Rubric design for authentic assessment: An exploration in co-designing a teacher resource as an artefact

Rachael Hains-Wesson
RMIT University, Australia
Lilia Mantai and James Tognolini
The University of Sydney, Australia

This paper outlines a co-creation process for designing a teacher-support resource in the form of a rubric as both an artefact and a product, using a Robust Design Method (RDM) informed by Design-Based Methods (DBM). We explore how these two approaches were combined to produce the rubric with clearly articulated descriptors and components to strengthen teachers' reflective and evaluative practices in authentic assessment critique. While DBM commonly focuses on developing and testing innovative educational interventions, RDM, originating in engineering, examines how a product performs under varied conditions to ensure reliability and adaptability. In this study, the rubric is conceptualised as a product and often referred to as an artefact. By integrating RDM and DBM across four iterative design phases, the authors reconceptualise the design process to co-develop a rubric artefact as both a process and an output. In this study, a rubric becomes not only a structured tool for teachers but also a mechanism for academic professional development, foregrounding teacher agency. The exploratory study focuses on highlighting RDM's value in generating a robust, teacher-centred resource, in the form of a rubric, to critique authentic assessment while positioning academic professional development as an iterative and inquiry-driven process.

Introduction

Authentic assessment design in higher education plays a pivotal role in shaping students' learning experiences and in developing the capabilities they need for real-world contexts (Ajjawi et al., 2024; Koretsky et al., 2022; Vlachopoulos, 2025). For instance, authentic assessment has been known to enhance skills, such as critical thinking, problem-solving, and collaboration (Ajjawi et al., 2024; Vlachopoulos & Makri, 2024), as well as incorporating aspects of students' lived experience, promoting meaningful learning experiences (Ravi & Besharat, 2026). Further, the shift towards competency-based learning and outcomes requires new types of assessment methods that evaluate practical skills and real-world application rather than traditional knowledge recall. Yet many institutions struggle to transition from traditional practices to authentic assessments (Corbin et al., 2026; Nieminen et al., 2023; Schultz et al., 2022; Vlachopoulos & Makri, 2024), especially in time of artificial intelligence (Kirsanov et al., 2026). For instance, higher education faces significant authentic assessment design challenges driven by societal demands, shifting educational paradigms, and technological advances in education, such as the rise and expansion of artificial intelligence (Corbin et al., 2026). Additionally, over the past two decades, teacher scholarship has increasingly emphasised the importance of authentic assessment, which is assessments that mirror professional tasks and challenges to better prepare graduates for the complexities of contemporary

work (Ashford-Rowe et al., 2014; Nallaya et al., 2025; Rahiman & Kodikal, 2023; Shishavan, 2024). Within this movement*, rubrics* have emerged as widely recognised tools to support consistent grading, enhancing transparency, and providing formative feedback to students (Buckley-Walker et al., 2017; Taylor et al., 2024.

In higher education, a rubric is generally understood as a structured set of criteria against which student work is judged, describing varying levels of performance for each criterion. A rubrics benefit is in how it communicates standards of achievement in a transparent and consistent manner (Ishikawa et al., 2021). Rubrics are therefore more than marking guides because they are pedagogical artefacts that support learning by making expectations explicit, guiding feedback, promoting self- and peer-assessment, and enhancing the reliability and fairness of assessment decisions (Askari, 2026). Further, there is significant evidence in the literature that well-designed rubrics enhance the overall learning experience for students, such as providing relevant feedback (Taylor et al., 2024; Tierney & Simon, 2004). Rubrics also provide efficiencies for teaching staff, especially when rubrics are used early in learning (Anglin et al., 2008; Tierney & Simon, 2004).

In addition, other research has noted that analytical rubrics are especially useful in assessments for learning because they often contain qualitative descriptions of performance criteria that work well within the process of formative evaluation (Tierney & Simon, 2004). Data sourced through the rubrics can also be used to support the continuous improvement and inform enhancement of courses (Taylor et al., 2024). Rubrics support course continuous improvement by providing structured and consistent data about student performance that can be analysed over time. Because rubric criteria are typically aligned to learning outcomes, the grading of a rubric can reveal not only the overall achievement of students' learning, but also which specific areas are being met successfully, and which require greater instructional focus (Anglin et al., 2008; Tierney & Simon, 2004). This allows teachers to move beyond looking at final grades alone and into identifying patterns in students' learning strengths and weaknesses. When trends emerge, such as consistently low performance in a particular skill or knowledge area, teaching strategies and assessment design can be adjusted in a targeted way to address those gaps. Over multiple course offerings, rubric data can also make it possible to track whether these changes are having the intended effect, creating an evidence-based feedback loop where adjustments are informed by clear performance patterns. This approach ensures that improvements are purposeful, measurable, and directly connected to the learning outcomes, supporting a cycle of continuous and inform curriculum enhancement for learning (de Barba, Oliveira & English, 2025).

This study, therefore, builds on authentic assessment design and its related broader body of work in teacher scholarship (Koretsky et al., 2022; Oo & Alonzo, 2023; Ravi & Besharat, 2026) while extending insights from a prior investigation into how to co-create an authentic assessment tool to critique teachers' authentic assessment design practice and validate standards (Hains-Wesson & le Roux, 2024). This is a timely study as authentic assessment remains a cornerstone of higher education, despite AI disruption (Rahiman & Kodikal, 2023; Shishavan, 2024), significantly influencing student learning outcomes, the

quality of learning experiences, and perceptions of learning. With this in mind, the research question posed to help guide the following exploration was:

> How does the use of Robust Design Methods (RDM) and Design-Based Research (DBR) support the building of an authentic assessment rubric tool as a teacher-self and peer review artefact in authentic assessment design?

Why RDM and DBR?

When designing and developing a particular teacher-support artefact, such as how-to guides or a white paper on good practice in teaching and learning, it is important to understand value and applicability in different contexts, as well as limitations when utilising RDM and DBM. In the field of engineering, design theory is commonly used to validate a product's (i.e., artefact's) usefulness and therefore, the value to end users (Bürdek, 2005). When creating a new product, designers may focus on mitigating "noise" factors, which are similar to variables that reduce reliability or performance, as well as improving the sustainability of a product's utility (Gremyr et al., 2014). Noise factors can also be in relation to enhancing an artefact's usability in response to user challenges. To achieve this, engineering manufacturers often employ structured processes such as RDM or DBR to iteratively refine artefacts for maximum functionality and user satisfaction.

