

Adaptive Comparative Judgement: A Method in Search of a Theory in Need of an Instrument

Jeffrey Buckley^{1,2} and Richard Kimbell³

¹Department of Technology Education, Technological University of the Shannon: Midlands Midwest

²Department of Behavioural Sciences and Learning, Linköping University

³Goldsmiths University of London

Jeffrey.buckley@tus.ie

Abstract

Adaptive comparative judgement (ACJ) is increasingly used in technology and design education to assess open-ended student work through holistic pairwise comparisons. Despite strong evidence of reliability, much ACJ research continues to treat the method “as it is”, focusing on algorithmic efficiency and technical feasibility while leaving the cognitive basis of judgement largely implicit. This paper proposes a Cognitive Theory of Comparative Judgment to reframe ACJ as a problem of human decision-making under constraint. The paper (i) deconstructs ACJ as a judgement environment, (ii) advances four initial postulates concerning computational constraint, heuristic strategy, algorithmic tension, and the normative outcome of disagreement, and (iii) introduces OpenACJ, an open-access, adaptable platform designed to operationalise and test these postulates through configurable choice architecture and process data capture. The implication is a shift from asking whether ACJ “works” to explaining how it works, enabling theoretically grounded improvements to validity in the assessment of complex design capability.

Key Words: Comparative Judgement, Decision Making, Cognitive Load, Heuristics, Assessment Validity

1 INTRODUCTION

Adaptive comparative judgement (ACJ) has significant potential to positively impact technology and design education research and practice. It is primarily used as a method of educational assessment, where its utility relates to the reliable assessment of open-ended pieces of student work, overcoming limitations in traditional rubrics. Such limitations include rubrics not adequately capturing the varying ways in which learners can demonstrate capability in open-ended activities and that rubric scores may be incongruent with holistic appraisals from expert, professional educators and assessors (cf. Sadler, 2009). Based on Thurstone’s (1927) Law of Comparative Judgement, ACJ was designed to overcome these challenges by fundamentally changing the question presented to assessors. Rather than having to assign absolute scores against

predetermined criteria, which humans are relatively unreliable at in open-ended scenarios (Laming, 2004; Thurstone, 1927), assessors are presented with two pieces of work and are asked to make a single, holistic comparison and select the “better” of the two. This is typically relative to an overarching criterion, such as which is more creative or which demonstrates more evidence of capability, and is a question type that is answerable more reliably. The extant literature in technology education, and indeed more broadly, illustrates that ACJ can achieve highly reliable assessment of the types of activity, e.g., design activity, characteristic of authentic technological learning (Bartholomew & Jones, 2021; Bartholomew & Yoshikawa-Ruesch, 2018).

Research into the use of ACJ in technology education was borne largely from the e-scape project in the United Kingdom (Kimbell, 2012a, 2012b). The impact of this project on the field was notable, resulting in a special issue in the *International Journal of Technology and Design Education* (Williams & Kimbell, 2012). Since then, research on or with ACJ within technology and design education appears to have taken four forms. First there are studies which explore the utility of ACJ as a valid method of measuring or assessing specific competencies, abilities, or outcomes. For example, Zhu et al. (2025) examined the potential of ACJ to measure student creativity. Second, there are a number of studies in which the reliability of ACJ is examined (Kimbell, 2022), typically in the assessment of work created in response to open-ended activities. Third are studies in which ACJ is used in an applied context for a purpose other than assessment or measurement. For example, in the Learning by Evaluating project, Thorne et al. (2025) expose learners to exemplars through ACJ, and examine the impact on their learning from making holistic comparative judgements. Finally, there is work aiming to use ACJ at scale, such as in the provision of a national rank against which teachers in local classrooms could compare their students work to get a sense of where they lie at a national level (Seery et al., 2022). This work has associated elements such as understanding the technical challenges in implementing ACJ this way (Buckley & Canty, 2022).

Each of these research programs is useful. There is value in understanding, scrutinising and continuously aiming to improve upon the validity, reliability, transferability, and feasibility of ACJ. However, they share an epistemological limitation in that they are asking about ACJ 'as it is', effectively treating the constraints of existing infrastructure as the boundaries of the method itself. This represents a missed opportunity. By narrowly conceiving ACJ as solely an efficient assessment tool (assessment of learning) or as a means to facilitate pedagogical comparisons (assessment as, or for, for learning), we overlook its potential as a powerful instrument for investigating the fundamental nature of human decision-making in educational assessment contexts.

