

# The Validity and Reliability of Adaptive Comparative Judgements in the Assessment of Graphical Capability

*Dr Niall Seery, Jeffrey Buckley, Andrew Doyle and Dr Donal Canty*

*Department of Design and Manufacturing Technology*

*University of Limerick*

## **Abstract**

*The valid and reliable assessment of capability is of paramount importance in education. Operationalizing assessment practices of divergent problems can be particularly challenging due to the variety of potential responses. This paper investigates the use of Adaptive Comparative Judgements (ACJ) in the assessment of graphical capability. A cohort of undergraduate Initial Technology Teacher Education (ITTE) students (N=128) participated in this study which involved completing a design task and subsequently assessing the work of their peers through ACJ and criterion referenced assessment. The performance from both methods was analysed and identified a high level of reliability for ACJ. Correlations between the criteria scores and ACJ parameter values suggest its validity as an assessment mechanism, however they also present the potential for additional variable to be influencing holistic judgements.*

## **Introduction**

The ultimate aim of graphical education is the espousal of graphical capability. In an educational context however this capacity is often externalized through the medium of design where the fluid nature of the design process often makes it difficult to explicitly identify criteria. It is therefore important that the operationalization of assessment practices considers the overarching principles of graphical capability. Delahunty, Seery and Lynch (2012), through a review of the pertinent literature, offer a variety of aptitudes associated with graphical education which include cognitive capacities such as spatial cognition and deductive reasoning, communication skills such as modelling and graphicacy, designerly proficiencies such as ideation and problem solving, and suggest consideration for the pertinent knowledge base. While these skills are not mutually exclusive, for example modelling could also be conceived as a designerly act depending on the intent of the model, the broad categories form a conceptual model which facilitates in framing the principles of graphical capability. These core principles appear to be graphicacy (both communication and interpretation), design (having an understanding of the stages and functions of design, being innovative, and being able to externalise ideas), and the pertinent knowledge base (having a conceptual understanding of graphical principles), which are all underpinned by an architecture of cognitive abilities such as fluid reasoning and spatial ability.

The assessment of graphical capability under these core principles requires a mechanism which can appropriately reward capacity despite the inherent difficulty in the explicit observation of criteria. Sadler (2009) highlights two critical problems with the use of criterion referenced assessment in this situation in that the sum of the criteria scores may not always reflect the intuitive or holistic mark of the assessor, and that there may be criteria missing from an assessment rubric that are important or alternatively may set the particular work aside as exemplary. Additionally, making a judgment about a piece of work based on abstract or generic criteria can be quite difficult.

The use of Adaptive Comparative Judgements (ACJ) (Pollitt, 2012) however affords a mechanism which has previously been identified as a reliable approach for the assessment of graphically orientated conceptual design tasks (Seery, Lane, & Canty, 2011). Based on Thurstone's (1927) law of comparative judgement, ACJ can alleviate the issues with criterion based assessment identified by Sadler (2009) as it is operationalized by judges making binary judgments between two pieces of evidence. Multiple judgements on pairs of work ultimately result in the generation of a rank order of the work. The issues identified with individual judgment are avoided by having multiple judges assessing work thus nullifying personal biases. The reliability in the ACJ method stems from the adaptive nature of the software, in that specific pieces of work are selected as pairs for adjudication when additional judgements are needed to reach a consensus on their rank position. The ACJ method relies on a holistic judgment with overarching criteria used to guide the assessor in making a professional judgment (Kimbell et al., 2009). Perhaps the most significant aspect of ACJ lies in its capacity to facilitate adjudications on varying criteria. While a judge may base an initial judgement on certain criteria, subsequent judgements may be subjected to different criteria depending on the nature of the work.

Therefore, considering the capacity of ACJ to incorporate professional and holistic judgements, the primary purpose of this study is to examine its validity and reliability in the assessment of graphical capability.

## **Method**

A cohort (N=128) of undergraduate Initial Technology Teacher Education (ITTE) students in the 3<sup>rd</sup> year of their degree programme participated in this study as part of a Design and Communication Graphics (DCG) module. All participants had previously completed three prerequisite graphics education modules prior to this study. The focus of these modules was on developing an understanding of plane and descriptive geometry with a particular emphasis on developing competencies related to freehand sketching, parametric CAD modelling, technical drafting and conceptual design.

