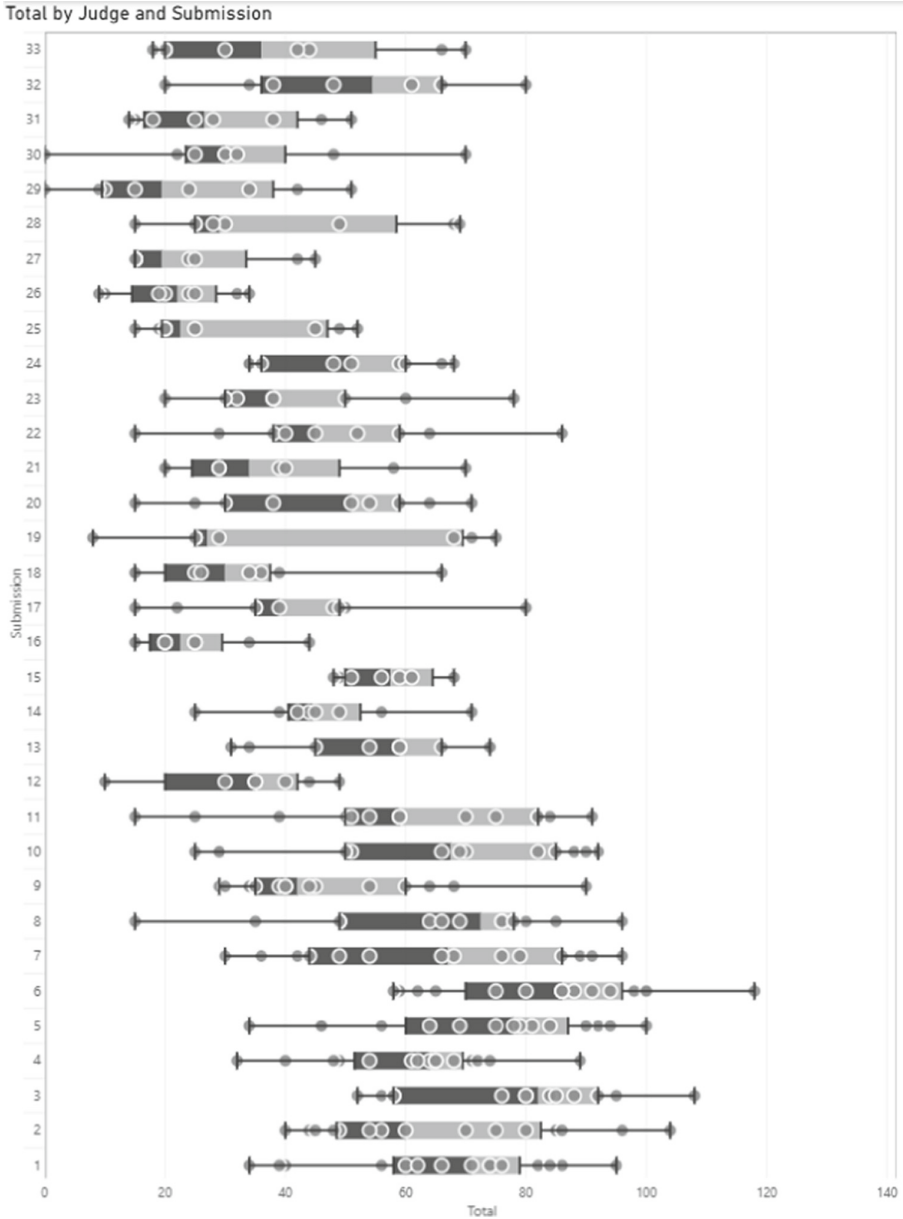


Fig. 3. Ratings by all judges on all submissions to the Hackathon

From these results it would indicate that although individual judges could be consistent in rating rubric elements the same between different submissions, there was still a large amount of subjectivity present in subcategories that were not defined precisely or with more granular and measurable detail. This also highlights the fact that even though

a rubric can enable a lecturer to lessen their subjectivity, when presenting the rubric to other markers it is important to also give training to the markers as to how each element should be interpreted. Certain elements are inherently more subjective than others. The idea behind the rubric was to remain as lean as possible for ease-of-marking.

