Quality Metrics Final Report

Quality Metrics National Test

CultureCounts | QUALITY METRICS NATIONAL TEST

by John Knell & Alison Whitaker

Supported using public funding by ARTS COUNCIL ENGLAND

CONTENTS

CONTENTS

EXECUTIVE SUMMARY

The Key Objectives of the Quality Metrics National Test (QMNT)

The key objectives of the Quality Metrics National Test (QMNT) were to:

- Recruit and support 150 National Portfolio Organisations (NPO) and Major Partner Museums (MPM) to use the quality metrics and Culture Counts platform to evaluate three events, exhibitions or performances between November 2015 and May 2016.

- Recruit and support 10 NPOs to refine and test a set of participatory metrics developed through a previous trial, assessing their alignment with the CYP Quality Principles

- Produce an anonymized aggregated dataset and public facing report providing an analysis of the information collected throughout the QMNT highlighting key trends, points of comparison across the participating organisations, and the most important insights from the project

- Produce a separate public facing report for the participatory metrics strand of the trial highlighting key insights from the project and documenting a revised set of participatory metrics in alignment with the CYP Quality Principles

With support from Arts Council England, Culture Counts put out a call for expressions of interest from organizations to take part. During October 2015, 150 National Portfolio Organisations (NPOs) and Major Partner Museums (MPMs) signed up to test the quality metrics.

The Evaluation Activity Undertaken

During the available timeframe, 418 evaluations were conducted by the NPOs / MPMS, of which 374 evaluations used the quality metrics (as defined in Table 1). The analysis of the 24 evaluations carried out using the participatory metrics is detailed in a separate report.[1] Twenty evaluations were excluded from the trial as they were either incomplete at the data cut-off date or had not used the quality metrics.

If the 150 participating NPOs and MPMS had each reached the targets set for them by the EOI conditions the cohort as a whole would have completed 450 successful quality metric evaluations; based on 13,500 public responses; 450 self assessments, and 2,250 peer assessments.

1 Knell and Whitaker. "Participatory Metrics Report." *Arts Council England (2016)*

The overall outcomes achieved by the participating NPOs and MPMs, against those aspirational targets, were as follows:

374 successful quality metrics evaluations:	(83% of target of 450)
1,358 self assessments:	(302% of target of 450)
921 peer assessments:	(41% of target of 2,250)
19.8K public responses:	(147% of target of 13.5K)

Throughout the life of the project we saw 137 'active' organisations within the Quality Metrics National Test. Eight of the organisations who signed up through the EOI process did not engage at all with Culture Counts and the process, with the remaining 5 'inactive' organisations proving unable to complete any successful evaluations. Taken as whole this means that 91% of the cohort of NPOs and MPMs fully and successfully engaged in the Quality Metrics National Test.

Table 1: Quality Metrics

DIMENSION	STATEMENT	RESPONDENT TYPE		
		SELF	PEER	PUBLIC
Concept	It was an interesting idea	✓	✓	✓
Presentation	It was well produced and presented	✓	✓	✓
Distinctiveness	It was different from things I've experienced before	✓	✓	✓
Captivation	It was absorbing and held my attention	✓	✓	✓
Challenge	It was thought-provoking	✓	✓	✓
Enthusiasm	I would come to something like this again	✓	✓	✓
Local Impact	It is important that it's happening here	✓	✓	✓
Relevance	It had something to say about the world in which we live	✓	✓	✓
Rigour	It was well thought-through and put together	✓	✓	✓
Risk	The artists/curators were not afraid to try new things	✓	✓	-
Originality	It was ground-breaking	✓	✓	-
Excellence	It is one of the best examples of its type that I have seen	✓	✓	-

The Evaluation Data

In addition to the core quality metrics data collected in this QMNT, basic demographic data (age, gender and postcode) was collected for public respondents. In addition, metadata was also assigned to responses or events accordingly.

Our overall approach to constructing metadata was designed with highly sensitive aggregate analysis in mind in order to tell a clear overall story of the top-line aggregated results which does not over simplify the findings. In order to manipulate each data point (i.e. one answer to one question – any individual answer to any individual question) individually with any other, and to create any combination of group comparisons that are meaningful, metadata needs to be assigned to each question response.

The Culture Counts platform does this by design. For this project, we collected additional metadata to that collected via the quality metrics surveys, specifically organisation data, event data and geomapping data.

Structured data for location was applied to each event so that events in particular regions could be viewed in the aggregate, with granularity maintained by postcode for geomapping enabling versatile groups for analysis on a national scale.

Event Choice and Location

Choosing events for the trial was in the hands of the participating cultural organisations. The freedom to choose which events would be suitable to test a new approach to evaluation was important; the nature of many cultural organisations is to approach their work in a unique way and one of the objectives of the QMNT was to see how well this evaluation methodology could support this instinct.

Key Findings

What do self, peer and public responses tell us about the overall quality of work in this evaluation?

The dimension scores for individual organisations, or in aggregate, are not a clapometer, in which a successful piece of work has to be seen to score highly on every single dimension. Where self prior scores (capturing their intentions / expectations for the work) are in close alignment with self post, peer and public responses then the work is delivering against the organisation's creative expectations. Are the cultural organisations in this study adept at making these judgements in alignment with peer and public response? What are the risk and originality profiles of the work they are producing?

Taken together the aggregate results suggest:

- The work presented and analysed in this study received a broadly positive response from peer and public respondents, and largely met the (quite high) prior creative expectations of the creative teams involved in its production (self assessors)
- When it comes to measuring the quality of a cultural experience (for self, peer and public respondents) three dimensions in particular - challenge, distinctiveness and relevance – in the aggregate tended to score lower than the other six dimensions
- The clustering of self, peer and public responses in relation to these metrics suggests that audiences are adept at assessing them, with their judgements showing broad alignment with self and peer responses.
- The participating cultural organisations largely met their creative intentions, as measured by the degree of alignment between their self prior scores for each dimension and the corresponding aggregate scores for peer and public respondents
- Peer responses (as we have seen in all previous evaluations) are consistently lower across all dimensions than self and peer responses

Risk, Originality and Excellence as measured through self and peer aggregate responses for all evaluations

Risk

The aggregated self responses across these three dimensions (risk, originality, and excellence) show that self assessors tend to score themselves more highly than peer assessors; a well-established trend in previous and ongoing evaluations using the quality metrics. Interestingly, at an aggregate level the self assessors perceive themselves to be taking quite high levels of risk (broadly supported by peer scores, which are highest for this metric out of the three). This is encouraging to the extent that it would suggest that taken as a whole the cultural organisations in this study are seeking to stretch themselves with the work they are producing, and that they have a well-developed appetite for creative risk. Another interesting finding is that there is a noticeable variation in self prior risk ratings by artform. This suggests that as more data is gathered across artforms about perceived risk, this could be used to provoke dialogue both within artforms, and across artforms, about what constitutes creative risk. One outcome might be that 'risk' as a dimension measure for self and peers is thickened up with additional metric components.

Originality

Originality is the lowest ranking dimension aggregate score for both peer and self respondents. This would suggest that at an aggregate level self and peer respondents did not consider the work being evaluated in this study to display high levels of originality. Is this a surprising assessment?

The bar is set high by the originality metric, with respondents asked to express their relative support for the notion that the work 'was ground-breaking.'

How Do The Regions Compare?

The analysis cross referenced the location data for each event analysed against the ONS classification[2] for rural and urban areas (which is based on a six-fold categorization moving from strongly rural (rural 1) to strongly urban (urban 6).

To what extent did evaluations in rural and urban areas attract different profiles in terms of dimension scores? Across six of the dimensions (concept, presentation, distinctiveness, rigour, relevance, challenge, and captivation) there is no significant variations in the profile of public dimension scores as you move from the most rural areas (rural 1) to the most urban (urban 6). In other words, distinctiveness is not being rated much higher in urban as opposed to rural areas.

For two of the dimensions, enthusiasm ('I would come to something like this again') and local impact ('it is important that it's happening here'), the differences between public responses in rural and urban areas are of particular interest. Given the differential access to cultural provision in rural as opposed to urban areas, one might expect public ratings for 'enthusiasm' in more rural areas to be high, and they were (higher than in other urban areas). Similarly, one would naturally hypothesise that local impact scores would also attract high public ratings in rural areas and where this was true in some rural areas compared with urban, the rural status alone does not have a specific influence on local impact scores in all cases.

Differences in the data by artform

As one might intuitively expect, different artforms do have distinctive dimension profiles, but this only becomes clear when detailed artforms are considered in the their own right. The variations existing for each artform could be for a variety or reasons, not least the particularities of the work evaluated in this QMNT. Exploring these differences requires dialogue and debate within and across artforms.

Immersive work presents a good example. Interestingly for work defined as 'immersive' in this study, the public and peer ratings were much higher than the aggregate average for challenge, distinctiveness and relevance (the 'lower' scoring aggregate dimensions across all evaluations). With more data it would be interesting to explore whether 'immersive' work is a consistent inflator in peer and public ratings across particular dimensions.

2 http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/2011-rural-urban/index.html

Cohort Engagement, Insights and Data Culture

The evaluative and data culture of the participating organisations as judged by support requests

In addition to Learning and Insight days, much of the contact with the cohort was via in dashboard, phone and email support. In broad terms, just under half the cohort required what we would label as 'high' levels of direct support (in other words the cultural organisations in this group initiating multiple calls and emails across the evaluation period). Around a third of the cohort required moderate levels of support (initiating some calls / emails). Just under a fifth required low levels of support (initiating a few calls / emails).

There are of course some subtleties here, in that some organisations made lots of contact with Culture Counts because they were trying to do imaginative things with their evaluations, as opposed to needing 'hand holding' support to carry out a basic evaluation. Nonetheless, overall it is true to say that the majority of organisations seeking the most help were those with lower capacity and / or evaluative expertise. When cross-referenced with well-configured and executed evaluations, these findings are consistent with our previous depiction of the cultural sector in England as being 80% data shy; 15% data ready; and 5% data driven.[2]

Engagement led to improved evaluation practice and outcomes

What also became very clear during the short span of this study (the majority of evaluations took place in a concentrated period between January and May 2016) was that as the participating organisations grew more familiar with the quality metrics; and the self, peer, public triangulation approach embedded within and facilitated by the Culture Counts platform; this in turn led to more accomplished evaluative practice and better outcomes, as evidenced by:

- Culture Counts observing more accurately configured evaluations (i.e. no mistakes in configuration or in URL attribution for self, peer and public responses) as organisations moved through the project
- A declining number of support calls on how to set up evaluations in the Culture Counts dashboard
- More interest from the cohort in 'value adding' activity, such as creating separate URLs for audience sub groups, and collaborating / sharing data with other organisations
- Improving data collection outcomes in terms of the total number of public responses being achieved by the participating organisations

3 http://artsdigitalrnd.org.uk/wp-content/uploads/2014/06/HOME-final-project-report.pdf

Challenges, Issues and Opportunities identified by the cohort

Culture Counts received feedback on the challenges being encountered by participating organisations both through their direct support request and calls; and through their contributions at the the Learning and Insight sessions. The key points raised were:

Peer Management, Engagement and Building a Peer Community

Planning and managing the peer process was a new experience for all of the participating organisations and represented the greatest challenge in successfully completing their evaluations. Not only do organisations have to select their peers, but they then have to invite them and secure their participation, and follow through with peers checking that they have attended their event and completed their evaluation. Some organisations also mentioned that depending on the location of the particular event the distance required to travel for some productions is a barrier to obtaining peer assessors[3].

The paradox around the peer review element was that whilst it was the most demanding element of the evaluation process, it was also seen as a very positive aspect of the Quality Metrics National Test. Participating organisations welcomed the opportunity to invite their own peers. Strong support was also expressed for the idea that as a result of this national test a peer database is formalised across the arts and cultural sector – in other words support is given to create an open searchable data base of peers for the sector to draw on – in which each peer could list their artform expertise and interests.

Enhancing Peer Continuing Professional Development

One clear potential deficit in the current process identified by some of the participating organisations was that having secured the engagement and participation of their peer evaluators, they had received feedback from peers that they had found the process 'too short'. Peers would have happily answered more questions and would have welcomed more discussion around the results. Clearly, the length of the quality metrics question schedules has no bearing on how the peer community is engaged around the results. Even with the current peer dimensions, and additional open questions, organisations could choose to bring their peer evaluators together on a conference call to discuss their opinions and evaluations, and their reactions to the triangulated self, peer, and public ratings.

3 For example, see http://www.qualitymetricsnationaltest.co.uk/new-blog/2016/5/3/royal-shakespeare-company

These observations notwithstanding, the feedback from the participating organisations suggests that the current evaluation process may be under utilising the insights that could come from the peer evaluation process, both for the organisations, but also in terms of critical reflection and continuing professional development for the peers.

Integration with other databases; online databases; and ticketing / CRM systems

At the Learning and Insight sessions the participating organisations asked a range of questions about how a system like Culture Counts could integrate with the other databases, tools, and CRMs they already use. The organisations were both interested in the future possibilities for integration with existing systems, and informed by the desire to ensure that their evaluation activity in the round is as efficient and as effective as possible both now, and in the future.

Integrating with existing evaluation practices – adaptation and innovation

In both this quality metrics strand, and the participatory metrics strand, those organisations with well developed evaluation frameworks and practices had to think carefully about how best to integrate their quality metrics evaluation work alongside other evaluation activity they already had planned or were committed to.

Sometimes these integration challenges concerned using the metrics within multi-stranded evaluation approaches; or focused on how to design surveys and manage their distribution through a range of URLs targeting different audience segments in ways that added value to existing evaluation activity.

In practical terms these issues of integration and complementarity need to be explored by users in real evaluation examples. For example, in both the quality metrics and participatory metrics strand, one response to this integration challenge saw participating organisations innovating in their survey designs, adding in bespoke questions or picking additional questions from the Culture Counts interface. In total, 485 custom questions were added to surveys in spite of no recommendation from Culture Counts encouraging organisations to do so. The appetite to innovate (but also ask audiences lots of questions) is clearly present.

Staff Turnover and Resource Challenges

As advised by Culture Counts, the majority of participating cultural organisations designated one member of staff to be a super-user of the Culture Counts system. In other words, on behalf of a participating organisation that super-user familiarised themselves with the Culture Counts system. From the outset of the project (in evaluation terms) on November 1st 2015 to the close of the project on May 31st 2016, 14% of the the originally designated super-users of the system either left their job role, or that role disappeared for resourcing issues inside the participating organisation.

Understandably, this was very challenging for the participating organisations in terms of the continuity of their engagement in the trial which definitely impacted on the ability of some organisations to complete the target of three evaluations. This turnover of roughly one seventh of the initially inducted users presented continuity challenges within the organisations evaluating, subsequently impacting on delivery of the overall project.