In educational contexts, an *artefact* refers to a tangible or digital object purposefully created to support learning, teaching, or assessment. In authentic assessment design, such artefacts might include rubrics, annotated exemplars, digital simulations, or interactive feedback tools that make expectations explicit, guide student performance, and improve the validity and reliability of grading. While the concept of the *artefact design process* has been widely applied in engineering and product design, its adaptation to higher education, particularly within the field of rubric assessment, is modest (Koretsky et al., 2022).

Consequently, by implementing RDM and DBM, the end users, in this instance teachers and students, work collaboratively in the design process to address challenges and ensure the product meets diverse user needs. This process can help highlight structured, research-informed approaches to artefact design that are often absent from traditional professional development and its processes. As Smith and Walker (2022) suggested,

> ... until a more precise and consistent understanding of the concept of scholarship is established and the developmental support infrastructure is more comprehensive, barriers for those pursuing teaching-focused career routes will remain (2022, p. 193).

Artefacts

Artefact creation in teaching and learning contexts is frequently learned through observation, trial, and error, which can result in inefficiencies and inconsistent quality standards (Staudt Willet, 2024). By introducing RDM and DBM as purposeful, complementary methods, this study provides early-career teachers in higher education with a framework to develop artefacts that are robust, contextually relevant, and

responsive to user needs. It offers practical guidance on navigating the complexities of *artefact* design while reducing the reliance on informal, *ad hoc* approaches, as well as formalising and acknowledging such a professional role and its activity.

Finally, the overall rationale for exploring RDM and DBR in rubric development does not rest solely on the existence of a gap in the literature. Rather, it emerges from a practical and conceptual problem in higher education: rubrics are widely used to guide assessment design and evaluate teaching practice, yet they are rarely subjected to systematic testing for reliability, adaptability, and performance across diverse disciplinary and contextual conditions, especially with the rise of AI being utilised as a form of rubric critique (Askari, 2026). In other words, rubrics function as quality-assurance artefacts, but their own design robustness is often assumed rather than examined through multiple lenses. RDM and DBR offer a structured way to human stress-test such artefacts, collaboratively, under variable teaching contexts, student cohorts, and institutional constraints, thereby strengthening their credibility and transferability.

Literature review

Defining the term "artefact" in teacher-support discourse

In the context of teacher-support design work, which often focuses on supporting teacher practice, the term *artefact* refers to a physical or virtual manifestation of the actions and procedures that take place in an organisation, in this case, rubric design. This manifestation, however, does not only passively represent organisational processes but can in turn shape and influence those processes (März et al., 2017) that leads to product formation. These artefacts are designed to enhance teaching and learning, facilitate effective classroom management, and/or provide guidance on various aspects of educational practice by teachers for teachers (Resende & Vasconcelos, 2012; Vermeir et al., 2017).

How-to-guides are a common example of teacher-support artefacts and products. They often contain step-by-step instructions and cover a wide range of topics, including lesson planning, classroom management, assessment techniques, using educational technology, and addressing the needs of diverse learners. Other examples of artefacts in teacher-support design may include lesson plans where objectives, content and teaching strategies are outlined. Similarly, assessment methods, and rubrics, as well as scoring guides, provide clear criteria for evaluating student work, helping teachers assess and grade assignments consistently. These are all examples of educational artefacts. The purpose of these artefacts is to provide teachers with valuable resources, such as creating a philosophy of practice for team-teaching (Hains-Wesson, 2022), authentic assessment guidance (Hains-Wesson & le Roux, 2024), and tools that can help teachers be more effective practitioners and enhance the learning experience for their students.

These artefacts can be, and often are, produced for teachers by teachers to help benchmark practice in teaching and learning. However, without artefact creators including systematic ways to evaluate their products, and as they are being developed, they run the

risk of developer bias, decreasing an artefact's robustness and wider application. In this study, the process of creating an artefact that meets user needs is viewed in parity to an artefact's final output. This approach highlights the importance of comprehensive testing during the artefact creation process, i.e. the artefact's validity.

Defining "authentic assessment"

The term and practice of *authentic assessment* is often promoted to improve quality of learning including deep learning, student engagement, relevance, and applicability of learning. Candy's (1991) influential work suggests that assessments that test memory or encourage transmissive learning are more likely to induce students to superficially learn. Brown's seminal work (1997) points out that given the primacy of assessments, transforming an assessment to an authentic task has a proven capacity to transform student learning and initiate a move away from surface approaches to deep learning (Chi & Ohlsson, 2005). More recent scholarship focuses on authentic assessment as a distributed leadership requirement to ensure teachers design assessments with authenticity in mind (Nallaya et al, 2025).

Authentic assessment design is, therefore, central to learning quality and can influence student learning and engagement. Consequently, and especially in the age of AI (Corbin et al., 2026) growing criticisms of transmissive assessment methods have led to an increased need for teachers to enhance the construction and design of their assessment tasks, including the implementation of more authentic assessments while growing teacher engagement into assessment design (Ajjawi et al., 2024; Rahiman & Kodikal 2023). Additionally, the value of authentic assessments reaches beyond mirroring work practice to also include psychological authenticity, ontological fidelity, and practice theory perspectives (Ajjawia et al., 2024). This further highlights the value that authenticity in assessment holds for the sector beyond gains in graduate employability (Geertshuis et al., 2022). Thus, authenticity cannot be completely unmoored from the reality of workplaces, disciplinary context and intended learning outcomes, paired with variability in how teachers perceive authenticity and authentic assessments.

Given the critical role of authentic assessment in higher education, institutions are well advised to focus on identifying and constructing appropriate artefacts to support authentic assessment design to meet diverse teacher and student needs, as well as technological disruption. In addition, such artefacts should be developed through collaboration and co-design with all stakeholders, such as teachers, students, and industry/community, to be effective and sustainable.