Multiple ‘lessons learnt’ were identified from this implementation, mainly revolving around the subjective nature of certain criteria and how dividing those criteria in ways that makes the criteria more objective would further decrease the subjective nature of an assessment. This does introduce the possibility of a longer rubric that delivers more objective evaluations.

Another observation is that due to the nature of a hackathon, very few clear outcomes are defined with the specific aim of alleviating the restriction of innovation and creativity. Modules that have defined and concise assessment, study unit and module outcomes may provide a better foundation for assessment criteria formulation. The flaws of this implementation led to the next iteration of the conceptual framework as discussed in Sect. 5.

5 Conceptual Framework

The challenge with sliding and categorical scales is maintaining consistency and objectivity while marking. Even when refining the criteria, the interpretation to degree of implementation allows subjective interpretation since there are more ‘choices’ available. In many university settings, lecturers deal with large groups of students for which they require assistants to assist with marking student assessments. In cases like these, some assessors may mark more strictly than others which not only jeopardise the consistency of the marks but also the objectivity of the marks. If marks were to be visualised and analysed, the marker would need to be taken into consideration as well.

The additional challenge with this approach is that students may attempt catering for a requirement simply to receive at least some marks for the attempt. With many sliding and categorical scales, the rubric criterion allows for the awarding of marks for an attempt even if the attempt was fruitless and did not contribute to a working solution. Although this is good for students, it does create a situation where students are not necessarily prepared for the kind of work (or criteria) they would be expected to execute on once they graduate which will require them to meet rigid requirements fully. In industry engagements in the ICT space contractual engagements tend to be of a nature that requirements need to be completely fulfilled before obtaining project sign-off and, in some cases, payment.

Although a pass mark for a module with a project-based assessment in it will still be 50%, it would be beneficial to the student to give them very detailed feedback on exactly which elements of the assessment they don’t pass. This is especially true if that feedback can be given more frequently throughout the course of the assessment rather than only at the end. Categorical and sliding scales may obfuscate this and create a sense that not meeting requirements fully is not a problem.

The concept that forms the foundation of using binary scales to evaluate project-based assessments in Computer Science, Information Systems or Information Technology modules needs to be instilled in all elements of the assessment, from the conceptual formation of the scope through to the weighting and rubric assessment criteria.

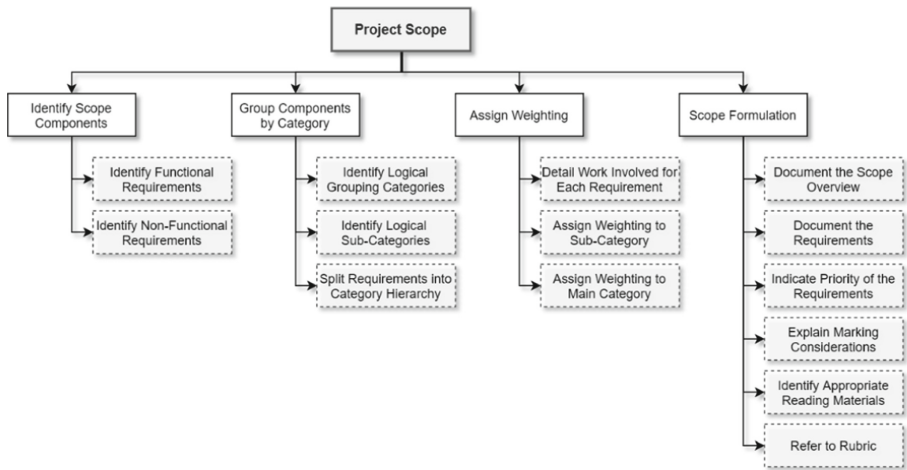


Fig. 4. Conceptual framework for creating a rubric with binary scaled elements for project based assessments in ICT programmes

There is a significant amount of prework that needs to take place in the preparation phases of the assessment before rubric criteria can be assessed using a binary scale. Figure 4 outlines the work involved and is further elaborated upon as the:

- **Identification of scope components:** The basis of any ICT project-based assessment is the student's ability to turn requirements into a working solution that fulfils those requirements. These requirements may be sub-divided between functional requirements and non-functional requirements. Functional requirements refer to the functionality that a system must have and how the functions should be performed whereas non-functional requirements refer to the aspects of a solution that have an impact on the quality attributes of a system (or platform). These non-functional requirements are deemed as supportive requirements to ensure that the functional requirements are implemented appropriately and according to good software practices.
- **Grouping of components by category:** Once the requirements have been identified, the requirements can be grouped through logical relation to form categories and sub-categories which may have an additional layer to evaluate the granular details of each requirement. The requirements would then need to be divided accordingly, into the appropriate categories and sub-categories.
- **Assigning weighting:** Before weighting can be applied to each requirement, the amount of work required to cater for the requirement would need to be considered. More work would mean higher weighting. It is important to also remain cognisant of the categories that are higher priority and ensure that the priority is indicated in the weighting. For example, in a core programming assessment where programmability and the use of patterns are being tested, visualisation should not contribute the highest weight to the overall project mark.
- **Formulating the scope:** The scope brief which should be provided to students should clearly state what is expected of them in order to achieve at least meritorious results.

The scope should provide a contextual problem which sets the tone of the assessment and helps the student understand the context for which they are trying to solve a problem. The requirements and the priority of each should be clearly defined within the scope. Marking considerations should be outlined and any additional learning materials should be provided.

Once the abovementioned work has been done, the rubric may be finalised with any additional varying factors (like group work division and assessment, bonus mark criteria, etc.) before the rubric is released to students.

5.1 Formulating a Scope

A project-based assessment within the ICT space would typically contain some form of solution design and development as well as a learning component which encapsulates any additional knowledge that students would need in order to complete the assessment, as seen in Fig. 5. These categories may differ based on the context and outcomes of the assessment.



Fig. 5. Project categories

Based on the project categories defined in Fig. 5, the content of the scope of the project-based assessment would need to be expanded to cover all of the categories in greater detail. The design of the solution could be covered through providing the functional and non-functional requirements, along with the prioritisation of each as well as the errors and exceptions that should be catered for, as seen in Fig. 6.

Figure 6 illustrates that the development category should be further divided to cater for the sub-categories that would make up development. The most common sub-categories that would typically be addressed in programming assessments would be infrastructure (what tools and technologies should be used and where should they be deployed, hosted or submitted), data access (what input data should be used and what output data is expected), functionality (what is the solution expected to do, with specific reference to the functional and non-functional requirements in more technical detail), presentation/user experience (how should the interface that the user is going to interact with the solution through look), standards (what architectural patterns, coding standards, design principles and design patterns should be used) as well as testing (what tests should be performed and how are they expected to be performed).