Accessibility Issues

The evaluation processes highlighted a range of accessibility challenges that need ongoing attention, and the participating organisations also innovated in trying to overcome some of these issues. The specific accessibility issues identified by the cohort were as follows:

i. Those with visual impairment would struggle to complete the survey alone with the Culture Counts interface as it currently stands

ii. Working with children and adults where English is a second language can in some cases pose difficulties in accurately understanding the questions (e.g. 'hard to decipher between some specific words e.g. 'produced' and 'presented')

iii. Specific groups, such as those with dementia, pose very specific challenges (from issues of informed consent to the appropriateness of a survey-based format)

v. The survey response scales are unlikely to be clear enough for participants with ' complex individual needs'

v. Elderly respondents (e.g. sometimes with less familiarity of touch screen interfaces and a greater chance of conditions such as Parkinson's)

vi. For 'early years' participants (0-8) the text base interface is not appropriate

The language of assessment versus evaluation

In a strong mirror of the Quality Metrics National Test work on the participatory metrics, the participating organisations discussed their attitudes to evaluating their work and sharing their findings with peers and other organisations.

Organisations acknowledged that the use of standardised metrics could create anxiety around particular pieces of work being 'judged' in particular ways. Clearly, these types of evaluation approaches will only thrive if cultural organisations are encouraged and supported to explore the resulting data in ways that put the emphasis on critical reflection and improvement, as opposed to a narrow emphasis on 'audit' and 'performance reporting.'

Conclusion

Self-driven scalability

The project has resoundingly confirmed that funded arts and cultural organisations, if offered the right tools and support, can self-drive large scale evaluation activity within a very short time frame, engaging in new ways with peers and audiences about the quality of their work.

This would suggest that the quality metrics and the sector's evident interest in being able to measure their creative intentions, allied to tools that help the arts and cultural sector to collect and analyse data easily and at scale, offers up the prospect of a much richer conversation about cultural value in the future informed by big data. The aggregated data set from this Quality Metrics National Test is available from Arts Council England.[4]

The future potential of this approach

The overall evaluation approach facilitated by the features of the Culture Counts system, mirrored in the design and analysis of the aggregate data set from this Quality Metrics National Test, allows the arts and cultural sector:

- To present a very clear story on quality which does not over simplify the findings

- To use co-produced metadata frameworks, for example relating to artform descriptions, to demonstrate both the variety and plurality of work being produced by the funded portfolio; and to allow a rich analysis of quality by artform and artform attribute.

The approach effectively unites data across the standardised quality metrics, artform, artform attributes, and other open data into a powerful prism through which to better understand quality. As we have seen this will deepen understanding of how artform and certain attributes of work influence quality, and offers up the potential to produce a very wide range of analytical and reflective insights.

Crucially, the interpretation of that data will be driven and widely discussed by the creative professionals that make the work, ushering in an era of co-produced quality metrics, co-produced analytical frameworks, and a co-produced conversation about cultural value informed by big data.

4 Published and managed by Arts Council England

2 Faced Dance, DREAMING IN CODE

Courtesy:
Lawrence Batley Theatre

1. CHAPTER ONE: Introduction

1.1. The Foundations of the Quality Metrics National Test

The foundations of this Quality Metrics National Test project were initiated in 2010 by the Department of Culture and the Arts in Western Australia, which commissioned Michael Chappell and John Knell[5] to work with arts and cultural organisations to develop a metrics system which uses a combination of self, peer and public assessments to capture the quality of cultural experiences.[6]

The resulting standardised quality metrics have the potential to offer art and cultural organisations greater insights into what people value about their work; allow them to gauge how far they are meeting their creative intentions; benchmark against similar organisations; and to help everyone talk about quality in a more consistent and confident way.

Dave Moutrey (CEO of the Cornerhouse in Manchester, now HOME) approached Arts Council England (ACE) in 2012 to ask if they would be prepared to support a group of arts/cultural organisations in England to develop a similar approach. The resulting work with a consortium of cultural organisations in Manchester underlined both the scale of the opportunity and the possible benefits of this approach.[7]

The Manchester organisations involved in these ACE pilots, working alongside Culture Counts and The University of Manchester, then successfully made a bid to the Digital R&D Fund (Big Data Strand), administered by NESTA, The Arts and Humanities Research Council, and Arts Council England.[8] As a direct result of working with an expanded group of cultural organisations the Digital R&D work prompted Re:Bourne Ltd to lead a piece of work designed to develop a set of participatory metrics to capture the quality of participatory work being produced by the cultural sector.[9]

1.2. The importance of co-production and an arts sector-led initiative

The key insight that came from these foundation pieces of work that preceded this Quality Metrics National Test was that the active involvement of the arts and cultural sector is fundamental to the creation of a credible and robust measurement framework for the quality of cultural experiences (whether as an audience member or participant). Without their input it was going to be difficult to build greater common language and currency about the value of the arts, and that this was going to be an ongoing iterative process.

5 Then representing Pracsys and Intelligence Agency Limited, now representing Counting What Counts Ltd.
6 http://www.dca.wa.gov.au/research-hub/public-value/
http://www.dca.wa.gov.au/Documents/New%20Research%20Hub/Research%20Documents/Public%20Value/DCA%20PVMF%20Valuing%20and%20Investing%20in%20the%20Arts%204.10.12_.pdf
7 http://www.artscouncil.org.uk/what-we-do/research-and-data/quality-work/quality-metrics/quality-metrics-pilot/
8 http://artsdigitalrnd.org.uk/projects/home-et-al/
9 See http://www.artscouncil.org.uk/developing-participatory-metrics

The attraction to the cultural organisations in being asked to frame new metrics on quality was the opportunity to shape a set of quality metrics that more fully reflected their creative ambitions and intentions. Therefore from the outset this has been a sector-led project that has sought to create a standardized and aggregatable metric system measuring what the cultural sector believes are the key dimensions of quality.

Following these successful pilot projects, Arts Council England invited Culture Counts to deliver a national test of the quality metrics to examine the validity and applicability of the framework across a diverse range of organisations in the Arts Council's National Portfolio.

1.3 The Key Objectives of the Quality Metrics National Test (QMNT)

The key objectives of the Quality Metrics National Test were to:

- Recruit and support 150 National Portfolio Organisations (NPO) and Major Partner Museums (MPM) to use the quality metrics and Culture Counts platform to evaluate three events, exhibitions or performances between November 2015 and May 2016.

- Recruit and support 10 NPOs to refine and test a set of participatory metrics developed through a previous trial, assessing their alignment with the CYP Quality Principles

- Produce an anonymized aggregated dataset and public facing report providing an analysis of the information collected throughout the QMNT highlighting key trends, points of comparison across the participating organisations, and the most important insights from the project

- Produce a separate public facing report for the participatory metrics strand of the trial highlighting key insights from the project and documenting a revised set of participatory metrics in alignment with the CYP Quality Principles

A key learning point from the project was to test the validity of the quality metrics framework across the diverse range of organisations in the Arts Council's National Portfolio. It was therefore very important to ensure that the cohort was broadly representative of the wider Portfolio in terms of artform, organisational size and location.

With support from the Arts Council, Culture Counts put out a call for expressions of interest from organisations to take part. During October 2015, 150 National Portfolio Organisations (NPOs) and Major Partner Museums (MPMs) signed up to test the quality metrics. As a group they are broadly representative of the wider portfolio in terms of artform, organisational size and location (see Appendix One).

1.4 The Expression of Interest Conditions

The 150 organisations that signed up to take part in the Quality Metrics National Test were committing to the following 'Expression of Interest' (EOI) aspirations and conditions:

- To trial the core quality metrics across 3 events, performances or exhibitions during the lifetime of the grant (which effectively meant a period between November 1st 2015 and May 31st 2016).
- For each evaluation the participating organisation should try to obtain a minimum of one self assessment; five peer respondents; a minimum of 30 audience responses.
- Access to self-guided use of the Culture Counts system from November 1st 2015 to July 31st 2016, to set up surveys, collect data, and interpret their results
- To share their data with other participating NPOs
- To have the opportunity to attend a series of 'Learning and Insight' sessions to share your experiences of the national test with other participants and to discuss your interpretation of the results
- Agree that their individual event evaluation data will be analysed at an aggregate level by Culture Counts to produce a public anonymised aggregated final report at the end of the project

1.5 The Quality Metrics

The core quality metrics chosen for this national test are outlined in Table 2. Nine dimensions were used in surveys designed for all respondent categories, with an additional three (risk, originality, and excellence) for self and peer respondents only.

Table 2: Quality Metrics

DIMENSION	STATEMENT	RESPONDENT TYPE SELF	PEER	PUBLIC
Concept	It was an interesting idea	✓	✓	✓
Presentation	It was well produced and presented	✓	✓	✓
Distinctiveness	It was different from things I've experienced before	✓	✓	✓
Captivation	It was absorbing and held my attention	✓	✓	✓
Challenge	It was thought-provoking	✓	✓	✓
Enthusiasm	I would come to something like this again	✓	✓	✓
Local Impact	It is important that it's happening here	✓	✓	✓
Relevance	It had something to say about the world in which we live	✓	✓	✓
Rigour	It was well thought-through and put together	✓	✓	✓
Risk	The artists/curators were not afraid to try new things	✓	✓	-
Originality	It was ground-breaking	✓	✓	-
Excellence	It is one of the best examples of its type that I have seen	✓	✓	-

1.6 The Culture Counts System and the Quality Metrics

Culture Counts is a cloud-based software system that captures artist, peer and public feedback on the quality of arts and cultural events, in this case combined with the quality metrics. Culture Counts captures feedback on the quality of a piece of work or an event from three different respondent groups:

- The artists, curators and/or cultural organisation that created the work or produced the event (self-assessment)
- Expert peers such as other artists, people working in cultural organisations in the same field (peer assessment)
- Audience members and visitors (public assessment)

Quality in this methodology is assessed by asking respondents to rate the work or event against the quality metrics dimensions. Respondents complete a short survey in which each quality dimension is presented as a statement or 'metric' and respondents record the extent to which they agree or disagree with the metric using a sliding scale. Respondents indicate agreement by moving the slider to the right, disagreement by moving the slider to the left and a neutral response by clicking on the slider once to leave it at the mid-point of the scale. Respondents record a 'don't know' response by not moving the slider at all. As well as rating the event against the quality dimensions, respondents are asked to provide their gender, age and postcode.

Self assessment is carried out both before (prior) and after (post) an event to explore how perceptions shift and the extent to which the event meets the creative intentions of the cultural organisation as measured by the alignment between self prior ratings and self post, peer and public ratings. Self and peer assessment takes place via an online portal, with each assessor given unique login details and emailed instructions on how to complete either 'before' and 'after' surveys, or both.

Public assessment takes place during or just after the event itself and captures real-time feedback on how the audience is responding to the work. Audience members record their ratings using a tablet computer (either through intercept interviewing or by using a fixed position tablet). Data can also be collected via post event email requests to ticket holders to complete the same survey online.

Data from all respondents for every event is stored in a single database and the individual cultural organisations can view the results in real-time in their Culture Counts dashboard or export their data as csv files for analysis.

1.7 Evaluations

The evaluation period commenced on November 1st 2015 and finished on May 31st 2016. This 'cut-off' point at the end of May was to allow the Culture Counts team to complete an analysis of the data by grant end (June 30th 2016). However, the participating NPOs and MPMs could continue to use the Culture Counts system through to the end of August 2016.

The original goal was to engage 150 NPOs and MPMs to complete three evaluations during the course of the evaluation period, completing at least one self assessment, aim to independently engage five nominated peer assessors, and obtain at least 30 public responses per evaluation.

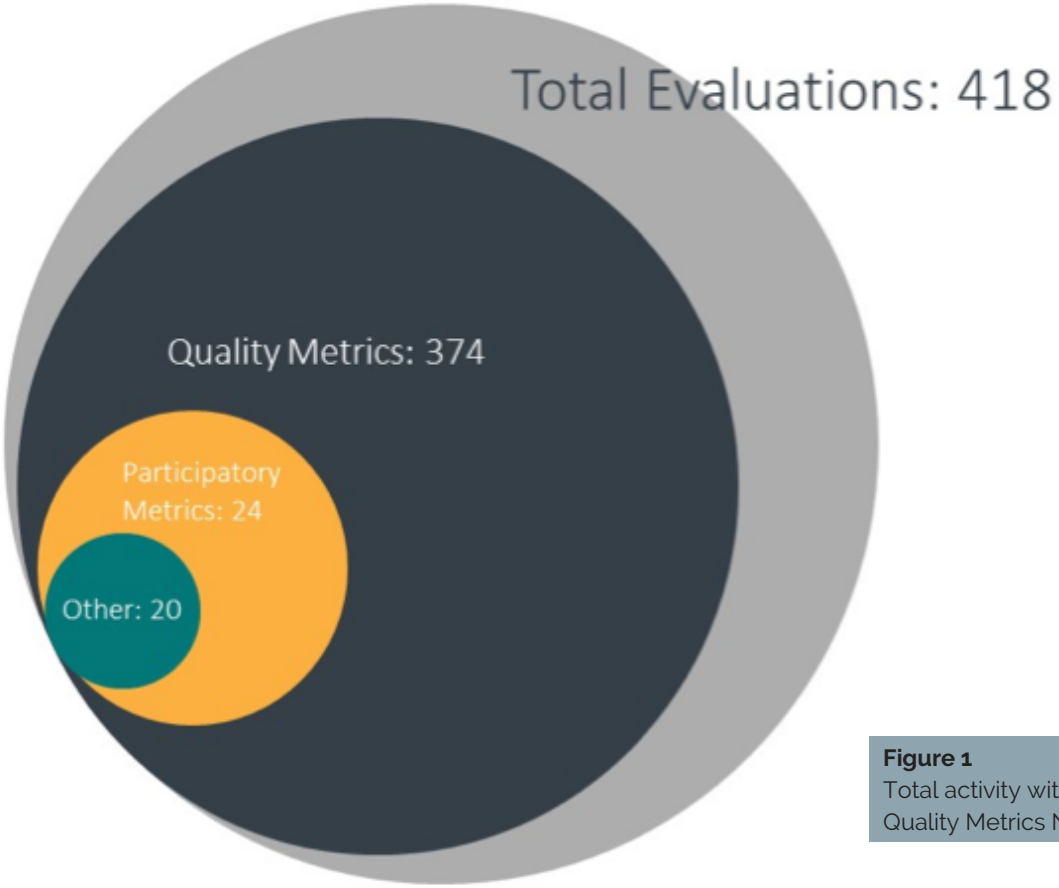


Figure 1
Total activity within the Quality Metrics National Test

Figure 1 summarises the total activity supported by the grant, including the participatory metrics strand evaluation activity. There were 418 evaluations in total during the available time frame, of which 374 evaluations used the quality metrics. The analysis of the 24 evaluations carried out using the participatory metrics is detailed in a separate report.[11]

11 Knell and Whitaker. "Participatory Metrics Report." *Arts Council England (2016)*

Figure 2
Overall Summary of Quality Metrics National Test Evaluation Activity

* includes 10 shared evaluations where data collected responsibilities were shared between two organisations, each contributing self and peer assessors

**Reasons: Collected self-responses only (12 organisations); didn't use dimensions (5 organisations); Submitted data after cut-off (2 organisations); data requested to be withdrawn (1 organisation)

Figure 2 summarises the status of the quality metrics strand of the Quality Metrics National Test on May 31st 2016 (excluding the evaluations testing the participatory metrics). For context here, if the 150 participating NPOs and MPMS had each reached the targets set for them by the EOI conditions the cohort as a whole would have completed 450 successful quality metric evaluations; based on 13,500 public responses; 450 self assessments, and 2,250 peer assessments.

