



Vulindlela – making new pathways

17th DEFSA conference – 21|22 September 2023

Flipping the script: Using artificial intelligence to design authentic assessment rubrics

Lizette Carstens, The Independent Institute of Education, Vega School
Maretha Geysler, The Independent Institute of Education, Vega School
Christiaan Graaff, The Independent Institute of Education, Vega School

Abstract

Generative artificial intelligence (AI) is a key driver of innovation across all sectors and in education it has the power to optimise teaching and learning to benefit educators and students alike. However, the increasing prominence and influence of AI in domains previously exclusive to humans, such as design, raises urgent questions about the assessment of learning in design education. Recent writings in the field of design education agree that in the age of AI, educators need to revisit existing assessment practices. Conversations about AI and assessment practices appear to revolve around upholding academic integrity and defining what should be assessed when students can create design outcomes using generative AI. This study flips the script: Instead of focusing on managing the use of AI by students to create design outcomes, this study asks a further question. How can design educators use generative AI to facilitate the design of authentic assessment rubrics?

An outcomes-based approach is common in South African design education, and practical projects are used as assessment instruments. The focus of the assessments is on helping students apply their knowledge and skills practically in a real-world context. As a result, design assessments lean towards assessing the design process rather than relying solely on the design outcome, therefore, assessing students' ability to articulate the reasoning behind their application of knowledge and skills. Assessments used for practical projects typically use assessment rubrics that feature clear evaluation criteria, and aim to provide comprehensive feedback to students. Using authentic assessment, a student-centred approach to assessment practices, curriculum designers can effectively mimic the real working environment, align to the principles and dimensions of authentic assessment, promote deeper learning, and ensure that holistic learning experiences for students are taken into account. However, defining assessment criteria and descriptors that are authentic, clear, well-formulated, and not overly restrictive demands careful consideration to avoid gaps, making it a time-consuming process.

This paper explores the potential of using AI by using ChatGPT in the design of authentic assessment rubrics, and presents a critical reflection on insights gained from action research. The research was conducted by three curriculum designers working across various design qualifications at a private higher education provider. The study contributes to the emerging broader discussion on the innovative use of AI to support teaching and learning within the field of design education.

Keywords: Artificial intelligence, assessment rubrics, assessment practice, authentic assessment.

Introduction

Artificial intelligence (AI) has been used in education for a while. Tools such as Turnitin, for example, are widely used to encourage honesty and foster academic integrity (Chen, Chen & Lin 2020). Recent advances have given AI the ability to generate or create new content such as code, images, audio, and text (Zhihan 2023), causing concern among educators about how we assess learning – especially in domains like design that were previously considered exclusive to humans (Siemens 2022). Even before the latest developments in generative AI, Grainger and Weir (2020) criticised traditional assessment practices for lacking transparency and fairness, and failing to align assessment criteria with learning outcomes. They argue that these practices provide only a snapshot of what students can do at a specific point in time, rather than a nuanced view of their learning (Grainger & Weir 2020).

Design education in most South African institutions follows an outcomes-based approach, and practical projects are used as assessment instruments (Giloi & Du Toit 2013). Design assessments emphasise the practical application of knowledge and skills in a real-world setting (Giloi & Du Toit 2013). However, if design education aspires towards more than skills development and technical mastery, assessments should also focus on process, and encourage students to actively engage in critical thinking, conceptualisation, and contextualising their work within the broader design practice (Giloi & Du Toit 2013). As the focus of design education becomes more learner-centred and outcomes-based, we are grappling with the challenge of providing students with clear learning outcomes that foster creativity, deep learning and understanding, yet accommodate the unique nature of a creative field; it should support transparent, reliable assessment of the person, the creative process, and the final artefact (Giloi & Du Toit 2013; Siemens, McGraw & Kelly 2022). Likewise, current perspectives on the impact of AI in education suggest that assessments should emphasise the process of learning, and shift its focus from what, who and when, to how and why (Academy of Science South Africa 2023; Carvalho, Martinez-Maldonado, Tsai, Markauskaite & De Laat 2022; Hardman 2023; Swiecki, Khosravi, Chen, Martinez-Maldonado, Lodge, Milligan, Selwyn & Gašević 2022).

Assessment practices typically emphasise three dominant areas of focus: The learning outcomes or skills that must be assessed, tasks students engage in to demonstrate those skills, and the instruments used to relate the evidence from tasks to the achievement of the desired outcomes (Swiecki et al. 2022). Giloi and Du Toit (2013) propose authentic assessment, a student-centred assessment approach that promotes deeper learning and focuses on the holistic learning experience of the individual by providing opportunities to engage with real-world problems in contexts mimicking a real-world work environment, as an alternative to traditional assessment practices. However, to support student engagement in the learning experience, such assessments should be guided by clear criteria that accurately describe the standards for mastery of the subject matter (Grainger & Weir 2020). Moreover, assessments with well-defined assessment criteria and clear rubrics foster more reliable assessments and support assessors, regardless of their level of experience or tacit knowledge, to assess work consistently (Grainger & Weir 2020). Developing such assessment rubrics can be arduous and time-consuming and requires a deep understanding of the discipline by the person designing the assessment (Chen et al. 2020).

So far, research has focused predominantly on the meaningful integration of AI as a tool for students to use during assessment activities. Chen et al. (2020) argue that both students and lecturers can potentially benefit from the use of AI to support assessment design (Chen, Chen & Lin 2020), but more research is needed on how AI can be integrated in determining the learning outcomes to be assessed

and in defining assessment criteria that accurately describe the standards for mastery of the subject matter (Swiecki et al. 2022).

The aim of this study was to explore how generative AI can be used in the design of authentic assessment rubrics, and to identify potential benefits and challenges. In this paper, we begin by providing a theoretical overview of the characteristics and principles of authentic assessment, which we employed as a framework to guide and evaluate our development of assessment rubrics using ChatGPT. Subsequently, we discuss the process and methods we used. We conclude with a critical personal reflection on the lessons we learnt and the benefits and challenges we identified during our investigation. The study contributes to an ongoing and broadening discussion on the innovative use of AI to support teaching and learning in design education.

Authentic assessment practices

The concept of authentic assessment first emerged in literature in the 1980s, spurred by the need to adequately prepare students for the world of work. It gave rise to more realistic and student-centred assessment approaches (Archbald & Newmann 1988; Frey, Schmitt & Allen 2012). Villarroel, Bloxham, Bruna, Bruna and Herrera-Seda (2018) systematically reviewed 112 articles on authentic assessment and instruction spanning more than 30 years. They identified 13 central characteristics, which they refined into three conceptual dimensions of authentic assessment (Figure 1).