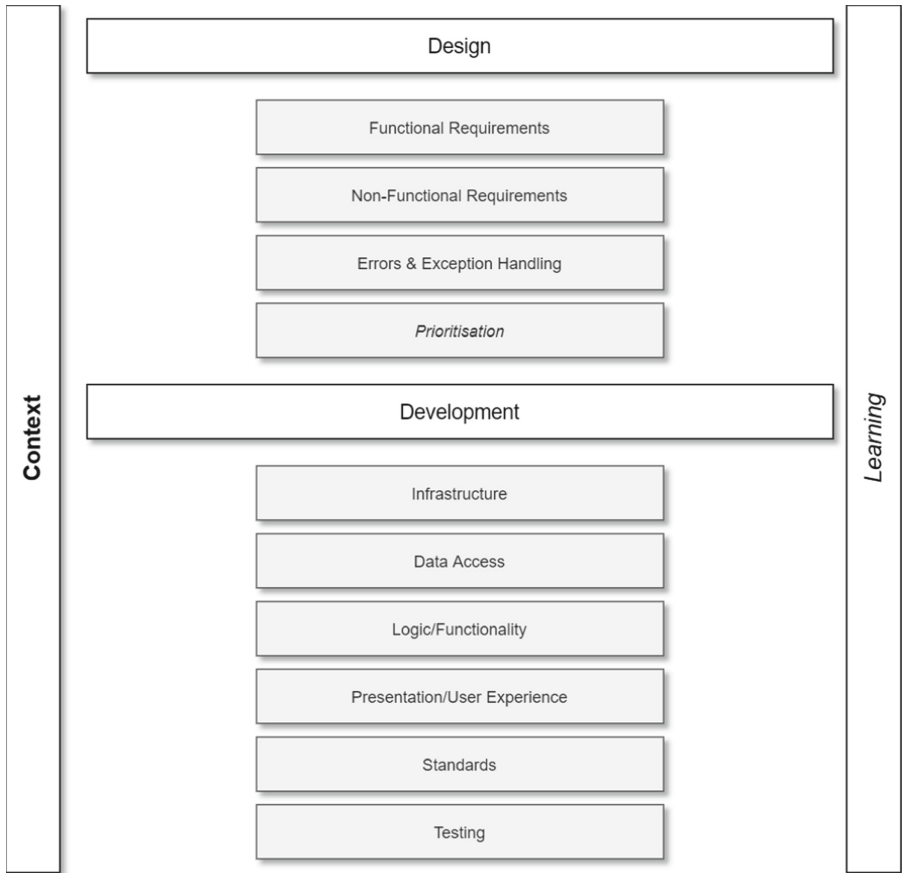


Fig. 6. Formulating a scope

The learning component, as represented in Fig. 6, would span across the design and development categories to ensure that any gaps in knowledge are bridged. The most important part of an assessment scope is the context that encapsulates all of these categories. It is important that the student understands what is expected of them, what they should achieve and what environment they are trying to achieve this in. A good way to apply context is through the use of practical and realistic industry problems. This also helps bridge the gap between industry expectation and student capability.

5.2 Formulating a Rubric

Once the scope has been defined, a rubric would need to be created to evaluate criteria which is adequately related to the content of the scope and can be evaluated using a binary scale. The content of a rubric can be broken down further, from Fig. 6, as seen in Fig. 7.