To advance beyond the current state-of-the-art, in this paper a Cognitive Theory of Comparative Judgment is proposed. The rationale is that ACJ primarily concerns decision making and must therefore be understood more broadly through a lens of cognition. Indeed, the educational application of Thurstone's (1927) Law of Comparative Judgement is relatively recent and small in scale with respect to applied evidence, whereas the fundamental study of decision making is a far more developed field.

The development and systematic refinement of such a theory would enable understanding of, for example, how assessors manage the cognitive load of comparison, how heuristics shape their

judgements, and where the limits of human processing constrain the validity of the method. However, building this theoretical understanding requires an iterative relationship between theory and tool. We also need an ACJ instrument that can be responsive to theoretical development, enabling researchers to configure the environment and react to generated evidence.

To this end, this paper simultaneously introduces OpenACJ, an open-access, adaptable platform designed specifically to support this iterative inquiry. The ultimate intention for OpenACJ is to serve as a research sandbox for the field, enabling the generation of fundamental insights and validated theoretical advancements that can subsequently inform and enhance assessment practices in general. In developing the Cognitive Theory of Comparative Judgment, hypotheses will need to be specified and tested, and OpenACJ presents an auspicious tool which can be adapted to enable this in a way commercial tools do not always have the flexibility to do. A useful byproduct of this activity is a tool which is also usable in practice.

The remainder of this paper proceeds in three parts. First, the ACJ method will be deconstructed and its development to date used to demonstrate a need for the Cognitive Theory of Comparative Judgment. Second, theoretical postulates are presented, such as relating to cognitive load theory and heuristics, to formulate a preliminary theory. Finally, the technical architecture of OpenACJ will be described, illustrating how its specific features and adaptability allow these theoretical postulates to be empirically tested.

2 ACJ AND THE NEED FOR A COGNITIVE THEORY OF COMPARATIVE JUDGEMENT

ACJ begins with the decision of whether to use ACJ in the first place. Comparative judgement (CJ) more broadly¹ has been used in assessment across several subject areas (Jones et al., 2015; Seery et al., 2019; Wheadon et al., 2020; Whitehouse, 2013), most often in higher education. The suitability of ACJ does not relate to subject area, but to the nature of what is being assessed. Specifically, ACJ is efficacious when the object of assessment is latent, i.e., a construct not directly observed but inferred. It is unnecessary for cases such as simple arithmetic, as the assessor can directly observe the learner's responses to declarative, closed problems. However, regarding broader constructs such as creativity, reasoning, or expertise, ACJ becomes very useful as the outcomes do not have absolute determinants of correctness but are instead manifestations of capability that exist on a continuum of quality. For this reason, some CJ researchers use the term "representations" to describe the work being assessed (Verhavert et al., 2022). This terminology posits that the pieces of work are representative of an inferred, unobservable construct, rather than constituting direct evidence of an observable competency. In this view, the assessment challenge is analogous to measuring a latent trait, and ACJ is efficacious as the observed assessor decisions which are subsequently discussed are modelled to capture the 'utility' or perceived value of the

¹ The distinction between comparative judgement (CJ) and adaptive comparative judgement (ACJ) lies in the pairing mechanism. In ACJ, there is some "adaptive" algorithm which dynamically selects which pieces of work to present as a pair for comparison, and in CJ this algorithm is static and does not change during the assessment session. The methodological principles are otherwise fundamentally identical.

trait (Yellott, 1977), without requiring the explicit articulation of criteria that is often necessary in rubric-based assessment.

The ACJ session itself involves selecting several people to act as the assessors or judges, and several pieces of work, often called portfolios or representations, to be assessed. For this paper, the terms “judge” and “portfolio” respectively will be used to correspond to ACJ literature in technology education. There are software solutions which then manage the process, such as RM Compare (2023), and for the purposes of this paper, OpenACJ (Buckley, 2026). From the judge perspective, each individual is shown several pairs of portfolios for comparisons in succession, and for each must select the “better” of the two relative to the assessment question, i.e., latent variable or utility, provided by the organiser of the session. There are several ways to determine the number of comparisons required for a single ACJ session. Importantly, it is a question of the number of judgements required, not the number of judges. The number of judgements can be associated with a desired level of reliability (Verhavert et al., 2019), but there is also a common heuristic of having the total number of judgements equal to 10 times the number of portfolios, with this amount then divided amongst the judges.

