ASCILITE 2026

The Future is Connection:

Collaboration, Creativity and Compassion in a Digital Society

Beyond the pretty graph: Human-centred assessment design for evaluative judgement

Malgorzata Korolkiewicz, Joshua Chopin
Adelaide University

Assessment in data visualisation courses too often rewards technical output over communicating substance. When visualisation is taught as a tool rather than a discipline, students learn to produce work that meets the brief but has little to say. This paper reports on a course redesign that addresses this gap through a human-centred assessment design grounded in evaluative judgement theory, treating evaluative judgement as a capability exercised by both students and markers. The redesign is structured around five elements addressing student ownership of project topics, portfolio assessment, descriptive capability levels, practitioner modelling, and creative freedom within a structured framework, implemented in a required course within a Master of Data Science program. Practitioner inquiry into the operation of this design in the presence of generative AI revealed that AI functioned not as a threat to academic integrity but as a mechanism that raised creative ambition, shifting the limiting factor from technical skill to evaluative judgement. Although situated in a data visualisation course, the contribution is broader. As generative AI makes technical production easier, helping students develop the capacity to recognise, justify, and communicate insight becomes the central challenge across higher education, and this paper suggests one way to address it.

Keywords: Evaluative judgement, portfolio assessment, process-based assessment, generative AI, flipped learning, creative freedom, practitioner inquiry

Introduction

Data visualisation has become a foundational capability across disciplines. As data increasingly shapes decisions in science, business, health, and public life, the ability to represent, interpret, and communicate data meaningfully is no longer the province of specialists but expected of graduates across many fields (Bach et al., 2023). Growing demand has driven a rapid expansion in visualisation education, yet the pedagogical challenges of teaching it well remain largely unsolved. Chief among these is the challenge of assessment. Data visualisation is a discipline built on judgement: which data matters, what does it mean, what does an audience need to understand, and what can be left out. Answering these questions well is not a technical skill; it requires effective communication. Yet in courses where visualisation is taught as a step in a data pipeline rather than a discipline in its own right, assessment is often built around what is easiest to specify: technical correctness, visual polish, and complexity as a proxy for effort. The result is work that can be impressive to look at while communicating very little.

Bach et al. (2023) identify the fair assessment of creative, project-based work as a significant ongoing challenge in data visualisation education. Literature on teaching data visualisation as a discipline is still developing, concentrated largely within the visualisation research community itself (Hedayati & Kay, 2025), with related perspectives emerging from analytics education more broadly (Camm et al., 2023), and distinct from the much larger literature on using data visualisation as a pedagogical tool within other subjects. This paper contributes to that literature by providing an account of what happened when evaluative judgement was enacted at scale under permitted generative AI use. The course discussed here is a required data visualisation course in a Master of Data Science program, also taken by undergraduate students from Information Technology, Computer Science, and Mathematical Sciences. Enrolments have grown substantially in recent years, reflecting broader demand for data science education, and currently sit at approximately 100 students per semester. Most arrive with reasonably strong quantitative backgrounds but relatively limited experience making and defending design decisions. The course is flipped: students engage with pre-class material, including progressively sequenced practitioner videos and TED Talks, before attending forums focused on discussion and critique. Assessment is portfolio-based, referenced against descriptive capability

levels rather than conventional grades. Although the course has been offered since 2016, the iteration reported here follows a substantial redesign undertaken during a broader university-wide curriculum renewal. The redesign was motivated both by the emergence of generative AI and by an opportunity to reconsider long-standing assessment structures through the lens of evaluative judgement. Additionally, the redesign aimed to address the question of what remains distinctively human once technical production is no longer a meaningful differentiator. Rather than responding with surveillance, such as detecting or restricting AI use (Bittle & El-Gayar, 2025), we relied on process documentation and found, unexpectedly, that permission to use AI in support of project work did not automatically become a threat to academic integrity. Instead, it appeared to raise students' creative ambition. We took risks in redesigning this course: what follows is an account of those risks, why we took them, and what happened, including outcomes that surprised us.

Conceptual framing

At the centre of the course design sits a deceptively simple question: *what does good look like?* In data visualisation, quality cannot be reduced to technical compliance. A visualisation may be accurate, well-structured, and visually appealing while failing to communicate anything meaningful. Developing the capacity to recognise and articulate that distinction is the primary learning goal of the course. Sadler (1989) described this capability as evaluative judgement: the ability to appraise the quality of work against standards that cannot be fully specified in advance. Tai et al. (2018) subsequently argued that evaluative judgement should be understood as a graduate capability in its own right, developed through repeated opportunities to engage with standards of quality. The emergence of generative AI has given this argument new urgency: Bearman et al. (2024) contend that the capacity to recognise and calibrate quality represents a distinctively human capability at a time of technological acceleration. That capability is the central ambition of our course redesign.

