Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Designing edtech-integrated authentic assessments for deep learning: insights from digital illustration tasks

Rachael Zoe Murray

rachael.murray@qut.edu.au

Queensland University of Technology https://orcid.org/0000-0002-4935-7794

Nadja Beck

Queensland University of Technology

Annalese B.T. Semmler

Queensland University of Technology

Francesca D. Frentiu

Queensland University of Technology

Lisa K. Chopin

Queensland University of Technology

Yasmin J. Antwertingere

Queensland University of Technology

Research Article

Keywords: authentic assessment, drawing to learn, active learning, digital illustration

Posted Date: March 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9109391/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

This study examines the impact of replacing a traditional short-answer examination with a BioRender-based scientific digital illustration assessment designed to integrate three established pedagogical frameworks: drawing to learn, active learning, and authentic assessment. We examined effects on student performance and student perceptions of learning and professional preparation in a longitudinal mixed-methods study (2015–2024; 436 students) in a second-year Cell Biology subject. Comparing five years of traditional examination data with five years of BioRender assessment data showed mean scores rose substantially from 49.7% to 83.1% with the same short-answer questions (Mann–Whitney U = 27,115, $p < 0.001$; Cliff's delta = 0.74). Institutional survey data and a purpose-designed questionnaire administered to current and graduating students provided insight into the mechanisms underlying this improvement: 95.5% of students felt the assessment enhanced their learning, 91.1% reported developing career-relevant skills, and 83% of graduating students confirmed it supported their transition to advanced study. Qualitative themes identified deeper conceptual understanding, reduced assessment anxiety, stronger engagement, and perceived professional preparedness as key drivers. These outcomes are interpreted through the convergence of all three frameworks, each compensating for limitations the others cannot address alone. The findings offer educators a theoretically grounded framework for edtech-integrated assessment design that extends beyond the specific tool examined here.

Introduction

Traditional written examinations have long occupied a central place in STEM assessment design. Their value in motivating broad curriculum coverage and providing standardised, comparable evidence of student knowledge is well documented (van der Vleuten, 2000). However, high-stakes exams elevate anxiety, impair performance, and can misrepresent capability (Chapell et al., 2005; DordiNejad et al., 2011; von der Embse et al., 2018). More fundamentally, traditional exams are poorly suited to activate three distinct but interrelated mechanisms that the contemporary educational literature has repeatedly linked to deep and enduring learning.

The first is the generative visual processing mechanism described by drawing-to-learn (D2L) theory. When students externalise their understanding by constructing visual representations, they are compelled to identify conceptual gaps, organise knowledge spatially and sequentially, and engage in the kind of higher-order cognitive work that passive recall tasks do not demand (Bobek & Tversky, 2016; Van Meter & Garner, 2005). The second mechanism is the agentic, iterative engagement at the heart of active learning (AL) theory: the conditions under which students exercise choice, receive timely feedback, and apply knowledge in increasingly sophisticated ways (Ainsworth & Scheiter, 2021; Black & Wiliam, 2009). The third is the motivation and knowledge transfer that authentic assessment (AA) theory predicts when students recognise that what they are being asked to do mirrors the work of their future profession (Villarroel et al., 2018). The one-time, text-bound format of traditional written examinations is poorly suited to activate any of these three mechanisms. The challenge is to design assessments that engage all three.

Digital visual communication in biomedical assessment

In STEM authentic assessment has predominantly been designed and evaluated through a research-laboratory lens, emphasising experimental design, data analysis, and evidence-based problem-solving (Schultz et al., 2021; Tan & Kidman, 2021). This overlooks visual communication, the ability to create clear, accurate, and professionally compelling illustrations of cellular pathways and biological processes, a competency increasingly demanded across all biomedical career pathways and identified as essential for preparing graduates for contemporary workplaces (Millar et al., 2023; Thelma et al., 2024; Zhang & Jenkinson, 2024). Yet assessment frameworks have been slow to respond in Biomedical Sciences.

User-friendly digital illustration platforms have opened new possibilities for authentic assessment. A challenge highlighted in the D2L literature is that computer-based drawing interventions often fail to match paper-based equivalents, largely because the cognitive effort required to learn complex software diverts attentional resources away from generative thinking (Cromley et al., 2020; Fiorella & Zhang, 2018). BioRender, launched in 2017 resolves this issue. It is widely used across biomedical research and industry, enabling scientists without specialist artistic training to create high-quality scientific illustrations using a library of over 50,000 pre-drawn biological icons and an intuitive drag, drop and scale interface (Perkel, 2020). Its low technical barrier means students can direct their cognitive resources toward organising, sequencing, and communicating biological content rather than toward tool mastery. For students preparing to enter biomedical research, clinical, or industry roles where BioRender is already standard practice, embedding it in assessment also gives students the opportunity to master this workplace tool. This study examines whether replacing a traditional written examination with a BioRender-based assessment can activate the cognitive conditions D2L theory predicts, the motivational conditions AL theory requires, and the professional alignment AA theory demands.

1.4 Research questions

This study is guided by three research questions, each of which is theoretically anchored to one or more of the three pillars:

RQ1

To what extent does implementation of a digital tool-based scientific illustration assessment impact Biomedical Science student performance and conceptual understanding of cellular pathways in undergraduate cell biology education, compared to traditional written examinations? [D2L; AL]

RQ2: How do students perceive the use of this digital graphics tool as a means of supporting their learning in Biomedical Science? [AL: engagement, agency, feedback]

RQ3: To what extent does the integration of an authentic digital tool into assessment influence students' confidence and preparedness for advanced studies and professional careers in the biomedical sciences? [AA: transfer, long-term learning, professional identity]

Theoretical background

Drawing-to-learn

D2L is an established pedagogical strategy with strong empirical support across multiple disciplines including science (Ainsworth et al., 2011; Ainsworth & Scheiter, 2021; Bobek & Tversky, 2016; Leopold & Leutner, 2012; Van Meter & Garner, 2005). Its theoretical foundation lies in the cognitive demands of visual externalisation: when learners construct a drawing, they are required to retrieve, organise, and structure their knowledge in ways that passive encoding strategies do not demand (Bobek & Tversky, 2016; Van Meter & Firetto, 2013). This process forces learners to confront the limits of their understanding, identifying gaps and inconsistencies in their mental models (Bobek & Tversky, 2016; Van Meter & Garner, 2005). In biological and biomedical contexts this externalisation process is particularly powerful as the learner must resolve not only what happens but in what order, at which location, and through what molecular mechanism.