The overall outcomes achieved by the participating NPOs and MPMs, against those aspirational targets, were as follows:

374 successful quality metrics evaluations:	(83% of target of 450)
1,358 self assessments:	(302% of target of 450)
921 peer assessments:	(41% of target of 2,250)
19.8K public responses:	(147% of target of 13.5K)

The other headline figure to highlight from Figure 2 is that throughout the life of the project we saw 137 'active' organisations within the Quality Metrics National Test. Eight of the organisations who signed up through the EOI process did not engage at all with Culture Counts and the process, with the remaining 5 'inactive' organisation proving unable to complete any successful evaluations. Taken as whole this means that 91% of the cohort of NPOs and MPMs fully and successfully engaged in the Quality Metrics National Test.

2. CHAPTER TWO: The Evaluation Data

In addition to the core quality metrics data collected, as described in Chapter 1, basic demographic data (age, gender and postcode) was collected for public respondents. In addition, metadata was also assigned to responses or events accordingly. Where organisations created custom questions, the questions have been recorded in the Culture Counts system for analysis to gain an understanding of what further questions were deemed complementary to the quality metrics dimensions, but any self, peer, or public responses to those additional custom questions have not been analysed and do not form part of the final dataset for the study.

2.1. Metadata and categorisation process

The metadata collected fell in to four categories:

METADATA	FIELD EXAMPLES	OBTAINED
Non-question Survey Data	Timestamp, Respondent ID	Automated via Culture Counts software
Organisation Data	Registered Area, Region, Organisational Size	Arts Council England Open Data
Event Data	Event artform(s) and artform attribute(s), event location	Supplied by organisations in the trial
Geomapping Data	Rural Status	Office of National Statistics Open Data

Our overall approach to constructing metadata was designed with highly sensitive aggregate analysis in mind. In order to manipulate each data point (i.e. one answer to one question – any individual answer to any individual question) individually with any other, and to create any combination of group comparisons that are meaningful, metadata needs to be assigned to each question response. The Culture Counts platform does this by design. For this project, we collected additional metadata to that collected via the quality metrics surveys, specifically organisation data, event data and geomapping data.

Structured data for location was applied to each event so that events in particular regions could be viewed in the aggregate, but granularity maintained by postcode for geomapping enabling versatile groups for analysis on a national scale.

12 For more information on this data standard visit: https://www.w3.org/RDF/
13 http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/2011-rural-urban/index.html

2.2. From metadata to artform categorisations

An ontology for artform and artform attributes was created from scratch with data sourced from each of the organisations in the trial. The artform categorisation of each event was defined by the organisation and transformed using principles underlined by the Resource Description Framework (RDF) data model .

This means that instead of a word being stored simply as a string of letters, it is also understood based on its *relationship* to the work it is attributed to.

For example, the data is computationally understood as:

The work has an attribute
The attribute type is specific artform
The specific artform value is ballet

This also enables analysis of all things which have the same attribute, but may not have the same artform. For example, grouping contemporary work together, or grouping immersive work together, regardless of its form or presentation. It also enables granular analysis of combined arts where the form and presentation can be defined both more flexibly and comprehensively.

A process of inference has been applied to artforms. This means that when a specific artform has been described such as ballet, it also gets attributed with a broad artform (dance) and a sensory artform (movement). Inference has not been applied to artform attributes at this stage. So for example, a piece of theatre could well be a performance (presentation attribute) but it has not been inferred as such. This is because the rules of such inference, with the data available, cannot be consistently applied across all metadata, therefore attributes are used exactly as they have been supplied by the organisations.

The development of these layered artform categorisations has enabled analysis far more sensitive than broad or ambiguous artform categories whilst also supporting analysis of work that fits in many different or atypical artform categories. Practical examples of the benefit of this approach are as follows:

The word 'film' on its own could mean, film as an artform, as presentation method (cinema broadcast), or as a subject (film music), or as a medium (archived materials). Instead of only keeping the pure artform term and disregarding the term as being useful in other contexts, assigning it as a type of artform attribute retained the term, but in its intended context for analysis (the assumption was made that when using the term e.g. film music, the film as a subject was important enough to use as a differentiator in describing the work, therefore it shouldn't be omitted or disregarded).

14 For more information on this data standard visit: https://www.w3.org/RDF/

2.3. Preparing the dataset for aggregation

The dataset was aggregated with standard fields based on the data collected through the surveys and metadata agreed at the start of the project. For a full list of fields, and their source, please see Table 3.

The raw data aggregate was used for monitoring and initial descriptive analysis. Each evaluation was then summarised with 5 numbers by respondent/time category (self prior, self post, peer and public) per dimension. This provided an accurate summary of the raw data using medians and interquartile ranges and also forms the basis of the anonymised dataset produced as part of the project. The data presented in this report use the median values from the raw data aggregate.

2.4. Triangulating the Data

A key element of this methodology is the triangulation of scores within respondent categories. Comparing and contrasting the respondent types provides a much more complete perspective on the work based upon the perspectives of the creators and production teams, other professionals, and the public audiences.

A dimension scoring high by one respondent category and low by another provokes a reflective conversation about the work. For example, here are some interpretative exchanges on the evaluation results from participating organisations at the Learning and Insight sessions:

- Commenting on similar scores across all three respondent types: *"our audiences are much savvier than we give them credit for"*

- Commenting on a high peer score and low public score: *"this would mean we created high quality theatre, but it may not be as assessable to the public as we hoped"*

Tabe 3: Raw Data Fields

DATA FIELD	SOURCE	DATA FIELD	SOURCE
Gender	Respondent	Organisation	Organisation
Age	Respondent	Evaluation Name	Organisation
Postcode	Respondent	Event Artform - raw	Organisation
Concept	Respondent	Location/ Event Notes	Organisation
Presentation	Respondent	Event Attendees - gross	Organisation
Distinctiveness	Respondent	Event Collaborators	Organisation
Challenge	Respondent	Event Postcode Event	Organisation
Captivation	Respondent	Town/City	Organisation
Enthusiasm	Respondent	Event County	Organisation
Local Impact	Respondent	User ID	Culture Counts
Relevance	Respondent	Timestamp	Culture Counts
Rigour	Respondent	URL	Culture Counts
Risk	Respondent	Delivery Method	Culture Counts
Originality	Respondent	Respondent Category	Culture Counts
Excellence	Respondent	Time Category	Culture Counts
Three words	Respondent	Evaluation Reference	Culture Counts
Organisation ACE Region	ACE Open Data	Evaluation Completion Status	Culture Counts
Organisation ACE Artform Discipline	ACE Open Data	Event Rural Status	ONS Open Data
Organisation ACE Area	ACE Open Data	Rural Description	ONS Open Data

2.5. Event Choice and Location

Choosing events for the trial was in the hands of the participating cultural organisations. The freedom to choose which events would be suitable to test a new approach to evaluation was important; the nature of many cultural organisations is to approach their work in a unique way and one of the objectives of the QMNT was to see how well this evaluation methodology could support this instinct.

In terms of the types of events organisations chose to evaluate we expected that organisations might choose the simplest way to test the metrics but a really large variety of events have been chosen (as confirmed by the meta data tags), which stretched the Culture Counts delivery team to support lots of different types of evaluative activity.

At the EOI stage, care was taken to ensure that the registered location of the organisations represented the breadth of the UK funded portfolio. As may be expected, dense areas with high cultural provision in the form of NPOs (in the trial) such as London, Manchester, Newcastle and Bristol had the most number of evaluations (see Figure 3: Event Location Map). The public respondent map (see Figure 4) also reflects these locations. The more interesting view of the following maps is of where cold spots lie.

It is reiterated here that whilst this sample is broadly representative of the whole funded portfolio it is inevitably incomplete, moreover some of the cold spots are less densely populated areas of England, such as the expansive national parks in the north of England. There are a few counties that seem to have far fewer public respondents than others which may not be explained by the baseline population density, further analysis of this however is out of scope for this study.

Figure 3: Event Location Map

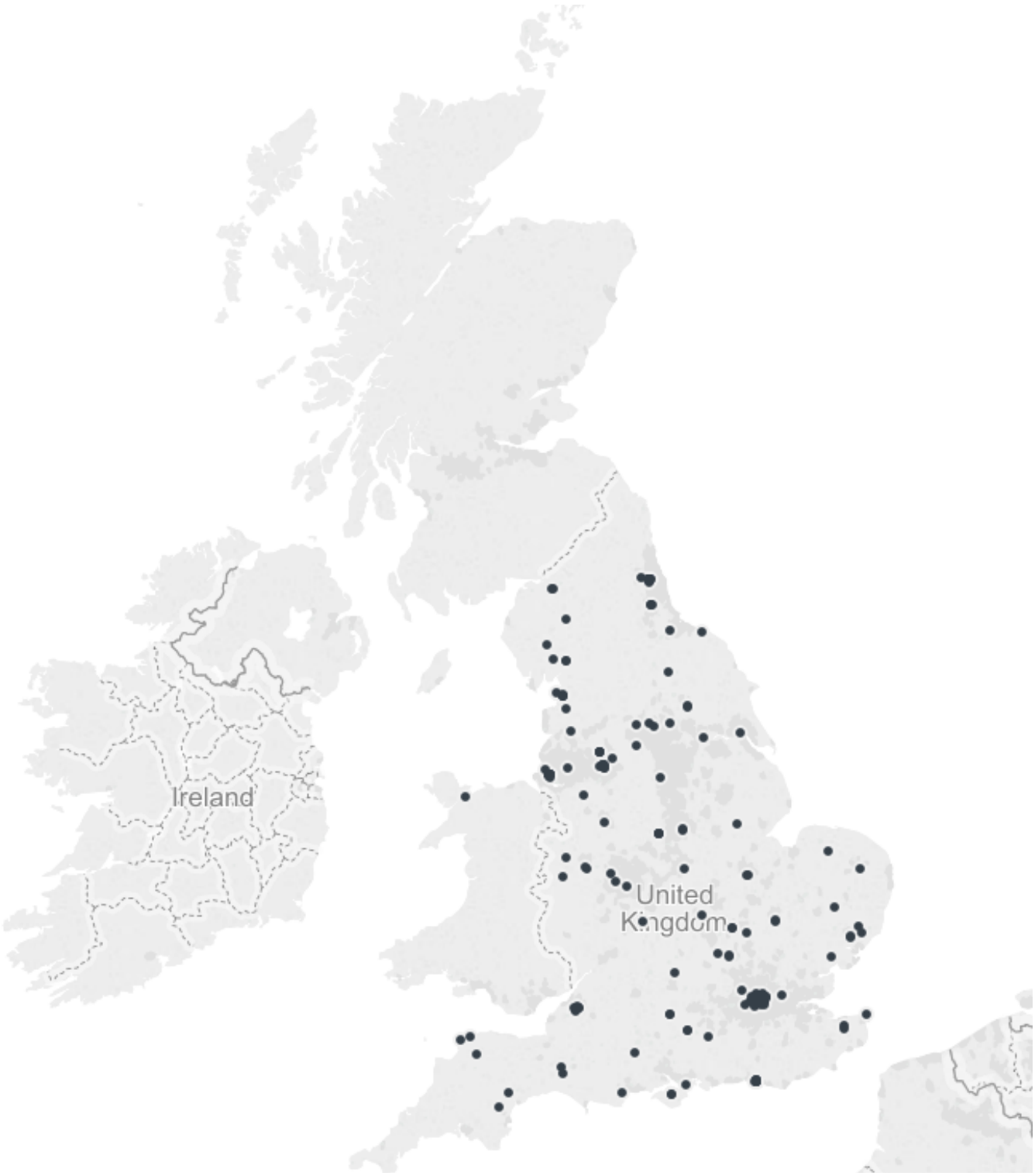
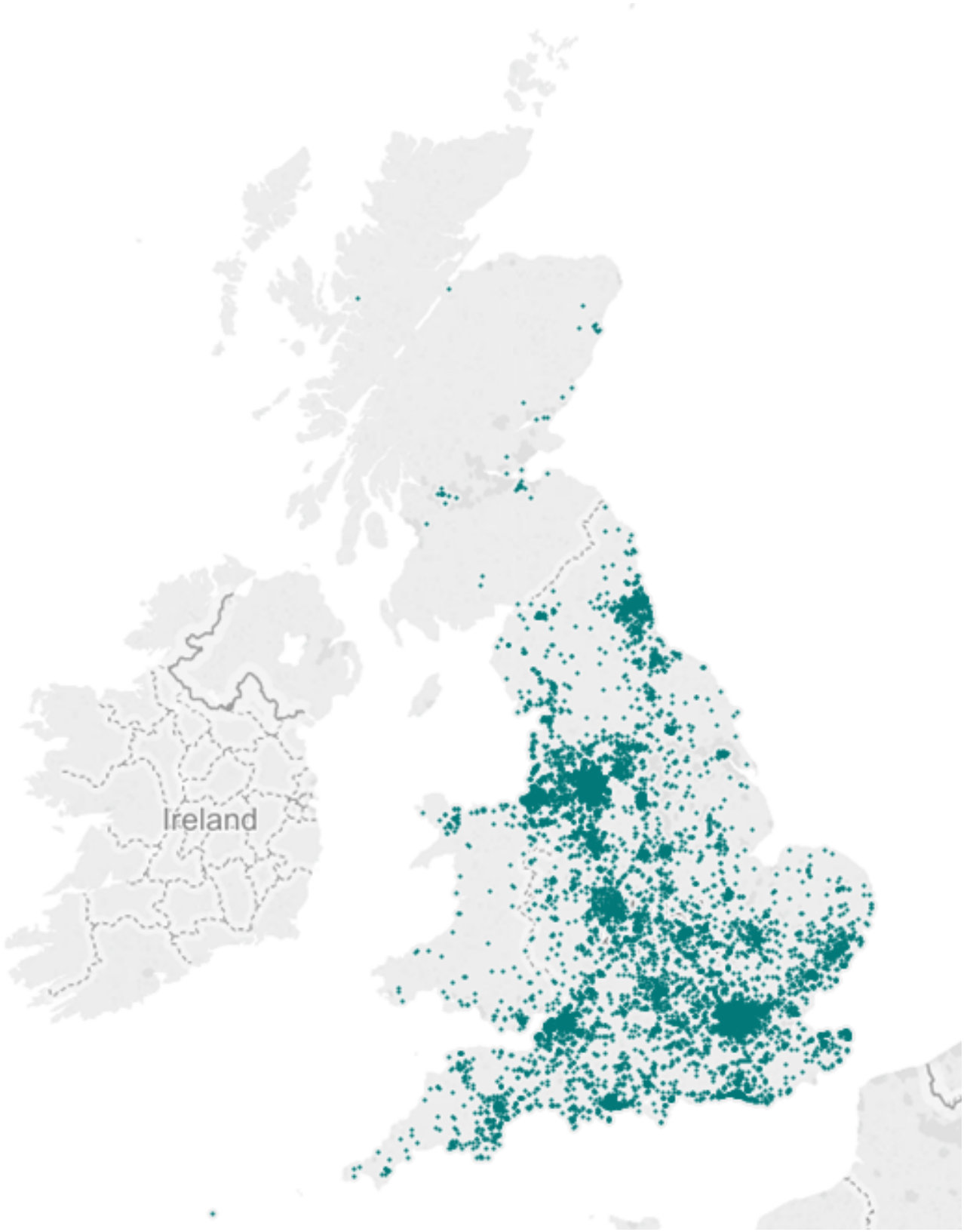


Figure 4: Public Respondent Map



3. CHAPTER THREE: The Core Aggregated Analysis

3.1. Introduction

As the previous chapter explained, the data set created from the Quality Metrics National Test evaluations has been constructed in a way that allows for a multitude of different types of cross-cutting analyses. This chapter, which presents the core aggregated analysis, has sought to confine attention to the most important questions and intuitive cross-cutting forms of analysis (e.g. dimension scores in aggregate by respondent; dimension scores by artform categorisation and respondent; dimension scores by region etc.)