Making that judgement visible to students before asking them to exercise it is the insight at the heart of cognitive apprenticeship (Collins et al., 1989). In this course, expert thinking is modelled deliberately through practitioner videos and TED Talks that show how working visualisation designers navigate trade-offs between clarity and complexity, completeness and argument, data and meaning. Students encounter practitioners making decisions and defending choices, not merely following rules, and begin to enact similar choices in their work. The same principle extends to marking: tutor guidance develops shared understanding of what each capability level actually means in practice, rather than enforcing rubric compliance. In this course, evaluative judgement is pursued simultaneously in students and markers.

Because students work on self-selected projects, assessment cannot be reduced to applying a checklist. Marking meetings therefore focus on discussing examples of student work, articulating the reasoning behind judgements, and collectively interpreting what the capability levels mean in practice. The aim is not simply consistency of grading but the development of a shared language of quality. Tai et al. (2018) argue that evaluative judgement functions most powerfully when it is treated as a collective capability, exercised throughout the community rather than solely by individuals. In this paper, we present a practitioner inquiry guided by three related questions: how assessment design could move students away from simply reporting what the data shows and using simple plot types that generative AI tools also default to; how students' own evaluative contribution could be distinguished from generic and potentially AI-generated output; and how that distinction could be captured meaningfully within descriptive capability levels. Our approach to course design that places evaluative judgement at the centre is summarised in Figure 1 below.

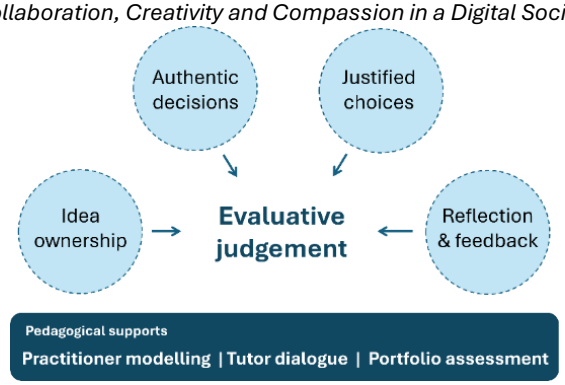


Figure 1. Conceptual design underpinning evaluative judgement supported through practitioner modelling, tutor dialogue, and portfolio assessment.

Design elements

The course is built around five components, each enacting the conceptual framing in a different way.

Student-owned projects

Students select their own topic and source their own dataset, then develop that project throughout the semester. Rather than completing a predefined exercise, they work on a sustained investigation that they shape through successive design decisions. Project ownership serves a purpose beyond motivation: it creates authentic situations in which evaluative judgement becomes necessary. Students must determine whether a dataset is fit for purpose, which questions are worth pursuing, what story the data supports, and how to effectively communicate that story using visualisations. Meaningful choice has been shown to increase students' engagement with assessment (Jopp et al., 2023; Ryan & Deci, 2020), but its role here is more fundamental. Working with authentic questions and datasets from the outset further strengthens this by introducing the ambiguity and uncertainty that real-world visualisation practice requires. Ownership of the project idea makes judgement unavoidable; the choose-your-own-adventure structure means the boundaries are defined, but the developmental path a student takes is not.

Assessing process, not artefact

The portfolio documents design decisions, iterations, and reasoning across the semester, including rejected choices with reasons why. Evaluative judgement becomes visible: the portfolio is intended as evidence of developing capacity to appraise one's own work, not merely a record of what was produced. It is also more robust in an AI-assisted environment, not because it detects AI use, but because it assesses the development of a point of view over time. This is something that is harder to outsource without outsourcing the learning itself (Bearman et al., 2024). AI use was explicitly permitted in this course at a level corresponding to AI-assisted ideation and technical support on the Artificial Intelligence Assessment Scale (Perkins et al., 2024), with students required to maintain a log of their AI use throughout the semester. The AI use log itself became part of that evidence: a record of where human judgement was exercised and where it was delegated. Making the final submission a portfolio documenting the process of arriving at a point of view over time was a bet that this is something students could not easily use AI to generate without sacrificing authenticity.

Descriptive capability levels, not grades

Student work is assessed against five levels: Exceptional, Competent, Developing, Emerging, and Foundation. Grade labels communicate position in a distribution; capability levels communicate position in a developmental trajectory. The shift appears to remove the grade-anxiety that often constrains creative risk-taking, and asks students to engage with the same evaluative question markers ask: how good is this work, and why? Detailed rubrics can orient students toward compliance rather than genuine appraisal of quality (Gyamfi et al., 2022); descriptive levels resist that orientation by preserving the interpretive work that evaluative judgement requires. Descriptive levels ask students to engage with the question of quality rather than the question of marks, and not all of them find that straightforward.