To complicate matters, there is neither a common definition nor guideline on what might constitute an authentic assessment task. Further, there are limited tools around authenticity that considers the design steps, nor clear direction for teachers to take when desiring to re-design authentic assessments (see for e.g., Ashford-Rowe et al., 2014), including AI considerations (Shishavan, 2024). Whereas for Archbald and Newmann (1988), formative work linked the term *authentic* to student achievement, requiring students to demonstrate learning outcomes in applied or work contexts beyond the school learning

environment. Following this, the term *authentic* has been used to denote the demonstration of higher-order thinking and the ability to solve problems (Newmann et al., 2001). Whereas others such as Wiggins (1989) have historically linked authenticity with student assessments, promoting authentic assessments as a process that requires students to perform at standards expected in the professional field (Gulikers et al., 2004). Thus, such long-term discourses around what an authentic assessment is and is not, as well as its theoretical and philosophical underpinnings (Ajjawia et al., 2023) continues to create diverse ways of defining the term, *authentic assessments*. For instance, the literature on authentic assessment provides numerous definitions that are often contextual and widely contested (Ajjawi et al., 2024; Savery & Duffy, 1995; Gore et al., 2004; Vlachopoulos, 2025). Thereby, creating confusion at the level of teacher involvement, how to define, measure and apply it to different contexts.

In view of the competing, diverse and context-specific definitions in the literature around authentic assessment, we choose to provide a definition to help contextualise the study. In this article, authentic assessment is viewed as a combination of tasks that mirror, replicate and "help students rehearse for the complex ambiguities of the 'game' of adult and professional life" (Wiggins, 1990, p.1). With the caveat that an authentic assessment task will also be clearly aligned to learning outcomes; coherent in structure with transparent instructions; fairly assessed via a robust rubric; and provide students with relevant and timely feedback that is focused on improvement in future learning (Schultz et al., 2022).

Design-based research (DBR) and robust design method (RDM) for rubric creation

Teacher-support resources, or "artefacts", require consistent tailoring to meet diverse needs. DBR is widely used for designing educational interventions and documenting their design process (Puntambekar, 2018; Oo & Alonzo, 2023). DBR involves iterative cycles of design and testing to create interventions that improve student learning by enhancing teacher practice (Wang & Hannafin, 2005). However, DBR does not always consider disruptions or variables, termed "noise," that can impede artefact creation (Plomp, 2009, p.21). In contrast, RDM focuses on understanding and mitigating "noise," such as user challenges, desires, and relevance in different contexts. RDM ensures user challenges remain central during design to create robust, context-sensitive artefacts (Clausing, 1994; Pahl & Beitz, 1996; Hasenkamp et al., 2007). This approach is crucial for developing teacher-support artefacts like rubrics to ensure their usefulness. As briefly mentioned, authentic assessments are central to higher education, significantly affecting student learning outcomes (Scouller, 1998; Sokhanvar et al., 2021), learning quality (Fergusson et al., 2022; Vlachopoulos, 2025), and students' perceptions of learning (James & Casidy, 2018; Jopp, 2020). Although assessment type influences learning depth, modest research has examined the use of artefact creation (Koretsky et al., 2022; Oo & Alonzo, 2023) to include both RDM and DBM to develop teacher-support rubrics for assessment design improvement (Pahl & Beitz, 1996; Ulrich & Eppinger, 1995; Ullman, 1997).

Responding to the institutional need for guidance on authentic assessment

The authors were tasked with designing a teacher-support artefact for evaluating authentic assessment. The request came from the University's executive leadership to support a teaching and learning strategy aimed at curriculum transformation and to improve student learning outcomes. As all teachers were expected to embed authentic assessment principles, the teaching community required a practical guide with agreed key elements, components, and signposts to support effective self- and peer review of assessments and provide clear directions for improvement. Ethics approval was obtained for the project (project number: 2021-2023/902). As part of the design process, eight university teachers in education-focused roles across various faculties and disciplines were invited to participate in co-creating a teacher-self-assessment rubric tool on authentic assessment critique. Participants were identified through the University's database of Higher Education Academy (HEA) and Higher Education Research and Development Society of Australasia (HERDSA) Fellows and contacted via email to join as key artefact co-creators. The first-named author facilitated several critical-friend meetings, roundtable discussions, and debates during which participants explored the concept of authentic assessment. This phase was crucial to the discovery process, as one of the project's main aims was to develop an artefact applicable across the university.

To avoid personal bias and decision making, a group of fifty interested members were also recruited over a twelve-month period by the authors, which included HEA and HERDSA Fellows. The members consisted of academics, professional staff, and higher degree by research students. Participants opted-in once per month to join facilitated meetings that were conducted via Zoom and in-person. Key feedback was provided to the authors, including from the initial eight HEA and HERDSA Fellows through an online chat platform. On average, a core group of five to seven university staff members participated in workshops via the online chat area with other participants joining *ad hoc* online or in-person, and at separate times.

Throughout the discussions and meetings, participants openly debated, listened, and documented pertinent ideas, thoughts, and opinions around authentic assessment. This was done in conjunction with examining the relevant literature, highlighting pertinent definitions, authentic assessment characteristics, measurement frameworks, models, and typologies (Ajjawi et al., 2023). Thus, participants were kept informed and updated via email and the online chat forum by the first named author on a consistent and regular basis. The final decision to create an artefact in the form of a measurement rubric, as depicted in the initial publication by Hains-Wesson and le Roux (2024), was made collectively by the authors and the HEA and HERDSA Fellows. The main reasons being, (1) a rubric includes key descriptors and components to test for learning soundness; (2) rubrics are widely used throughout teacher practice; and (3) rubrics are a familiar evaluation artefact in both teacher practice and student learning.

The next section outlines each design phase and key outcomes that informed the artefact's creation. The phases were deliberately structured to integrate RDM and DBR methods,

combining the strengths of both approaches. This ensured that future user needs remained central, while minimising noise factors, to enhance the artefact's robustness.