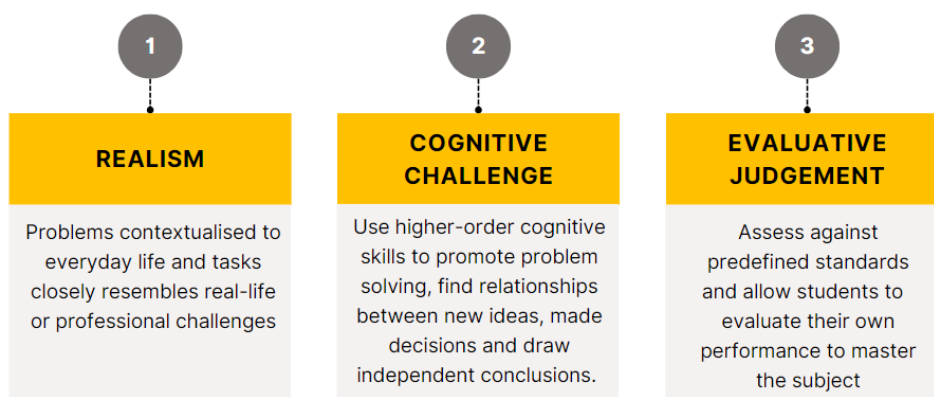


Figure 1: Three conceptual dimensions of authentic assessment (adapted from Villarroel et al. 2018)

The first dimension, Realism, refers to the incorporation of a real-world context in the assessment, where the task replicates challenges encountered in professional work settings (Villarroel et al. 2018). Realistic assessments are performance-based, requiring students to produce work that closely aligns with the demands of the respective profession, accurately representing or simulating the knowledge, understanding, and skills required in that field of employment (Villarroel et al. 2018).

The second dimension, Cognitive Challenge, relates to the cognitive complexity of learning objectives and tasks (Anderson & Krathwohl 2001) to ensure that an assessment is cognitively challenging and includes higher-order thinking skills (Villarroel et al. 2018). Instead of simply regurgitating information, cognitively complex assessments promote problem solving, decision making, the identification of relationships between new ideas, connection of theoretical concepts to real-life experiences, and the analysis of existing data to draw independent conclusions (Villarroel et al. 2018).

The third dimension, Evaluative Judgement, involves the practise of assessing against predefined standards such as those included in assessment rubrics (Villarroel et al. 2018). Setting clear and explicit assessment criteria at the start of an assessment promotes students’ comprehension of the required standard, empowers them to reflectively evaluate their performance and quality of work, and supports them in mastering the subject (Villarroel et al. 2018). In this study, the dimensions of authentic assessment were used as the guiding theoretical framework to define the assessment tasks and criteria included in ChatGPT prompts.

The process used is based on the work of Jon Mueller, Professor of Psychology at North Central College in Naperville. Mueller (2005) developed the Authentic Assessment Toolbox, a tool to design authentic standards, tasks, and assessment rubrics that measure and improve student learning.

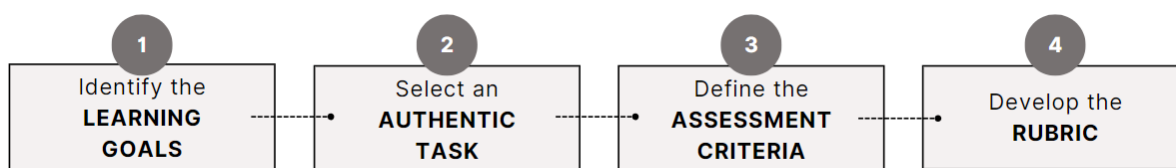


Figure 2: Mueller’s (2018) Authentic Assessment Toolbox

The toolbox presents a four-step process to guide educators in designing authentic assessments (Figure 2). In the first step, the learning outcomes to be assessed are defined (Mueller 2005). These outcomes inform the assessment tasks and criteria. In the second step, an appropriate assessment task that allows students to authentically demonstrate the extent to which they have mastered the defined learning outcomes is selected. In the third step, criteria for assessing mastery of the specified learning outcomes are defined (Mueller 2005). The first three steps in the process lay the foundation for the final step, where an assessment rubric is developed (Mueller 2005). This toolbox supports curriculum designers in considering the alignment of learning outcomes, assessment tasks and criteria, and clearly articulating the expected outcomes and standards to students (Mueller 2005). Clear outcomes and standards support students’ understanding of expectations and support them in achieving learning goals (Mueller 2005).

Methodology and process

This study was born from the curiosity of three curriculum designers responsible for three distinct design qualifications at a private higher education institution. We investigated the potential use of generative AI to support the development of quality, authentic assessment rubrics within tight institutional deadlines. The aim was to explore the use of ChatGPT, a generative AI tool, to support the development of authentic rubrics for design assessments. Action research was employed because the cycles of observation, reflection, planning, and action inherent to this mode of enquiry are uniquely suited to the purpose of improving the practice of curriculum design (Cresswell & Guetterman 2019). Moreover, action research aligns with the everyday practices of design (Cole, Purao, Rossi & Sein 2005), presenting a mode of enquiry that extends the everyday practices of curriculum designers into a vehicle for academic enquiry.

The launch of ChatGPT early in December 2022 caused concern as educators and institutions wondered if and how students should be taught to harness it as a tool, while maintaining academic integrity (Buckingham-Schum 2023). The notion of a tool capable of generating quality content faster and more efficiently, intrigued us. What if we flipped the script and instead of considering how

students could use it to complete assessment, we explored its usefulness in designing assessment rubrics?

Initially, we randomly prompted ChatGPT to generate assessment rubrics for various existing design projects. Although the ChatGPT rubrics left much to be desired, some descriptors hinted at learning outcomes, assessment criteria, or assessment tasks that could be added to enhance the authenticity of our assessments. Our observations suggested that ChatGPT could be used as a generative tool to support our design of meaningful assessment tasks and rubrics that provide clear guidelines of expected standards to both students and markers.

We decided that our initial observations warranted a deeper investigation by following a more rigorous approach. From a review of the literature on authentic assessment, we identified appropriate criteria for evaluating the authenticity of assessment rubrics and a process that we could draw on to reliably conduct our investigation. We combined the Authentic Assessment Toolbox (Mueller 2005) and the Dimensions of Authentic Assessment (Villarroel et al. 2018) to create an Authentic Assessment Planning worksheet (Appendix A) that could be used to evaluate the alignment between learning outcomes, assessment tasks, and criteria.