Beyond gap-detection, D2L engages the simultaneous activation of visual, spatial, and semantic memory systems, which significantly enhances recall and longer-term retention compared to verbal or textual study strategies (Fernandes et al., 2018). The value of constructing biological pathway diagrams affords is established in the biochemistry and biology education, where visual representations are recognised as essential to conceptual understanding (Prain & Tytler, 2012; Quillin & Thomas, 2015). Student-generated drawings, rather than passive viewing of instructor diagrams, have also been shown to improve conceptual understanding and recall in biology (Heideman et al., 2017). A key distinction is generative versus reproductive drawing. Generative drawing, in which students construct their own representations from conceptual understanding, produces substantially stronger learning outcomes than reproductive drawing (Ainsworth & Scheiter, 2021; Fiorella & Zhang, 2018). The BioRender assessment tasks used in this study are explicitly generative: students independently determine how to represent a cellular pathway, what elements to include, how to sequence and spatially arrange them, and what labels are necessary.

The specific challenge that D2L literature identifies for digital drawing contexts is the extraneous cognitive load imposed by complex software interfaces (Cromley et al., 2020). BioRender addresses this directly through its drag-and-drop interface and intuitive tools, allowing students to focus on the conceptual task rather than tool mastery. Evidence supports this, Ha et al. (2024) (Ha et al., 2024) found that percentage completion of BioRender assignments predicted academic outcomes more strongly than visual-spatial ability or learning style preference.

Agency, engagement, and higher-order cognition in AL

Active learning, broadly defined, encompasses instructional approaches in which students engage in meaningful cognitive activity during the learning process (Ainsworth & Scheiter, 2021; Doolittle et al., 2023; Freeman et al., 2014). Its theoretical grounding reflects constructivist views that learners build understanding through generative activity, alongside extensive evidence that engagement, feedback, and student agency markedly improve learning outcomes across disciplines (Driessen et al., 2020; Freeman

et al., 2014). AL environments that incorporate student choice, iterative task structures, and opportunities for higher-order analysis, synthesis, and application consistently produce deeper engagement and more transferable understanding than traditional didactic approaches (Arneson & Offerdahl, 2018; Patall et al., 2008; Ryan & Deci, 2020). AL treats feedback as an ongoing process rather than a one-off event, enabling learners to identify gaps between current and desired performance and refine their work across iterative cycles (Carless, 2019; Winstone et al., 2017). When feedback is accessible, personalised, and embedded within repeated opportunities to apply it, student engagement and learning gains are significantly enhanced (Henderson et al., 2019; Lunt & Curran, 2010).

Authentic assessment, disciplinary transfer, and professional identity

AA is defined by the alignment between assessment tasks and the conditions of real-world professional practice (Villarroel et al., 2018; Zhan et al., 2025). Its theoretical foundations emphasise that tasks should be professionally relevant, require the integration and application of knowledge, and mirror the tools, standards, and decision-making processes used in the workplace (Schultz et al., 2021; Tan & Kidman, 2021; Villarroel et al., 2018; Zhan et al., 2025).

In the biomedical sciences, visual communication is a core professional competency: researchers, clinicians, and industry scientists must produce accurate, compelling illustrations of biological processes for publications, grant proposals, health communication, and collaboration (Fairman & Phelps, 2023; Millar et al., 2023; Perkel, 2020). Yet this skillset is rarely assessed explicitly in undergraduate curricula, creating a misalignment between what students are evaluated on and what the profession demands.

AA seeks to close this gap by designing assessments in which students undertake the kinds of tasks biomedical professionals actually perform, using the same tools and adhering to the same standards. When students engage in tasks that genuinely mirror scientific practice, AA theory predicts stronger motivation, deeper learning, better transfer, and improved long-term retention (Boud & Falchikov, 2007). AA also supports the development of professional identity: by working with authentic tools and expectations, students begin to see themselves as emerging biomedical practitioners rather than learners completing tasks solely for assessment (Vlachopoulos & Makri, 2024).

Methodology

Mixed-methods design was selected to capture both the cognitive outcomes predicted by D2L theory, measurable through objective performance data, and the experiential and motivational dimensions predicted by AL and AA theory, best captured through student perception data. Longitudinal performance records, institutional survey data, and a purpose-designed instrument together allow the convergent framework to be evaluated across multiple dimensions.

Context and background

This study, conducted in the School of Biomedical Sciences at a large Australian university under Human Research Ethics Committee approval (#5248), centred on a second-year Cell Biology subject that is part of the Bachelor of Biomedical Science curriculum. Between 2015 and 2019, the subject included a traditional end-of-semester short-answer examination worth 30% of the final grade (increased to 40% in 2019). Analysis of student performance data revealed a mean examination score of 49.6% (median 49; 345 students) from 2015–2018, rising modestly to 53.9% (median 50.5; 89 students) in 2019 when weighting increased. Informal feedback indicated strategic behaviour: once passing, students shifted effort to other subjects, aiming for the minimum needed rather than deep engagement, in line with surface learning and cramming under high-stakes final exams (French et al., 2024; Kusurkar et al., 2023). In 2020, the disruptions from COVID-19 prompted comprehensive re-evaluation of assessment, resulting in the BioRender-based redesign described here. The assessment draws on D2L, AL, and AA. D2L provides the cognitive mechanism, AL structures agency, iteration, and feedback, and AA anchors the work in professional practice; this integration underpins the design rationale mapped in Table 1.

Table 1
Mapping of BioRender assessment design elements to the three theoretical pillars

Assessment design element	D2L mechanism	AL mechanism	AA mechanism
Create a scientific illustration (not write)	Externalises mental model; forces gap identification and repair	Generative, not reproductive; activates higher-order cognition	Mirrors professional scientific communication task format
Student selects from question pools		Promotes agency and ownership of learning	
Three submissions	Iterative refinement of mental model across semester	Creates sustained feedback loop; metacognitive reflection	
Personalised audio feedback via digital platform	Supports identification of conceptual errors for revision	Closes AL cycle	
BioRender as the illustration platform	Low extraneous load enables generative cognitive focus	Interactive, intuitive scaffolding supports active engagement	Industry-standard professional tool; used in research and publication

Assessment structure

The revised assessment was built from the short-answer questions used in 2015–2019. Integrative, pathway-level questions were re-set as illustration tasks to compare understanding of the same content across formats. The formats differ: the illustration task also assessed visual communication, and the

exam included some less complex items. Accordingly, the comparison addresses content understanding on equivalent questions, and results should be read with this distinction in mind.