From the system perspective, there is an algorithm dictating which pairs of portfolios to present to each judge and in what order based on some criteria. The decisions made are then recorded as wins/losses for each comparison, and these are modelled using a ranking model, typically the Bradley-Terry-Luce (BTL) model, providing outputs including a rank from “best” to “worst” of the portfolios, with relative differences, along a single latent continuum, and misfit statistics.

Research into ACJ (and CJ) aiming for improvement has typically been concerned with increasing algorithmic efficiency and transparency (Buckley & Lin, 2024; Gray et al., 2024), or in standardising the outputted relative ranks across sessions to enable merging or contrasting (Benton, 2021; Buckley & Canty, 2022; Verhavert et al., 2022). It has been less concerned with understanding the theoretical basis (Kelly et al., 2022). There is a need to connect both ideals, and hence the need for a Cognitive Theory of Comparative Judgement. For example, Pollitt (2012) spoke about an “improvement” to ACJ, being able to create a “chain” in the comparisons received by each assessor. Specifically, according to Pollitt (2012):

This brings a further improvement since the system is able to ‘chain’ comparisons once the Rasch method is in use. Suppose a judge has compared portfolios P77 and P93: the next decision they are asked to make may be between P93 and P21, and then P21 and P67—each comparison only involves one new portfolio rather than two. This may make little difference when comparing oriental rugs, but the time saved when comparing complex portfolios is substantial. (p. 163)

While Pollitt (2012) acknowledges a potential concern in chaining bias, he defends the gain in efficiency as worth the risk, stating that the bias would apply in a small and random way to each portfolio. However, this is an entirely statistical perspective relying on conditions not certain in practice. It assumes a very large number of judges and sufficient randomness in the pairing algorithm to leverage the central limit theorem. Otherwise, there could be too few chains that error would actually be systematic rather than stochastic.

Pollitt himself has suggested moving away from randomness (Kimbell et al., 2009), though others prefer to retain randomness initially (Buckley, 2024), and given that there are several possible adaptive algorithms (cf. Verhavert, 2018), the required number of judges to ensure minimal chaining bias is unknown. Consequently, Pollitt's (2012) goal to make ACJ more feasible at a judge level may actually make the process less feasible at a holistic level. Overlooking the cognitive dimension to ACJ in favour of statistical optimisation risks the creation of a feasible and reliable, but not valid, method.

3 TOWARDS A COGNITIVE THEORY OF COMPARATIVE JUDGEMENT

To work towards a Cognitive Theory of Comparative Judgment, it is necessary to move beyond a procedural description and statistical improvement and examine ACJ as a complex cognitive interaction. Drawing on the decision-making framework established by Bell, Raiffa, and Tversky (1988), this theory is positioned at the intersection of three perspectives: the descriptive (how judges actually make comparisons), the normative (how judges ought to make comparisons), and the prescriptive (how can judges be supported to make better comparisons whilst respecting their cognitive concerns). As described by Baron (2012), the descriptive lens lies in the domain of psychology, the normative in philosophy, and the prescriptive in engineering, or educational intervention (Larrick, 2004). By deconstructing the decision-making process inherent to ACJ through these lenses, we can identify specific postulates regarding the limitations of human processing. Such postulates can be translated into specific research questions or hypotheses and be examined through the manipulation of the assessment environment in tools like OpenACJ. Below are four initial postulates to exemplify the establishment of this theory, though several more are derivable and others would be established as the theory is developed.

3.1 *The computational constraint*

The cognitive demands of ACJ are inherent to the presented comparisons. Whilst the normative assumption of ACJ is that a holistic judgment is superior to criterion marking for assessment, the descriptive reality is affected by the information type, amount, and interactivity relative to the decisions to be made. Drawing on Computational Complexity Theory (Bossaerts & Murawski, 2017), the assessment of complex portfolios can be framed as a problem of tractability. In educational psychology, this concept is congruent with element interactivity within Cognitive Load Theory (Sweller, 2010).

Design portfolios typical of technology education are characterised as having high element interactivity, usually containing several related elements such as a deconstruction of the initial brief, iterative sketches, and technical specifications. When a judge is asked to compare 'Portfolio A' against 'Portfolio B', they are attempting to process two distinct webs of interacting elements. If the cognitive load resulting from negotiating the combined interactivity of these portfolios exceeds the working memory capacity of the judge, the decision problem could become intractable. Consequently, the first postulate suggests that when the cognitive load generated by the element interactivity of the portfolios exceeds the bounds of human processing, the judge cannot engage in the holistic judgment that ACJ purports to facilitate.