We applied purposive criterion sampling to select assessments with sufficient similarities and differences to compare and contrast the findings (Palinkas et al. 2015). We drew a sample of nine assessments to include one assessment per NQF Level across levels 5, 6 and 7, each from three distinct design disciplines. While each discipline has its unique theories, skills, and learning outcomes, they share common philosophical paradigms, design processes, and creative practices (Cross 2007; Dorst 2017). All assessments in the sample evaluate core design competencies such as research, abductive thinking, creativity, and iterative prototyping, as well as the application of discipline-specific knowledge and skills. As NQF levels increase, assessments are progressively more cognitively demanding, with stricter mastery criteria. Competencies assessed across the three qualifications and the three NQF levels allowed sufficient similarity for meaningful comparisons, while discipline-specific and level-specific knowledge and skills in each assessment provided sufficient differences for contrast. We regarded a small sample as adequate because the goal of the study was only to discover possible ways in which generative AI might support authentic assessment development and improve our practice as educators, not to derive heuristics or generalisable guidelines.

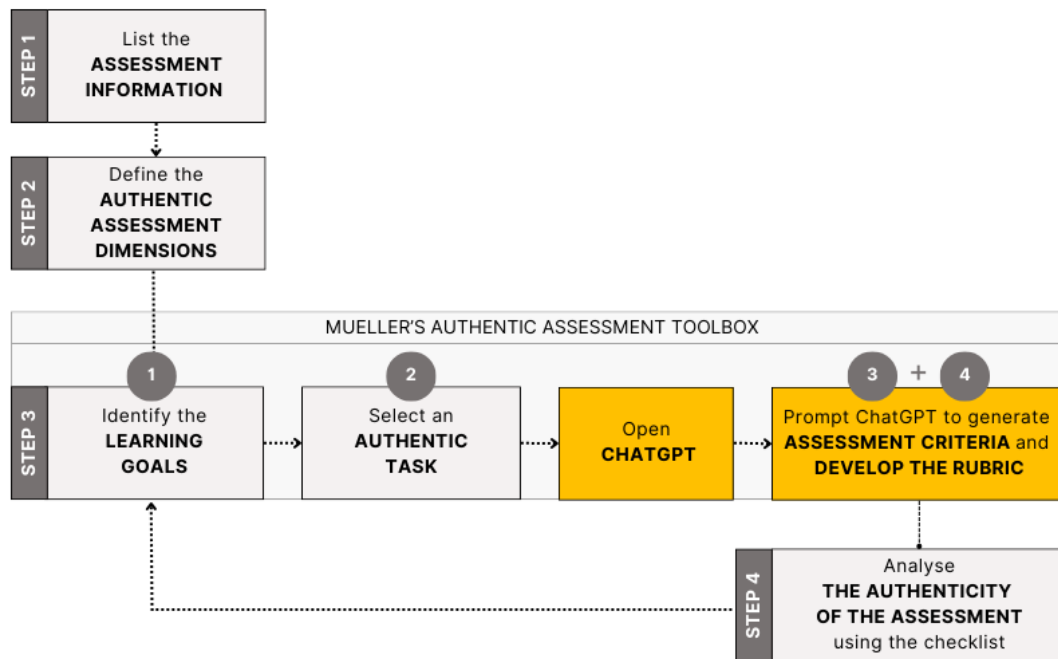


Figure 3: Process to plan, generate, and evaluate assessment rubrics with ChatGPT (authors' own)

Each researcher independently followed the process outlined in Figure 3 (above). In the Authentic Assessment Planning worksheet (Appendix A), we used learning outcomes and assessment tasks from our existing assessments as a departure point. We employed these defined learning outcomes and assessment tasks as prompts for ChatGPT to generate assessment criteria and rubric descriptors. We systematically evaluated the generated rubrics against the Authentic Assessment Checklist (Appendix B) and iteratively refined the learning outcomes and assessment tasks included in the ChatGPT prompts after each evaluation until the generated rubrics met the criteria in the Authentic Assessment Checklist.

Findings

Although the design qualifications we manage all share certain broad design competencies, each qualification also includes discrete, discipline-specific knowledge and skills. The ability to apply such knowledge and skills must also be assessed. In addition, the real-world problems and world-of-work contexts of the three design disciplines differ, resulting in different nuances in the assessment approaches of each curriculum designer. In order to facilitate comparison, each of us reported on our discrete findings in this study.

Design Qualification 1

To simulate real-world contexts, the scope of projects in this qualification is quite big, and assessments are in the form of practical Portfolios of Evidence with discrete parts building on one another. I used the assessment descriptions, learning outcomes and assessment tasks of existing NQF Level 5, 6 and 7 assessments for the purpose of this study. Following the process outlined in Figure 3, I included a description and the learning outcomes and assessment tasks in my initial ChatGPT prompt to generate a rubric (Figure 4). ChatGPT used the assessment tasks to generate criteria. An evaluation of the generated rubric revealed that the rubric only assesses whether tasks were completed or not, and

that the descriptors provide little or no guidelines about the expected standard of mastery of learning outcomes (Figure 5).

Create a five-column rubric to assess a concept for a basic board game where students had to demonstrate the ability to conduct research to inform and original idea for a game that will appeal to players, the ability to generate multiple original game ideas and select the best one and the ability to develop the idea into a unique board game concept through considered application of game play and core game mechanics.

Figure 4: The first prompt to generate a rubric included assessment description, learning outcomes, and assessment tasks

Criterion	Unsatisfactory (1)	Developing (2)	Proficient (3)	Advanced (4)	Exceptional (5)
Conduct, Analyze, and Interpret Research	The student did not conduct research or failed to analyze and interpret findings.	The student conducted some research but struggled to analyze and interpret findings effectively.	The student conducted research and provided basic analysis and interpretation.	The student conducted thorough research, and the analysis and interpretation were insightful.	The student conducted extensive research, offering deep insights through analysis and interpretation.
Affinity Diagram for Analysis of Findings	The student did not use an affinity diagram for analysis, or it was used incorrectly.	The student attempted to use an affinity diagram but did so with errors or limited effectiveness.	The student used an affinity diagram effectively, but the application was somewhat basic.	The student used an affinity diagram skillfully to analyze findings, demonstrating a strong understanding.	The student used an affinity diagram masterfully, resulting in a comprehensive and insightful analysis.
Creation of Player Persona	The student did not create a player persona or did so with significant inaccuracies.	The student created a player persona, but it lacked depth or relevance to the research findings.	The student created a relevant player persona with reasonable depth and insights.	The student created a well-crafted player persona that effectively demonstrated insights from research.	The student created an exceptionally detailed and insightful player persona, showcasing a deep understanding of research findings.

