



RM[™]



Transforming Assessment

Using adaptive comparative judgement to assess KS3 art & design at scale across school groups

Case Study | E-ACT, Merrick Ed

Background

Victoria Merrick is a leading UK based school improvement consultant who has worked with school groups in the UK and beyond for many years.

The E-ACT Academy Trust consists of 29 primary and secondary schools covering a large geographical area in England. They have a particularly strong focus on innovation and enabling its academies to collaborate and share ideas in ways that standalone schools may not be able to.

“When Vicky told me we were going to use software to assess artwork, I thought - this is going to awful! But you were right, this is a game-changer. I cannot believe how simple, quick and intuitive it was. And the reliability score is better than we'd expected!”

Art & Design teacher

Introduction

The UK Key Stage 3 National Curriculum offers some direction for mid-key stage assessment of creative and performance-based subjects, such as Art & Design. Programmes of Study and assessment objectives provide a framework for evaluating student progress towards the key stage's goals.

However, the National Curriculum's focus lies primarily on content, offering limited guidance on specific assessment practices. As such, schools and MATs across the country are independently seeking to address this challenge, looking beyond the National Curriculum to develop bespoke assessment processes, establish clear criteria and employ time-consuming standardisation and moderation practices.

With competing improvement priorities, challenges with pupil absence, behaviour and recruitment, as well as limited time and resources afforded to schools to effectively resolve these challenges, the reality is that reliable assessment of creative and performance subjects, at scale, slips off the radar.



The Challenge

Assessing Key Stage 3 Art & Design reliably and at scale presents a unique set of challenges for all schools. Unlike subjects with objective answers, creative and performance subjects, such as Art & Design, rely on judgements of subjective qualities, reliant upon individual interpretation of pre-determined criteria. This inherent subjectivity can lead to inconsistencies in assessment, a problem which, when working across groups of schools (such as Multi-Academy Trusts (MATs)), is amplified immeasurably.

As cohorts move through from Key Stage 3 to Key Stage 4 and consider embarking upon GCSE courses, they refer to data and feedback from assessments to help inform their decision making. Teachers use data and feedback from assessments to guide and support their pupils in this process. Parents and carers rely on data and feedback on pupil progress reports to gain insight into achievements, aptitude and potential within a subject area. Some might suggest that mid-key stage assessments are not high stakes, but to those navigating their way through, this is all the information they have and they feel the pressure of making informed and confident decisions.

Moderation plays a critical role in Art & Design education at Key Stages 4 and 5 as a mechanism to ensure both fairness in assessment and the maintenance of national standards. But to replicate this time-intensive (and somewhat flawed) process, using limited internal resources at Key Stage 3 is not a realistic approach. Schools have no option but to reluctantly accept that, until now, there has simply been no valid and reliable way to measure, monitor and track standards in Key Stage 3 Art & Design.

What if there was another way?

In March 2024, Art teachers from E-ACT, a large multi-academy trust of academies spread geographically across the UK, sought to rewrite this narrative as part of a wider Key Stage 3 Assessment Strategy redevelopment project. Working with Victoria Merrick, Merrick-Ed Limited, they devised a small-scale pilot of RM Compare, an Adaptive Comparative Judgement (ACJ) tool.

After an introduction to comparative judgement as an assessment methodology, and some time to explore its implications, the E-ACT Art teachers agreed to have a sample of Year 7 pupils complete a set piece of work from a centrally defined stimulus. The stimulus provided pupils with the opportunity to demonstrate their understanding and application of four of the formal elements of art: line, shape, tone and texture, which were found to be common across all academies' Year 7 curricula.

