

Illustrating Educational Development through Ipsative Performance in Design based Education



Dr Niall Seery
Athlone Institute of Technology
KTH Royal Institute of Technology
nseery@ait.ie



Dr Thomas Delahunt
University of Nebraska-Lincoln
tdelahunt2@unl.edu



Dr Donal Canty
University of Limerick
donal.canty@ul.ie



Jeffrey Buckley
KTH Royal Institute of Technology
jeffbuckley1992@gmail.com

Abstract

The type of assessment implemented within an educational course has profound effects on the nature and depth of learning that students engage in. Typically there are two core types discussed within the pertinent literature. These are criterion and norm referenced assessment. The nature and impact of these modes of assessment have been explored within a variety of learning contexts. However, another and often overlooked form of assessment is ipsative enquiry. This refers to the comparison, by the student, of current performance to previous performance within a course of learning.

Ipsative assessment has been shown to increase motivational effects among students and promotes inclusivity in the learning process. Despite the potential benefits, its role in technology education contexts is under researched. This paper gives an overview of an ipsative approach to assessment that serves two functions. Firstly, to facilitate an opportunity for each student to develop a personal construct of what it means to be capable and secondly to provide a capacity to track their level of competence based gain both normatively and ipsatively.

The study tracks the performance of a cohort of student teachers ($N = 128$) in a core graphics module during a year three semester of an Initial Technology Teacher Education (ITTE) programme. Four consecutive graphical design tasks, focusing on the application of graphical principles, were designed to elicit core graphical skills and knowledge. An adaptive comparative judgment method (see Pollitt (2012) and Kimbell (2012)) was employed by the students to rank the responses to each task.

The paper highlights the potential of this approach in developing students' epistemological understanding of graphical and technological education, while tracking competence based gain through ipsative enquiry within the collective performance of their peers. Significantly, this approach demonstrates the capacity to track performance over time. The paper concludes with a discussion around the benefits of utilising ipsative assessment in design and technology education.

Keywords: Ipsative assessment, Personal constructs, Design education.

Introduction

This paper looks specifically at the need to synthesise perspectives on assessment as learning that enhance the depth of comprehension and development of Initial Technology Teacher Education (ITTE) students with respect to their discipline expertise. With the focus on discipline competency, there is a rationale and justification for enhancing feedback as defined by Nicol (2007) and its relationship with the expectations of higher education students as postulated by Sadler (2009).

The work of Nicol (2007) supports the view of actively engaging students in the assessment process through engagement with feedback. Nicol (2007) outlines the importance of active participation that supports engaging with diverse views on evidence of learning and the benefit of evaluating and rehearsing internalised knowledge and understandings when reviewing and preparing feedback. The responsibility is placed on the learner to comprehend the qualities and standards associated with expected performance. This enables a self-regulatory capacity that is critical to developing teacher expertise. Together with the benefits of engaging with feedback and formative assessment, students must become inducted into the assessment practices (Sadler, 2009). Sadler (2009) argues that the development of a conceptualisation of quality is critical to consistency in high achievement. Building on the work of Sadler (2009), this study was cognisant that the prescribed task design was meaningful and pre-planned to support students in evaluating the quality of evidence created, building appraisal skills, relating directly to learning experiences, and learning how to make judgments about the quality of emerging and finished works holistically rather than analytically.

Assessment as Learning

Without oversimplifying views on learning as either passive or constructed, what is important is that there is a shared view between the learner and the teacher. Agreeing to view learning as being constructed affirms not only the importance of the nature of feedback but also the way that it is received. Hughes (2011) highlights that teaching staff claim to give good feedback and yet students disagree. Involving the learner in creating feedback as much as they engage in constructing knowledge can satisfy the alignment of this action. Nicol and Macfarlane-Dick's (2006) seven principles of effective feedback support this participatory view of learner involvement. The function of assessment is multi-faceted and as such the discourse around defining the functions of assessment in response to a targeted agenda become critical. Assessment as a diagnostic tool supports direct and meaningful feedback in a way that is a pedagogical feed-forward. Within the scope of this paper, a macro view of assessment is adopted in an attempt to push the boundaries of implementations and interpretations so as to actively include the learner in the process of assessment (Sadler, 2009; Black & Wiliam, 1998; Yorke, 2003; Orsmond et al. 2000). This focus is critical within design and technology as the disposition of enquiry is central to developing capability. The value of assessment 'as' learning is not argued but there is a challenge in implementing an engaging and sustainable model that encourages self and peer appraisal as a self-regulatory act.