Figure 5: In the rubric generated with ChatGPT, assessment tasks were turned into assessment criteria, and descriptors are superficial

For the second attempt, I only included learning outcomes as defined in existing assessments but added a purpose at the end of each outcome (Figure 6). ChatGPT turned the information in the prompt into assessment criteria. The generated rubric descriptors assessed the completion of a task, but also considered to what extent the purpose had been met (Figure 7). The descriptors provided better guidelines for the standard of mastery (see Column 5), but were still superficial and open to interpretation, using words like creativity and diversity.

Create a five column rubric to assess the following learning outcomes: Conduct research to inform game ideas
 Generate multiple game ideas
 Evaluate ideas to select the best one
 Apply the principles of game design to develop a game idea into a game concept

Figure 6: Prompt based on learning outcomes and purpose

Criterion	Unsatisfactory (1)	Developing (2)	Proficient (3)	Advanced (4)	Exceptional (5)
Conduct Research to Inform Game Ideas	The student did not conduct research or failed to use it to inform game ideas.	The student conducted minimal research, but its application to game ideas was weak.	The student conducted research and applied it somewhat effectively to game ideas.	The student conducted thorough research and used it proficiently to inform game ideas.	The student conducted extensive research, resulting in a rich source of insights for game ideas.
Generate Multiple Game Ideas	The student struggled to generate multiple game ideas, producing only a few, or they lacked diversity and creativity.	The student generated some game ideas, but the quantity or creativity was limited.	The student generated a variety of game ideas, demonstrating creativity and diversity.	The student generated numerous creative game ideas, showcasing a high level of creativity and diversity.	The student produced a wealth of exceptionally creative game ideas, reflecting remarkable diversity and innovation.

Figure 7: In the ChatGPT-generated rubric, learning outcomes were turned into assessment criteria, descriptors do not consider assessment tasks, and lack clear guidelines

On critical reflection, I realised that the learning outcomes in the existing assessments were, in fact, assessment tasks and not learning outcomes at all. One task often assessed several learning outcomes. Consequently, I redefined the learning outcomes. I included the new learning outcomes and brief descriptions of associated assessment tasks in my next prompt (Figure 8). The generated rubric assessed each discrete learning outcome and included guidelines for the standards of mastery (Figure 9, Column 6). However, the descriptors in the columns to the left still lacked depth or accurate guidelines for improvement (Figure 9, Columns 2 and 3)

Please create a five column rubric to assess the following outcomes: Analyse board games to identify what makes them fun, Gather and analyse player information to identify their preferences, interpret the findings to gain insight and inform unique and fun game ideas , create a player persona to personify the target audience

Figure 8: A prompt including the assessment task and the discrete outcomes to be assessed

Criteria	Unsatisfactory (1)	Basic (2)	Proficient (3)	Advanced (4)	Exceptional (5)
Analyze Board Games for Uniqueness and Fun	Limited or no analysis of board games	Basic analysis with some depth	Thorough analysis with depth	Comprehensive analysis with exceptional depth and insight	Exceptional analysis with exceptional depth, insight, and innovation
Analyze Player Interviews for Preferences and Dislikes	Limited or no analysis of player interviews	Basic analysis with some depth of interview responses	Thorough analysis with depth and insight	Comprehensive analysis with exceptional depth and insight	Exceptional analysis with exceptional depth, insight and innovation
Interpret Findings on Game Uniqueness and Fun	Limited or no interpretation of findings	Basic interpretation with some depth	Thorough interpretation with depth and insight	Comprehensive interpretation with exceptional insight	Exceptional interpretation with innovative insight

Figure 9: In the ChatGPT-generated rubric, the learning outcomes were used as assessment criteria, and the descriptors included more detailed guidelines

I was curious about the ability of ChatGPT to identify assessment criteria, and in my next prompt, asked what it would consider good evidence of research in my project (Figure 10).

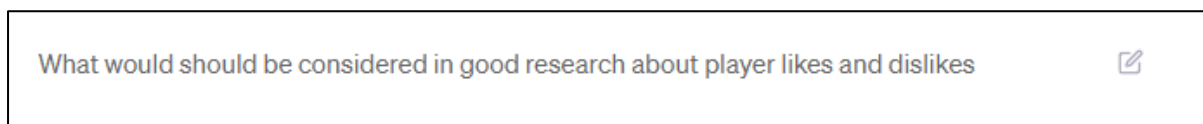


Figure 10: Prompt to ask ChatGPT what it would consider good evidence

ChatGPT returned 15 considerations. Several were too advanced for the level of the assessment (Figure 11), but Game Selection, Player feedback and motivations, Ethical considerations, and Actionable recommendations presented criteria that were covered in the learning content, appropriate for the assessment level, and could meaningfully assess learning. I prompted ChatGPT to generate rubric descriptors using these four considerations as assessment criteria. An evaluation of the generated rubric against the Authentic Assessment Checklist confirmed that this rubric included appropriate assessment criteria and clear, detailed descriptors to guide students and markers on the standards of mastery, and to provide meaningful feedback for learning (Figure 12).