Four phases of robust design methodology

As previously mentioned, a Robust Design Methodology (RDM) was chosen, which was combined with Design Based Research DBR. The methodology of choice included four main phases (refer to Figure 1). The first phase was based on *planning and clarifying the needs of the user*. For instance, this phase of the RDM is where artefact creators (i.e. participants) agreed to consider various artefact ideas based on the end user requirements, which in this case were teachers and students. The chosen idea was explored during each participant meeting and via a set of agenda items that included (1) design specifications; (2) functions; (3) properties and authentic assessment components and (4) relevancy to the user. These key points have also been suggested as an appropriate method in artefact creation in the product design literature (Ulrich & Eppinger, 1995; Pöhlsen et al., 2026).

The second phase was termed *concept design and improvement*. In this phase, a focus on describing the form, function, and features that the artefact would take and why, including how users would ensure that they could continually and iteratively improve upon practice, was the key interest. For example, discussions focused on the platforms participants might use, so that key stakeholders (i.e., Fellows) could provide continuous feedback and recommendations as the artefact began to take shape.

According to design-based practice (Thornton, 2004; Oo & Alonzo, 2023) the design phase should include rough design layouts, simple prototypes with key components and technical choices. It is also important during this phase, that diverse ideas and concepts are considered, solutions generated, and all concepts are evaluated by the artefact creators. This is achieved by implementing regular opportunities for participants to discuss, debate, self- and peer reflect, resulting in less useful concepts being eradicated through consensus making. It was during this phase that the participants decided that the artefact would take the form of a rubric.

The third phase was *embodiment design*. This is where the artefact creators define the arrangements of each component, ensuring that each descriptor articulated specific levels of competency with examples. This phase involves many iterations before artefact creators feel confident that an agreed-to design emerges (Biddy et al., 2026; Hasenkamp et al., 2007). This phase is also cyclic in nature, encouraging analysis, synthesis, evaluation, and improvement, which occurs simultaneously until collective agreement through consensus is achieved.

The fourth phase is called *detail design, test, improvement, and evaluation*. This phase includes all the detail, concepts, parts, and components that will be included in the artefact. It is a design phase that consists of decisions being made on the materials needed, production effort and dissemination possibilities (Thornton, 2004; Hasenkamp et al., 2007). Costs are a key consideration for this phase too, as well as the short- and long-term management requirements for the long-term outcomes for the artefact's use, its purpose, and the

requirement to undertake continuous evaluation. This phase is where a prototype is created and initially tested. This was achieved by inviting future users, such as members and additional staff through word-of-mouth. Participants were invited to bring current assessment examples to undertake self- and peer-evaluation using the rubric. This enabled the artefact creators to locate areas of product strengths and weaknesses, including noise factors and a lack of robustness, for example.

Finally, a process of continuous improvement was embedded into the project. The participant community tested and re-tested the rubric artefact using assessment tasks drawn from a broad range of disciplines, including health, education, social sciences, and business. Within these disciplines, participants contributed varied assessment types, such as essays, reflective journals, case studies, practical demonstrations, and, in some instances, teacher-identified non-authentic formats like quizzes. Peer review of these tasks using the rubric enabled the core development team to examine the artefact's performance across different disciplinary contexts and task formats. This iterative feedback process informed targeted refinements, including clarifying key components, recalibrating descriptor levels, and strengthening the rubric's applicability and robustness for diverse teaching and assessment scenarios.

Applying RDM and DBR to rubric design

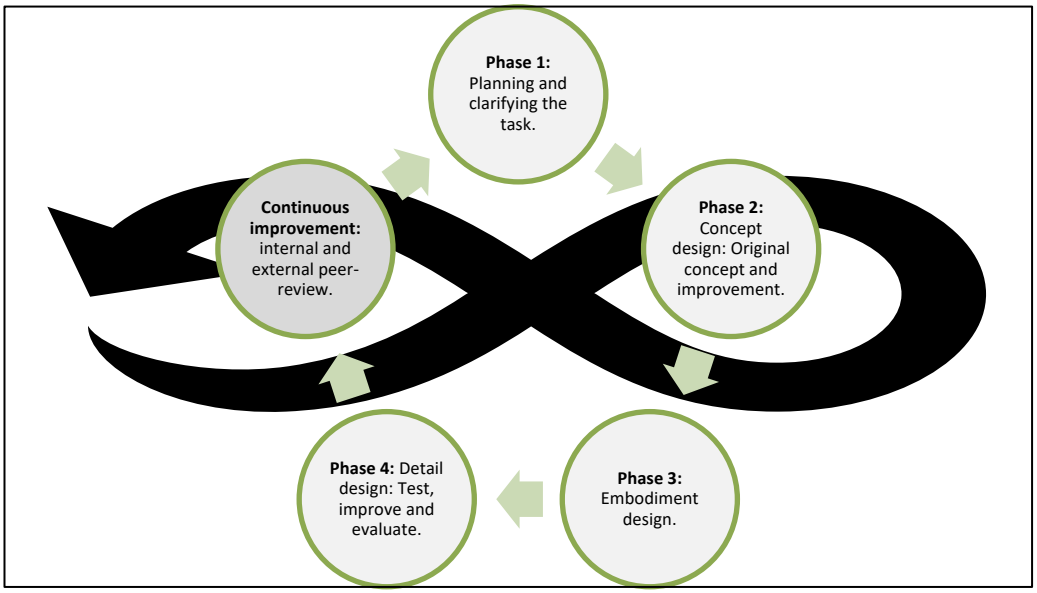

Figure 1: Artefact creation process: A design cycle (image created by the authors, 2026)

Combining RDM and DBR methods assists artefact creators to evaluate each phase collectively while routinely revisiting the original intention, which was to ensure that the artefact met user needs and was robust (see Figure 1). An important key requirement in RDM (Thornton, 2004; Hasenkamp et al., 2007) is that a product is shared more broadly

and externally validated, for the developers to then revisit the first phase of a product creation intention, asking the question: *does the product meet user needs, and if not, why not?*

In the next section, authors' experiences of incorporating each phase of the design cycle as depicted in Figure 1 are presented. This further tested the design-based mixed method approach in the domain of teacher-support artefact creation and where process is just as important as any output.