Translating this into an inquiry, one could investigate where the theoretical threshold of element interactivity (e.g., portfolio complexity) lies such that the reliability of comparative judgment degrades due to intractability. Having this information then could translate into a task design implication for OpenACJ, wherein portfolios are created which are smaller, lower-interactivity chunks to increase transitivity compared to overly complex comparisons.

3.2 *The cognitive strategy*

When the cognitive limits of a judge during a comparison are exceeded, a strategy must be adopted to facilitate reaching a decision. While the normative expectation is that judges engage in exhaustive information search and integration, the descriptive reality is framed by bounded rationality (Simon, 1990), suggesting an adaptation to the constraint. To manage the associated cognitive load, it is postulated that processing shifts from 'System 2', characterised by slow, analytical evaluation, to 'System 1', characterised by fast, intuitive, and heuristic-based judgments.

The Cognitive Theory of Comparative Judgment therefore requires the categorisation of these heuristic adaptations. For instance, an overloaded judge may employ Elimination by Aspects (Tversky, 1972), where the decision is reached by identifying a single negative or absent feature in one portfolio to justify rejection rather than weighing the total utility of both portfolios. Alternatively, satisficing may occur (Simon, 1990), wherein the first option that meets a minimum acceptability threshold is selected to terminate the cognitive effort.

Consequently, the second postulate suggests that high-cognitive load inducing comparisons bring a reliance on fast and frugal heuristics. Viewed through the lens of ecological rationality (Gigerenzer, 2024), these heuristics are not seen as leading to error, but as adaptive strategies which enable normatively correct decisions within the constraints of the comparative task.

Translating this into an inquiry, it is necessary to determine the behaviours that indicate a shift from holistic assessment to heuristic decision making, and in such cases, which heuristics are used. Specifically do ACJ judges work from an adaptive toolbox (Gigerenzer, 2001). This informs a research implication for OpenACJ, questioning whether the capture of granular process data (e.g., decision latency, navigation patterns) can empirically differentiate between holistic judgements and heuristic shortcuts.

3.3 *The algorithmic tension*

A theoretical tension can exist between the ACJ algorithmic logic and the sequential cognition of the judge. Normatively, adaptive algorithms may assume transitivity, that if $A > B$ and $B > C$, then $A > C$. This transitivity axiom assumes the independence of utility, implying that the relationship between two portfolios according to a judge holds regardless of the context in which they are presented.

However, the descriptive reality of decision-making suggest that utility is emergent and subject to path dependency. Evidence illustrates that judges do not utilise a static set of criteria across all

comparisons, but rather, the criteria are elicited based on the specific comparison presented (Buckley et al., 2022; Zhu et al., 2025). This path dependency is particularly relevant to the chains of portfolios described earlier by Pollitt (2012). Judging a portfolio immediately after a highly exceptional one may fundamentally reformulate the judge's internal standard of quality via the contrast effect.

The Cognitive Theory of Comparative Judgment reframes this path dependency as a mechanism for calibration. It is postulated that the sequence of comparisons serves a pedagogical function, allowing the judge to test and refine their internal criteria and develop their construct of quality framed in Polanyi's (1958) connoisseurship. Consequently, the third postulate suggests a shift from intransitivity to transitivity parallel to the crystallisation or stabilisation of the judge's own view, establishing within-person normativity.

Translating this into an inquiry, it is necessary to determine at what point in a session a judge's susceptibility to the contrast effect diminishes. This informs a prescriptive implication for OpenACJ, suggesting that adaptive algorithms must account for portfolio order and potentially utilise "warm-up phases" to allow for internal stabilisation before judgments significantly weigh on the final rank.

3.4 *The normative outcome*

Finally, for this paper, the theory addresses the nature of disagreement between judges. Having established that individual judges operate within their cognitive bounds (Postulate 1), utilise ecologically valid heuristics (Postulate 2), and stabilise their internal standards over time (Postulate 3), any remaining divergence cannot be dismissed as simple error.

The normative assumption of is that convergence, typically operationalised as high inter-rater reliability or scale separation reliability in ACJ, indicates validity. However, the descriptive reality of complex domains, such as technology and design education, is that they invite multiple legitimate interpretations of capability (Kimbell, 2011). The Cognitive Theory of Comparative Judgment therefore posits that, conditional on two judges being internally stable and adequately calibrated, disagreement may reflect distinct but coherent interpretations of a multidimensional construct rather than random noise.

Consequently, the fourth postulate suggests that the rank positions or ability scores of portfolios in ACJ are not always or necessarily representative of a single underlying consensus. Instead, the validity of the assessment lies in its capacity to capture and aggregate distinct judge perspectives. Disagreement is thus framed epistemologically as a feature of the complexity of the construct being assessed or measured.