Adaptive Comparative Judgement (ACJ)

Boud and Falchikov (2007) identify a fundamental problem with the dominant discourse in assessment being the positioning of the learners as passive subjects to be measured or classified by the assessment acts of others. Seeing the learner as a 'passive subject' does not subscribe to the philosophy of design education. Within the context of design and technology education a participatory notion aligns well with a discipline area governed by

the synthesis of critical and speculative characteristics.

Kimbell et al. (1991) believe that due to the complex and integrated nature of design based activities, a model of holistic assessment that takes account of learning processes and interactions is the most effective in assessing the overall capability of students. Holistic assessment (as made possible by the ACJ pairs engine - see Kimbell 2012, Pollit, 2012) is the judgement of value of 'the whole' rather than the sum of a set of individual components of a task or assessment. The system has three elements; a set of portfolios of work from the students that is a response to the assessment task, a community of judges, and a 'pairs engine' – a software solution that dynamically selects pairs of portfolios and presents them to judges for adjudication on the quality of the work. As judges make decisions on pieces of evidence, the process begins with a rough sorting mechanism to establish broad categories of quality and evolves into an adaptive system that refines the position of the work on a rank order of capability as collectively determined by the group of judges.

The binary decision required when making pairwise comparison allows judgements to be made on an unarticulated recognition of qualities. This is an important feature of this study. It is the immediate and unqualified appraisal that requires the students to internalise their position, with direct reference to their submitted work and its comparison to the work of peers.

The flexibility in the ACJ approach requires that the judge call on a hazy, yet emerging 'construct of capability' to make a value judgement rather than being bound by fixed and predetermined criteria (Hager & Butler, 1996). For a novice (learner), the judging process forces the learner to build a conception of what it means to be capable and this can be achieved (at least in part) by exposure to exemplars and the breadth of interpretations and submissions. These qualities must be internally processed by the judge, based on personally set (externally influenced) criteria and standards.

The comparative judgment approach (ACJ) presents a mechanism to support the tracking of a learners performance gain over a series of tasks/activities and has a number of significant advantages when considered within a series of learning tasks:

- It allows the tracking of performance over time
- Can directly compare performance to the norm (for reference)
- Requires active participation
- Supports the development of appraisal skills
- Links appraisal to defining capability
- Directly links feedback with learning task

While ACJ has been applied in several contexts and empirical benefits have included enhanced participation among students, it has not been applied to observe an individual's personal gain over time. As such, there is a need to explore its implementation within education from a perspective of Ipsative enquiry.

Ipsative Development

The important characteristic of ipsative assessment is that it focuses on a learner's performance relative to a previous individual performance. This concept of reference to the self makes explicit the learners progress. This form of feedback qualifies the development over time in response to a target or goal. In practice, the learners'

performance in an initial task is then compared to a subsequent task in a sequential process of feedback and action. Developmental tasks can be intermediate and also examined through an Ipsative perspective. The benefits of an ipsative approach are multifaceted. From a learner perspective there is a motivational value in clearly seeing progress and development relative to a previous achievement or understanding (Hughes, 2011). From a teacher's perspective, it illustrates a macro view of performance over time and allows a diagnostic scheme to structure learning with defined expectations and/or pathways for individual learners.

By its nature, ipsative assessment supports a cumulative understanding of performance developed over sequential tasks and activities. Tasks that support ipsative assessment tend to be designed to reflect governing principles and a macro view of capability, therefore allowing the learner to compare performances, while honing associated attitudes, skills and knowledge. Ultimately the approach supports a formative agenda that is largely self-regulated (Hughes et al., 2014). The constructivist view of learning further supports an ipsative approach that facilitates the unmediated feedback that compares performance to previous performance utilising a "feed-up" lens on the part of the students (Hughes, 2011). Here students are facilitated in reflecting on their own personal goals in reference to their cumulative performance to date.

Ipsative assessment also facilitates an effective synthesis of feedback and feedforward processes (Hughes, 2011). It increases the likelihood of the learner acting on feedback, as it is personally relevant and benchmarked with previous performance. However, the capacity of the individual to unpack performance deficits may be outside the capacity of a novice learner, especially if there is no reference to expected progress either absolute or normative. Therefore, the ipsative agenda must be critically managed and articulated. As outlined by Hughes (2011) one of the challenges with ipsative assessment is in operationalising it. The requirement to design and implement a sequence of learning tasks or activities can be difficult within a modularised and semester based system, usually constrained by 13 weeks of teaching. Additionally, providing feedback that is timely, relevant and bespoke to each learner on all tasks is impractical and further inhibits its implementation.