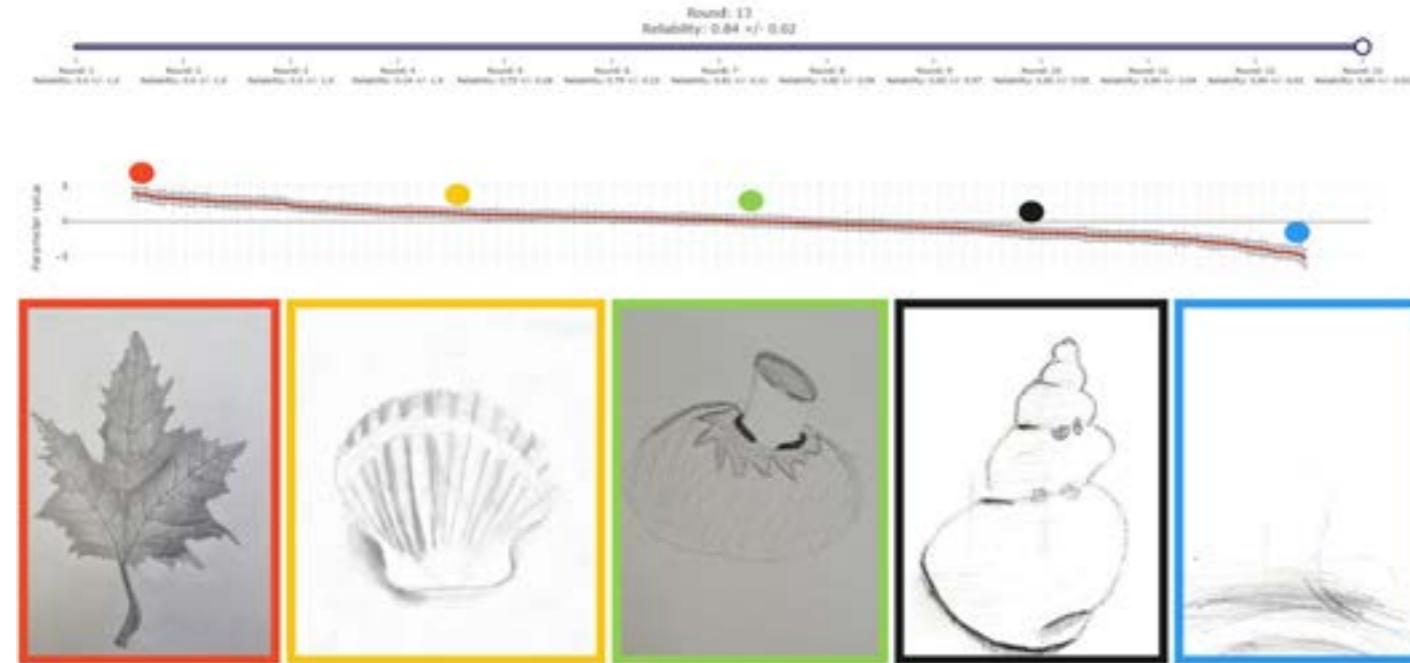
6 academies took part in the trial, submitting 272 pieces of work (items) produced by pupils across 6 mixed ability Art & Design groups.

Using the RM Compare online platform, 15 Art teachers (judges) were presented with two pieces of work, side-by-side, and asked to compare the pieces, against a holistic statement. The judges selected which was the better piece and were then presented with two new pieces to compare. This process was repeated multiple times by multiple judges to ensure a high level of reliability.



Rapid reliable assessment: Key Stage 3 Art Trial

Once all artwork had been completed and submitted, the session opened. Across a 5-day window, teachers were accessing the platform at a time that suited them to complete their judgements. Each were allocated 119 judgements to complete, from across the sample. Items were anonymised, meaning teachers did not know which pupil or which academy the artwork belonged to.



After 13 rounds of assessment (meaning each item was judged 13 times, by multiple judges) the reliability score exceeded expectations, 0.84 (+/- 0.02), which indicates a very good reliability (a high level of repeatability between measurements). The average decision time for comparing a pair of items was 15.4 seconds, equating to an average total judgement time of 31 minutes per judge - with no additional time required for standardisation and moderation, as this is all baked into the process.

Benefits

The trial confirmed, without doubt, that Adaptive Comparative Judgement offers academies and groups of schools a solution to the previously accepted problem of achieving reliable assessment of creative and performance subjects, at scale.

It offered teachers the opportunity to work together as one pool of judges, combining judgements and opinions to reach one statistically reliable rank order, from which inferences about pupil performance in that aspect of the curriculum can be formed. As well as including valuable standardisation and moderation practices, without the onerous and time consuming meetings before and after an assessment window, the process developed teachers' tacit knowledge.

Teachers reported increased confidence in assessment outcomes and their own assessment practice and felt assured that issues of bias and unconscious bias were being addressed through the anonymised assessment process.



The future

E-ACT intends to scale up its trial to complete a whole cohort (~2000) end of Year 9 assessment in summer 2024, using the ACJ methodology.

There is further development work to do to review how and when to schedule trust-wide benchmarking assessments, and which components of the curriculum to sample for assessment.

Beyond Art & Design, there is scope and ambition to scale this approach further, to incorporate other creative and performance subjects, such as Drama, Music, PE, Design & Technology as well as exploring potential for reliable assessment of wider learner attributes and skills (collaboration, problem solving etc).

“At RM, we take pride in working with our customers to innovate. This initial research provides an exciting demonstration of how to tackle some of our education systems challenges. We look forward to driving this work forwards”

John Baskerville, Managing Director,
RM Assessment



**RM and Merrick Ed are working
with educators around the world**

**Interested in finding
out more?**

Contact us for a conversation

Or visit compare.rm.com