4. **Game Selection:** Include a variety of board games in your research, covering different genres, complexity levels, and player counts. This allows for a more holistic understanding of player likes and dislikes across various game types.
5. **Player Feedback:** Consider both explicit feedback (what players say they like or dislike) and implicit feedback (observed behaviors, emotional reactions) to paint a complete picture.
6. **Player Motivations:** Explore why players enjoy certain aspects of board games. Understand their motivations, whether it's for social interaction, competition, strategy, storytelling, or other factors.
7. **Contextual Factors:** Recognize that player likes and dislikes can be influenced by contextual factors such as the gaming group, setting, and duration of play. Consider these variables in your research.
8. **Comparative Analysis:** Compare and contrast player preferences and dislikes across different board games. Identify patterns and common themes that emerge.
9. **Player Segmentation:** Analyze data to segment players into groups with similar preferences. This can help in tailoring game design to specific target audiences.
10. **Longitudinal Studies:** Consider conducting longitudinal studies to track changes in player preferences over time. This can be valuable for understanding trends in the gaming community.
1. **Ethical Considerations:** Ensure that your research adheres to ethical guidelines. Obtain informed consent from participants and protect their privacy.
2. **Expert Opinions:** Supplement player feedback with insights from board game experts, designers, and reviewers. Expert opinions can provide additional context and validation.
3. **Feedback Channels:** Explore various feedback channels, including online forums, social media, and game review platforms, to gather player opinions and sentiments.
4. **Data Analysis Tools:** Utilize appropriate data analysis tools and techniques to identify patterns, correlations, and trends in player likes and dislikes.
5. **Actionable Recommendations:** Conclude your research with actionable recommendations for game designers, developers, and publishers. Provide insights on how to enhance

Figure 11: Criteria for evidence of good player research generated in ChatGPT

Criterion	Unsatisfactory (1)	Developing (2)	Proficient (3)	Advanced (4)	Exceptional (5)
Game Selection (variety and relevance of board games included in the research)	Limited game selection with no diversity in genres, complexity, or player counts. Games chosen lack relevance to the research objectives.	A basic selection of board games, covering only a few genres or player counts. Some games may not align well with research objectives.	Offers a diverse selection of board games, including various genres, complexity levels, and player counts. Games chosen are relevant to the research objectives.	An extensive and well-balanced selection of board games that represents a wide range of genres, complexity levels, and player counts. Games chosen are highly relevant to the research objectives.	Exceptional game selection that showcases a comprehensive array of board games from various genres, complexity levels, and player counts. Games chosen are exceptionally relevant to the research objectives.
Depth and relevance of player feedback and motivations	Limited player feedback gathered. Fails to explore player motivations or delve into the "why" behind preferences.	Collects player feedback, but it lacks depth or is predominantly quantitative. - Provides minimal insights into player motivations. - Some feedback may be tangential to the research objectives.	Collects player feedback, but it lacks depth. Provides minimal insights into player motivations.	Gathers comprehensive player feedback. Identifies and explores player motivations related to likes and dislikes. Feedback is largely relevant.	Collects rich player feedback from diverse sources. Offers deep insights into player motivations, delving into the underlying reasons behind preferences. Feedback is highly relevant.
Ethical considerations in the research	Fails to address ethical considerations in the research. - Does not obtain informed consent from participants or protect their privacy.	Addresses some basic ethical considerations, such as informed consent, but with shortcomings. Privacy protection measures may be insufficient or inconsistent.	Adequately addresses ethical considerations, including informed consent and privacy protection.	Demonstrates a strong commitment to ethical principles in research. Obtains informed consent and protects participant privacy effectively.	Exhibits an exceptional commitment to ethical standards, with meticulous attention to detail. Exceeds expectations in obtaining informed consent and safeguarding participant privacy.

Figure 12: The rubric generated in ChatGPT based on generated assessment criteria

With the remaining two assessments, I refined my process. Before prompting ChatGPT to generate a rubric based on learning outcomes in existing assessments, I prompted it to generate assessment criteria and then checked the alignment of those criteria against existing learning outcomes. I then selected those criteria appropriate to the assessment level and aligned with the outcomes and used them in my ChatGPT prompts to generate assessment rubrics. I evaluated the generated rubrics against the Authentic Assessment Checklist. The generated rubrics required only minor adjustments, such as the simplification of the language used in the descriptors. Using the refined process, I was able to generate a quality, authentic rubric in an hour, compared to the two to three hours that I required without ChatGPT.

Design Qualification 2

I started by providing ChatGPT with detailed prompts, including the learning outcomes, the module purpose, the tasks, and so on. The generated rubric was very superficial. Consequently, for the second attempt, I used short instructions in the prompts (Figure 13). The rubrics generated in ChatGPT

provided more detailed guidelines (Figure 14). Additional conversational prompts (Figure 15) generated more detailed assessment criteria and descriptors (Figure 16).

Design a 5-column marking rubric to assess the following learning outcomes:

- Analyse an existing Healthy Space and what principles you can apply in your design;
- Design and develop an original space that intends to facilitate healing and mitigate stress and is inclusive and accessible to all
- Use advanced industry appropriate technology to effectively convey a design to the client.
- Apply relevant building regulations pertaining to the built environment relative to interior design solutions with the use of working drawings and specifications
- Design an overall creative and visually appealing client presentation to industry standards

Figure 13: A simplified prompt used to generate an assessment rubric on ChatGPT

Criteria	Excellent (5)	Proficient (4)	Satisfactory (3)	Needs Improvement (2)	Poor (1)
Outcome 1: Analyze Healthy Space					
Demonstrates deep understanding of principles in existing healthy spaces and effectively applies them in design.	Exceptional insight and application.	Proficient understanding and application.	Adequate understanding and application.	Limited understanding and application.	Minimal understanding and application.

Figure 14: The results of using a simplified prompt to generate descriptors in ChatGPT

Provide more details under level 1-4 not just 5

Figure 15: Additional prompt to refine rubric created using ChatGPT

Criteria	Level 5 (Excellent)	Level 4 (Proficient)	Level 3 (Satisfactory)	Level 2 (Needs Improvement)	Level 1
Outcome 1: Analyze Healthy Space	Exceptional insight and application. Student demonstrates a deep understanding of principles in existing healthy spaces and effectively applies them in the design.	Proficient understanding and application. Student has a strong understanding of principles in existing healthy spaces and successfully applies them in the design.	Adequate understanding and application. Student demonstrates an understanding of principles in existing healthy spaces and applies some of them in the design.	Limited understanding and application. Student has a basic understanding of principles in existing healthy spaces but struggles to apply them effectively in the design.	Minimal understanding and application. Student has an understanding of principles in existing healthy spaces but fails to apply them effectively in the design.
Outcome 2: Design Healing	Innovative and comprehensive design that	A well-thought-out design with	A design with some elements, but not fully	A design with limited consideration	Poorly

Figure 16: The rubric generated in ChatGPT after refining the assessment criteria

I wanted to investigate what other assessment considerations ChatGPT might include in the rubric. Instead of using all five learning outcomes in the first prompt, I experimented by only including one learning outcome (Figure 17) per prompt. ChatGPT generated five subcategories to be assessed. The first rubric (Figure 16) generated in ChatGPT assessed insight and application, but the second rubric (Figure 18) deconstructed the requirements to evaluate insight and application. The generated rubric considered understanding, identification of principles, integration into the design, clarity of communication, and critical thinking. Combining these assessment criteria resulted in clear descriptions of what is required to demonstrate mastery of the learning outcome (Figure 19).