Method

Approach

The method was designed to explore students' development over time by synthesising feedback and appraisal through an Ipsative model for assessment as learning. The study employed four tasks over a 12 week semester, all equispaced to observe students' gain over time. Students responded to similar design tasks and then using the ACJ method produced a rank order of performance and generated peer feedback on the quality of the work observed. Although there is significant qualitative data that supports the value of an ipsative approach, the focus of this paper was to initially explore a quantifiable variance (if one existed) in performance, as a precursor to exploring the evidence from a qualitative perspective.

Participants

The study engaged Year 3 ITTE students ($N = 136$). Natural attrition accounted for 8 students and therefore the final study cohort consisted of 128 students. The cohort was divided into quartiles ($n = 32$) based on performance on the first task to support the statistical analysis of the data by providing an initial baseline of performance for which

subsequent tasks can be compared to from an ipsative perspective.

Design

The four design tasks were governed by core graphical design principles (graphical communication, innovation, stages of design, and functions of design), common to all tasks and the conceptual design agenda mirrored the nature of the tasks used for the national assessment in the participants subject discipline of Design and Communication Graphics (DES, 2007). This allowed for comparison with respect to overarching skills and competencies.

Implementation

Following instruction on core descriptive geometry, students were given a design task and two weeks to complete. The design task was an opportunity to apply the newly acquired knowledge to resolve a design problem. Following submission, students used the ACJ engine to complete a pair-wise comparison of peer submissions. This immediately exposed students to the work of their peers and as such saw a breadth of interpretation and solutions. On average, students made 17 comparisons of work during the ACJ assessment activity. This activity was immediately followed by the next design task and this approach was repeated for the final tasks. For all assignments the rank to grade conversion was validated by the module leaders as an independent task to ensure validity of performance.

Findings

The peer assessment of each of the four assignments recorded interrater reliability scores of 0.974, 0.973, 0.965 and 0.971 respectively. This demonstrated that there was a shared understanding of what was of value when appraising the evidence from each of the tasks.

The mean score for Assignment 1 (54.93%) compared to Assignment 4 (76.56%) illustrate an increase in performance over the course of study. Although this is welcome, it is important to consider how this incremental improvement manifested with varying ability levels. To explore the relationship between the performances in each of the four tasks across the quartiles, a multivariable analysis of variance (MANOVA) was employed using the assignment as independent variables and the quartiles as dependant variables. The results indicated a statistically significant difference in assignment scores based on quartiles, $F(12, 320.427) = 21.362$, $p < .000$; Wilk's $\Lambda = .211$, partial $\eta^2 = .405$. To determine how the dependant variables (quartiles) differ for the independent variable (assignment performance), Tukey's Post Hoc test was used. Figure 1 illustrates the consistent statically significant difference that existed between assignments 1 and 2 (with the exception of quartile 3) and between 3 and 4, with no difference in performance between assignment 2 and 3 recorded for all baseline quartiles. Although the design of the assignments in principle and the associated transposition of the rank to grade were constant, task design could account for the differences between assignments. However, the pattern was reflected in all quartiles.