To give students meaningful choice consistent with AL principles, the assessment used three thematic blocks across the semester. To provide agency over what they choose to focus on, students selected one of three questions in each block and created a BioRender scientific illustration for that topic. Blocks aligned with the subject's learning progression and had set submission dates, creating a distributed structure that promoted sustained engagement rather than end-of-semester cramming. See Table 2 for an example of task type and complexity.

Table 2
Representative example of an illustration assessment task, highlighting the depth and complexity of required work

Topic	Example question
Nuclear import	The tumour suppressor p53 is a transcription factor that regulates numerous target genes involved in cell cycle control, apoptosis, differentiation, and DNA repair. To perform these critical functions, p53 must be transported from the cytoplasm to the nucleus through nuclear pore complexes. This transport is regulated by nuclear localisation signals (NLS), with p53 containing several such signals. A key step in this process is the removal of ubiquitin molecules that mask an essential NLS on p53. Task: Create a scientific illustration of the nuclear import pathway of p53, including: • The initial deubiquitination of p53 • The complete nuclear import process • The release of p53 in the nucleus • The recycling of transport components for subsequent import cycles Your illustration should demonstrate your understanding of the general nuclear import machinery as covered in lectures, specifically applied to p53 transport.

Workshop scaffolding

At semester start, a BioRender workshop prepared students to use the platform before assessment. It first contextualised scientific visual communication within biomedical careers, then shifted to visual communication principles rather than software skills. Using exemplars and counterexamples, students identified what makes an effective scientific illustration: clear visual flow and layout, stepwise segmentation, accessible colour use, readable labels, and a text-free visual format. Students then set up free for education BioRender accounts and learned core functions (icon library, drag and drop, basic editing), followed by a short practice task to recreate a simple image. The scaffolding aimed to reduce extraneous cognitive load identified in D2L research so students could focus on conceptual construction from their first submission (Cromley et al., 2020).

Assessment criteria and feedback process

Submissions were graded with a purpose-designed rubric comprising three weighted criteria. 1) Accuracy of topic content (6 marks): coverage of the full biological process and molecules as taught; accurate representation and correct labelling; and a strictly visual-only format with no explanatory text. Full marks required complete, accurate coverage; lower grades reflected missing, mislabelled, or inaccurate content. 2) Effective use of visual material (2 marks): clarity and engagement of visuals; clear separation of sequential steps; and a layout that guides the viewer logically. 3) Appearance (2 marks): overall layout, spacing, label size, colour use, and absence of spelling errors, meeting professional presentation standards. The rubric aligns with the convergent framework: the accuracy criterion targets the conceptual understanding D2L predicts, while visual material and appearance reflect the professional communication standards specified by AA.

Using Canvas, we provided rubric grades plus personalised audio commentary for every submission. Audio feedback addressed accuracy, professional visual standards, and next-step improvements in sequence to close the AL cycle between tasks, consistent with evidence that timely, actionable feedback improves student performance (Black & Wiliam, 2009; Hattie & Timperley, 2007).

Data collection and analysis

This study used an explanatory sequential mixed-methods design in which quantitative data guided the interpretation of qualitative data. Longitudinal performance data from 2015 to 2024 were used to determine whether the change in assessment format affected student outcomes. Open-ended survey responses were then analysed to explain the mechanisms behind any differences by capturing student experiences and motivations. The qualitative insights were interpreted in relation to the quantitative results to provide a fuller understanding of the impact of the BioRender-based assessment.

Three data sources informed the evaluation. Performance data from Blackboard, and later Canvas, enabled comparison of grades from short-answer exams and illustration-based assessments. Student feedback came from institutional Student Experience of Teaching surveys in 2021 and 2022 and from a targeted 2022 Qualtrics survey completed by both the 2022 cohort and now near-graduating 2021 students. Open-ended comments were thematically analysed by two researchers who coded themes such as satisfaction, perceived learning, workload, engagement, and feedback, with a third reviewer resolving any differences. Theme frequency provided an overview of student perspectives. Performance differences between exam-based cohorts (2015 to 2019) and illustration-based cohorts (2020 to 2024) were tested using a Mann-Whitney U test, and effect sizes were calculated using Cliff’s delta.

Table 3
The SET and BioRender-specific Qualtrics survey questions relevant to this study

Survey	Question/Statement	Response type
Student evaluation of teaching survey	U4 - I received useful feedback in this unit U5 - This unit provided opportunities to develop skills and knowledge I need in my career U6 - What aspects of this unit were done well? U7 - What aspects of this unit could be done better?	U4-5 Likert scale: Strongly agree, agree, neutral, disagree, strongly disagree U6-7 Open ended comment box*
BioRender digital graphics tool-specific Qualtrics survey	Q1. Using the online BioRender graphics tool enhanced my learning Q2. The online BioRender graphics tool is preparing/prepared me for my final year of learning Q3. Which would you prefer? (exam or illustration) Q4. Please specify why you selected the above choice? Q5. Do you have any comments/feedback you would like to share?	Q1-2 Likert scale: Strongly agree, agree, neutral, disagree, strongly disagree Q3 Choice: BioRender assessment or exam Q4-5 Open ended comment box*

* Comments were coded prior to analysis.

Results

Academic performance analysis

Replacing the traditional short-answer examination with a BioRender-based scientific illustration assessment led to a substantial and statistically significant improvement in student performance (RQ1). Illustration scores (mean 83.7%) were markedly higher than exam scores (mean, 50.9%; Mann-Whitney U = 27115, $p < 0.001$), with a very large effect size (Cliff's delta = 0.74). This indicates a meaningful shift in learning outcomes following the change in assessment format (Fig. 1). Assessor observations supported these results. Common conceptual errors, especially incorrect sequencing or localisation of events in biological pathways, were far less frequent in visual submissions than in written responses to the same questions. This aligns with D2L theory, which proposes that diagram construction requires learners to resolve spatial and sequential ambiguities that written recall does not surface.

The illustration format also produced a wide range of accurate and creative representations of cellular processes (Fig. 2), demonstrating deep conceptual understanding and students' ability to synthesise information through diverse visual communication strategies. This diversity enhanced the authenticity of submissions and made marking more engaging by reducing the monotony associated with grading large sets of similar written responses, a factor linked to evaluator fatigue and scoring drift (Myford & Wolfe, 2009). In addition, the inherently individualised nature of each illustration, shaped by students' spatial reasoning and stylistic choices, reduced opportunities for academic misconduct compared with text-based responses.