Translating this into inquiry, one could ask whether the final unidimensional rank order masks distinct schools of thought. Specifically, can clusters of judges be identified who value specific dimensions (e.g., creativity vs. technicality) differently, but consistently? This creates a research implication for OpenACJ, asking whether the binary comparison data can be modelled in ways

that support multidimensional representations of judgement with better explanatory fit than a single latent continuum.

4 OPENACJ AS TOOL TO SUPPORT THE DEVELOPMENT OF THE COGNITIVE THEORY OF COMPARATIVE ASSESSMENT

With an agenda of inquiry established, the necessity for a bespoke instrument becomes apparent. Refining and validating the Cognitive Theory of Comparative Judgment requires access to the granular process data that underpins decision-making. OpenACJ was developed to enable this.

OpenACJ is an open access, adaptable platform designed to function as a research sandbox. It allows researchers to manipulate the choice architecture of the assessment environment to directly test the theoretical postulates. Currently, the platform operates through three distinct beta configurations.

Version 1 (Single-Device) and Version 2 (Multi-Device) share a core engine but serve distinct experimental needs. Version 1 isolates the judge on a single device, ideal for controlled laboratory settings where variables need to be strictly regulated. Version 2 supports distributed cohorts, mirroring the ecology of typical asynchronous ACJ sessions. As shown in Figure 1, the interface has a neutral presentation to minimize extraneous load. Both versions can be customised to capture psychometric data. For instance, upon making a decision, the judge can be prompted to respond to the Paas Cognitive Load Rating Scale (Paas, 1992) and provide explanatory comments (Figure 2). This feature directly addresses Postulate 1, allowing researchers to empirically identify the "tractability threshold" by correlating perceived mental effort with portfolio complexity.

Figure 1

OpenACJ Version 1 interface. Note that the “portfolios” are AI generated images in response to a prompt to design a bird house, intentionally showing a single idea versus multiple ideas, to support the creation of Figures for this paper

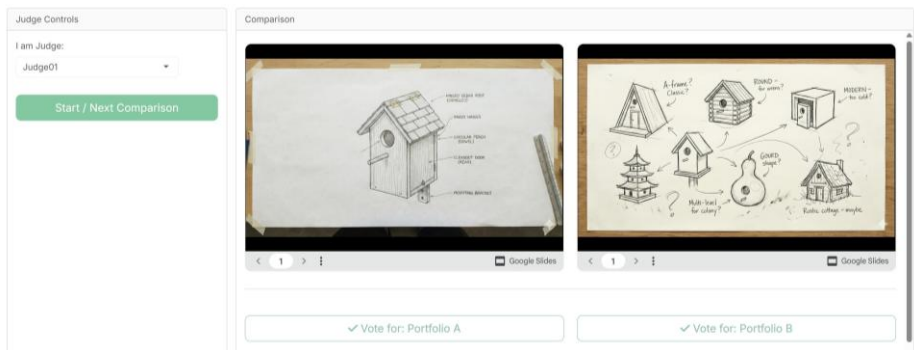
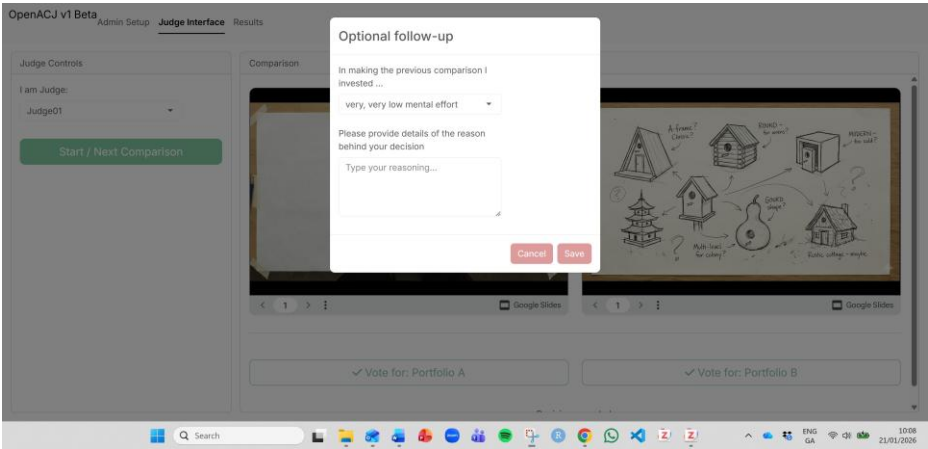