Collaborative approach to designing an authentic assessment rubric tool (AART)

Phase 1: Planning and clarifying the task

The authors, HEA and HERDSA Fellows, as well as participants collectively gathered and interpreted relevant information about the university's context and situation in the domain of authentic assessment design. This was achieved by purposefully exploring what had already been experienced, as well as potential gaps in the literature. This process is similar to what Ullman (1997) originally described when utilising RDM; as the time required to understand the market and establish what the customer wants from a particular product. Relative to Ullman's (1997) process, the artefact creators were viewed as product developers and the end users as being teachers and/or students. Thus, the main aim of the artefact was to help make a real impact on the working lives of teachers and students as co-partners in their learning while being context sensitive. To assist in meeting this goal, the following question became a key focus during meetings and debates, which was: *what kind of evidence has proven useful in building the case for the validity of an artefact?*

Phase 2: Concept design and concept improvement

This part of the design phase required participants to be creative and to explore their experiences while questioning previous ways of identifying threats to validity and robustness. For instance, scholars have suggested that concepts are best conceived by individuals while the selection of concepts is optimally done in groups (Biddy et al., 2026; Hasenkamp et al., 2007; Pugh, 1991). To initiate the process highlighted here and via individual and group think, brainstorming sessions proved to be an ideal and highly productive strategy. Additionally, each participant had time to undertake individual and collective reflective processes about the artefact formation, coming together post-discussion to further reflect and to unpack issues, challenges, and non-collective understandings to reach consensus. This practice, at times was uncomfortable and took time because it meant that the probing and debating of ideas was needed until collective agreement occurred.

Phase 3: Embodiment design

Bandurek et al. (1990) argued that RDM includes the setting up of efficient experimental plans to help identify the environmental factors that result in the acute stress levels of an artefact not working effectively. The experiments can then be compared with the

anticipated environment that the artefact under development is likely to be exposed to, resulting in key areas for improvement and re-design while ensuring that the original intention is still clear and apparent. When undertaking this part of the artefact creation, via the preferred design method, the instigation of purposely planned workshops with key stakeholders was achieved, such as students and teachers to use the artefact. In the workshops, participants collaboratively experimented with the artefact and explored how it would be used in a real-life situation, noting its strengths and weaknesses to increase robustness. This was also a time where the identification of problems was highlighted, which were then addressed in Phase four.

Phase 4: Detail design: test, improvement, and evaluation

This phase is concerned about the design and its components that together will make up the artefact in its entirety. For instance, it was during this phase that further testing, validation, and tweaking of the artefact was completed. Certain questions to those who were testing and improving the artefact were also of a focus. The key questions asked were: (1) does the artefact meet the needs of all stakeholders and if not, why; (2) does the design allow for variations of use; (3) who will be responsible for future testing, evaluation, and long-term maintenance of the rubric; and (4) how will the rubric be shared and used to meet the initial goals of the intent? The answers to these questions helped to provide good examples of *noise* factors, highlighting the way different disciplines might view and value authenticity in an assessment.

Consequently, during phase four of the design process more noise factors became apparent. The noise factors indicated the need for refinement in terms of the vocabulary used within the rubric artefact. For instance, some terms were refined or removed altogether, such as *learning styles,* due to such a term being disputed in the literature. This was replaced with *learning preferences*, which is viewed as more appropriate because it better highlights the importance of student choice in learning (Dinsmore et al., 2022).

Summary

Applying the RDM and DBM methodologies, to conduct a cross-institutional and co-design process to create a teacher self-assessed rubric tool to critique authentic assessment design, proved critical in boosting staff engagement. Not only did the authors arrive at an artefact that would help measure growth in authentic assessment design but by the end of the process, a culture was created where positive engagement and ownership of the artefact collectively occurred. Arguably developing authentic assessment that is meaningful, relevant, and engaging for learners forms a critical part of education excellence (Ajjawi et al., 2024). In times where students and teachers alike have access to various artificial intelligence (AI) tools and technologies to help create and respond to assessment requirements, it is more important than ever. Therefore, teachers need to come together more regularly and as a teaching collective to better understand what *authentic* means (Nieminen et al., 2023), especially during the rise of AI (Salinas-Navarro et al., 2024). This includes how and why teachers design assessments to ensure academic integrity, inclusivity, and equity (Nieminen et al., 2023). The process and the methods

presented here offers an alternative way to rethink artefact creation to influence teacher practice.

For early career teachers who undertake artefact creation, looking to engage in design as a process, and as part of teacher support offerings, the following considerations and recommendations are worth noting. First, it is vital to spend enough time in the early design work and phases to allow for enough open and authentic disagreement for artefact creators and users to buy into the parameters and assumptions underlying the product. If the initial phase is rushed or there is no true consensus, the discussion in later phases will be about underlying fundamental issues rather than details. This might include budgeting enough time to sit with uncomfortable discussions that might seem unproductive at the time, instead will lay the groundwork for more productive future discussions and artefact creations.

Second, having a problem statement, such as 'what is authentic assessment?' to return to can help serve as a grounding principle in the initial phases. Further, record-keeping and the sharing of those records are instrumental in keeping the project focused and minimising time spent on repeating and revisiting previous discussions. As the project took place over a twelve-month period, there were periods of a month or more which passed without the participants directly engaging in the project. Thus, the third point is an obvious one, yet vital, which is to create a shared online channel where all notes, meeting minutes, recordings, and readings are housed. This means that participants can refresh their memories between workshops and share relevant insights to be considered at future workshops.

Fourth, it is important to avoid rushing through the phases or skipping directly to phase four due to limited timelines or competing obligations. Instead, utilising a staged approach via the design framework presented here can assist with managing expectations and competing deadlines as each participant will better understand (upfront) what will occur at each development stage, for instance. This needs to be resourced sufficiently by higher education leadership.

Lastly, thorough consideration of how the selection of participants will be determined is required because this will impact the progression and direction of the creation of the artefact. In this study, participants came from diverse disciplines and used a variety of distinct types of assessments. However, participants were mostly of the same level of experience and accomplishment in their careers as teachers, including PhD candidates. That is to say, the recruitment of participants with a solid level of knowledge, understanding, and experience in designing assessments via teaching and learning Fellow memberships can increase the quality of the final artefact but not necessarily its utility. Subsequently, it will be important to recognise that more extensive testing of the artefact will be needed, including the input from less experienced teachers.