Design a 5-column marking rubric to assess the following learning outcome: Analyse an existing Healthy Space and what principles you can apply in your design 

Figure 17: Simplified prompt using only one learning outcome

Criteria	Excellent (5)	Very Good (4)	Good (3)	Fair (2)	Number of Responses (1)
Understanding of Healthy Space	Exceptional understanding of what constitutes a Healthy Space, demonstrated through detailed analysis and clear articulation of key elements and their impact.	Strong understanding of Healthy Space, with clear analysis of its key elements and their relevance.	Adequate understanding of Healthy Space, with some analysis of key elements.	Limited understanding of Healthy Space, with minimal analysis of key elements.	Percentage of responses
Identification of Principles	Thorough identification and articulation of multiple design principles that can be applied, with a clear rationale for each.	Identification of several design principles that can be applied, with a rationale for most.	Identification of some design principles that can be applied, but rationale may lack depth or clarity.	Limited identification of design principles, with weak or unclear rationale.	Number of design principles
Integration into Design	Demonstrates a comprehensive and creative integration of Healthy Space principles into the design proposal, showcasing an	Effectively integrates Healthy Space principles into the design proposal, showcasing an	Partially integrates Healthy Space principles into the design proposal, but the connection	Limited integration of Healthy Space principles, with a tenuous link to the design	Depth of integration

Figure 18: The rubric generated in ChatGPT when using one learning outcome in the prompt

Combine the text and shorten: Exceptional understanding of what constitutes a Healthy Space, demonstrated through detailed analysis and clear articulation of key elements and their impact. Thorough identification and articulation of multiple design principles that can be applied, with a clear rationale for each. Demonstrates a comprehensive and creative integration of Healthy Space principles into the design proposal, showing a deep understanding of their applicability. Communication is exceptionally clear, well-organized, and free of errors, enhancing the reader's understanding of the analysis and design principles. Demonstrates exceptional critical thinking skills by providing insightful and original perspectives on Healthy Space analysis and design principles.

Exceptional comprehension of Healthy Space, evident in detailed analysis and clear articulation of key elements and their impact, along with a thorough identification and rationale for multiple design principles. Demonstrates a comprehensive and creative integration of Healthy Space principles into the design proposal, enhancing reader understanding. Exceptional critical thinking is evident through insightful and original perspectives on Healthy Space analysis and design principles.

Regenerate

Figure 19: The final assessment criteria in the Excellent column of the rubric generated in ChatGPT

Upon critical reflection, I realised that the previous rubrics that I created provided superficial guidelines and focused heavily on deliverables. These rubrics considered the possibility for students to include elements beyond the basic requirements of the brief but lacked guidelines for students to achieve mastery. Using ChatGPT to generate the assessment rubric supported me in unpacking the assessment criteria and in providing detailed guidelines on what is required to master a learning outcome and reliable means to evaluate their own performance.

The Authentic Assessment Planning Worksheet supported my ability to define clear learning outcomes and align these with authentic tasks. Using ChatGPT enabled me to design rubrics in less time than doing so completely by myself, but I still needed to draw on my detailed subject matter knowledge to refine the generated rubrics into authentic assessment tools.

Design Qualification 3

My initial attempts to generate assessment rubrics with ChatGPT using the process described in Figure 3 resulted in superficial rubrics. Only when I refined my prompts did ChatGPT generate detailed rubric descriptors. I explored using prompts with learning outcomes and associated assessment tasks listed in brackets behind each outcome (Figure 20). This approach generated detailed descriptors that provide guidelines for the standards required as evidence of mastery (Figure 21).

Design a five-column marking rubric to assess the following outcomes: 1. Textual and Visual research (on your chosen brand, visual identities, competitors, and visual research to be placed on a mood board). 2. Strategic direction (problem statement, target audience, findings, insights and message) 3. Concept development and Ideation (various ideation techniques, sketches, and prototypes to develop an original concept and creative solution) 4. Final executions (fundamentals and principles of design and visual communication). 5. Creative rationale (the why and how of the design decisions made) 6. Personal reflection.

Figure 20: Learning outcomes used as prompts with assessment tasks included in brackets

Criteria	Novice	Developing	Proficient	Accomplished	Exemplary
Textual and Visual Research	Research was conducted on the chosen brand, visual identities, competitors, and mood board.	Basic research was conducted but lacked depth and variety of sources. The mood board is incomplete or lacks cohesion.	Sufficient research was conducted on the chosen brand, visual identities, competitors, and mood board, incorporating a range of relevant sources. The mood board demonstrates a clear understanding of the visual landscape.	Comprehensive research was conducted on the chosen brand, visual identities, competitors, and mood board, demonstrating a thorough understanding of the subject matter. The mood board is well-curated, visually appealing, and effectively communicates the desired aesthetics.	Extensive and insightful research was conducted on the chosen brand, visual identities, competitors, and mood board, showcasing a deep understanding of the subject matter. The mood board is exceptionally well-crafted, visually compelling, and thoughtfully conveys the desired visual direction.
Strategic Direction	Problem statement, target audience, findings, insights, and message are missing or unclear.	Problem statement, target audience, findings, insights, and message lack clarity and coherence.	Problem statement, target audience, findings, insights, and message demonstrate a clear understanding of the project's objectives and provide a basic strategic direction.	Problem statement, target audience, findings, insights, and message are well-defined, cohesive, and support the overall creative strategy.	Problem statement, target audience, findings, insights, and message are highly articulated, showcasing a deep understanding of the project's objectives and providing a compelling and innovative strategic direction.
Concept Development and Ideation	Limited or inappropriate ideation techniques employed. Sketches and prototypes lack creativity and originality.	Some appropriate ideation techniques are employed but lack depth and variety. Sketches and prototypes show limited exploration of concepts.	Appropriate ideation techniques are employed, generating creative and original concepts. Sketches and prototypes demonstrate a solid exploration of ideas and present viable design solutions.	Well-chosen ideation techniques employed, producing innovative and imaginative concepts. Sketches and prototypes showcase a thorough exploration of ideas and demonstrate refined design solutions.	Highly effective and diverse ideation techniques were employed, resulting in exceptionally creative, innovative, and original concepts. Sketches and prototypes exhibit a comprehensive exploration of ideas, displaying refined design solutions and a strong sense of visual communication.
Final Executions	Final executions lack understanding and application of design fundamentals and principles of visual communication. The visual identity solutions do not effectively solve the design problem.	Final executions show some understanding and application of design fundamentals and principles, but overall execution is weak or inconsistent. The visual identity solutions partially address the design problem.	Final executions demonstrate a good understanding and application of design fundamentals and principles of visual communication. The visual identity solutions effectively solve the design problem.	Final executions showcase excellent understanding and application of design fundamentals and principles, displaying strong visual communication and attention to detail. The visual identity solutions seamlessly and cohesively solve the design problem.	Final executions exhibit exceptional understanding and application of design fundamentals and principles, displaying a mastery of visual communication. The visual identity solutions are executed flawlessly and work exceptionally well to solve the design problem with a high