Figure 2

OpenACJ version 1 optional post-comparison data collection



Version 3 is modular, with the currently available module introducing a process tracing to investigate Postulate 2, The Cognitive Strategy. Drawing on decision science methodologies, this configuration overlays the standard interface with mouse-tracking functionality (Figure 3). As illustrated in Figure 4, this generates high-resolution temporal data on where and for how long a judge focuses their attention. By visualizing these trace maps and heatmaps, researchers can empirically differentiate between holistic and heuristic processing, and infer heuristic adoption. While this mouse tracing module is an example, future modules could include, for example, think aloud audio capture or eye tracking.

Figure 3

OpenACJ version 3 mouse tracing module interface

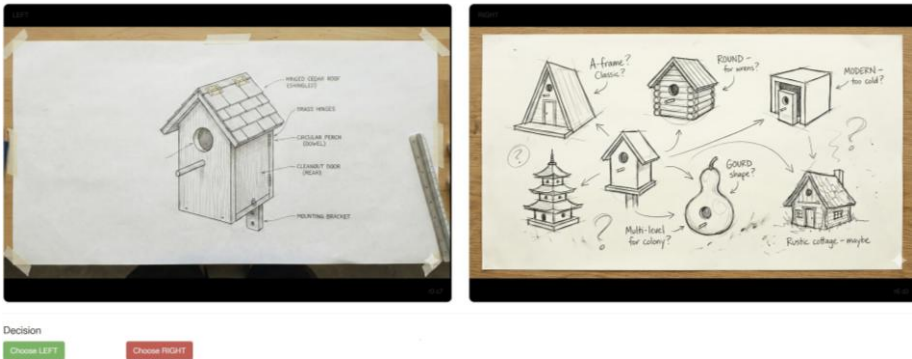
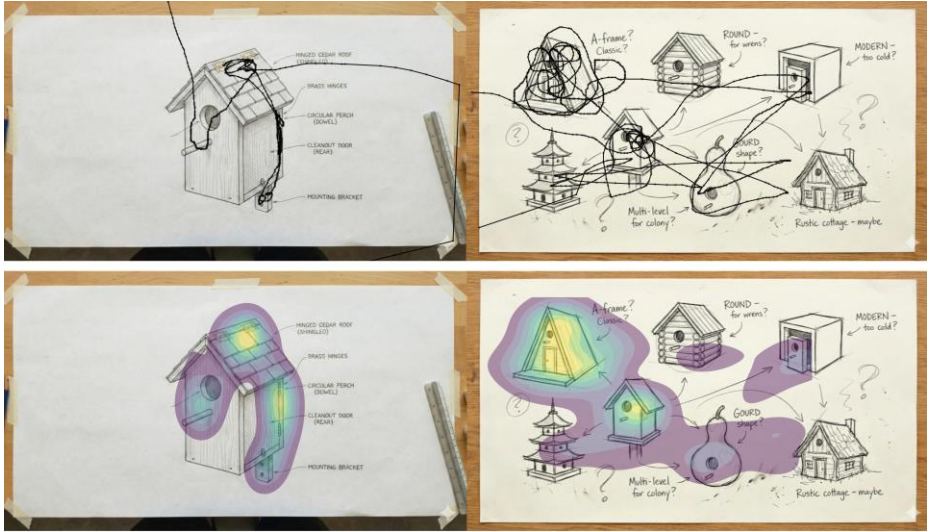


Figure 4

OpenACJ version 3 mouse tracing data output includes the mouse trace (top) and a derived heatmap (bottom)



5 CONCLUSION

This paper presents an initial proposal of the Cognitive Theory of Comparative Judgment to offer a framework to guide continued research into ACJ which merges ongoing algorithmic focus with a fundamental lens of human cognition. However, as this is just an initial presentation, it is inherently limited in potential scope. Further, theory development requires empirical verification. OpenACJ was thus developed to provide the necessary instrumentation to capture the granular process data required to test theoretical postulates. Ultimately, the synergy between this theory and instrument can move the field beyond asking if ACJ works, to understanding how it works, ensuring the method is not just robust, but valid.

6 AI DISCLOSURE

Artificial Intelligence (AI) technologies were used in the generation of images for use in the manuscript as noted. They were further used to correct any spelling errors in the final submission. All scholarly contributions, theoretical perspectives, and analyses were produced by the authors.