3.2. The Core analysis presented here

Our presentation of the aggregate analysis has the following aims:

- To tell a clear overall story of the top-line aggregated results which does not over simplify the findings.
- To reveal granular insights from the findings through the use of metadata tags that bring richness to the aggregated results
- To use the metadata frameworks we have created from the inputs of the participating organisations, for example relating to artform descriptions, to demonstrate both the variety and plurality of work being produced by the funded portfolio; and to allow us to dig a bit deeper into the profile of dimensions scores for different categorisations (artform; artfrom attributes) to see whether the patterns that emerge are insightful and suggest interesting lines of ongoing research.

To achieve these aims we have grouped together a set of linked questions that reveal a very clear story about the data and which we hope will trigger significant discussion and interest across the whole cultural sector. The data is, as ever, the starting point for a rich conversation about quality and the value of the work being produced.

In Appendix 2 (Supplementary Data Charts) we have provided a range of charts for those readers that wish to look in more detail at some of the data patterns emerging out of our early analysis of the data set. If we were to include all of these analyses in the main body of this report it would be much harder to frame a clear narrative presenting the key findings from the study.

We have also produced an open data set of the QMNT data stemming from the purpose of critically exploring the potential of this overall evaluation approach, both for the cultural sector, and for the research and policy making communities.[15] Clearly, an important element of the potential of this overall approach concerns whether the resulting evaluation data at self, peer and public respondent level is useful and insightful for each of the participating cultural organisations. Does it allow them to gain insights into their creative practice, and to build a better understanding of their work and audience reaction?

15 One of the public value outcomes of this study is the resulting data set, which appropriately anonymised, has been lodged with Arts Council England.

The other important public value element of this approach relates to whether using sector produced standardised metrics statements, allied to sector produced metadata that allow for fine layered analytical frameworks around artform or mode of presentation, enables the cultural sector to explore in detail some of the dynamics of how cultural value is being created – whether that be the risk profile of the work or the variety and plurality of work being produced.

This chapter presents what we regard as the 'core' analysis of the data. Chapter 4 presents some additional lines of enquiry that we think are suggestive of the further potential of building cultural value data sets of this kind.

3.3. What do self, peer and public response tell us about the overall quality of work in this evaluation

Unwrapping what the data tells us about the overall quality of work being produced by the participating organisations is a multi-layered task. At the outset it is important to remember a number of things about this study and the data.

Firstly, the dimension scores for individual organisations, or in aggregate, are not a clapometer, in which a successful piece of work has to be seen to score highly on every single dimension. If you imagine the work of a producing theatre over the course of twelve months, it will feature new writing and challenging work which the artistic director would expect to score highly on originality, risk or challenge. The same theatre might feature a traditional show, or indeed a seasonal show, for which the artistic team would be predicting a different profile of dimension scores. Therefore, one very important element of the results presented in this study is how far the creative intentions for the work, as captured by the self prior scores (i.e. creative practitioners' expectations of how the work will be received), are in close alignment with public and peer responses. Where they are in close alignment the work is delivering against creative expectations. Are the cultural organisations in this study adept at making those judgements in alignment with peer and public response? What are the risk and originality profiles of the work they are producing? This is something the data can help us explore.

Secondly, care needs to be taken when interpreting the data that inappropriate aggregate comparisons are not being drawn – to stretch that well-worn analogy a little - comparing apples, pears and pineapples (substitute your chosen artform or genre terms) should only be done in ways that are appropriate, such as where similarities between evaluation events can be closely aligned and therefore comparison provides meaningful insight. Our analysis of the dimensions by artform (and by other attributes (see Chapter 4)) will explore this dynamic.

Thirdly, it is important, therefore, when interpreting the aggregate results to consider the dimension profile, in terms of the shape of responses to the dimensions for any given set of respondents (self, peer and public) and variables (e.g. organisation, artform, region etc.), and acknowledging that a variety of work with different individual profiles make up that aggregate.

3.3.1 Aggregated self, peer and public scores by dimension

Self respondent scores

In order to effectively represent each respondent's response from the raw data for a number of aggregate analyses, each respondent type (self prior, self post, peer post, public) was summarised with 5 numbers for each evaluation (20 numbers per evaluation). These numbers represent the median and interquartile ranges for each respondent category.

Figure 6 shows the average self prior scores for the nine quality metrics statements answered by all self, peer and public respondents, with the dimensions labelled at the bottom of the chart. Why start with this chart? One of the key benefits of the quality metrics is that they allow cultural organisations and their creative teams to gain feedback on the extent to which they are meeting their creative intentions for a piece of work (as measured by the proximity between their prior self assessment scores, and their post event self assessment scores and peer and public respondent scores). The closer the match the more successfully they have met their creative intentions.

Therefore, all the participating cultural organisations in this study were encouraged to carry out a self assessment both prior to their event taking place, and then to conduct another self assessment after their event. Figure 5 provides a summary of average self scores by dimension (self, peer and public) carried out prior to the event.

Figure 5: Average self prior ratings by dimensions

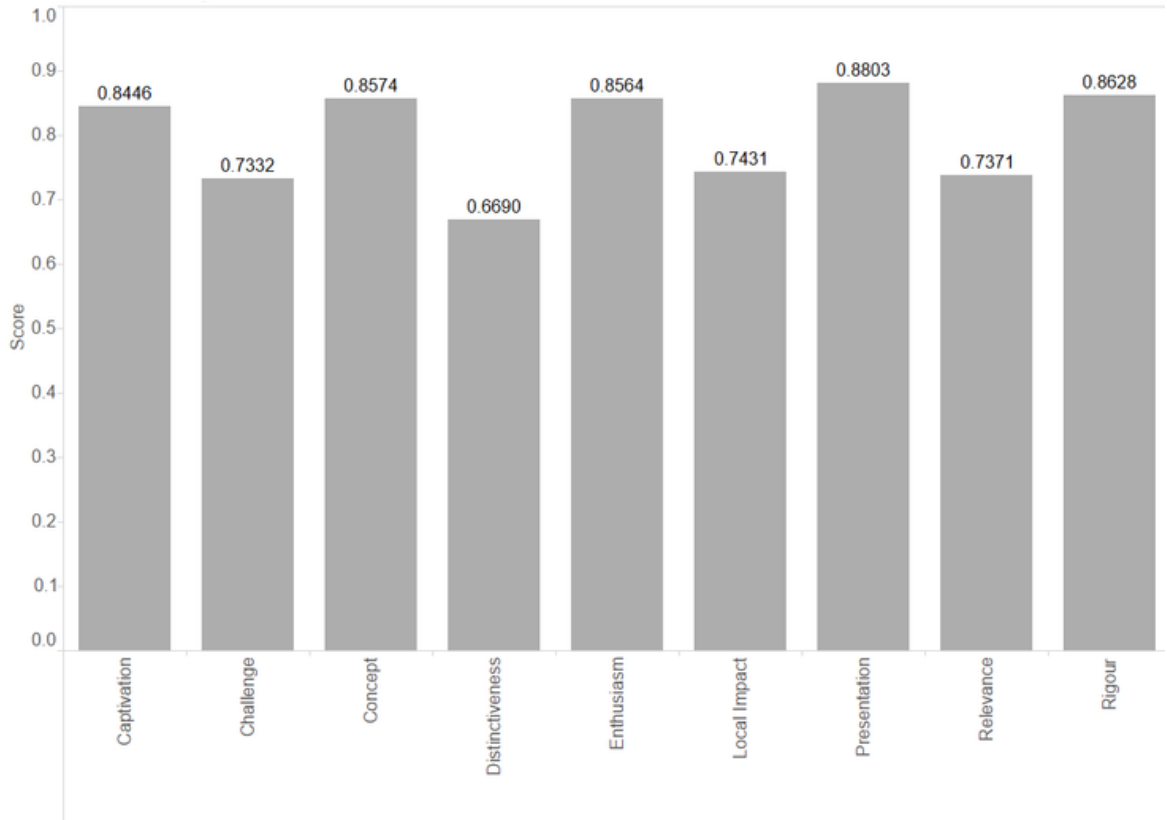
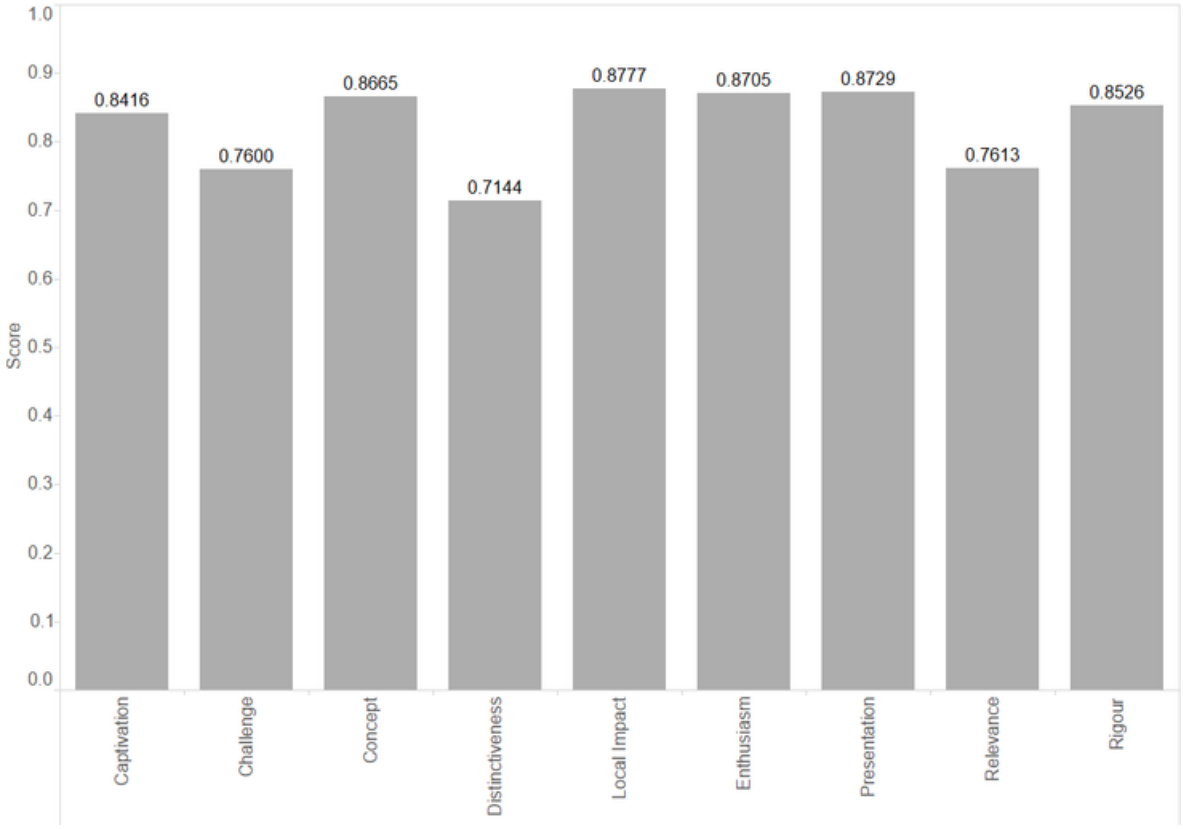


Figure 6 shows the average self post scores, evaluated by self respondents after the event.

Figure 6: Aggregate self post (after the event) ratings by dimensions



Comparing these self prior and post scores, the only significant movement in their dimension scores post event was for local impact, which self respondents rated noticeably higher after the event.

Looking at the dimension profile as a whole for self respondents, the scores suggest a strong degree of self confidence in the quality of the work being produced by the participating organisations, and that on the whole the work lived up to their creative intentions and expectations (in other words self respondents were not giving harsher scores, compared to their self prior assessments, after they had seen their work presented to an audience). The charts that follow in this paper will only use the self prior scores when presenting self assessor information unless otherwise specified.

The other noticeable finding from the aggregated self responses is that three dimensions (challenge: it was thought-provoking, distinctiveness: it was different from things I've experienced before, and relevance: it has something to say about the world in which we live) attracted significantly lower average self scores than the other six higher scoring dimensions (captivation, concept, enthusiasm, local impact, presentation and rigour), as Figure 6 demonstrates.

What are the likely explanations for this outcome? Firstly it could reflect the profile of the work that featured in this QMNT, with the dimension profiles for challenge, distinctiveness and relevance possibly shifting if more of the ACE funded portfolio's work was represented in the data. Secondly, it could be because that in evaluation terms these three dimensions offer a sterner sentiment test of the quality of a piece of work as evidenced by audiences being more divided on these dimensions than any of the other quality metrics (see Figure 18). Or it could be be a combination of both of these factors.