					level of creativity and innovation.
Creative Rationale	The rationale is missing or lacks coherence and clarity. The explanation of design decisions is absent or insufficient.	Rationale provides a basic explanation of design decisions but lacks depth and insight. The connection between design decisions and the overall creative strategy is unclear.	Rationale effectively explains design decisions, demonstrating a clear understanding of how they support the overall design solution. The connection between design decisions and the overall creative strategy is evident.	Rationale demonstrates a strong connection between design decisions and the overall creative strategy, offering insightful analysis and explanation. The rationale effectively communicates the motivations behind design choices.	Rationale is exceptional, showcasing a highly articulate and persuasive discourse on design decisions, supported by critical analysis, and demonstrating a deep understanding of the project's objectives and outcomes. The rationale eloquently communicates the significance and impact of design choices.
Personal Reflection	Limited or no personal reflection provided.	Some personal reflection is provided but lacks depth and critical insights.	Personal reflection demonstrates a thoughtful and considered assessment of the project, highlighting personal growth, challenges overcome, and lessons learned.	Personal reflection provides a comprehensive evaluation of the project, offering insightful analysis of personal strengths, weaknesses, and areas for improvement. Reflection showcases self-awareness and critical thinking skills.	Personal reflection is highly perceptive and demonstrates a profound understanding of personal growth, challenges, and growth opportunities. Thoughtful insights are provided, and future recommendations are outlined based on the reflection conducted.

Figure 21: More detailed rubric descriptors generated in ChatGPT

Upon critical reflection, I realised that my existing rubrics included high cognitive thinking skills, as well as exposing students to a variety of assessment tasks. They assessed problem solving, knowledge application, and decision making, while also evaluating the effectiveness with which students performed various tasks. However, the descriptors were somewhat superficial and did not provide clear guidelines to help students grasp what is expected and the standard for mastering each task. Similarly to the findings of Game Design, I observed that the descriptors generated in ChatGPT showed the scaffolding between different assessment tasks. This is an essential component of graphic design assessments since each task has an impact on the next, and a thread\link from one task to the next is required for students to perform well in their practical projects.

Lessons learnt

Although we all achieved slightly different results in the various iterations of our rubrics, the lessons we learnt from reflection were very similar. After our first attempts, we realised the importance of critically considering and making sure that learning outcomes are accurately defined and aligned with assessment tasks. Poorly defined learning outcomes resulted in poor assessment criteria that focus primarily on task completion and superficial descriptors that provide little guidance to students. Moreover, the lack of standards in such descriptors left them open to interpretation, providing limited

guidance for marking. The Authentic Assessment Checklist provided a good framework for evaluating assessment criteria and rubric descriptors; it sometimes forced us to critically reconsider our defined learning outcomes.

ChatGPT proved useful for identifying assessment criteria and generating clear and detailed descriptors that provide students and markers with guidelines in terms of standards for mastery. Prompting ChatGPT to generate assessment criteria also helped us identify potential learning goals we had previously overlooked, sometimes obliging us to revisit them, and reconsider learning content in our curricula. However, we still had to apply personal knowledge and experience as curriculum designers and educators to critically select criteria appropriate to the level of assessment aligned with the defined learning goals of those generated by ChatGPT. We also had to critically review the generated descriptors and adjust them to enhance clarity, and ensure that the level of language used is appropriate for our students. We found that quality, authentic rubrics cannot be generated by ChatGPT alone, but are co-created in an iterative, collaborative effort between the curriculum designer and generative AI. Although the use of ChatGPT significantly reduced the time required, the knowledge, critical analysis, and experience of the curriculum designer were crucial for prompting and for evaluating and refining generated content.

Conclusions

This study explored the potential role of generative AI in the design of authentic assessment rubrics. Findings from the study suggest that authentic assessment rubrics cannot be generated, but are co-created, and require a critical dialogue between the curriculum designer and the AI.

Using generative AI for rubric creation has distinct benefits, such as speeding up the process of writing clear, detailed descriptors once learning goals have been defined and appropriate assessment criteria identified. Speedier writing at the tail end of the process potentially affords curriculum designers more time during the initial stages to critically consider and refine learning goals, assessment tasks, and criteria.

More rigorous research with larger samples is required to identify stronger themes and general rules of thumb. However, the findings of this study suggest that there is room for reflection on traditional assessment practices and for the innovative use of AI to support curriculum and assessment design.

References

- Academy of Science South Africa 2023, 'The implications of ChatGPT for assessment in higher education', *Academy of Science South Africa Webinar*, viewed 19 June 2023, <<https://research.assaf.org.za/handle/20.500.11911/275>>.
- Anderson, LW & Krathwohl, DR 2001, *Taxonomy for learning, teaching and assessing: a revision of Bloom's taxonomy of educational objectives*, New York, Addison Wesley Longman.
- Archbald, DA & Newmann, FM 1988, *Beyond standardized testing*, National Association of Secondary School Principals.
- Buckingham-Schum, S 2023, *ChatGPT: What do we know now? What must we learn next?*, Australian Association for Research in Education, viewed 26 June 2023, <<https://www.aare.edu.au/blog/?p=17029>>.
- Carvalho, L, Martinez-Maldonado, R, Tsai, Y, Markauskaite, L & De Laat, M 2022, 'How can we design for learning in an AI world?', *Computers and Education: Artificial Intelligence*, vol. 3, p. 100053, <<https://doi.org/10.1016/j.caeai.2022.100053>>.