7 REFERENCES

- Baron, J. (2012). The point of normative models in judgment and decision making. *Frontiers in Psychology*, 3. <https://doi.org/10.3389/fpsyg.2012.00577>
- Bartholomew, S., & Jones, M. (2021). A systematized review of research with adaptive comparative judgment (ACJ) in higher education. *International Journal of Technology and Design Education*. <https://doi.org/10.1007/s10798-020-09642-6>
- Bartholomew, S., & Yoshikawa-Ruesch, E. (2018). A systematic review of research around adaptive comparative judgement (ACJ) in K-16 education. In J. Wells (Ed.), *CTETE - Research Monograph Series* (Vol. 1, pp. 6–28). Council on Technology and Engineering Teacher Education.
- Bell, D., Raiffa, H., & Tversky, A. (1988). Descriptive, normative, and prescriptive interactions in decision making. In D. Bell, H. Raiffa, & A. Tversky (Eds.), *Decision making: Descriptive, normative, and prescriptive interactions* (pp. 9–30). Cambridge University Press. <https://doi.org/10.1017/CBO9780511598951.003>
- Benton, T. (2021). Comparative Judgement for Linking Two Existing Scales. *Frontiers in Education*, 6. <https://www.frontiersin.org/article/10.3389/feduc.2021.775203>
- Bossaerts, P., & Murawski, C. (2017). Computational Complexity and Human Decision-Making. *Trends in Cognitive Sciences*, 21(12), 917–929. <https://doi.org/10.1016/j.tics.2017.09.005>
- Buckley, J. (2024). An Open-Source Adaptive Comparative Judgement App for Technology Education Research and Practice: Alpha Version. *Journal of Technology Education*, 36(1). <https://doi.org/10.21061/jte.v36i1.a.4>
- Buckley, J. (2026). *The OpenACJ Project*. <https://jeffbuckleyphd.com/openacj.html>
- Buckley, J., & Canty, D. (2022). Assessing performance: Addressing the technical challenge of comparing novel portfolios to the “ACJ-Steady State.” *PATT39: PATT on the Edge - Technology, Innovation and Education*, 523–537.
- Buckley, J., Canty, D., & Seery, N. (2022). An exploration into the criteria used in assessing design activities with adaptive comparative judgment in technology education. *Irish Educational Studies*, 41(2), 313–331. <https://doi.org/10.1080/03323315.2020.1814838>
- Buckley, J., & Lin, T.-J. (2024). Simulating adaptive comparative judgement sessions to determine the effect of judge disagreement on reliability. *PATT41: K-12 Technology and Engineering Education and Student Development*, 10–17.
- Gigerenzer, G. (2001). The adaptive toolbox. In G. Gigerenzer & R. Selten (Eds.), *Bounded Rationality: The Adaptive Toolbox* (pp. 37–50). The MIT Press.
- Gigerenzer, G. (2024). From bounded rationality to ecological rationality. In *Elgar Companion to Herbert Simon* (pp. 149–175). Edward Elgar Publishing. <https://www.elgaronline.com/edcollchap/book/9781800370685/book-part-9781800370685-17.xml>