The initial phase of the study involved each of the participants engaging with a thematic conceptual design brief (Table 1). The brief required the participants to design an aid for an elderly person(s) to enhance their quality of life. No explicit criteria except for a size limitation on the final portfolio were incorporated into the brief. Instead students were required to evidence their own understanding of graphical capability.

**Table 1. Design brief utilized in the study**

---

Brief:

Population pyramids for many developed countries highlight the reality of an aging population. The inevitability of growing older brings with it many challenges to everyday activities. This calls for new and innovative thinking to enrich the lives of our elderly and ensure facilitation of the emotional, physiological, and social needs that guarantee an independent, dynamic and stimulated life.

Reinforcing the link between technology and society;

Design and model a personal device/artefact that will enhance the quality of life for an elderly person.

Criteria:

From a culmination of your knowledge and experience to date demonstrate evidence of graphical capability

---

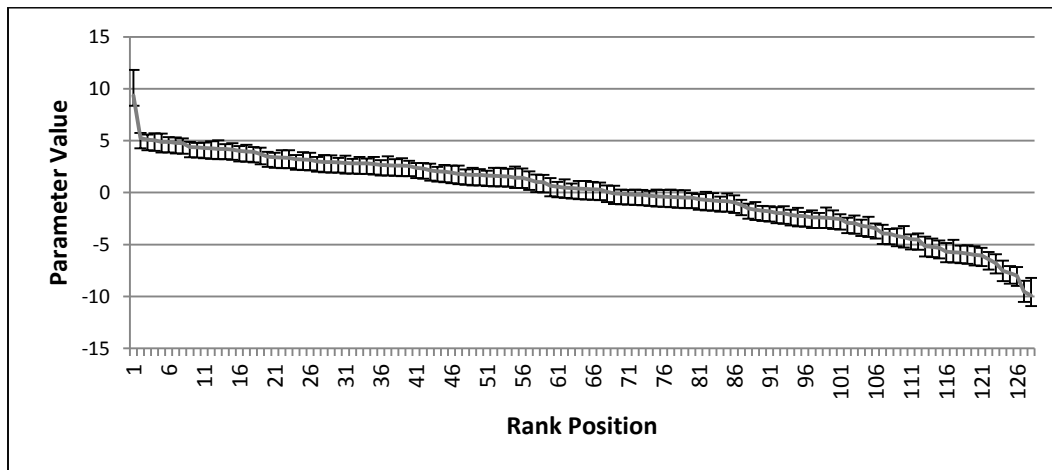
Upon completion of the design task, the second phase of the study required the participants to assess the portfolios using two methods. Initially, all participants assessed the work in an ACJ session. For this, participants each made 10 judgements on unique pairs of coursework. Participants were instructed to make judgements based on evidence of graphical capability. Finally, subsequent to the ACJ session each participant then graded a randomized selection of portfolios (mean = 14.67) on a ten point scale (1 = lowest, 10 = highest) under criteria aligning with the core principles of graphical capability previously discussed (Table 2). The average grades received for each portfolio under the individual criteria were derived as well as an average total score across all criteria to support comparisons with the ACJ data.

**Table 2. Grading system and codex used for data analysis**

Code	Criteria
Communication	Overall rate how effective the portfolio was communicated
Creativity	Rate how innovative or creative the design solution was
Stages	How well did the student define the stages of the design approach
Functions	Rate the selection of appropriate functions (i.e. was the use of CAD/sketching/etc. appropriate for the stage of the design that the student used them in?)
Principles	Rate the evidence that supports the level of the knowledge displayed of graphical principles