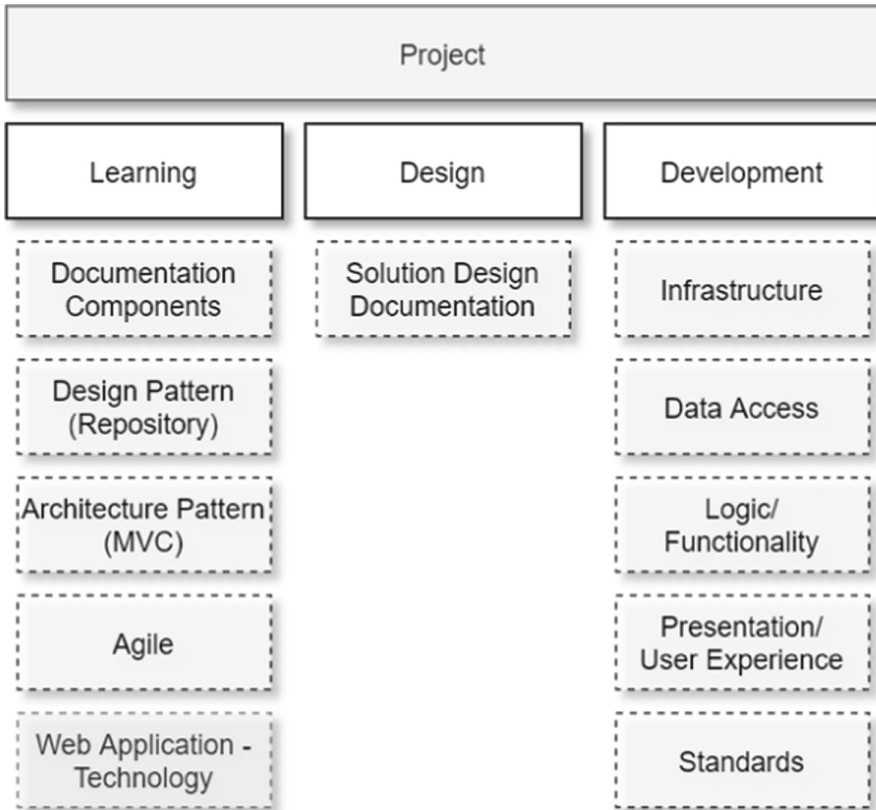


Fig. 7. Formulating a rubric

Figure 7 represents the scope categories broken down into sub-categories. Each of the sub-categories may be divided further to show how each requirement fits in and is evaluated as part of its relative sub-category and category. Criteria can then be attached to sub-categories' descriptions which clearly outline and define what is being evaluated, as seen in Fig. 8.

Once the categories, sub-categories, criteria and descriptions have been defined, weightings can be assigned and rolled up per sub-category and category. Effort and priority should be taken into account when assigning the weighting.

5.3 Assigning Weighting

Typically, projects or assessments are assigned an overall weighting to a module. The project is then further weighted on a criterial level, as seen in Fig. 9.

Category	Sub-Category	Criteria	Description
Learning	Agile	[Online test] Agile	Agile concepts tested via online test
	Design Pattern	[Online test] Design Pattern	Design pattern knowledge tested via online test
	Architecture Pattern	[Online test] Architecture Pattern	Architecture pattern knowledge tested via online test
	Document Management	[Online test] Documentation Components	Documentation components knowledge tested via online test
	Web-Application Technology	[Practical assignment] Web-Application Technology	Web Application technology to be tested via online test/assignment and practically applied as part of the project
Design	Solution Design Documentation	ERD	ERD is provided and accurately describes the data landscape
		Data Flow Diagram	Data flow diagram is present and accurately defines the flow of data between processes and agents, using the correct syntax
		Process Flow Diagram	Process flow diagram is present and accurate. It defines the flow of logic between the process and decision steps, using the correct syntax
		User Manual	User manual is present and appropriately defines how a user would use the system, with enough detail to show basic and core functions of the solution
		Website Hosted on Cloud	The website should be accessible from any computer without having to compile and run the code on that computer
Development	Infrastructure	File Storage Hosted on Cloud	The files should be stored on an online server or file storage mechanism that is secure. Files stored on a local computer and cannot be accessed from a different computer do not qualify as cloud storage
		Database Hosted on Cloud	The database should be accessible, through the site, from any computer without having the database created on that computer.