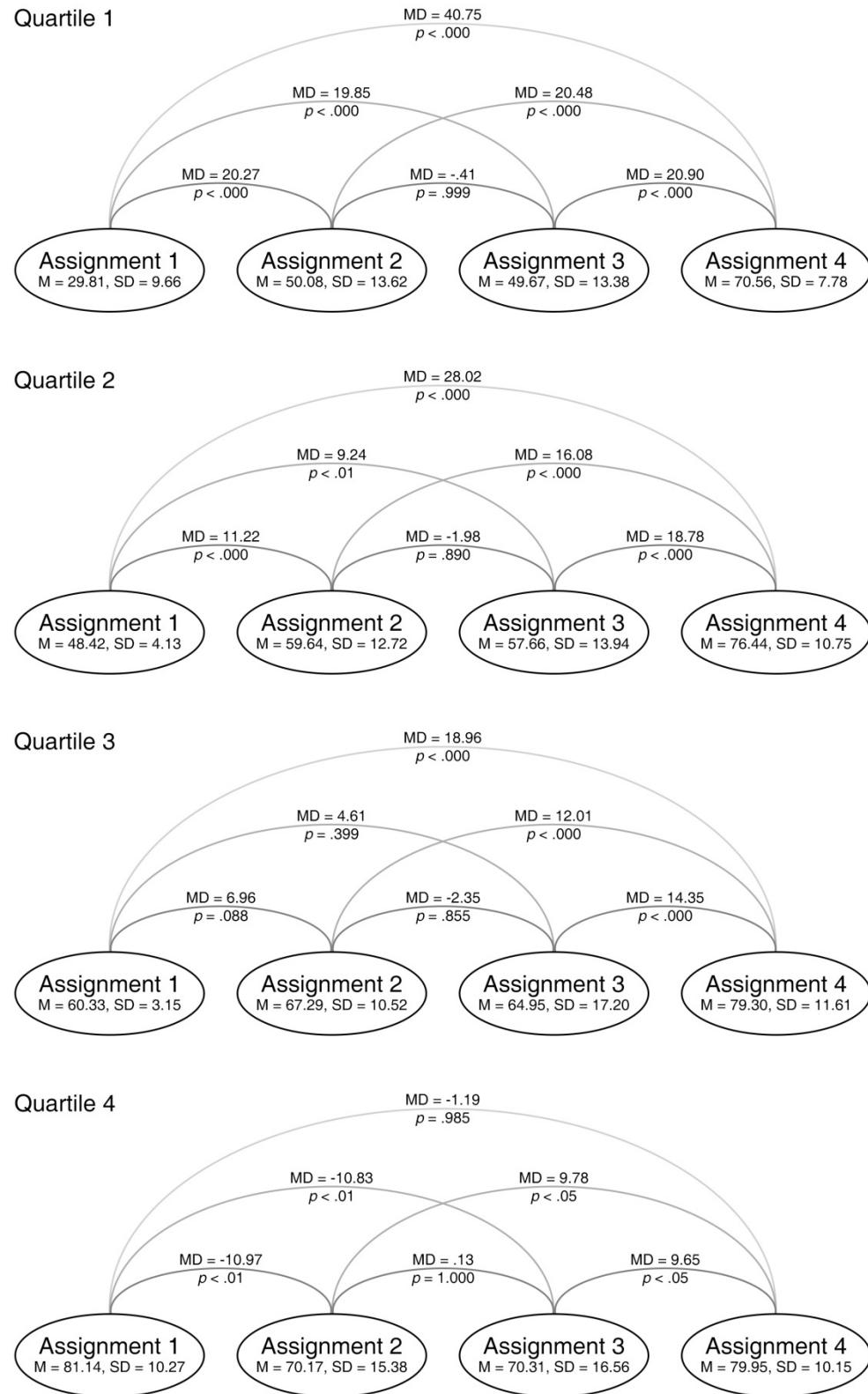


Figure 1. Tukey's HSD post hoc test results

The tracked improvement for the quartile 1 group shows a gain of 40% between assignment 1 and 4. While quartile 4 decreased by 1% over the same tasks. It is important to note that for quartile 4, the initial mean score of 81.14% meant there was limited capacity to increase relative to the other quartiles. From Figure 1 the comparison within groups (Horizontal) demonstrates an ultimate gain for quartiles one, two and three. While quartile 4 saw a decrease in mean score, it is posited that this may still represent development as relative to Assignment 1, Assignment 4 would have had inherently higher expectations. The comparison between groups (Vertical) also suggests that the gain for all performance levels was also relatively constant.

From an ipsative perspective, what is interesting is that from the baseline definition of quartiles, how much variance occurred by definition of quartiles as the module progressed – i.e. did students move in to other quartiles? To examine this, the quartiles determine from Assignment 1 were compared with quartiles derived from the results of Assignment 4. The results of this are presented in Table 1. Further analysis is required to examine the statistical significance of the movement however it is interesting from the overview to see the scope of the movement. Of particular interest is Q1 which saw substantial movement to Q2 and Q3 however only one student ultimately moved to Q4. A similar trend is seen in reverse for Q4 whereby most movement is into Q2 and Q3 with only three students ending in Q1.

Table 1. Student movement between quartiles

Initial Quartile	Final Quartile			
	Q1	Q2	Q3	Q4
Q1	10	14	7	1
Q2	6	10	11	5
Q3	4	9	9	10
Q4	3	9	9	11

Discussion

Reflecting much of the literature on ACJ (Kimbell, 2012; Seery & Canty, 2017; Seery et al., 2012) the data here once again demonstrated the value of the ACJ strategy as a powerful means of establishing what is of value as determined by a community. In this study, the high level of consensus consistently across the four tasks demonstrated the capacity of the community to agree on qualities and standards in an unarticulated way. The alignment of the high interrater reliability scores with the increasing performance across the tasks also highlights the incremental development of the students as a cohort on the sequence of assignments.