- Gray, A., Rahat, A., Crick, T., & Lindsay, S. (2024). A Bayesian active learning approach to comparative judgement within education assessment. *Computers and Education: Artificial Intelligence*, 6, 100245. <https://doi.org/10.1016/j.caeai.2024.100245>
- Jones, I., Swan, M., & Pollitt, A. (2015). Assessing mathematical problem solving using comparative judgement. *International Journal of Science and Mathematics Education*, 13(1), 151–177. <https://doi.org/10.1007/s10763-013-9497-6>
- Kelly, K. T., Richardson, M., & Isaacs, T. (2022). Critiquing the rationales for using comparative judgement: A call for clarity. *Assessment in Education: Principles, Policy & Practice*, 29(6), 674–688. <https://doi.org/10.1080/0969594X.2022.2147901>
- Kimbell, R. (2011). Wrong ... But right enough. *Design and Technology Education: An International Journal*, 16(2), 6–7.
- Kimbell, R. (2012a). Evolving project e-scape for national assessment. *International Journal of Technology and Design Education*, 22(2), 135–155. <https://doi.org/10.1007/s10798-011-9190-4>
- Kimbell, R. (2012b). The origins and underpinning principles of e-scape. *International Journal of Technology and Design Education*, 22(2), 123–134.
- Kimbell, R. (2022). Examining the reliability of Adaptive Comparative Judgement (ACJ) as an assessment tool in educational settings. *International Journal of Technology and Design Education*, 32(3), 1515–1529. <https://doi.org/10.1007/s10798-021-09654-w>
- Kimbell, R., Wheeler, T., Stables, K., Shepard, T., Martin, F., Davies, D., Pollitt, A., & Whitehouse, G. (2009). *E-scape portfolio assessment: Phase 3 report*. Goldsmiths, University of London.
- Laming, D. (2004). *Human judgment: The eye of the beholder*. Cengage Learning.
- Larrick, R. (2004). Debiasing. In D. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 316–338). Blackwell Publishing. <https://doi.org/10.1002/9780470752937.ch16>
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving Skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429–434.
- Polanyi, M. (1958). *Personal knowledge: Towards a post-critical philosophy*. University of Chicago Press.
- Pollitt, A. (2012). Comparative judgement for assessment. *International Journal of Technology and Design Education*, 22(2), 157–170. <https://doi.org/10.1007/s10798-011-9189-x>
- RM Compare. (2023). *RM Compare*. RM Compare. <https://compare.rm.com/>
- Sadler, D. R. (2009). Transforming holistic assessment and grading into a vehicle for complex learning. In G. Joughin (Ed.), *Assessment, Learning and Judgement in Higher Education* (pp. 45–63). Springer.
- Seery, N., Buckley, J., Delahunty, T., & Canty, D. (2019). Integrating learners into the assessment process using adaptive comparative judgement with an ipsative approach to identifying competence based gains relative to student ability levels. *International Journal of Technology and Design Education*, 29(4), 701–715. <https://doi.org/10.1007/s10798-018-9468-x>

- Seery, N., Kimbell, R., & Buckley, J. (2022). Using Teachers' Judgments of Quality to Establish Performance Standards in Technology Education Across Schools, Communities, and Nations. *Frontiers in Education*, 7. <https://www.frontiersin.org/article/10.3389/feduc.2022.806894>
- Simon, H. A. (1990). Bounded Rationality. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *Utility and Probability* (pp. 15–18). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-20568-4_5
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22(2), 123–138. <https://doi.org/10.1007/s10648-010-9128-5>
- Thorne, S., Mentzer, N., Bartholomew, S., Strimel, G. J., & Ware, J. (2025). Learning by evaluating as an interview primer to inform design thinking. *International Journal of Technology and Design Education*, 35(3), 1013–1030. <https://doi.org/10.1007/s10798-024-09933-2>
- Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Review*, 34(4), 273–286. <https://doi.org/10.1037/h0070288>
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79(4), 281–299. <https://doi.org/10.1037/h0032955>
- Verhavert, S. (2018). *Beyond a mere rank order: The method, the reliability and the efficiency of comparative judgment* [Doctoral thesis, Universiteit Antwerpen]. <https://repository.uantwerpen.be/desktop/irua>
- Verhavert, S., Bouwer, R., Donche, V., & De Maeyer, S. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 26(5), 541–562. <https://doi.org/10.1080/0969594X.2019.1602027>
- Verhavert, S., Furlong, A., & Bouwer, R. (2022). The accuracy and efficiency of a reference-based adaptive selection algorithm for comparative judgment. *Frontiers in Education*, 6. <https://www.frontiersin.org/article/10.3389/feduc.2021.785919>
- Wheadon, C., Barmby, P., Christodoulou, D., & Henderson, B. (2020). A comparative judgement approach to the large-scale assessment of primary writing in England. *Assessment in Education: Principles, Policy & Practice*, 27(1), 46–64. <https://doi.org/10.1080/0969594X.2019.1700212>
- Whitehouse, C. (2013). *Testing the validity of judgements about geography essays using the adaptive comparative judgement method*. Centre for Education Research and Policy.
- Williams, P. J., & Kimbell, R. (Eds.). (2012). Special issue on e-scape [Special issue]. *International Journal of Technology and Design Education*, 22(2).
- Yellott, J. (1977). The relationship between Luce's Choice Axiom, Thurstone's Theory of Comparative Judgment, and the double exponential distribution. *Journal of Mathematical Psychology*, 15(2), 109–144. [https://doi.org/10.1016/0022-2496\(77\)90026-8](https://doi.org/10.1016/0022-2496(77)90026-8)
- Zhu, C., Buckley, J., Klapwijk, R., Spandaw, J., & De Vries, M. (2025). A holistic look at creativity: Evaluating pupils' creative design ideation and prototypes through comparative judgment. *International Journal of Technology and Design Education*. <https://doi.org/10.1007/s10798-025-10027-w>