Fig. 8. Section of rubric categories, sub-categories, criteria and description

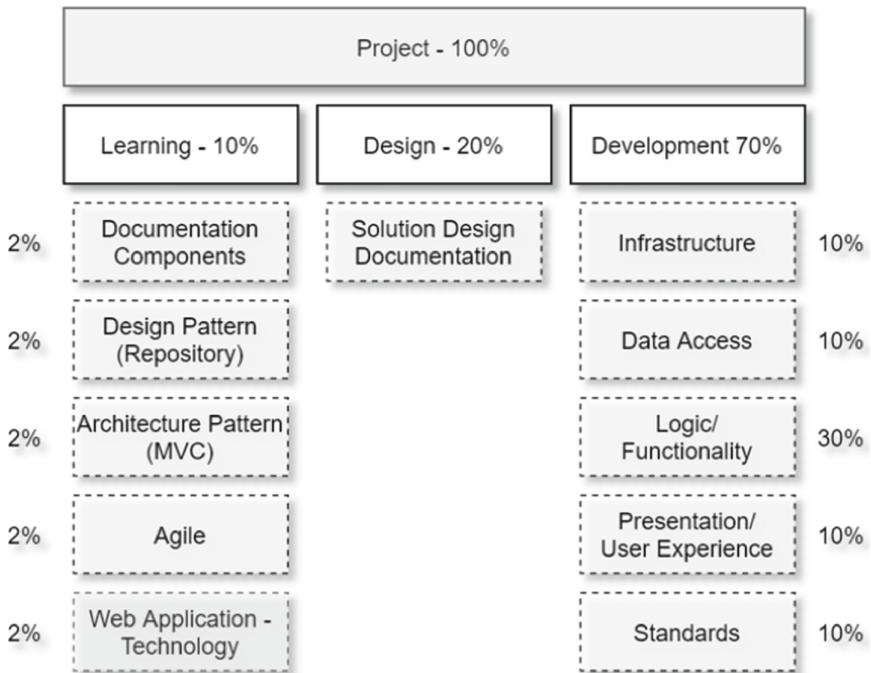


Fig. 9. Assigning weighting

Figure 10 elaborates further on the weightings defined in Fig. 9, with the binary indicator which impacts the overall mark. In a case where the weighting of a criteria is 2% and the binary indicator is '1', meaning that the criterion has been adequately

fulfilled, the mark contribution increases by 2. In a case where the binary indicator is ‘0’, meaning that the criterion has not been adequately fulfilled, the mark contribution does not increase.

Category	Sub-Category	Criteria	Description	Weight	Binary Indicator	Mark
Learning	Agile	[Online test] Agile	Agile concepts tested via online test	2%	1	2
	Design Pattern	[Online test] Design Pattern	Design pattern knowledge tested via online test	2%	0	0
	Architecture Pattern	[Online test] Architecture Pattern	Architecture pattern knowledge tested via online test	2%	0	0
	Document Management	[Online test] Documentation Components	Documentation components knowledge tested via online test	2%	1	2
	Web-Application Technology	[Practical assignment] Web Application Technology	Web Application technology to be tested via online test/assignment and practically applied as part of the project	2%	1	2
Design	Solution Design Documentation	ERD	ERD is provided and accurately describes the data landscape	6%	1	6
		Data Flow Diagram	Data flow diagram is present and accurately defines the flow of data between processes and agents, using the correct syntax	4%	1	4
		Process Flow Diagram	Process flow diagram is present and accurately defines the flow of logic between the process and decision steps, using the correct syntax	6%	0	0
		User Manual	User manual is present and appropriately defines how a user would use the system, with enough detail to show basic and core functions of the solution	4%	1	4

Fig. 10. Section of the example weighted rubric

6 Future Work

In future iterations on this work the subjective nature of assessment criteria would need to be addressed and refined. The use of dialogistic studies could be used to formulate specific questions with very specific intent, suitable for binary assessment without any room for subjectivity.

The comparison of the approach with other, more traditional, ways of assessing students would also need to be taken into consideration as part of future work once the approach provides guidelines for less subjective criteria.

The rubrics will be expanded to consider assessment not just at the end of the project when students hand in their work at the end, but instead to measure success earlier on in projects as milestones are reached so that lecturers can establish early on when there are problems in students’ abilities to meet the requirements of the project scope. This way a more data driven approach can be taken to assessment which could lead to a more dynamic teaching approach which will serve to address students’ needs as soon as they become relevant.

Additionally, the binary nature of the rubric elements lend themselves to automation which could allow for some of the marking in project based work to be done using suitable automation tools. This way students don’t need to wait for the lecturer to mark a batch of students’ work and can instead submit partially completed work for assessment and get immediate feedback. Automating this process will also allow the lecturer to track the progress of students in real time rather than in batches.

Through the combined use of automation and analytics, this study has the potential to create a level of transparency, traceability, understanding and insight into students performance and the mapping to very specific outcomes (assessment, study unit and module).

7 Conclusion

The expectation of Computer Science, Information Systems and Information Technology degrees is that a student who completes and passes the degree has the knowledge and ability to execute on each of the module outcomes throughout the degree. These module outcomes are a rolled-up version of study unit and assessment outcomes. In the event that students cannot perform certain outcomes, feedback should be clear and immediate, which categorical and sliding scales tend to obfuscate. The use of binary scales to evaluate project-based assessments allows marking (or rubric) criteria to be specifically identified as fulfilled fully or not at all. The concept that forms the foundation of using binary scales to evaluate project-based assessments needs to be instilled in all elements of the assessment, from the conceptual formation of the scope through to the weighting and rubric assessment criteria. This approach to evaluating project-based assessments would lead to more effective, granular feedback to students that could be automated and implemented in such a way as to allow students to get feedback as they progress through their assignments rather than at the end when they have completed them leading to a more dynamic form of teaching.