Public Respondent Scores

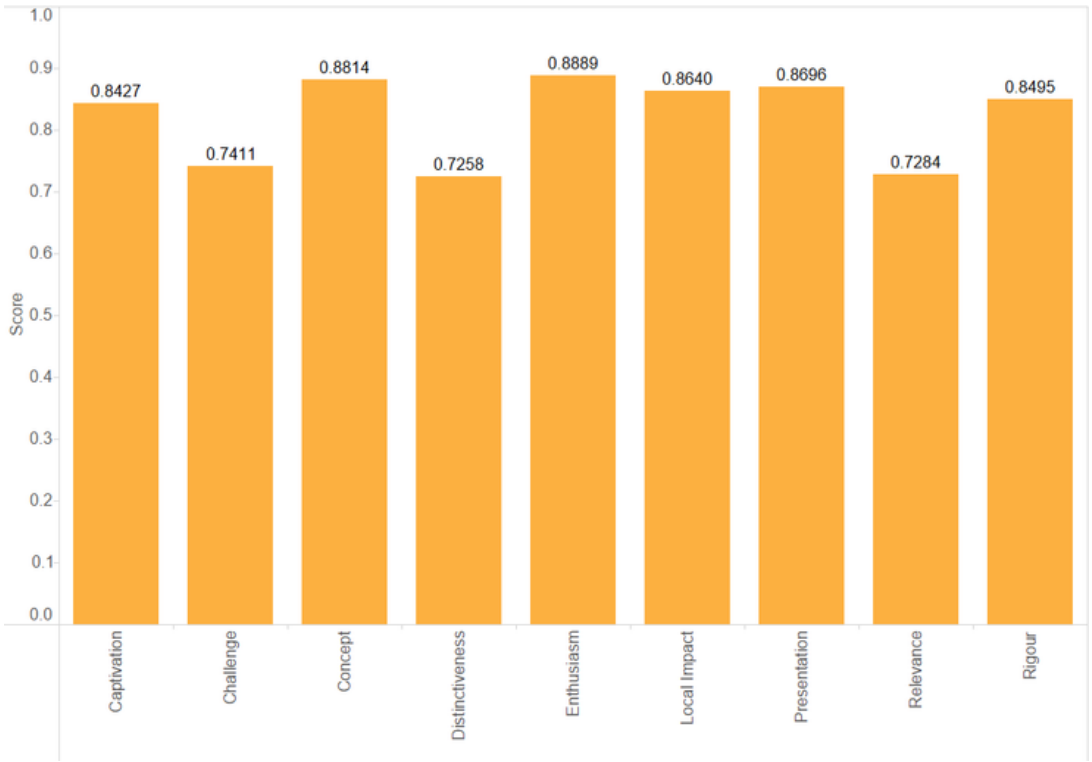
Figure 7 shows the average public scores by dimensions for all of the evaluations in this study, with the dimensions labelled at the bottom of the chart.

Overall, the scores show a strong positive response to the work the public experienced as part of this national test.

As with the dimension profile for self respondent scores, three dimensions (challenge, distinctiveness, and relevance) attracted significantly lower average public scores than the other six higher scoring dimensions (captivation, concept, enthusiasm, local impact, presentation and rigour).

These scores for the lower ranking dimensions are however still positive rankings (a value of 0.5 represents a neutral response to the metric statement) in relation to the metrics statements concerned.

Figure 7: Aggregated public ratings by dimensions



Peer Respondent Scores

Figure 8 shows the average peer scores across the nine self, peer, public dimensions for all of the evaluations in this study, with the dimensions labelled at the bottom of the chart.

The dimension profile for peer responses follows the same pattern as for self and public responses, with peers giving significantly lower scores for lower scores for distinctiveness, challenge and relevance, than for the six more highly scoring dimensions (captivation, concept, enthusiasm, local impact, presentation and rigour).

Figure 8: Aggregated Peer ratings across the nine (self, peer, public) dimensions

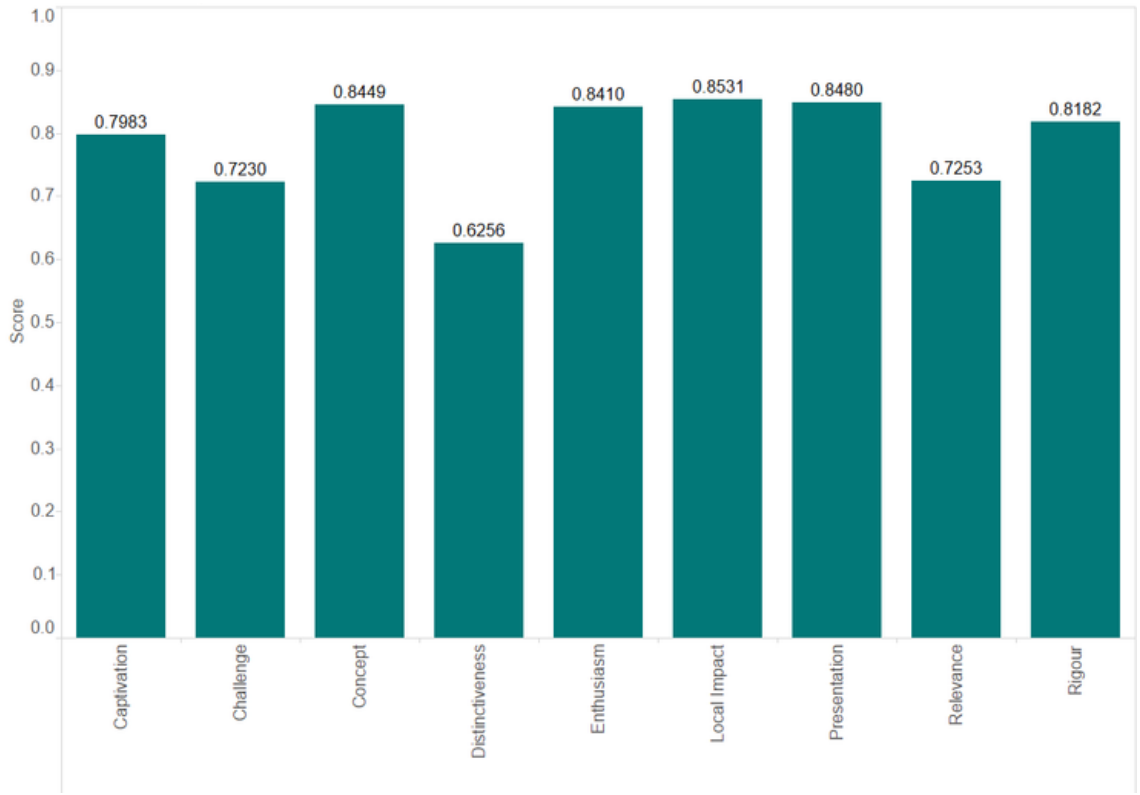
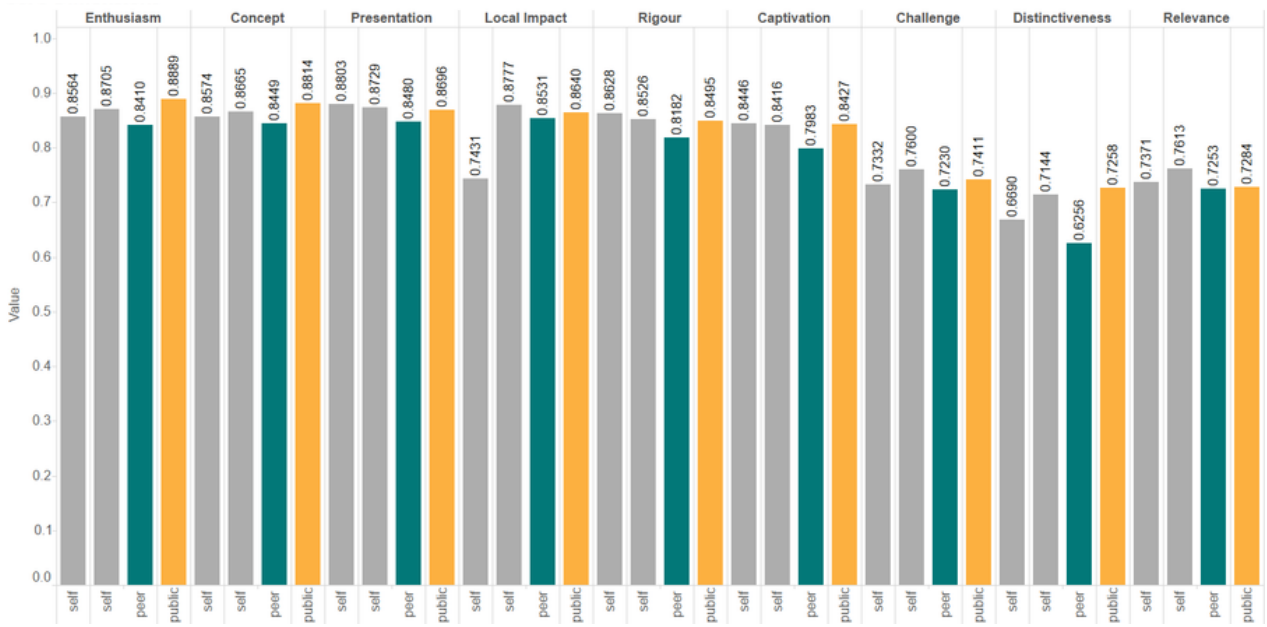


Figure 9 brings together the aggregate scores for self/peer and public responses for their shared nine dimensions. The two grey columns for each dimension are the self prior and post scores, alongside the orange (public) and teal (peer) scores for each dimension. This chart gives a clear visual demonstration of self, peer and public scores clustering at a much lower level for challenge, distinctiveness and relevance.

Figure 9: Aggregated self, peer and public ratings by the 9 shared quality dimensions

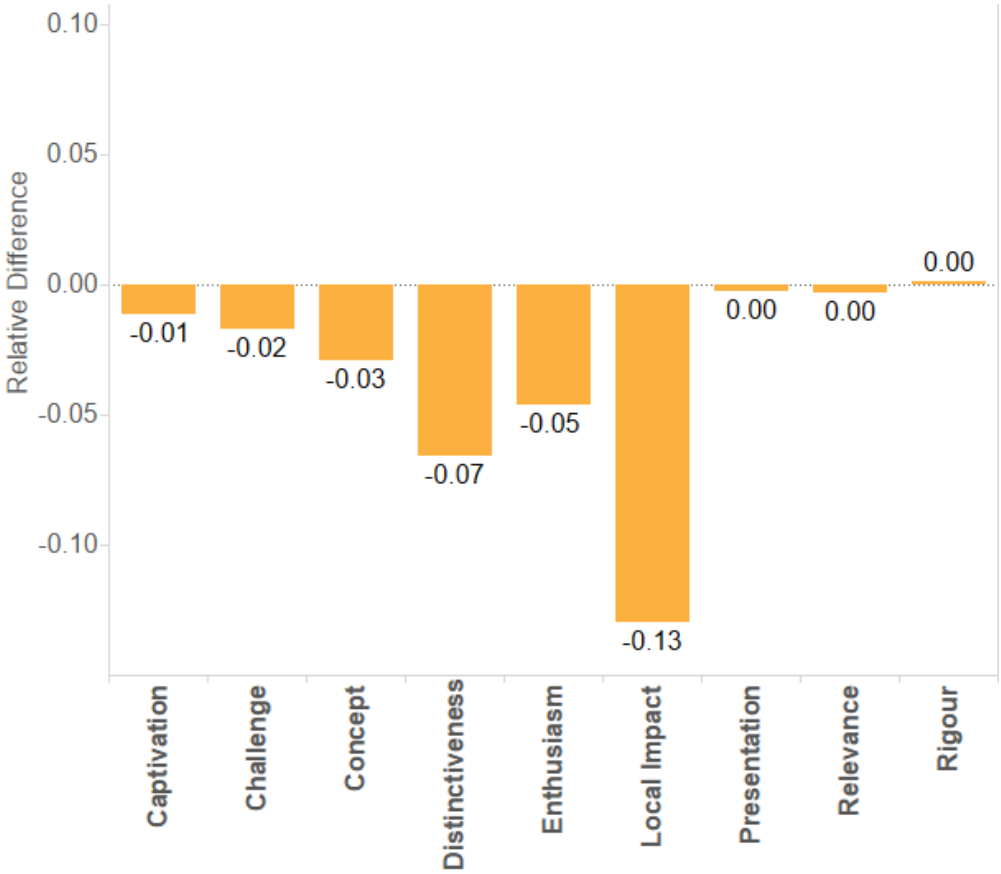


3.4. Meeting creative intentions through the prism of triangulation

As we have noted, one of the key benefits of standardised quality metrics, and the triangulation of self, peer and public feedback, is that it allows creative teams to gain insight into how far they have met their creative intentions for a piece of work, as measured by the strength of alignment between self assessor prior ratings and peer and public ratings for the given event.

Figure 10 shows the variance between average self prior quality dimension ratings for the events evaluated in this study, and corresponding public responses. Overall, there is a very strong fit between self prior and public responses (the relative variance by dimension is low, at 5% or less across 7 of the 9 dimensions). Therefore, when it comes to public audiences, the self prior ratings suggest that the work evaluated in this trial was broadly meeting the creative intentions (as measured through audience response) of the participating cultural organisations.

Figure 10: Differences between self prior scores and public scores across the dimensions

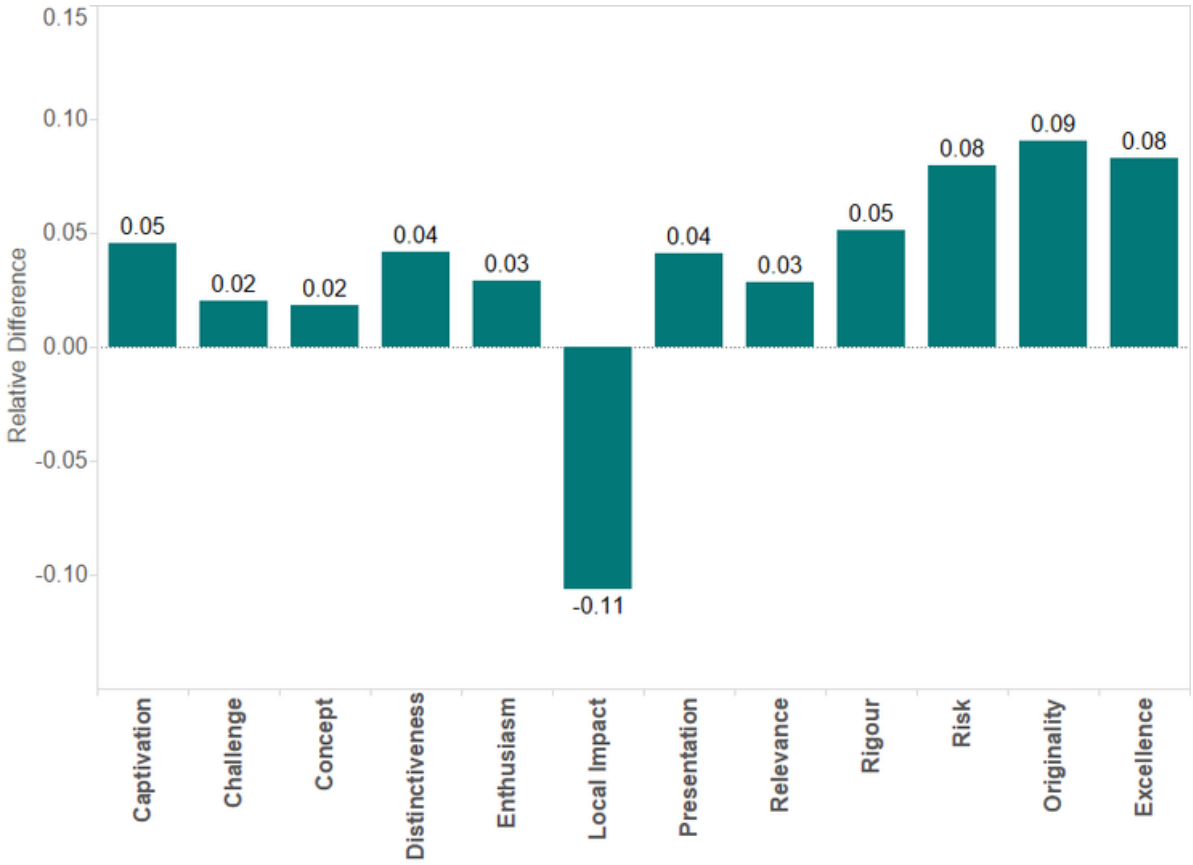


As the chart clearly demonstrates, generally, self assessors are most likely to underestimate the local impact and distinctiveness of work in comparison to the public responses. We also know that on local impact these scores realign post-event (in other words self assessor ratings post event are much closer to public ratings on local impact). Interestingly, for challenge and relevance, the differences between self and public are negligible, indicating that when work has been rated by the public as not particularly challenging or relevant that is indeed in line with the balance of the creative intentions of the self respondents. In other words, work regarded as challenging by cultural organisations, is not being judged unchallenging as audiences.