Further qualitative evidence of cumulative learning emerged from the oral feedback records. Initial submissions commonly required three to ten minutes of audio feedback because of the range of conceptual and visual errors present early in the assessment cycle. By the third submission, most students required only one to three minutes, reflecting substantial improvements in scientific accuracy and illustration quality. This steady reduction in corrective feedback is consistent with the iterative learning gains predicted when students repeatedly engage with feedback-rich assessment structures.

Thematic coding of student evaluation surveys

Student perceptions of the BioRender assessment provided qualitative insight into the AL and AA mechanisms underlying the performance gains. SET surveys from 2021 and 2022 showed strong support for the revised assessment, with 91.1 percent of respondents agreeing that it developed career-relevant skills and knowledge. Open-ended comments (80 students) offered further detail on student experiences and learning outcomes (Table 4). Students (30%) reported that the visual format improved their understanding of complex processes, and 31.3%found it more effective for deep learning than traditional exams. A smaller proportion (2.5%) described increased engagement with the content. Overall sentiment toward the task was highly positive: 37.5% expressed clear enjoyment when using BioRender. Students also described reduced stress (13.8%) and recognised the opportunity to build industry-relevant skills (21.3%).

Feedback processes were well regarded, with 80.9% agreeing they received useful feedback and 10% specifically praising the personalised audio format for helping them identify knowledge gaps and make improvements across submissions. These perceptions align with the cognitive and affective conditions that active learning and authentic assessment theory identify as drivers of deeper engagement and more durable learning. The following section explores whether these perceived benefits persisted beyond the immediate subject context.

Table 4
SET survey response themes and supporting quotations

Thematic category	Theme	% (n)	Representative quotations
Learning and cognitive impact	Conceptual understanding	30% (n = 24)	"...creating the infographics helped me understand the concepts a lot easier." "...allow not only a creative input, but I will now forever know that process intimately."
	Assessment effectiveness	31.3% (n = 25)	"...much more helpful in gaining an in-depth understanding of the content than an exam would be." "I feel it made me understand more than having a big final exam at the end of semester."
	Content engagement	2.5% (n = 3)	"Not only did I find myself thoroughly understanding topics and processes, but it engaged me more in the unit."
Professional development	Career-relevant skills	21.3% (n = 17)	"...assessments really helped to develop useful skills which I am sure I will use later in my career." "...skill is necessary for our careers, and I found creating infographics helpful for skill and knowledge development."
Student experience	Enjoyment of learning	37.5% (n = 30)	"...genuinely the best assessment item I have ever experienced. I learn so much more and also have so much fun." "...fun and a good way to get some creativity flowing. As a visual learner, I loved applying what I learnt into a visual medium."
	Stress reduction	13.8% (n = 11)	"...reasonable assessment delivery without huge stress factors." "...assessment schedule was fantastic – integrating learning opportunities with satisfying challenges in a way that minimised stress to the students."
Pedagogical support	Feedback quality	10% (n = 8)	"Verbal feedback ... very helpful and made understanding much easier. Tonality was also much nicer than written feedback..." "The feedback we received for the infographics was well detailed and I loved that it was as a voice over. It was very easy to understand how we could improve for further assessments."

Longitudinal impact

Student perceptions of the BioRender assessment offered clear insight into the AL and AA mechanisms driving the performance gains. To examine whether these benefits persisted over time, a purpose-designed 2022 survey gathered perceptions from both the then current students and graduating students from the previous year, allowing immediate and longer-term views to be compared (22 students). This comparison aligns with AA theory, which predicts that the value of professionally aligned tasks becomes increasingly evident as students move closer to practice. The survey sample represented diverse demographics, and all respondents (100%) preferred the BioRender assessment over a traditional examination. Support for the approach was strong, with 95.5% stating it enhanced their learning (Fig. 3). Among second-year students, 81% felt it was preparing them for their final year, while 83% of near graduating students reported that it supported their transition to upper-level study.

Qualitative comments indicated sustained learning benefits (Table 5). Most students (86.4%) reported improved conceptual understanding and greater confidence in explaining complex processes, and 31.8% highlighted the visual and interactive aspects as central to this deeper comprehension. Students also valued the creative nature of the task (36.4%), reported lower stress compared with exams (27.3%), and recognised the development of career-relevant skills (22.7%). Together, these findings suggest that the benefits of the BioRender assessment extended beyond the immediate subject, supporting the transfer and emerging professional identity predicted by authentic assessment theory.

Table 5
Identified themes and example supporting quotations from the targeted 2022 survey

Thematic category	Theme	% (n)	Representative quotations
Learning and cognitive impact	Deep understanding	86.4% (n = 19)	"... leaves me feeling like I could explain the process to someone...With exams all I do is cram as much info in as possible for that 2-to-3-hour block and then the majority of that information is lost to the void." "... a great tool to understand what is happening in the cellular process. In an exam I know I would be relying on regurgitation not understanding...I would be able to explain the concepts I have done to non-biomedical friends confidently."
	Content engagement	40.9% (n = 9)	"It facilitated extended learning of associated topics, actually enhanced learning and interest..." "...a more interactive way to understand a concept...helped me evaluate my understanding of a concept and allowed me to research further aspects of a topic."
	Visual learning process	31.8% (n = 7)	"Visualising the topics also helps to understand them in another way, rather than just memorising passages." "The aspect of creativity and thinking through how to put together a sequence of images logically really consolidated the processes in my brain in a way which exams do not."
Student experience	Learning enjoyment	36.4% (n = 8)	"I enjoyed making the infographics because they were a lot more creative than an end of semester exam." "I find the tasks challenging and stimulating as the content we are taught doesn't directly tell you how to create the infographic, so you need to think out each step in detail which is engaging."
	Stress reduction	27.3% (n = 6)	"...much less stressful and is quite fun to make." "...forces students to learn throughout the semester instead of cramming content for the end of semester."
Professional development	Career relevance	22.7% (n = 5)	"... provides a more realistic representation of the skillset and abilities required beyond university." "I feel that assessments where you have to really think about a concept rather than regurgitate information set us up for success in our professional lives."

Discussion

Visual communication through digital tools remains an underused but increasingly important assessment approach in STEM education (Dechamps & Skulmowski, 2025). This study shows that embedding a digital drawing tool within an assessment intentionally aligned to D2L, AL, and AA, the result is substantial gains in conceptual understanding, engagement, and professional preparation.