Acknowledgement. The full scope of the hackathon, all data used and the template rubric for use in other projects can be obtained at <https://github.com/JacquiM/The-Use-of-Binary-Scales-in-Rubrics-to-Evaluate-Computer-Science-Assessments>.

References

- da Alves, N.C., von Wangenheim, C.G., Hauck, J.C.R., Borgatto, A.F.: A large-scale evaluation of a rubric for the automatic assessment of algorithms and programming concepts. In: Proceedings of the 51st ACM Technical Symposium on Computer Science Education, pp. 556–562. Association for Computing Machinery, New York (2020)
- Dawson, P.: Assessment rubrics: towards clearer and more replicable design, research and practice. *Assess. Eval. High. Educ.* **42**(3), 347–360 (2017)
- Dimitrov, D.M.: An approach to scoring and equating tests with binary items: piloting with large-scale assessments. *Educ. Psychol. Meas.* **76**(6), 954–975 (2016)
- Department of Basic Education: Curriculum Assessment Policy Statements (CAPS) (2021). [https://www.education.gov.za/Curriculum/CurriculumAssessmentPolicyStatements\(CAPS\)/tabid/420/Default.aspx](https://www.education.gov.za/Curriculum/CurriculumAssessmentPolicyStatements(CAPS)/tabid/420/Default.aspx). Accessed 09 Feb 2022
- Haghdoost, A.A.: Complexity of the socioeconomic status and its disparity as a determinant of health. *Int. J. Prev. Med.* **3**(2), 75–76 (2021)
- Hevner, A.R.: A three cycle view of design science research. *Scand. J. Inf. Syst.* **19**(2), 1–6 (2007)
- Hevner, A.R., March, S.T., Park, J.: Design science in information systems research. *Manag. Inf. Syst. Q.* **28**(1), 75–105 (2004)

- Imbault, C., Shore, D., Kuperman, V.: Reliability of the sliding scale for collecting affective responses to words. *Behav. Res. Methods* **50**(6), 2399–2407 (2018). <https://doi.org/10.3758/s13428-018-1016-9>
- Jones, D., Tadros, A.: Successful use of rubrics to assess student performance in capstone projects. In: 2010 Annual Conference & Exposition, pp. 15.1144.1–15.1144.8. ASEE, Louisville, Kentucky (2010)
- Janse van Rensburg, J.T., Goede, R.: Reflecting on the use of project-based learning for 21st century competencies in an IT extended programme. In: Proceedings of the 10th Annual UNISA ISTE Conference on Mathematics, Science and Technology Education, pp. 76–84. UNISA (2019)
- Lukka, K.: The constructive research approach. In: Case Study Research in Logistics. Publications of the Turku School of Economics and Business Administration, Series B, vol. 1, pp. 83–101 (2003)
- Morolong, P.: Teachers' views on the challenges they experience when implementing the curriculum and assessment policy statement in the foundation phase. (Dissertation), Central University of Technology, Free State (2009)
- Mustapha, A., Samsudin, N.A., Arbaiy, N., Mohamed, R., Hamid, I.R.: Generic assessment rubrics for computer programming courses. *Turk. Online J. Educ. Technol.* **15**(1), 53–68 (2016)
- Novita, L.: The design of assessment instruments in English research. *J. Engl. Educ.* **3**(1), 27–30 (2017)
- Park, H., Yan, X.: An investigation into rater performance with a holistic scale and a binary, analytic scale on an ESL writing placement test. *Pap. Lang. Test. Assess.* **8**(2), 34–64 (2019)
- Simon, H.A.: *The Sciences of the Artificial*, 3rd edn. MIT Press, Cambridge (1996)
- Plomp, T., Nieveen, N.M.: *Educational Design Research*, 1st edn. SLO, Enschede (2013)
- Zano, K.: Exploring teachers' enactment of the Curriculum and Assessment Policy Statement (CAPS) in selected Free State Province schools. (Dissertation), UNISA, Pretoria (2015)