It is also noteworthy regarding ratings on distinctiveness that self assessors think the work less distinctive than the public. There might be a number of reasons for this pattern in the data. For example, it is likely that at the point a self assessor comes to complete a prior survey they will have lived with that piece of work for a number of months, their familiarity acting as a dimension score deflator as compared to the view of seeing something the work with fresh eyes.

Figure 11 examines the same variance dynamic for self prior ratings and peer ratings across the dimensions. The pattern is visibly very different from Figure 10, with the local impact dimension being the only outcome on which self prior ratings are lower than the peer response. With all of the other dimensions, self prior ratings are higher than than the peer response, although as with the self prior / public comparison, the degree of variance was quite low (5% or less across eight out of the twelve dimensions). There was a significant over-rating of risk, originality and excellence by the self assessors as compared to peer response, which we examine in more detail below.

Figure 11: Differences between self prior scores and peer scores across the dimensions



Overall interpretation for these aggregate self, peer and public scores

Taken together these aggregate scores suggest:

- The work presented and analysed in this study received a broadly positive response from peer and public respondents, and largely met the (quite high) prior creative expectations of the creative teams involved in its production (self assessors)
- When it comes to measuring the quality of a cultural experience three dimensions in particular - challenge, distinctiveness and relevance – overall, score lower in all respondent categories but may offer a sterner sentiment test than the other six dimensions

- The clustering of self, peer and public responses in relation to these metrics suggests that audiences are adept at assessing them, with their judgements showing broad alignment with self and peer responses.
- The participating cultural organisations largely met their creative intentions, as measured by the degree of alignment between their self prior scores for each dimension and the corresponding aggregate scores for peer and public respondents Peer responses (as we have seen in all previous evaluations) are consistently lower across all dimensions than self and peer responses
- The participating cultural organisations largely met their creative intentions, as measured by the degree of alignment between their self prior scores for each dimension and the corresponding aggregate scores for peer and public respondents
- Peer responses (as we have seen in all previous evaluations) are consistently lower across all dimensions than self and peer responses

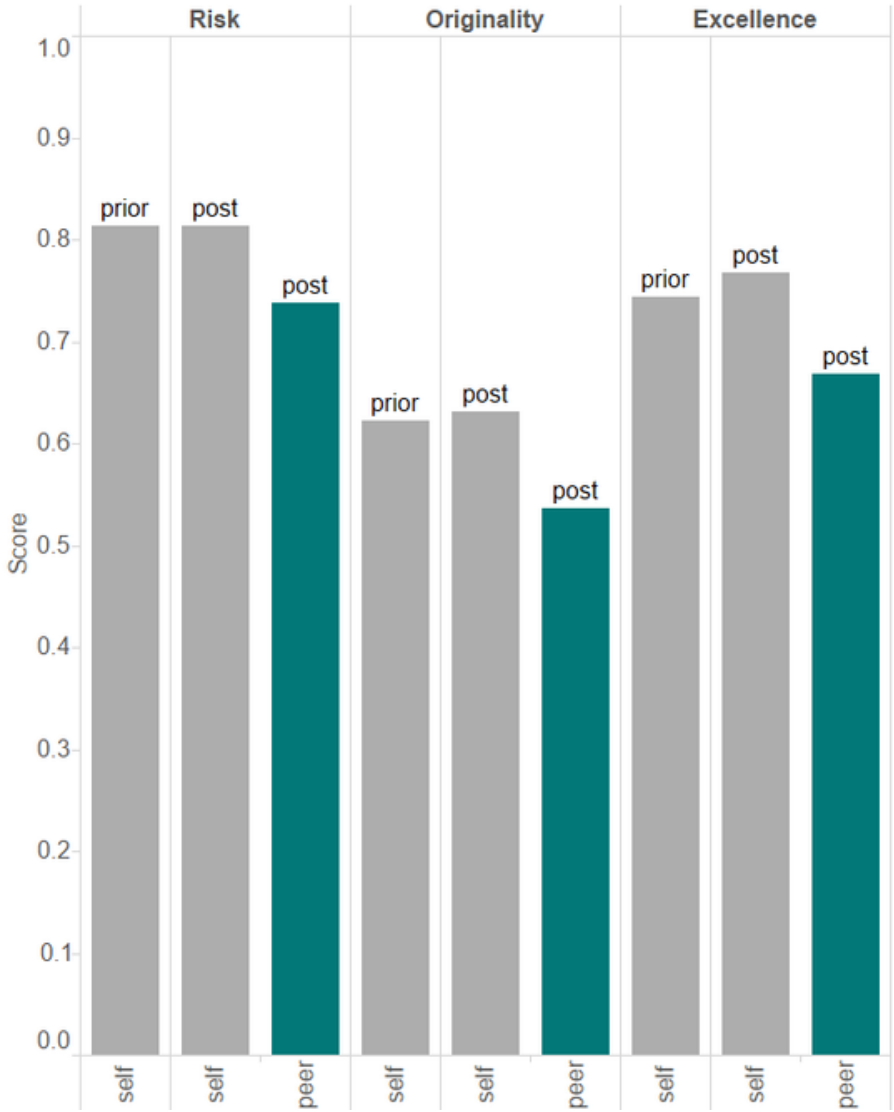
3.5. Risk, Originality and Excellence as measured through self and peer aggregate responses for all evaluations

Figure 12 below presents the aggregated self (prior and post) and peer scores for all evaluations across the three quality metrics that are for self and peer completion only, namely:

Risk: 'The artists / curators were not afraid to try new things'
Originality: 'It was ground-breaking'
Excellence: 'It was one of the best examples of its type I have seen'

The aggregated self responses across these three dimensions show that self assessors tend to score themselves more highly than peer assessors; a well-established trend in previous and ongoing evaluations using the quality metrics. Interestingly, at an aggregate level the self assessors perceive themselves to be taking quite high levels of risk (broadly supported by peer scores, which are highest for this metric out of the three). This is encouraging to the extent that it would suggest that taken as a whole the cultural organisations in this study are seeking to stretch themselves with the work they are producing, and that they have a well-developed appetite for creative risk.

Figure 12: Aggregated Self prior and post and peer ratings for risk, originality & excellence



Originality is the lowest ranking dimension aggregate score for both peer and self respondents. This would suggest that at an aggregate level self and peer respondents did not consider the work being evaluated in this study to display high levels of originality. Is this a surprising assessment? The bar is set high by the originality metric, with respondents asked to express their relative support for the notion that the work 'was ground-breaking.'

As Figure 12 also shows, the aggregate self and peer ratings for excellence suggest that some of the work in this evaluation is regarded as amongst the best of its type experienced by the self and peer respondents.

Let's examine these patterns in a little more detail, both in terms of self and peer rating behavior, and through a more detailed consideration of risk appetite and originality by artform.

Self and Peer rating ranges for risk

Figure 13 shows the distribution of self and peer responses for the risk dimension, broken down by our broad art form categorisations[16] (placing visual sensory artform alongside the broad artforms of dance, literature, music and theatre). In broad terms the charts show:

- There are significant variations in the risk scores within artforms
- Peers gave a noticeably high, and narrower, range of risk scores for literature
- Peers show a greater willingness to use all of the respondent scale as compared to self respondents (the self score range stops at 0.2. whereas the peers use the whole range.)

Figure 13: Distribution of Self and Peer risk dimension scores by broad artform

16 Please see Chapter 4 and Appendix 2 for further exploration of the results cut by artform and artfrom attribute.

Within this overall picture, it is worth commenting on the observed differences between how some of ACE's Artistic Quality Assessment (AQA) assessors scored the work they evaluated as opposed to the sector nominated peers undertaking evaluations. Across 23 events within the overall sample both cultural organisation nominated peers and AQA assessors completed a quality metrics peer assessment.

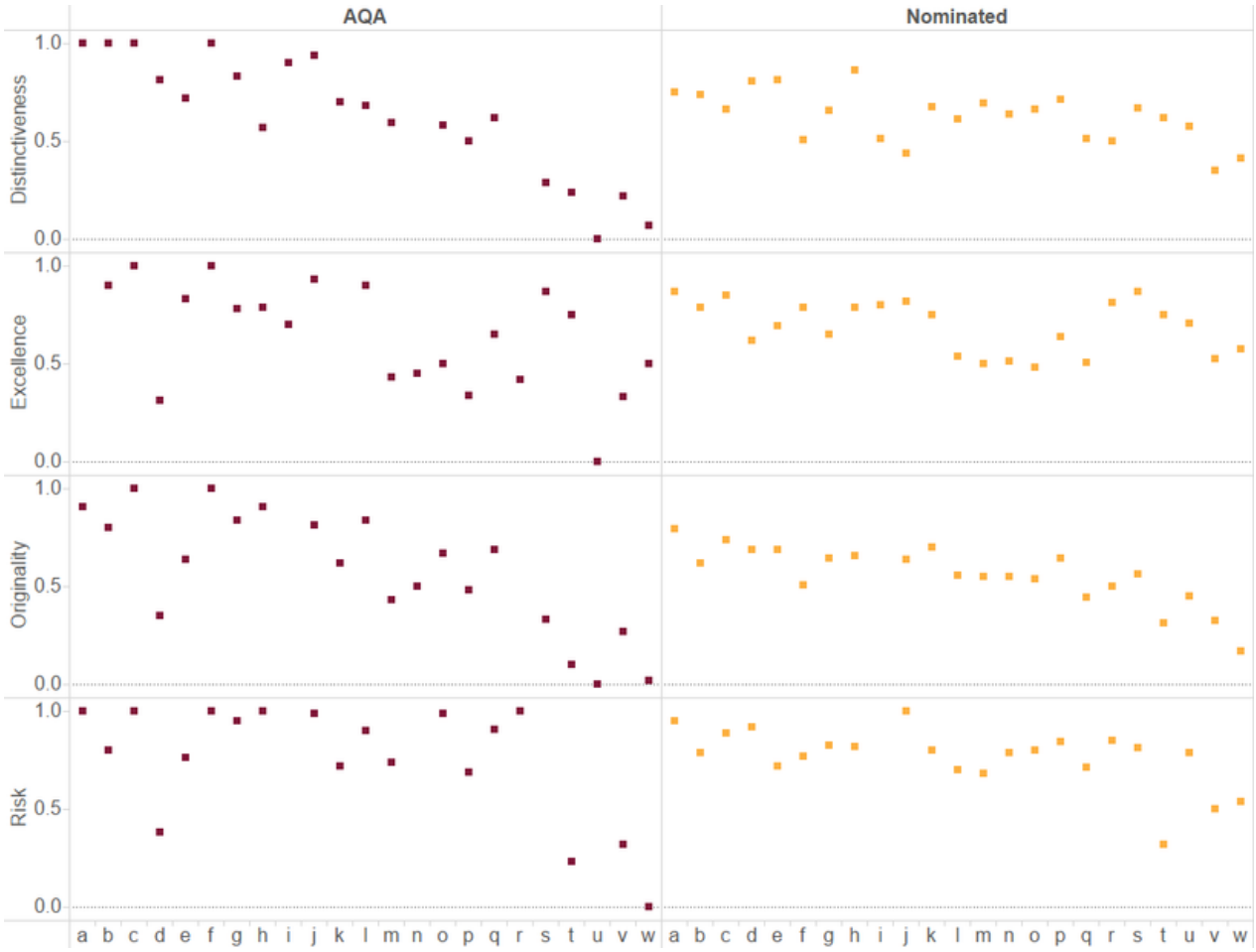
Initial analysis indicates that on most dimensions there are no significant differences between these groups of assessors. We are unable to tell whether this observed pattern in the scores are significant due to the small sample size of assessors evaluating each event.

To truly measure differences between these groups of assessors, we would need to look at numerous AQA assessors and numerous nominated peers all evaluating the same event, rather than different assessors evaluating different events.

An interesting finding, however, has emerged when looking at the general scoring profile of AQA assessors and nominated peer assessors. The distribution of scores is significantly more varied for distinctiveness, originality, risk, and excellence for the AQA assessors compared with the nominated assessors. The averages are overall not significantly different (although the sample doesn't enable us to truly measure this as noted above) but for these four dimensions, nominated assessors stay within a smaller range in contrast to AQA assessors who are more likely to use the full range of scores to evaluate the work.

Figure 14 shows the dimensions with significant variation between peer groups. Each letter along the x axis denotes a single evaluation. (the maroon colour = ACE AQA peers; the orange colour = nominated peers).