Performance gains cannot be explained by format change alone. D2L provides the stronger account: constructing illustrations requires learners to externalise mental models, make spatial and sequential decisions, and surface knowledge gaps that written recall can conceal (Bobek & Tversky, 2016; Van Meter & Garner, 2005). Assessor observations of fewer sequencing and localisation errors in visuals than in written responses align with this gap-detection and repair mechanism (Leopold & Leutner, 2012). The progressive, distributed submissions in this study also strengthened D2L processes. Providing time and guidance supports the gap-detection and repair cycle (Ainsworth & Scheiter, 2021; Fiorella & Zhang, 2018), and the three-task structure did so in a more demanding way than revising a single drawing. Students applied feedback from one task to create a new illustration on a different topic, requiring them to transfer visual reasoning and scientific communication skills across content areas. This cross-topic transfer, combined with iterative feedback, fostered the deeper and more flexible understanding predicted in the D2L literature.

High engagement with feedback provided direct evidence of AL mechanisms. In AL, feedback works when learners actively seek and apply it; iterative designs create those conditions (Black & Wiliam, 2009; Winstone et al., 2017). The three-submission structure gave students both motive and opportunity to use feedback, consistent with sustainable feedback models (Carless, 2019). Audio feedback was specifically valued and widely accessed, echoing findings that audio can enhance emotional connection and comprehension relative to written comments (Henderson et al., 2019; Lunt & Curran, 2010).

Emotional responses are also an important part of the learning effect. Students' enjoyment and creativity reports are not incidental; positive emotions predict academic engagement and performance, supporting the cognitive effort needed for generative processing (Carmona-Halty et al., 2021; Xu et al., 2023). Lower stress relative to exams is consistent with research on non-invigilated, distributed assessment and with evidence that artistic creation reduces physiological stress markers, including cortisol (Graf et al., 2023; Kaimal et al., 2016; Sion et al., 2023). Students who are less anxious have greater cognitive resources available for the generative processing that produces understanding.

Students recognised the career relevance of scientific illustration and reported that the assessment supported progression to advanced study, matching AA's predictions about motivation and transfer (Boud & Falchikov, 2007; Vlachopoulos & Makri, 2024). Qualitative comments distinguished their deep, retained understanding from the short-term recall typical of exam cramming. By using an industry-standard platform to produce publication-style visuals, the task made a discipline-specific professional competency visible and assessable. Individualised outputs also limited opportunities for

misconduct and mitigated rater fatigue associated with homogeneous written responses (Myford & Wolfe, 2009).

The central insight is that outcomes arose from the convergence of the three pillars, not from the tool alone. D2L supplied the cognitive engine for understanding, AL supplied the iterative feedback structure for refinement, and AA supplied the professional context that sustains motivation and supports transfer. Remove any one pillar and the effect weakens. This has practical implications for tool selection and assessment design. As educational technology (EdTech) options proliferate, tools should be chosen for their capacity to activate generative work, iterative engagement, and authentic alignment, not for novelty or efficiency alone (Klug & Pietsch, 2024; Cromley et al., 2020). The framework is not BioRender-specific. Concept-mapping platforms, virtual laboratory simulations, data-visualisation environments, and modelling tools used in computational biology or public health can all apply the same principles when tasks require students to generate their own representations, when submission structures are iterative and feedback-rich, and when activities mirror authentic disciplinary practice (Fiorella & Zhang, 2018; Villarroel et al., 2018).

Limitations

Several limitations should be noted. This study was conducted within a single undergraduate cell biology subject, a discipline well suited to visual representation, although the convergent framework is likely applicable across many STEM fields where diagrammatic reasoning and visual modelling are central. Future research should therefore explore its use in disciplines both within and beyond STEM. The influence of the non-invigilated, distributed assessment format also cannot be separated from the effects of the illustration task itself. As AI-based text-to-image tools advance and BioRender introduces new AI-supported features, further work is needed to understand how these developments will shape learning with digital drawing. Whether prompt construction and iterative refinement engage cognitive processes similar to manually placing and organising visual elements remains an open question. Together, these limitations highlight the need for continued research across diverse educational contexts to establish the broader applicability and impact of the convergent framework.

Conclusion

This study demonstrates that a thoughtfully designed, digitally mediated scientific illustration assessment can produce substantial improvements in STEM student learning. Evidence from performance data, student perceptions, and longer-term reflections supports this conclusion. Drawing to learn prompts students to externalise and organise their mental models; the iterative structure and targeted feedback across staged submissions activate active learning; and authentic assessment situates tasks in professionally meaningful contexts that enhance motivation and support transfer. Together, these findings indicate that educators should prioritise digital tools and assessment designs that activate cognitive, motivational, and contextual mechanisms simultaneously.

Future research should examine AI-supported creation directly, investigating how prompt construction and iterative refinement influence gap detection and repair, cognitive load, and academic integrity. The aim is to identify assessment designs and evidence requirements that preserve the learner's generative work while still allowing educators to leverage AI productively.

Declarations

Acknowledgements

The authors gratefully acknowledge all participants involved in the study, especially those who produced the BioRender images used as student examples in this publication.

Availability of data and material

The datasets generated and/or analysed during the current study are not publicly available due to ethics requirements, but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no external funding.

Author contributions

RZM: conceptualisation, project administration, methodology, data collection and analysis, manuscript writing, figures and tables. NB: project administration, review and editing ABTS: conceptualisation, review and editing. FDF: conceptualisation, review and editing. LKC: conceptualisation, review and editing. YJA: conceptualisation, data analysis, review and editing. All authors read and approved the final manuscript.

References

1. Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to Learn in Science. *Science*, *333*(6046), 1096-1097. https://doi.org/doi:10.1126/science.1204153
2. Ainsworth, S. E., & Scheiter, K. (2021). Learning by Drawing Visual Representations: Potential, Purposes, and Practical Implications. *Current Directions in Psychological Science*, *30*(1), 61-67. https://doi.org/10.1177/0963721420979582
3. Arneson, J. B., & Offerdahl, E. G. (2018). Visual Literacy in Bloom: Using Bloom's Taxonomy to Support Visual Learning Skills. *CBE—Life Sciences Education*, *17*(1), ar7. https://doi.org/10.1187/cbe.17-08-0178

4. Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability(formerly: Journal of Personnel Evaluation in Education)*, *21*(1), 5-31. https://doi.org/10.1007/s11092-008-9068-5
5. Bobek, E., & Tversky, B. (2016). Creating visual explanations improves learning. *Cogn Res Princ Implic*, *1*(1), 27. https://doi.org/10.1186/s41235-016-0031-6
6. Boud, D., & Falchikov, N. (2007). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, *31*(4), 399-413. https://doi.org/10.1080/02602930600679050
7. Carless, D. (2019). Feedback loops and the longer-term: towards feedback spirals. *Assessment & Evaluation in Higher Education*, *44*(5), 705-714. https://doi.org/10.1080/02602938.2018.1531108
8. Carmona-Halty, M., Salanova, M., Llorens, S., & Schaufeli, W. B. (2021). Linking positive emotions and academic performance: The mediated role of academic psychological capital and academic engagement. *Current Psychology*, *40*(6), 2938-2947. https://doi.org/10.1007/s12144-019-00227-8
9. Chapell, M. S., Blanding, Z. B., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005). Test Anxiety and Academic Performance in Undergraduate and Graduate Students. *Journal of Educational Psychology*, *97*(2), 268-274. https://doi.org/10.1037/0022-0663.97.2.268
10. Cromley, J. G., Du, Y., & Dane, A. P. (2020). Drawing-to-Learn: Does Meta-Analysis Show Differences Between Technology-Based Drawing and Paper-and-Pencil Drawing? *Journal of Science Education and Technology*, *29*(2), 216-229. https://doi.org/10.1007/s10956-019-09807-6
11. Dechamps, T., & Skulmowski, A. (2025). The Effective Design of Tasks Involving Learning by Drawing: Current Trends and Methodological Progress in Research on Drawing to Learn. *Educational Psychology Review*, *37*(2). https://doi.org/10.1007/s10648-025-10026-2
12. Doolittle, P., Wojdak, K., & Walters, A. (2023). Defining Active Learning: A Restricted Systemic Review. *Teaching and Learning Inquiry*.
13. DordiNejad, F. G., Hakimi, H., Ashouri, M., Dehghani, M., Zeinali, Z., Daghighi, M. S., & Bahrami, N. (2011). On the relationship between test anxiety and academic performance. *Procedia - Social and Behavioral Sciences*, *15*, 3774-3778. https://doi.org/https://doi.org/10.1016/j.sbspro.2011.04.372
14. Driessen, E. P., Knight, J. K., Smith, M. K., & Ballen, C. J. (2020). Demystifying the Meaning of Active Learning in Postsecondary Biology Education. *CBE—Life Sciences Education*, *19*(4), ar52. https://doi.org/10.1187/cbe.20-04-0068
15. Fairman, J. E., & Phelps, T. (2023). Editorial Medical and Scientific Illustration. In M. Roughley (Ed.), *Approaches for Science Illustration and Communication* (pp. 59-93). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-41652-1_3
16. Fernandes, M. A., Wammes, J. D., & Meade, M. E. (2018). The Surprisingly Powerful Influence of Drawing on Memory. *Current Directions in Psychological Science*, *27*(5), 302-308. https://doi.org/10.1177/0963721418755385
17. Fiorella, L., & Zhang, Q. (2018). Drawing Boundary Conditions for Learning by Drawing. *Educational Psychology Review*, *30*(3), 1115-1137. https://doi.org/10.1007/s10648-018-9444-8

18. Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410-8415. https://doi.org/10.1073/pnas.1319030111
19. French, S., Dickerson, A., & Mulder, R. A. (2024). A review of the benefits and drawbacks of high-stakes final examinations in higher education. *Higher Education*, *88*(3), 893-918. https://doi.org/10.1007/s10734-023-01148-z
20. Graf, A., Adama, E., Afrifa-Yamoah, E., & Adusei-Asante, K. (2023). Perceived Nexus Between Non-Invigilated Summative Assessment and Mental Health Difficulties: A Cross Sectional Studies. *J Acad Ethics*, 1-15. https://doi.org/10.1007/s10805-023-09472-w
21. Ha, J., Afana, D., Moghaddam, K. N., & Nicholas, A. (2024). Using BioRender for Active Learning: Exploring Learning-Style Preference and Visual-Spatial Ability in Undergraduate Students. *J Undergrad Neurosci Educ*, *22*(3), A289-A295. https://doi.org/10.59390/QWMZ9919
22. Henderson, M., Phillips, M., Ryan, T., Boud, D., Dawson, P., Molloy, E., & Mahoney, P. (2019). Conditions that enable effective feedback. *Higher Education Research & Development*, *38*(7), 1401-1416. https://doi.org/10.1080/07294360.2019.1657807
23. Kaimal, G., Kendra, R., & and Muniz, J. (2016). Reduction of Cortisol Levels and Participants' Responses Following Art Making. *Art Therapy*, *33*(2), 74-80. https://doi.org/10.1080/07421656.2016.1166832
24. Kusurkar, R. A., Orsini, C., Somra, S., Artino, A. R., Jr., Daelmans, H. E. M., Schoonmade, L. J., & van der Vleuten, C. (2023). The Effect of Assessments on Student Motivation for Learning and Its Outcomes in Health Professions Education: A Review and Realist Synthesis. *Acad Med*, *98*(9), 1083-1092. https://doi.org/10.1097/ACM.0000000000005263
25. Leopold, C., & Leutner, D. (2012). Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learning and Instruction*, *22*(1), 16-26. https://doi.org/10.1016/j.learninstruc.2011.05.005
26. Lunt, T., & Curran, J. (2010). 'Are you listening please?' The advantages of electronic audio feedback compared to written feedback. *Assessment & Evaluation in Higher Education*, *35*(7), 759-769. https://doi.org/10.1080/02602930902977772
27. Millar, B. C., Tarasov, A., Ternan, N., Moore, J. E., & Murphy, C. (2023). Embedding Scientific Communication and Digital Capabilities in the Undergraduate Biomedical Science Curriculum [Original Research]. *British Journal of Biomedical Science*, *Volume 80 - 2023*. https://www.frontierspartnerships.org/journals/british-journal-of-biomedical-science/articles/10.3389/bjbs.2023.11284
28. Myford, C. M., & Wolfe, E. W. (2009). Monitoring Rater Performance Over Time: A Framework for Detecting Differential Accuracy and Differential Scale Category Use. *Journal of Educational Measurement*, *46*(4), 371-389. https://doi.org/10.1111/j.1745-3984.2009.00088.x

29. Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: a meta-analysis of research findings. *Psychol Bull*, *134*(2), 270-300. https://doi.org/10.1037/0033-2909.134.2.270
30. Perkel, J. M. (2020). The software that powers scientific illustration. *Nature*, *582*(7810), 137-138. https://doi.org/10.1038/d41586-020-01404-7
31. Prain, V., & Tytler, R. (2012). Learning Through Constructing Representations in Science: A framework of representational construction affordances. *International Journal of Science Education*, *34*(17), 2751-2773. https://doi.org/10.1080/09500693.2011.626462
32. Quillin, K., & Thomas, S. (2015). Drawing-to-Learn: A Framework for Using Drawings to Promote Model-Based Reasoning in Biology. *CBE—Life Sciences Education*, *14*(1), es2. https://doi.org/10.1187/cbe.14-08-0128
33. Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, *61*, 101860. https://doi.org/https://doi.org/10.1016/j.cedpsych.2020.101860
34. Schultz, M., Young, K., K. Gunning, T., & Harvey, M. L. (2021). 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
35. Sion, A., Czamanski-Cohen, J., Halbrecht-Shaked, O. C., Galili, G., & Cwikel, J. (2023). Self-correction, digital art making and stress reduction. *The Arts in Psychotherapy*, *85*, 102060. https://doi.org/https://doi.org/10.1016/j.aip.2023.102060
36. Tan, H., & Kidman, G. (2021). Authentic assessment in stem education: an integrative review of research. In T. B. a. T. McLaughlin (Ed.), *Authentic Assessment and Evaluation Approaches and Practices in a Digital Era* (1st ed., pp. 24-52). Brill. https://doi.org/10.1163/9789004501577_002
37. Thelma, C. C., Sain, Z. H., Shogbesan, Y. O., Phiri, E. V., & Akpan, W. M. (2024). Digital Literacy in Education: Preparing Students for the Future Workforce. *International Journal of Research*, *11*(8), 327–344. https://doi.org/10.5281/zenodo.13347717
38. van der Vleuten, C. (2000). Validity of final examinations in undergraduate medical training. *BMJ*, *321*(7270), 1217-1219. https://doi.org/10.1136/bmj.321.7270.1217
39. Van Meter, P., & Firetto, C. M. (2013). Cognitive model of drawing construction: Learning through the construction of drawings. In *Learning through visual displays.* (pp. 247-280). IAP Information Age Publishing.
40. Van Meter, P., & Garner, J. (2005). The Promise and Practice of Learner-Generated Drawing: Literature Review and Synthesis. *Educational Psychology Review*, *17*(4), 285-325. https://doi.org/10.1007/s10648-005-8136-3
41. Villarroel, V., Susan, B., Daniela, B., Carola, B., & and Herrera-Seda, C. (2018). Authentic assessment: creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, *43*(5), 840-854. https://doi.org/10.1080/02602938.2017.1412396

42. 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*, 101425. https://doi.org/https://doi.org/10.1016/j.stueduc.2024.101425
43. von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders*, *227*, 483-493. https://doi.org/https://doi.org/10.1016/j.jad.2017.11.048
44. Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting Learners' Agentic Engagement With Feedback: A Systematic Review and a Taxonomy of Recipience Processes. *Educational Psychologist*, *52*(1), 17-37. https://doi.org/10.1080/00461520.2016.1207538
45. Xu, X., Shi, Z., Bos, N. A., & Wu, H. (2023). Student engagement and learning outcomes: an empirical study applying a four-dimensional framework. *Medical Education Online*, *28*(1), 2268347. https://doi.org/10.1080/10872981.2023.2268347
46. Zhan, Y., Boud, D., & Du, Z. (2025). Designing for authentic assessment: a scoping review. *Higher Education*. https://doi.org/10.1007/s10734-025-01588-9
47. Zhang, K. E., & Jenkinson, J. (2024). The Visual Science Communication Toolkit: Responding to the Need for Visual Science Communication Training in Undergraduate Life Sciences Education. *Education Sciences*, *14*(3).