Synthesising the creation of evidence with its appraisal engaged students in a double looped system of reflection in action. Not only were students critiquing their own work relative to the breadth of work presented by their peers, they were also engaged in the ‘first principles’ critique that considered the purpose of the design task with respect to core competency development. The identification of qualities that aligned with capability could be observed during the judging process. As students were engaged with the recurring evidence as presented by their peers, it is posited that they developed a better capacity to discriminate standards. This hypothesis is supported by the increase in performance

between Assignment 1 and Assignment 4 for the entire cohort. This engagement is often only visible to a grader at a summative stage of the process.

Two significant elements of this approach are noteworthy. Firstly, students by virtue of their ranked position got a normative indication of their performance relative to the entire cohort. Additionally, there was an immediate context for this performance as students apprised a number of their peers work. The immediate feedback was therefore both situational and contextual. Secondly, the focus on appraisal (which possibly also became more sophisticated over the four tasks) engaged students in determining the critical aspects of capability and formed a dialogic creation of individual constructs of capability. This discussion with self was mediated by authentic evidence that exemplified varying qualities and standards. The exposure to peer work was immediate and unqualified, requiring that each student critiqued evidence as the basis of constructing or refining his or her own construct of capability. In essence this unmediated feedback acted as an explicit feed-forward.

Interestingly, the impact of the process was variable across the cohort. Students who initially performed poorly (bottom quartile) showed a significant gain over the course of the four assignments. This is also evident by their movement into higher quartiles by the end of the semester. The middle quartiles (Q2 and Q3) had a more general dispersal however still saw an overall increase in performance. Q4, while initially the top quartile, was the only group to see a ultimate decrease in performance. Many students also ended up in lower quartiles. It is posited that this is a result of having less capacity for improvement and therefore this group may have required a different education experience which offered more of a challenge than the other groups.

Conclusion

The study presents evidence of a performance gain that was both ipsative and normative. The unmediated feedback formed by the exposure and appraisal of peers work supported a reference for ipsative development. The identification of qualities and standards provided explicit exemplars as targets. The capacity to interpret unmediated evidence of capability and discriminate between qualities and standards although beneficial may have proven more difficult for the less able students and as a resulted limited their capacity to perform in the upper quadrants of performance. Although, beyond the scope of this paper further research is required to explore the individual level ipsative gain and unpack the associated qualitative data.

References

- Boud, D. & Falchikov, N. (2007). *Rethinking Assessment in Higher Education: Learning for the longer term*. London: Routledge.
- Black, P. & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education*, 5(1), 7-74.
- DES. (2007). *Leaving Certificate Design and Communication Graphics Syllabus*. Dublin: The Stationery Office.
- Hager, P. & Butler, J. (1996). Two Models of Educational Assessment. *Assessment and Evaluation in Higher Education*, 21(4), 367-378.
- Hughes, G. (2011). Towards a personal best: a case for introducing ipsative assessment in higher education. *Studies in Higher Education*, 36:3, 353-367.
- Kimbrell, R. (2012). The origins and underpinning principles of e-scape. *International*

- Journal of Technology and Design Education*, 22(2), 123-134.
- Kimbell, R., Stables, K., Wheeler, T., Wosniak, A. & Kelly, V. (1991). *The Assessment of Performance in Design and Technology*. London: School Examinations and Assessment Council/Central Office of Information.
- Nicol, D.J., and MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*, 31:2, 199-218.
- Orsmond, P., Merry, S. & Reiling, K. (2000). The Use of Student Derived Marking Criteria in Peer and Self-assessment. *Assessment & Evaluation in Higher Education*, 25(1), 23-38.
- 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). The Netherlands: Sense Publishers.
- Seery, N., and Canty, D (2017). Assessment and Learning: The Proximal and Distal Effects of Comparative Judgment. *Handbook of Technology Education*, Springer. DOI: 10.1007/978-3-319-38889-2_54-1.
- Seery, N., Canty, D., & Phelan, P. (2012). The validity and value of peer assessment using adaptive comparative judgement in design driven practical education. *International Journal of Technology and Design Education*, 22(2), 205-226.
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education*, 45(4), 477-501.