Conclusion

Creating a universal teacher-support artefact for a contested concept like critiquing authentic assessment requires collaborative design methodologies and a refocus on stakeholder diversity (Ajjawi et al., 2023). Our paper outlines the application of RDM and DBR in four phases to co-create a rubric that guides the design and evaluation of authentic assessments. These methodologies emphasise the value of process, enhancing both the final product and the co-design experience itself. Collaborating with teachers to define authentic assessments through exploration and co-creation has been shown to increase the sustainability, usability, and impact of the rubric while fostering a sense of psychological ownership during the co-design process (Juel et al., 2023).

The project was led by eight university teachers and involved fifty staff (HEA and HERDSA Fellows) over one year. Participants explored teacher-support mechanisms for authentic assessment design, ultimately converging on the rubric. User-centred questions guided development, and potential "noise factors" were addressed. Ensuring clear, consistent language in rubric descriptors proved challenging across diverse teaching contexts, but was essential for cross-disciplinary use. Once the detailed design was established, the artefact's strengths and weaknesses were evaluated and validated against the original objective, which was to create a user-centric, relevant, and impactful teacher-support artefact. The process not only achieved its intended outcomes but also demonstrated the importance of choosing methodologies that support collaborative processes to produce sustainable and meaningful artefacts for teachers.

Limitations
This study has several limitations, such as self-selection bias being likely, as participants were volunteers with existing expertise in assessment or a strong motivation to further develop their practice.

Future research
Further research that tests the artefact while incorporating participants' perspectives through interviews and surveys would be valuable. Building on the method-driven process outlined in this study, future work should explore how to support early career teachers in artefact creation and foster cross-institutional collaborations to design and evaluate tools that enhance teaching practice.

Obtaining the rubric artefact
The rubric artefact may be obtained from https://www.mdpi.com/2227-7102/14/8/894 (Hains-Wesson & Le Roux, 2024).

References

Ajjawi, R., Tai, J., Dollinger, M., Dawson, P., Boud, D. & Bearman, M. (2024). From authentic assessment to authenticity in assessment: Broadening perspectives. *Assessment & Evaluation in Higher Education*, 49(4), 499-510. https://doi.org/10.1080/02602938.2023.2271193

Anglin, L., Anglin, K., Schumann, P. L. & Kaliski, J. A. (2008). Improving the efficiency and effectiveness of grading through the use of computer-assisted grading rubrics. *Decision Sciences Journal of Innovative Education*, 6(1), 51-73. https://doi.org/10.1111/j.1540-4609.2007.00153.x

Archbald, D. A. & Newmann, F. M. (1988). *Beyond standardized testing: Assessing authentic academic achievement in the secondary school.* Reston, Va.: National Association of Secondary School Pffincipals. http://files.eric.ed.gov/fulltext/ED301587.pdf

Ashford-Rowe, K., Herrington, J. & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205-222. https://doi.org/10.1080/02602938.2013.819566

Askari, M. (2026). Reliable but supervised: Evaluating a generative AI-rubric model for consistent and fair assessment in postgraduate education. *Assessment & Evaluation in Higher Education*, 51(2), 213-230. https://doi.org/10.1080/02602938.2025.2537774

Bandurek, G. R., Hughes, H. L. & Crouch, D. (1990). The use of Taguchi methods in performance demonstrations. *Quality and Reliability Engineering International*, 6(2), 121-131. https://doi.org/10.1002/qre.4680060209

Biddy, Q., Jacobs, J., Bhaduri, S., Recker, M., Nixon, J., Bush, J. & Sumner, T. (2026) From co-design to co-adaptation: The evolution of teacher professional learning across a long-term research practice partnership. *Frontiers in Education,* 11, article 1761033. https://doi.org/10.3389/feduc.2026.1761033

Brown, S. & Knight, P. (1994). *Assessing learners in higher education.* Psychology Press. https://doi.org/10.4324/9780203062036

Brown, A. L. (1997). Transforming schools into communities of thinking and learning about serious matters. *American Psychologist*, 52(4), 399-413. https://doi.org/10.1037/0003-066X.52.4.399

Buckley-Walker, K., Tognolini, J., Lockyer, L., Brown, I. & Caputi, P. (2017). Evaluating the validity of the online multiliteracy assessment tool. *Australian Journal of Education*, 61(3), 305-327. https://doi.org/10.1177/0004944117729056

Bürdek, B. E. (2005). *Design: History, theory and practice of product design.* Birkhäuser. https://link.springer.com/book/10.1007/3-7643-7681-3

Candy, P. C. (1991). *Self-direction for lifelong learning: A comprehensive guide to theory and practice.* Jossey-Bass.

Chi, M. T. H. & Ohlsson, S. (2005). Complex declarative learning. In K. J. Holyoak & R. G. Morrison (Eds.), *Cambridge handbook of thinking and reasoning* (Ch.16). Cambridge University Press. https://www.cambridge.org/au/universitypress/subjects/psychology/cognition/cambridge-handbook-thinking-and-reasoning#contents

Clausing, D. (1994). *Total quality development: A step-by-step guide to world-class concurrent engineering.* ASME Press. https://doi.org/10.1115/1.800695

Corbin, T., Bearman, M., Boud, D. & Dawson, P. (2026). The wicked problem of AI and assessment. *Assessment & Evaluation in Higher Education*, 51(4), 736-752. https://doi.org/10.1080/02602938.2025.2553340

De Barba, P. G., Oliveira, E. A. & English, N. (2025). Development and validation of a learning analytics rubric for self-regulated learning. *Education Technology Research and Development,* 73, 3223-3245. https://doi.org/10.1007/s11423-025-10521-x

Dinsmore, D. L., Fryer, L. K. & Parkinson, M. M. (2022). The learning styles hypothesis is false, but there are patterns of student characteristics that are useful. *Theory into Practice*, 61(4), 418-428. https://doi.org/10.1080/00405841.2022.2107333

Fergusson, L., van der Laan, L., Imran, S. & Danaher, P. A. (2022). Authentic assessment and work-based learning: The case of professional studies in a post-COVID Australia. *Authentic Assessment and Professional Studies*, 12(6), 1189-1210. https://doi.org/10.1108/HESWBL-03-2022-0074