Figures

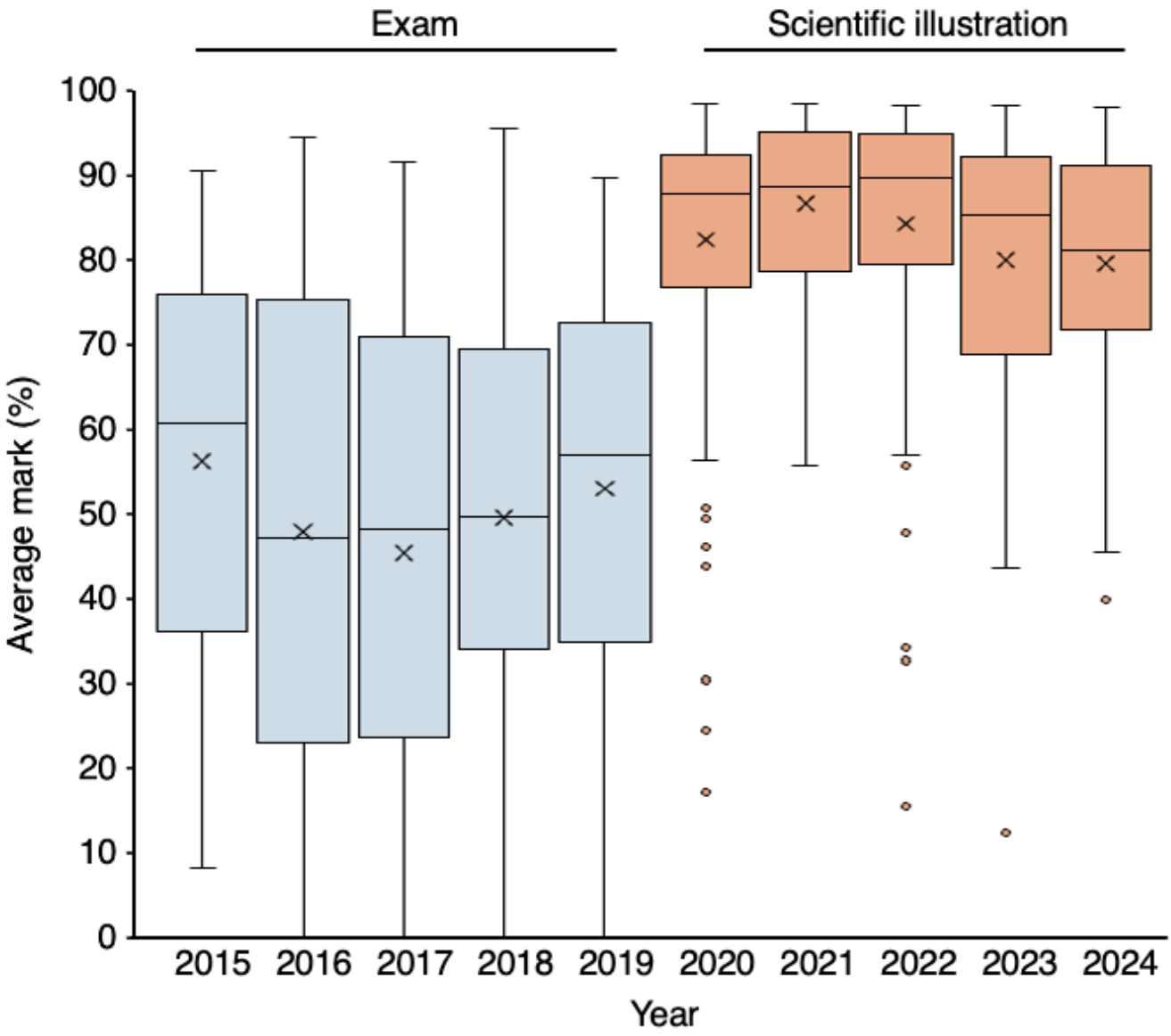


Figure 1

Box and whisker plot illustrating the exam scores (2015-19) and the scientific illustration scores. The horizontal line at the centre of each box corresponds to the median student score (n= 56, 87, 101, 101. 89, 102, 82, 95, 100, 106 students), the x is the mean score, while the lower and upper edges of the box correspond to the first and third quartiles, respectively. The whisker length is the furthest data point within 1.5 times the interquartile range from the edge of the box. Outliers are indicated with open black circles.

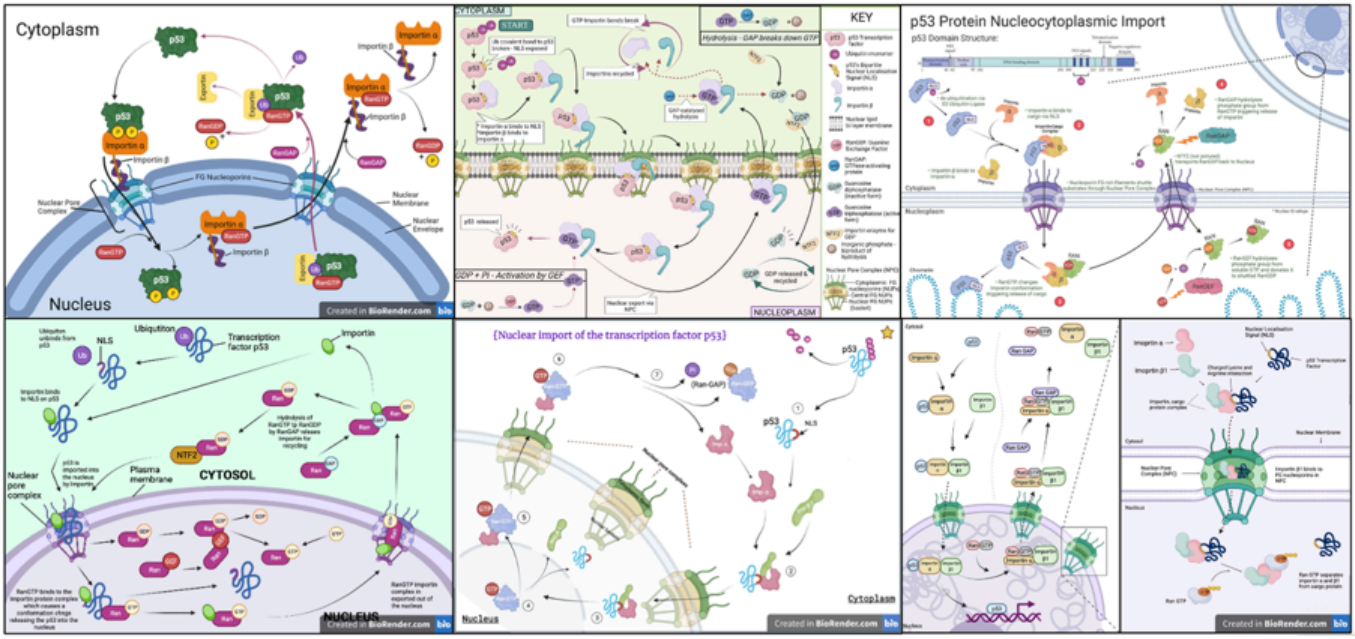


Figure 2

Six student-created scientific illustrations depicting the nuclear import of the transcription factor p53, shared with permission from both the students and BioRender. All images were created using BioRender.com.

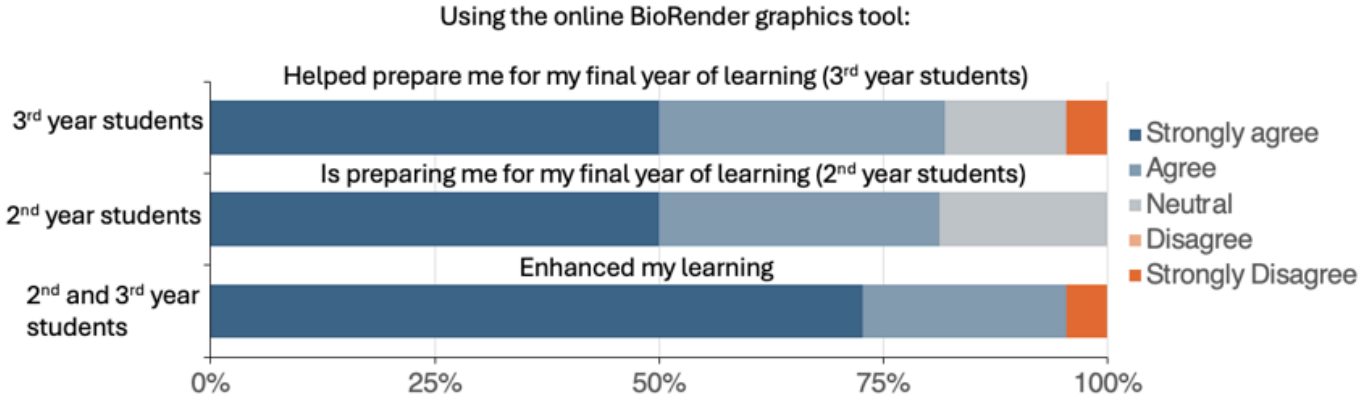


Figure 3

Student perceptions of BioRender's learning impact, based on Likert scale responses from current (2022) and past (2021) students.