- Chen, L, Chen, P & Lin, Z 2020, 'Artificial Intelligence in education: a review', *IEEE Access*, vol. 8, pp. 75264–75278, <<https://doi.org/10.1109/ACCESS.2020.2988510>>.
- Cole, R, Purao, S, Rossi, M & Sein, MK 2005, 'Being proactive: where action research meets design research', in DE Avison, DF Galetta & JI DeGross (eds), *Proceedings of ICIS International Conference on Information Systems*, Las Vegas, Association for Information Systems.
- Cresswell, JW & Guetterman, TC 2019, *Educational research: planning, conducting, and evaluating quantitative and qualitative research*, 6th edn, Pearson.
- Cross, N 2007, *Designerly ways of knowing*, Basel, Birkhauser.
- Dorst, K 2017, *Notes on design: how creative Practice works*, Amsterdam, BIS.
- Frey, BB, Schmitt, VL & Allen, JP 2012, 'Defining authentic classroom assessment', *Practical Assessment, Research, and Evaluation*, vol. 17, no. 2, pp. 1–18, <<https://doi.org/10.7275/sxbs-0829>>.
- Giloi, SL & Du Toit, PH 2013, 'Current approaches to assessment of graphic design in higher education', *International Journal of Art and Design Education*, vol. 32, no. 2, pp. 256–268, <<https://doi.org/10.1111/j.1476-8070.2013.01758.x>>.
- Grainger, P & Weir, K 2020, 'Introduction: the significance of rubrics', in P Grainger & K Weir (eds), *Facilitating student learning and engagement in higher education through assessment rubrics*, Newcastle upon Tyne, Cambridge Scholars Publishing.
- Hardman, P 2023, 'Teaching & learning in a post-AI world how to leverage the *real* power of AI to drive learner outcomes, today', *Learning Science Newsletter*, viewed 19 June 2023, <<https://drphilippahardman.substack.com/p/doms-edu-mega-prompt-for-c21st-lesson>>.
- Mueller, J 2005, 'The authentic assessment toolbox: enhancing student learning through online faculty development', *Journal of Online Learning and Teaching*, viewed 25 June 2023, <https://jolt.merlot.org/documents/vol1_no1_mueller_001.pdf>.
- Palinkas, LA, Horwitz, SM, Green, CA, Wisdom, JP, Duan, N & Hoagwood, K 2015, 'Purposeful sampling for qualitative data collection and analysis in mixed method implementation research', *Administration and Policy in Mental Health*, vol. 42, no. 5, p. 533, <<https://doi.org/10.1007/S10488-013-0528-Y>>.
- Siemens, G 2022, 'The biggest challenges facing education now and ways to meet them', viewed 10 February 2023, <<https://www.aare.edu.au/blog/?p=15047>>.
- Siemens, G, McGraw, K & Kelly, N 2022, 'The AI future for lesson plans is already here', Australian Association for Research in Education, viewed 10 February 2023, <<https://www.aare.edu.au/blog/?tag=george-siemens>>.
- Swiecki, Z, Khosravi, H, Chen, G, Martinez-Maldonado, R, Lodge, JM, Milligan, S, Selwyn, N & Gašević, D 2022, 'Assessment in the age of artificial intelligence', *Computers and Education: Artificial Intelligence*, vol. 3, <<https://doi.org/10.1016/j.caeai.2022.100075>>.
- Villarroel, V, Bloxham, S, Bruna, D, Bruna, C & Herrera-Seda, C 2018, 'Authentic assessment: creating a blueprint for course design', *Assessment & Evaluation in Higher Education*, vol. 43, no. 5, pp. 840–854, <<https://doi.org/10.1080/02602938.2017.1412396>>.
- Zhihan, L 2023, 'Generative artificial intelligence in the metaverse era', *Cognitive Robotics*, vol. 3, pp 208–217, <<https://doi.org/10.1016/J.COGR.2023.06.001>>

APPENDIX A: AUTHENTIC ASSESSMENT PLANNING WORKSHEET

Based on the authentic assessment toolbox by Jon Mueller and the components of authentic assessment identified by Villarroel et al (2018)

STEP 1: LIST THE ASSESSMENT INFORMATION

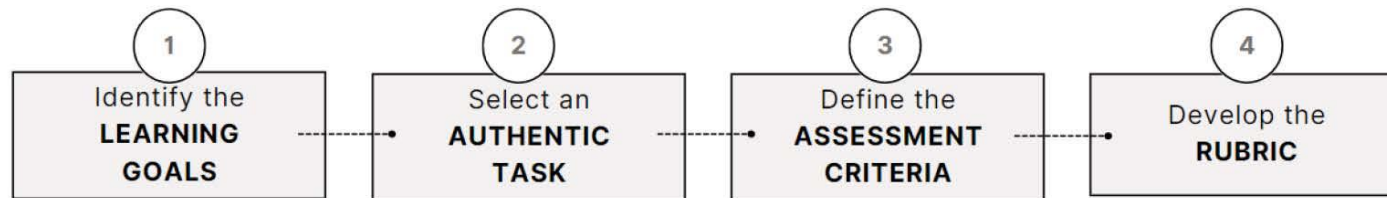
Module Name	
Audience	
NQF level	
Project title	
Assessment Description	
Module Purpose	
Module Outcome	

STEP 2: DEFINE THE AUTHENTIC ASSESSMENT DIMENSIONS

(you will need these to ensure your outcomes are pitched correctly and tasks are authentic for the next step)

Realism: Describe the real context and/or how the task addresses a real-world problem similar to those faced in real life	
Cognitive Challenge List the higher order cognitive skills to be assessed You can use Bloom's Revised Taxonomy for this	
Evaluative Judgement	Table below to include criteria and standards about what a good performance means and include feedback on how students may improve (feedforward feedback)

STEP 3: AUTHENTIC ASSESSMENT PLANNING TABLE:



Note: You will fill in the learning outcomes and tasks (1 and 2) , ChatGPT to generate the criteria and rubric (3 and 4)

	1. Identify the learning outcomes	2. Select an authentic task to address the learning goal	3. Define the criteria	4. Rubric
TIPS	<i>Identify what you want students to know/able to do in this assessment This needs to align with the module outcomes and purposes (assessment goals) Ensure that the verb used aligns to the task selected Use higher order skills (create, evaluate, analyse, apply)</i>	<i>Define the task and how it relates to the real world Align the task to the learning outcomes Ensure the task is complex in nature Include a variety of tasks</i>	<i>What does good performance on this task look like? What will students need to do in order to master the task?</i>	<i>The combination of the criteria and the levels of performance for each criterion will be your rubric for that task</i>
1				
2				
3				
4				
5				
6				
7				