Figure 14: AQA peer responses contrasted with cultural organisation nominated peers



3.6. Risk, Originality and Excellence as measured through self and peer aggregate responses by detailed art-form categorisation

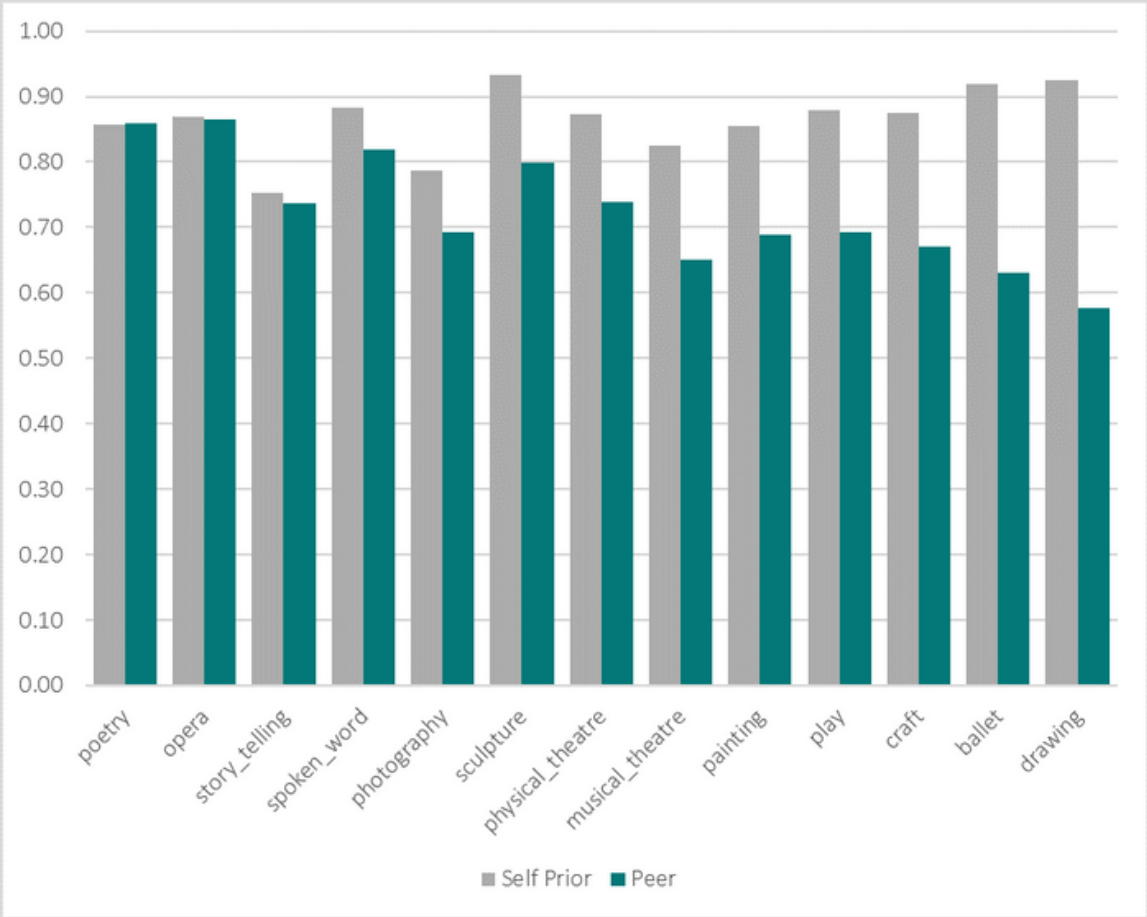
Thus far we have reported aggregate responses across all of the evaluations in the sample. We can explore the variation that sits beneath these aggregated results by exploring self and peer scores by detailed artform categorisations. Figure 15 below presents self prior scores for risk ('the artists / curators were not afraid to try new things') against peer scores. There is a noticeable variation in self prior risk ratings by artform. This suggests that as more data is gathered across artforms about perceived risk, it would become fruitful to facilitate some cross art form conversations about these results and how different artforms interpret creative risk.

Remember, as reported above, that at an aggregate level for all evaluations the self prior rating for risk (0.813) was somewhat aligned to the peer rating for risk (0.737) Figure 15 underlines that as with the self prior ratings, there is a considerable variation in the peer risk ratings by detailed artform.

In comparing the self and peer results, self assessors generally attribute more risk than peers to their work. What is also noticeable is a greater degree of misalignment between some sets of self prior and peer risk ratings by artform. So for example, with drawing, ballet, musical theatre, painting and play there are quite significant differences between self and peer ratings, with the peer ratings being much lower for risk in all those artform categories.

Terms are only included if there were more than 5 evaluations of that form evaluated in the study.

Figure 15: Average Self Prior and Peer Risk Score by Detailed Artform

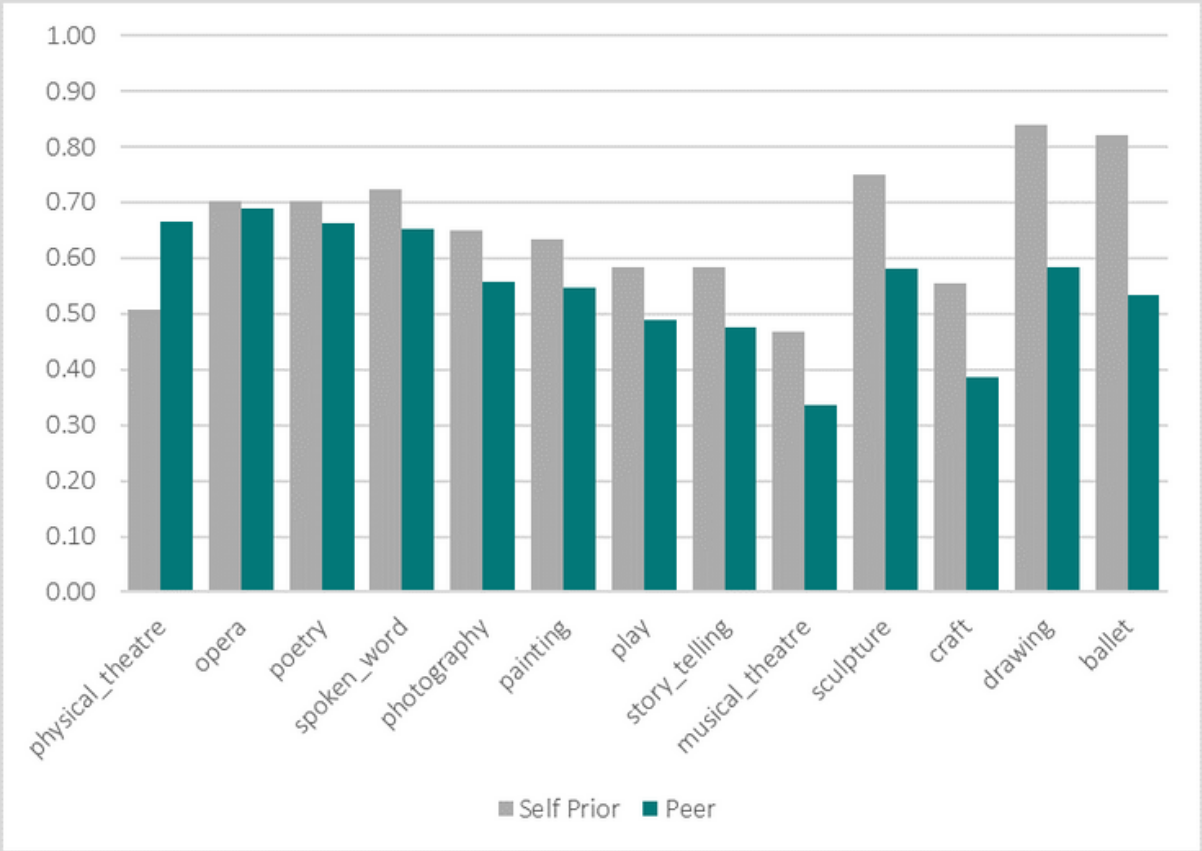


These results highlight how data of this kind could be used to provoke dialogue both within artforms, and across artforms, about what constitutes creative risk. For example, if the metrics become widely used across the cultural sector, and as we move from mid scale to big data across these types of dimensions, any persistent and marked variations in say risk ratings by artform; or between self and peer ratings on risk and originality by artform; could then by explored in detailed dialogue. One outcome might be that 'risk' as a dimension measure for self and peers is thickened up with additional metric components. Helpfully, the data will inevitably drive consideration of what is notable and discussion worthy for the sector, and where insight can be gained by further development and detailed enquiry.

Turning to originality, Figure 16 presents self prior and peer ratings on originality by artform. The aggregate self prior score for originality across all evaluations was 0.744. The aggregate peer rating for originality across all evaluations was 0.536.

As Figure 16 demonstrates that aggregate figure hides significant variations in how self assessors rated the 'originality' of the work being evaluated in this national test. This is what you would expect. The trial took place in a short six-month window in terms of evaluations, and cultural organisations were not necessarily 'picking' their best, or the most original work, in their current programme of current and planned activities. Moreover, no one could reasonably expect all work being presented to be judged as ground-breaking.

Figure 16: Average Self Prior and Peer Originality Scores by Detailed Artform

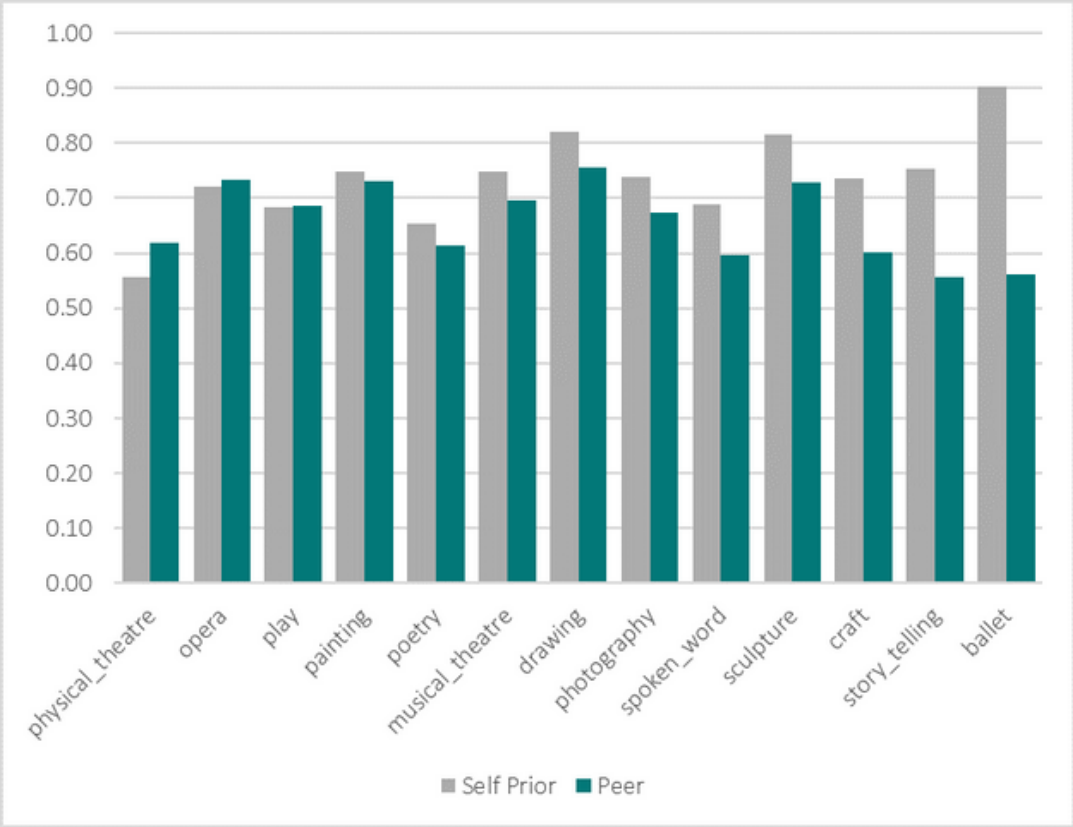


In addition Figure 16 demonstrates that the aggregate figure hides significant variations in how peers rated the 'originality' of the work by artform. In comparing the self and peer results, self assessors generally attribute greater originality to their work than peers. What is also noticeable is a greater degree of misalignment between some sets of self prior and peer originality ratings by artform. So for example, with drawing, ballet, sculpture, and craft there are quite significant differences between self and peer ratings, with the peer ratings for originality being much lower in all those artform categories.

The only artform in which the peer scores for originality were noticeably higher than the self prior scores for originality was physical theatre.

Finally, what of self prior and peer scores on excellence – 'it is amongst the best of its type I have seen'. Figure 17 presents the self prior and peer scores for excellence by artform. The aggregated self prior rating for excellence across all evaluations was 0.744. As with risk, and originality, there is significant variation in the self prior ratings for excellence by artform.

Figure 17: Average Self Prior and Peer Excellence Scores by Detailed Artform



There is a stronger overall alignment between self and peer ratings on excellence by artform, as compared to the self and peer profiles by artform on risk and originality. This is an interesting finding. As we have discovered throughout the course of the Quality Metrics National Test, which has helped support the peer review process central to this evaluation work, creative practitioners are constantly going to see each other's work to both be inspired but also to understand the state of practice in their artform and to equip themselves with a well informed perspective of what good looks like.

When judging the excellence dimension, they are making a relative judgement about where their own work, or if they were acting as a peer the work of others, sits in the national ecology in relation to being the 'best of its type I have seen'.

It is therefore an intuitive finding that given the frequency of exposure to 'benchmarks', and the breadth of context they have to make those judgements, self and peer scores have shown closer alignment on excellence by artform, than on risk and originality.

More broadly, the subtleties of the interplay between self and peer opinion suggests that over time if the metrics were used at scale, thought would have to be given to the process of supporting a bank of peers, maintaining a directory of their skills and expertise, and highlighting their relationship and experience to any given artform. This type of resource base would help cultural organisations to understand which types of perspectives they wish to draw on for their peer assessments, providing further context for them to interpret the peer ratings they receive for any given piece of work.

We noted under our reporting of self and peer ratings on originality, if some of these patterns were observed at a larger scale, they should be the trigger for dialogue and further research. For example, in this study ballet proved to be the outlying artform in terms the biggest difference between self and peer ratings on excellence.

3.7. How 'beige' are the results?

The term 'beige results' was coined by Nick Merriman of Manchester Museum in the original quality metrics pilot. A set of beige results occurs when the average dimension scores for public respondents are all grouped around the 0.65 / 0.70 mark, and in addition the standard deviation for the public scores on each dimension are also low (in other words display very little variation amongst respondents). The original Manchester group regarded 'beige results' as the audience equivalent of 'meh.'

The power of the 'beige results' description is that it highlighted that a successful piece of work does not mean that everybody has to 'like it' or agree about it. Indeed, a piece of work that sharply divides opinion across the dimensions is clearly having a significant impact on its audiences which may fully reflect the creative intentions behind the work.

How much variation did public respondents display in their reaction to the dimensions in this study? Clearly where respondents have given high scores in response to the dimensions this reflects genuine appreciation for the work they have seen, whereas some of the dimensions have received consistently lower scores, and all dimensions have received low scores from some respondents.