Geertshuis, S., Wass, R. & Liu, Q. (2022). Equipping graduates with future-ready capabilities: An application of learning theories to higher education. *Teaching in Higher Education*, 29(5), 1302-1321. https://doi.org/10.1080/13562517.2022.2137399

Gore, J. M., Griffiths, T. & Ladwig, J. G. (2004). Towards better teaching: Productive pedagogy as a framework for teacher education. *Teaching and Teacher Education*, 20(4), 375-387. https://doi.org/10.1016/j.tate.2004.02.010

Gremyr, I., Siva, V., Raharjo, H. & Goh, T. N. (2014). Adapting the robust design methodology to support sustainable product development. *Journal of Cleaner Production*, 79, 231-238. https://doi.org/10.1016/j.jclepro.2014.05.018

Gulikers, J. T. M., Bastiaens, T. J. & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67-86. https://doi.org/10.1007/BF02504676

Hains-Wesson, R. (2022). A philosophy of practice to inform team-teaching: A blended auto-ethnographical account. *Issues in Educational Research*, 32(4), 1403-1420, http://www.iier.org.au/iier32/hains-wesson.pdf

Hains-Wesson, R. & le Roux, S. (2024). Bridging teacher knowledge and practice: Exploring authentic assessment across educational levels. *Education Sciences*, 14(8), article 894. https://doi.org/10.3390/educsci14080894

Hasenkamp, T., Adler, T., Carlsson, A. & Arvidsson, M. (2007). Robust design methodology in a generic product design process. *Total Quality Management and Business Excellence*, 18(4), 351-362. https://doi.org/10.1080/14783360701231294

Ishikawa, Y., Umemoto, T., Tsubota, Y., Fujiwara, S., Suto, A. & Nishiyama, K. (2021). Achieving student engagement in learning: Utilizing a rubric-based assessment system for visualizing learners' self-, peer, and teacher assessments. In P. Zaphiris & A. Ioannou (Eds), *Learning and collaboration technologies: New challenges and learning experiences*. HCII 2021. Lecture Notes in Computer Science, vol 12784. Springer, Cham. https://doi.org/10.1007/978-3-030-77889-7_20

James, L. T. & Casidy, R. (2018). Authentic assessment in business education: Its effects on student satisfaction and promoting behaviour. *Studies in Higher Education*, 43(3), 401-415. https://doi.org/10.1080/03075079.2016.1165659

Jopp, R. (2020). A case study of a technology enhanced learning initiative that supports authentic assessment. *Teaching in Higher Education*, 25(8), 942-958. https://doi.org/10.1080/13562517.2019.1613637

Juel, A., Berring, L. L., Erlangsen, A., Larsen, E. R. & Buus, N. (2023). Sense of psychological ownership in co-design processes: A case study. *Health Expectations*, 27(1), article e13886. https://doi.org/10.1111/hex.13886

Kirsanov, O., Kushwah, L. & Selvaretnam, G. (2026). Beyond detection: How students use—and hide—AI in online assessments and what authentic tasks can do about it. *Journal of Academic Ethics*, 24, article 14. https://doi.org/10.1007/s10805-025-09691-3

Koretsky, M. D., McColley, C. J., Gugel, J. L. & Ekstedt, T. W. (2022). Aligning classroom assessment with engineering practice: A design-based research study of a two-stage exam with authentic assessment. *Journal of Engineering Education*, 111(1), 185-213. https://doi.org/10.1002/jee.20436

März, V., Kelchtermans, G. & Vermeir, K. (2017). Artifacts as authoritative actors in educational reform: Routines, institutional pressures, and legitimacy in student data systems. *Journal of Educational Change*, 18(4), 439-464. https://doi.org/10.1007/s10833-017-9309-9

Nallaya, S., Kenyon, K., Gentili, S., Dinmore, S. & Janssen, A. (2025). Authentic assessment enacted: Academic developers and curriculum transformation through distributed leadership. *Issues in Educational Research*, 35(4), 1644-1663. http://www.iier.org.au/iier35/nallaya.pdf

Newmann, F. M., Bryk, A. S. & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests: Conflict-or-coexistence?* Consortium on Chicago School Research. https://consortium.uchicago.edu/publications/authentic-intellectual-work-and-standardized-tests-conflict-or-coexistence

Nieminen, J. H., Bearman, M. & Ajjawi, R. (2023). Designing the digital in authentic assessment: Is it fit for purpose? *Assessment & Evaluation in Higher Education*, 48(4), 529-543. https://doi.org/10.1080/02602938.2022.2089627

Oo, C. Z. & Alonzo, D. (2023). Developing pre-service teachers' skills in assessment for learning (AfL): A plan using design-based research. *Issues in Educational Research*, 33(2), 693-712. http://www.iier.org.au/iier33/oo.pdf

Pahl, G. & Beitz, W. (1996). *Engineering design: A systematic approach*. Bell & Bain. https://doi.org/10.1007/978-1-4471-3581-4

Plomp, T. (2009). Educational design research: An introduction. In T. Plomp & N. Nieveen (Eds.), *An introduction to educational design research* (2nd ed., pp. 9-35). SLO Netherlands Institute for Curriculum Development. https://ris.utwente.nl/ws/portalfiles/portal/14472302/Introduction_20to_20education_20design_20research.pdf

Pöhlsen, L., Inselmann, P. & Krause, D. (2026). Towards an impact model of modular product strategies. *Proceedings of the Design Society*, 6, 767-776. https://doi.org/10.1017/pds.2026.10435

Pugh, W. (1991). The Omega test: A fast and practical integer programming algorithm for dependence analysis. In *Proceedings of the 1991 ACM/IEEE conference on Supercomputing* (pp. 4-13). https://doi.org/10.1145/125826.125848

Puntambekar, S. (2018). Design-based research (DBR). In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman & P. Reimann (Eds.), *International handbook of the learning sciences* (pp. 383-392). Taylor and Francis. https://doi.org/10.4324/9781315617572

Rahiman, H. U. & Kodikal, R. (2023). Revolutionizing education: Artificial intelligence empowered learning in higher education. *Cogent Education*, 11(1), article 2293431. https://doi.org/10.1080/2331186X.2023.2293431