Modelling expert judgement

Expert thinking is made visible through a deliberately layered set of pre-class resources: purpose-built concept videos introducing key ideas, curated videos from industry practitioners as well as TED Talks, showing how working visualisation designers navigate communicating with data. Students arrive at workshop-style forums having already seen expert judgement in action. Forum activities then move through the core sequence of cognitive apprenticeship (Collins et al., 1989): students propose approaches, plan design decisions, assess examples, interpret data, and construct and defend a point of view. Computer labs serve a different but equally important function: introducing students to what is technically possible, including interactivity early in the semester. The technical sequence moves at pace, trusting that exposure to what is possible would carry more than repeated consolidation of basics.

Creative freedom within a structured framework

Data visualisation courses typically teach students to select and produce appropriate chart types for a given data structure. This course asks for more: students are expected to find a story in their data, decide how to tell it, and choose the form, including interactive and dynamic visualisations, that best serves that purpose. Students are given frameworks, practical guidance, and explicit discussion of design trade-offs. Through those they learn what not to do, and why. Within those boundaries, the decisions are theirs: what story to tell, what form the visualisation takes, what interactive features to include, and how to justify those choices, thus enacting structured creative agency (Brauer et al., 2025) that makes assessment feel genuinely interesting rather than a requirement to receive a grade. The design creates the conditions for ambition, but whether those conditions are enough to spark creativity in students' work is something only they can answer.

What we saw

The observations reported here are based on teaching-team reflection during and after the delivery of the redesigned course; no systematic collection or analysis of student data was undertaken. Across the semester, the teaching team's overriding impression was one of surprise. Students who expected technical content found something different: a course that asked them to have a point of view and treated that point of view as the actual object of assessment. For many, the freedom to pursue a topic they genuinely cared about appeared to change how the assessment task was experienced, shifting it from a university requirement toward something closer to a personal investigation. This shift was not simply a new skill acquired but a threshold crossed once a student had internalised enough of the quality framework to recognise what their own work is capable of becoming (Sadler, 1989).

For many students, the early introduction of interactive visualisation, starting with maps, proved to be a turning point. The teaching team observed something shift when students realised that interactivity was a viable option. This realisation came during a computer lab when they saw not only that interactive visualisation was possible, but also that *they* could build it. That moment changed what students believed they were able to and wanted to produce. These were not just technical discoveries; they were moments of creative realisation, which the teaching team saw, as another surprise, carried into portfolio work.

At times, student enthusiasm for interactive techniques appeared to outpace the intended pacing of the course, requiring the teaching team to defer some of that ambition to later in the semester. One further observation was genuinely unexpected: for some students, AI seemed to remove the technical ceiling on what they could produce. Rather than diminishing learning, the additional, on-demand technical support of AI-based tools shifted ambition upward. Visualisations that would previously have been out of reach became possible, and the questions evolved from whether something could be built towards whether it should be built that way, and for whom. The quality of the work produced under these conditions suggested that, for at least some students, the ceiling had genuinely moved and the assessment design gave that ambition somewhere to go.

Not every student took up the invitation to do something different. Some continued to equate complexity with quality; some produced technically polished work that had little to say; others produced traditional data reports that were careful not to take a point of view. In a large predominantly international cohort, this is unsurprising as the aim was not to guarantee creativity, but to create conditions in which it could emerge. The marker dimension also mattered as teaching team discussions were often as much about interpreting quality

as they were about assigning grades, extending well beyond grade compliance into genuine engagement with what the work was trying to achieve.

Conclusion

Designing assessment for evaluative judgement is not a matter of removing structure; it is a matter of placing structure where it is most useful. Student ownership coupled with real data creates conditions for genuine decision making. Process portfolios make that choice visible, descriptive levels keep the focus on quality rather than compliance, practitioner modelling establishes a shared vocabulary of what good looks like, and creative freedom gives students somewhere worth going with it all. Remove any one element and the others weaken.