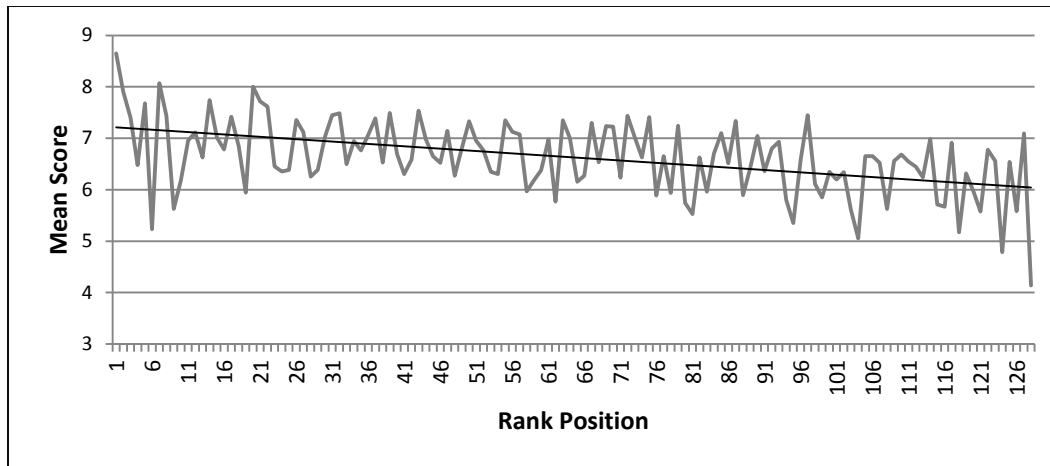
**Findings**

To analyse the data it was first necessary to elicit the performance rank created from the ACJ session. Each portfolio attained a specific parameter value based on the outcomes of the judgements it was involved in. The rank (Figure 1) illustrates a very high level of interrater reliability of 0.961.



**Figure 1. Portfolio parameter values and standard error bars indicating ACJ rank position**

Subsequent to this, a preliminary graphical analysis was conducted to observe any underlying relationships between the portfolios ACJ rank position and the performance on the grading criteria. This involved graphing the mean score achieved for each criterion against the rank positions. An example of this is shown below in Figure 2 which illustrates a positive relationship between the portfolios rank position and the average score achieved across all grading criteria. A similar positive trend emerged in all cases.



**Figure 2. Mean 'average score' score and ACJ rank position**

To examine these relationships more explicitly, a correlational analysis was conducted between average scores for all criteria and the parameter values achieved by each portfolio. All observable correlations were statistically significant at the  $p < 0.001$  level with moderate correlations ( $r = .403$  to  $r = .507$ ) emerging between the parameters values and grading criteria. Correlations between each of the grading criteria range from high ( $r = .760$ ) to very high ( $r = .956$ ).

**Table 3. Correlation matrix of performance variables**

	ACJ Parameter	Communication	Creativity	Stages	Functions	Principles	Average
ACJ Parameter	–						
Communication	.493**	–					
Creativity	.403**	.772**	–				
Stages	.484**	.863**	.760**	–			
Functions	.465**	.854**	.735**	.817**	–		
Principles	.504**	.872**	.764**	.847**	.933**	–	
Average	.507**	.940**	.867**	.923**	.943**	.956**	–

\*\* . Correlation is significant at the 0.001 level (2-tailed).

### Discussion and Conclusion

The results of this study are of particular interest in the assessment of graphical capability. The use of ACJ proved highly reliable through the achievement of an interrater reliability score of 0.961. This result corroborates the findings of Seery et al., (2011) who achieved a similar score.

With respect to the validity of ACJ, the high correlations amongst all of the grading criteria suggest that they are all aspects of the same construct which is posited to be graphical capability. However as only moderate correlations are observable with the parameter value, this presents a degree of misalignment which suggests that additional variables are contributing to rank position. As no criteria correlated excessively highly with the parameter relative to the others, this suggests one single criterion was not the sole focus of the judging cohort which aligns with the holistic nature of ACJ. It is posited that the grading criteria list is omitting critical elements associated with the task which would strengthen the correlation between the ACJ parameter and the average criteria score. This could take the form of additional variables or a bifurcation of the current variables. Ultimately it appears that ACJ has the capacity to validly measure the construct of graphical capability as biases towards specific elements are not present, however the question regarding the nature of additional variables impacting on its adjudication has now emerged.

## References

- Delahunty, T., Seery, N., & Lynch, R. (2012). The Growing Necessity for Graphical Competency. In T. Ginner, J. Hallström, & M. Hultén (Eds.), *PATT26* (pp. 144–152). Stockholm, Sweden: PATT.
- Kimbell, R., Wheeler, T., Stables, K., Shepard, T., Martin, F., Davies, D., ... Whitehouse, G. (2009). *E-scape Portfolio Assessment: Phase 3 Report*. London: Goldsmiths College.
- Pollitt, A. (2012). Comparative Judgement for Assessment. *International Journal of Technology and Design Education*, 22(2), 157–170.
- 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). Netherlands: Springer.
- Seery, N., Lane, D., & Canty, D. (2011). Exploring the Value of Democratic Assessment in Design Based Activities of Graphical Education. In *118th Annual American Society of Engineering Education Conference*. Vancouver, British Columbia: American Society for Engineering Education.
- Thurstone, L. L. (1927). A Law of Comparative Judgement. *Psychological Review*, 34(4), 273–286.