Ravi, M. & Besharat, M. (2026). A holistic consideration of authentic assessments: Student perception of assessment design, delivery, flexibility and creativity. *European Journal of Engineering Education*, 51(1), 25-42. https://doi.org/10.1080/03043797.2025.2480116

Resende, A. E. & Vasconcelos, F. H. (2012). Design of educational artifacts as support to learning process. *WORK: A Journal of Prevention, Assessment & Rehabilitation*, 41(S1), 61-68. https://doi.org/10.3233/WOR-2012-0136-61

Salinas-Navarro, D. E., Vilalta-Perdomo, E., Michel-Villarreal, R. & Montesinos, L. (2024). Designing experiential learning activities with generative artificial intelligence tools for authentic assessment. *Interactive Technology and Smart Education*, 21(4), 708-734. https://doi.org/10.1108/ITSE-12-2023-0236

Savery, J. R., & Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational Technology*, 35(5), 31-38. https://www.jstor.org/stable/44428296

Schultz, M., Young, K., Gunning, T. K. & Harvey, M. L. (2022). Defining and measuring authentic assessment: A case study in the context of tertiary science. *Assessment & Evaluation in Higher Education*, 47(1), 77-94. https://doi.org/10.1080/02602938.2021.1887811

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, 35(4), 453-472. https://doi.org/10.1023/A:1003196224280

Shishavan, H. B. (2024). AI in higher education: Guidelines on assessment design from Australian universities. *ASCILITE 2024 Conference Proceedings*, 118-126. https://doi.org/10.14742/apubs.2024.1205

Smith, S. & Walker, D. (2022). Scholarship and teaching-focused roles: An exploratory study of academics' experiences and perceptions of support. *Innovations in Education and Teaching International*, 61(1), 193-204. https://doi.org/10.1080/14703297.2022.2132981

Sokhanvar, Z., Salehi, K. & Sokhanvar, F. (2021). Advantages of authentic assessment for improving the learning experience and employability skills of higher education students: A systematic literature review. *Studies in Educational Evaluation*, 70, article 101030. https://doi.org/10.1016/j.stueduc.2021.101030

Staudt Willet, K. B. (2024). Early career teachers' expansion of professional learning networks with social media. *Professional Development in Education*, 50(2), 386-402. https://doi.org/10.1080/19415257.2023.2178481

Taylor, B., Kisby, F. & Reedy, A. (2024). Rubrics in higher education: An exploration of undergraduate students' understanding and perspectives. *Assessment & Evaluation in Higher Education*, 49(6), 799-809. https://doi.org/10.1080/02602938.2023.2299330

Tierney, R. & Simon, M. (2004). What's still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research, and Evaluation*, 9(1), 2. https://doi.org/10.7275/jtvt-wg68

Thornton, A. C. (2004). *Variation risk management: Focusing quality improvements in product development and production*. Wiley. https://www.wiley.com/en-cn/shop/general-introductory-industrial-engineering/variation-risk-management-focusing-quality-improvements-in-product-development-and-production-p-9780471446798

Ullman, D. G. (1997). *The mechanical design process*. McGraw-Hill. https://archive.org/details/mechanicaldesign0000ullm/mode/2up

Ulrich, K. T. & Eppinger, S. D. (1995). *Product design and development*. McGraw-Hill. [8th ed.] https://www.mheducation.com/highered/product/product-design-and-development-ulrich.html

Vermeir, K., Kelchtermans, G. & März, V. (2017). Implementing artifacts: An interactive frame analysis of innovative educational practices. *Teaching and Teacher Education*, 63, 116-125. https://doi.org/10.1016/j.tate.2016.12.006

Vlachopoulos, D. (2025). Authentic assessment: Bridging higher education and real-world skills through diverse methodologies. In *INTED2025 Proceedings* (pp. 2225-2234). IATED. https://doi.org/10.21125/inted.2025.0628

Vlachopoulos, D. & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*, 83, article 101425. https://doi.org/10.1016/j.stueduc.2024.101425

Wang, F. & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development,* 53(4), 5-23. https://doi.org/10.1007/BF02504682

Wiggins, G. (1989). The futility of trying to teach everything of importance. *Educational Leadership,* 47(3), 44-59. https://files.ascd.org/staticfiles/ascd/pdf/journals/ed_lead/el_198911_wiggins.pdf

Authors' statements

Disclosure statement

The authors report there are no competing interests to declare.

Funding

The authors received no financial support for the research, authorship and/or publication of this article.

Professor Rachael Hains-Wesson (corresponding author) is the Associate Dean, Learning and Teaching in the School of Global, Urban and Social Studies, RMIT University. Rachael's research interests reside in authentic assessment, work-integrated learning, and teacher scholarship.
ORCID: https://orcid.org/0000-0002-1549-4504
Email: rachael.hains-wesson@rmit.edu.au

Dr Lilia Mantai is a Senior Lecturer at the Business School, The University of Sydney, NSW, Australia. Her work and research focus on graduate skill development, curriculum and assessment, and AI impact on jobs and skills. She is a Senior Fellow of AdvanceHE, Executive Member for HERDSA, Vice Chair of the Australasian Council for Undergraduate Research [http://www.acur.org.au/], and Associate Editor for the journal *Higher Education Research and Development* [https://www.tandfonline.com/journals/cher20]
ORCID: https://orcid.org/0000-0002-4603-3831
Email: lilia.mantai@sydney.edu.au

Professor James (Jim) Tognolini is Director of the Centre for Educational Measurement and Assessment (CEMA) at the University of Sydney and Managing Director of JT Educational Consulting Pty Ltd [https://www.janison.com/about/]. In his current position he is responsible for establishing and expanding the Centre of Assessment and Educational Measurement at the University of Sydney. His research areas include educational assessment, educational measurement and psychometrics.
ORCID: https://orcid.org/0000-0002-3481-0234
Email: jim.tognolini@sydney.edu.au

Please cite as: Hains-Wesson, R., Mantai, L. & Tognolini, J. (2026). Rubric design for authentic assessment: An exploration in co-designing a teacher resource as an artefact. *Issues in Educational Research*, 36(3), 994-1011. http://www.iier.org.au/iier36/hains-wesson.pdf