This paper reported practitioner inquiry from the redesign of a single course rather than formal educational research. The observations illustrate how one coherent assessment design operated in practice, but they are not intended to establish causal claims or generalise across contexts. Future work could examine how similar design principles function in other disciplines, institutions, and teaching teams. Although situated in a data visualisation course, the contribution is broader. The discipline makes particularly visible a challenge emerging across higher education as generative AI becomes more capable: helping students develop the capacity to recognise, justify, and communicate quality rather than merely produce artefacts. Bach et al. (2023) identify the fair assessment of creative, project-based work as an unsolved challenge in the field, one likely mirrored in other contexts and echoed in recent literature on what data visualisation courses should aim to teach (Camm et al., 2023; Hedayati & Kay, 2025). The course described here suggests the gap between technical competence and communicative judgement can be bridged by course design, but only if the design commits to closing that gap structurally. The growing presence of generative AI makes solving this challenge more urgent. Rather than asking how assessment can prevent students from using AI (Bittle & El-Gayar, 2025), educators might ask what forms of learning remain valuable when such tools are readily available (Bearman et al., 2024). One answer is the capacity to develop and defend a position, and designing explicitly for it may be the most useful response available to educators right now.

There is a final important implication of what we did in the course, which is that the joy of discovery is a legitimate design outcome. When students move from seeing data to seeing stories, when markers are surprised by the quality of work they encounter, something significant has happened. Moments like these are not incidental and, in our case, they were part of the course design. The approach was not without risk. That most of those risks produced outcomes beyond what was anticipated is, perhaps, the most rewarding outcome of all.

Acknowledgement

Claude (Anthropic) was used to support initial drafting, structural development, reference verification, and editorial critique. All ideas, theoretical framing, evidence, and final decisions are the authors' own; substantial rewriting was undertaken throughout.

References

Bach, B., Keck, M., Rajabiyazdi, F., Losev, T., Meirelles, I., Dykes, J., Laramee, R. S., AlKadi, M., Stoiber, C., Huron, S., Perin, C., Morais, L., Aigner, W., Kosminsky, D., Boucher, M., Knudsen, S., Manataki, A., Aerts, J., Hinrichs, U., … & Carpendale, S. (2023). Challenges and opportunities in data visualization education: A call to action. *IEEE Transactions on Visualization and Computer Graphics*, *30*(1), 1–12. https://doi.org/10.1109/tvcg.2023.3327378

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education*, *49*(6), 893–905. https://doi.org/10.1080/02602938.2024.2335321

Bittle, K., & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information (Basel)*, *16*(4), 296. https://doi.org/10.3390/info16040296

Brauer, R., Ormiston, J., & Beausaert, S. (2025). Creativity-fostering teacher behaviors in higher education: A transdisciplinary systematic literature review. *Review of Educational Research*, *95*(5), 899–928. https://doi.org/10.3102/00346543241258226

Camm, J. D., McCray, G. E., & Roehm, M. L. (2023). More than just charts and graphs: What to teach in a data visualization course. *Decision Sciences Journal of Innovative Education*, *21*(3), 112–122. https://doi.org/10.1111/dsji.12282

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Lawrence Erlbaum Associates.

Gyamfi, G., Hanna, B. E., & Khosravi, H. (2022). The effects of rubrics on evaluative judgement: A randomised controlled experiment. *Assessment & Evaluation in Higher Education*, *47*(1), 126–143. https://doi.org/10.1080/02602938.2021.1887081

Hedayati, M., & Kay, M. (2025). What university students learn in visualization classes. *IEEE Transactions on Visualization and Computer Graphics*, *31*(1), 1072–1082. https://doi.org/10.1109/TVCG.2024.3456291

Jopp, R., Pallant, J. L., & Russell, H. (2023). Choose your own adventure: Understanding why students prefer certain types of assessment. *Journal of University Teaching and Learning Practice*, *20*(7). https://doi.org/10.53761/1.20.7.11

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, *21*(6). https://doi.org/10.53761/q3azde36

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/10.1016/j.cedpsych.2020.101860

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*(2), 119–144. https://doi.org/10.1007/BF00117714

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make judgements about their own work. *Higher Education*, *76,* 467-481. https://doi.org/10.1007/s10734-017-0220-3

Korolkiewicz, M., & Chopin, J. (2026). Beyond the pretty graph: Human-centred assessment design for evaluative judgement. In Cochrane, T., Narayan, V., Poole, T., Lai, P., Lake, P., Moore, O., & Li, L. (Eds.), *The Future is Connection: Collaboration, Creativity and Compassion in a Digital Society*. Proceedings ASCILITE 2026. Brisbane (pp. 672–677). https://doi.org/10.65106/apubs.2026.3259

Note: All published papers are refereed, having undergone a double-blind peer-review process. The author(s) assign a Creative Commons by attribution license enabling others to distribute, remix, tweak, and build upon their work, even commercially, as long as credit is given to the author(s) for the original creation.

© Korolkiewicz, M., & Chopin, J. 2026