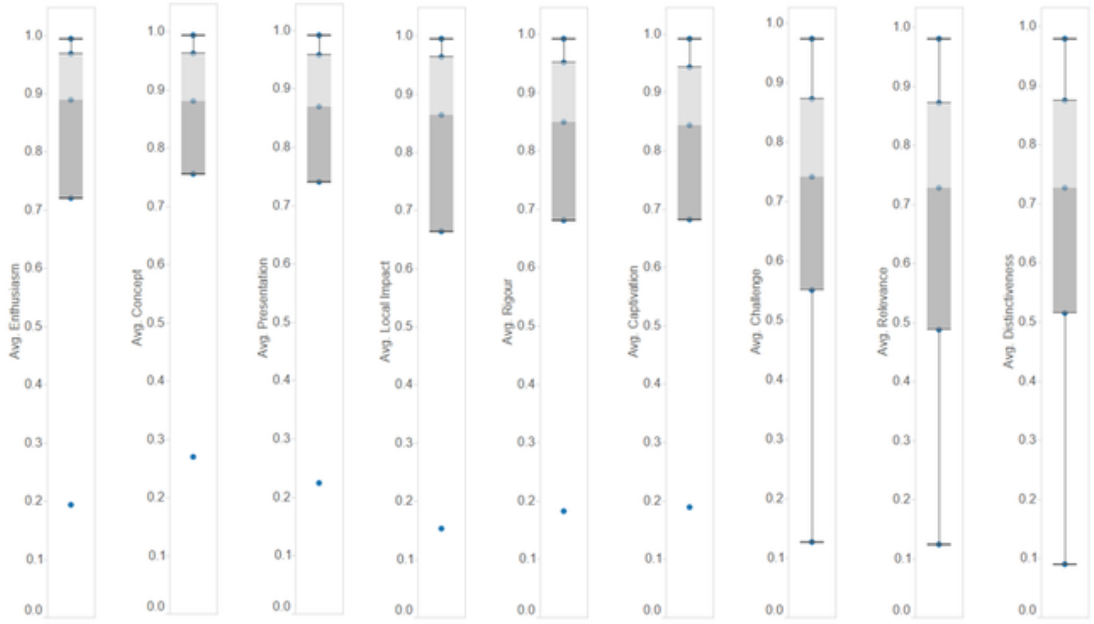
In analysing the data we have obviously asked ourselves whether some element of the maximum scores may be due to a growing familiarity with swipe mechanisms facilitated by touch technology coupled with the known behaviour of survey 'box ticking' rather than measured responses. For example, where we see a maximum score of 1.0 from a respondent we might suspect that to be a result where the emotional reaction to the question is a resolute 'yes' rather than a measured proportion of 'yes.'

In response, maintaining an accurate representation in the dataset for manipulation was done through calculating medians and interquartile ranges.

This means that the range of results is represented by five numbers for each individual evaluation's respondent category (self prior, self post, peer, public): the lowest value, the 25th percentile, the median, the 75th percentile, and the maximum value.

Figure 18 diagrammatically presents the variation in public response. It presents box and whisker' plots[17] for the public responses to each dimension. So for example, reading the plot for the dimension of relevance – second plot from the right, it plots the lowest public rating at the bottom of the chart area (a value of 0.1241); the highest at the top of the chart area (a value of 0.9801), and then the three values for the lower quartile (0.4881), median (0.7284), and upper quartile (0.8713).

Figure 18: 'Box and Whisker' Plot for public ratings by dimension



The box and whisker plots for each dimension are ordered by median – with the highest median to the far left (the enthusiasm dimension). An even distribution in each part of the plots is depicted with a box (with a centre-point of the median) and two lines (whiskers) which reach to the maximum and minimum values. In the first six plots, the downward line does not reach to the minimum values. This is because the minimum values in these lower quartiles are outliers. This tells us that a score on these dimensions near that minimum value is unusual and that when it comes to enthusiasm, concept, presentation, rigour, local impact and captivation although there are varying degrees of agreement, generally the work is positively scored.

What this chart shows very clearly is that on challenge, relevance and distinctiveness, audiences are divided. In other words, on the basis of the evaluations in this study, it is not unusual for a piece of work to be judged irrelevant, unchallenging or similar to previous work.

17 Box and Whisker plots are a way of mapping the five points in the interquartile range, representing the full range of results in a given set. http://www.basic-mathematics.com/box-and-whiskers-plot.html

Taken as a whole, considering the average public scores by dimension, and the variation in public response, the evaluation results across this sample of work are not 'beige.'

3.8. How Do the Regions Compare?

Introduction

In Chapter 2 we discussed event location and choice across the country. In this segment we present a more detailed regional analysis. We have generated that analysis by:

- Using the metadata for the location of each event evaluated
- Mapping the organisation level and event level data against ACE Regions
- Then cross referencing that location data for each event against the ONS classification for rural and urban areas (which is based on a six-fold categorization moving from strongly rural (rural 1) to strongly urban (urban 6)[18]

The ONS classification defines these rural-urban classifiers as follows:

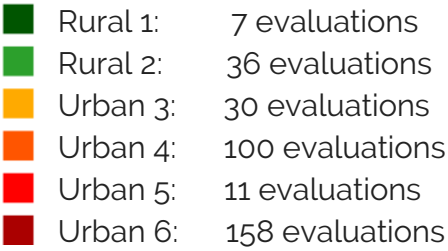
- Rural 1: Mainly rural
- Rural 2: Largely rural

These two categories together define those areas of the country that are 'predominantly rural.'

- Urban 3: Urban with significant rural
- Urban 4: Urban with city and town
- Urban 5: Urban with minor conurbation
- Urban 6: Urban with major conurbation

The urban 4, 5, and 6 categories together define those areas of the country that are 'predominantly urban.' Figure 19 shows the colour coding we are using for rural 1 to urban 6 and the number of evaluations carried out in each of these areas across the country. Clearly, the great majority of our evaluations were carried out in those areas of the country defined as 'predominantly urban.' Of the 374 evaluations in our data set, 43 evaluations took place in areas defined as 'predominantly rural', approximately 12% of the total evaluations. The 10 evaluations classed as shared were only counted once (as they are throughout the analysis). Twenty-two evaluations are not mapped here due to either data collected in multiple locations or they existed as works broadcast on the world wide web.

Figure 19: Rural / Urban evaluation frequencies and colour key



18 http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/2011-rural-urban/index.html

Figure 20 shows the detailed location breakdown of evaluations by the region in which the organisation is registered in. So for example, organisations based in the South West, produced 53 evaluations in total, of which 12 (light green segment) were in related to work presented to the public in Rural 2 locations; 3 in Urban 3; 34 in Urban 4; reflecting the distinctive geography of the South West as represented by the NPOs taking part in this study.

Figure 20: Total number of evaluations by rural / urban classification and ACE area

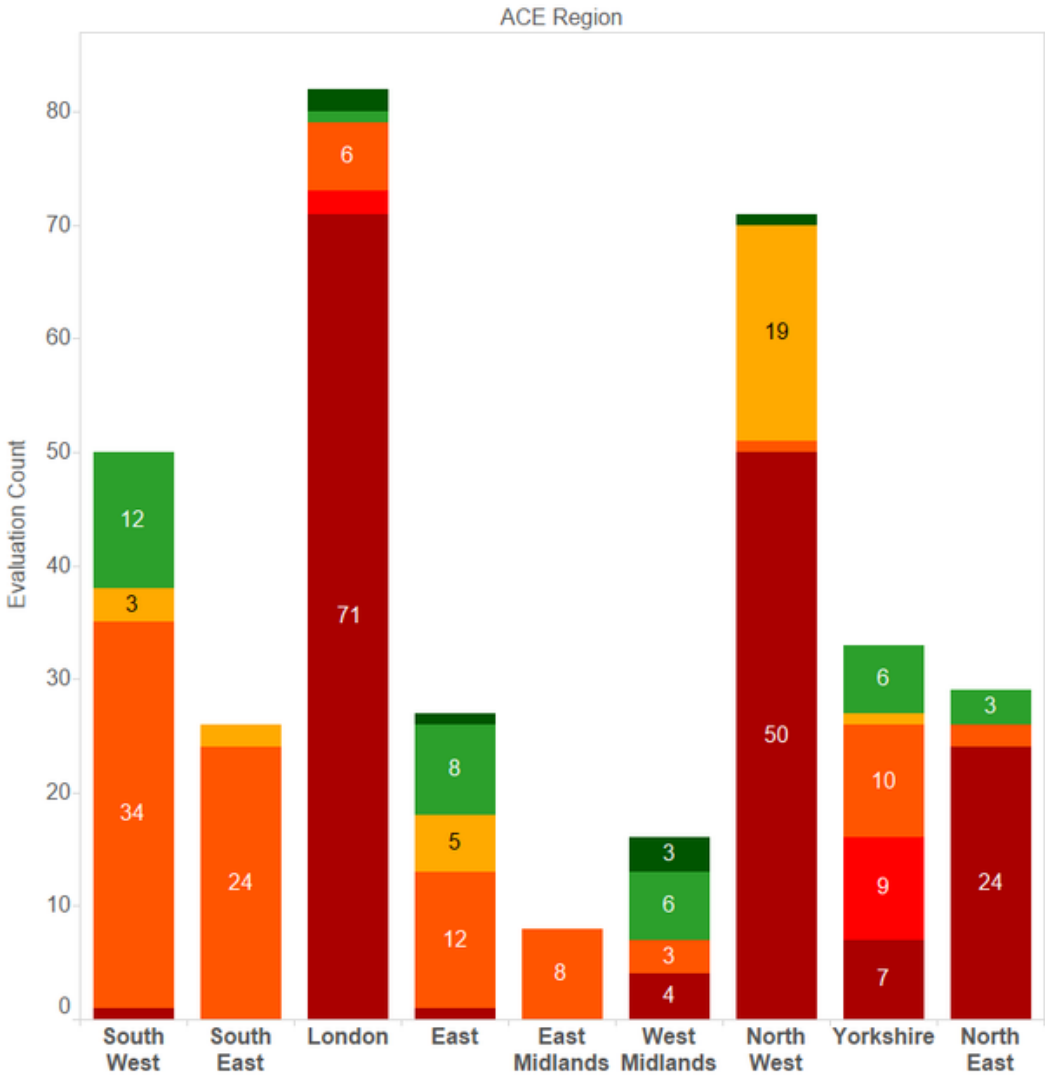
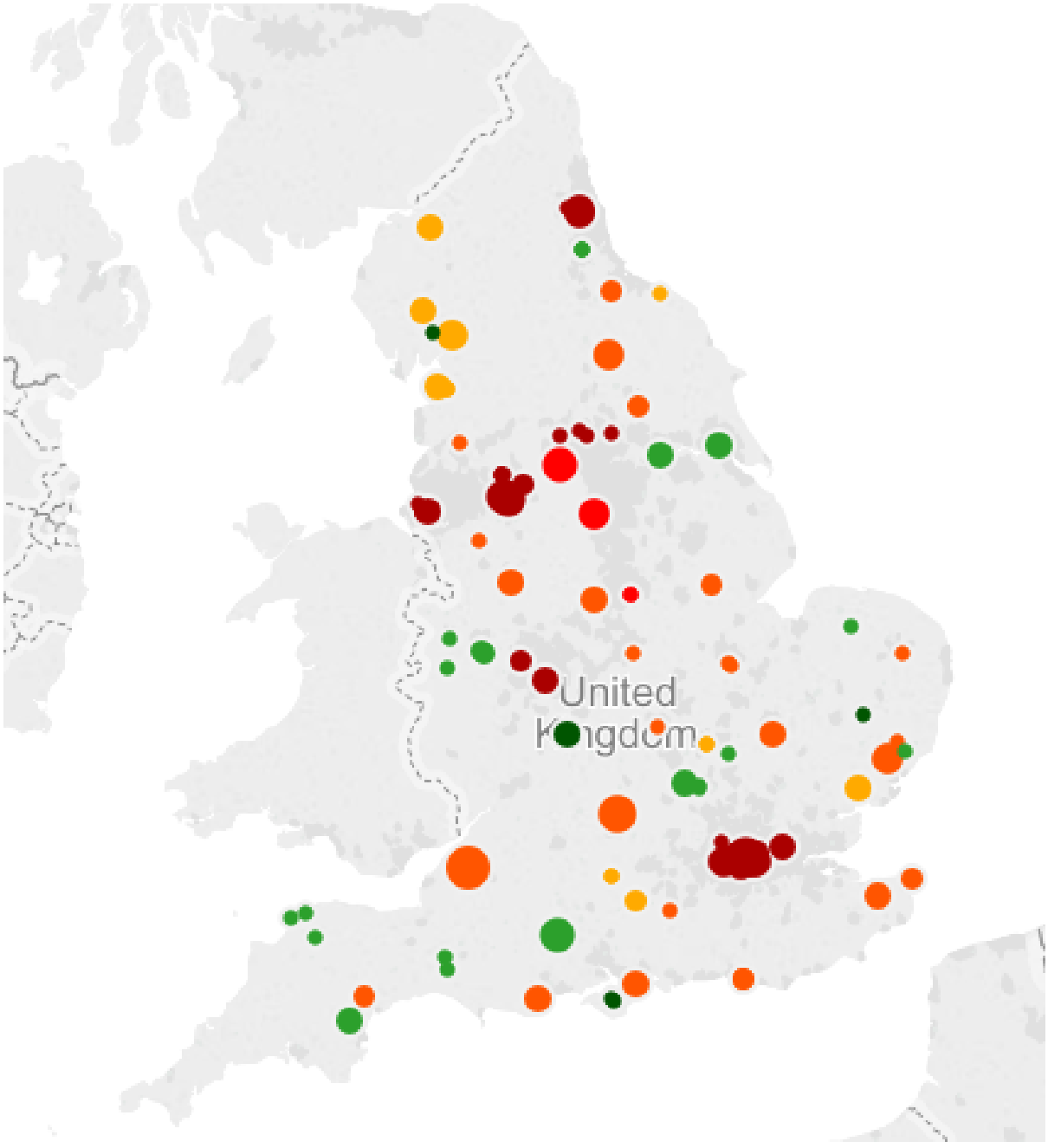


Figure 21 presents a geographical representation of that same pattern of evaluations across the country. The colour codes represent the 6 rural / urban classifiers, with the size of the circles representing the numerical concentration of evaluations in that geographical location. Remember that this Quality Metrics National Test was conducted through an Expression of Interest process, in which at the shortlisting process Culture Counts and ACE did their best to create a sample of participating organisations that was as representative as possible of both artform and geography.

Figure 21: Location of evaluations by rural / urban classifiers and frequency



3.8.1 Regional Analysis

The following charts present public respondent scores by the region in which the work was presented to the public. They are presented as three figure which are arranged purely by region: North (North East, North West, Yorkshire); Midlands (West Midlands, East Midlands, and East of England); South (South West, London, South East). It is noticeable that the East Midlands has higher scores on many of the dimensions, although it does have the smallest sample size as show in Figure 20.

Figure 22 - Aggregated public response dimensions across the North of England

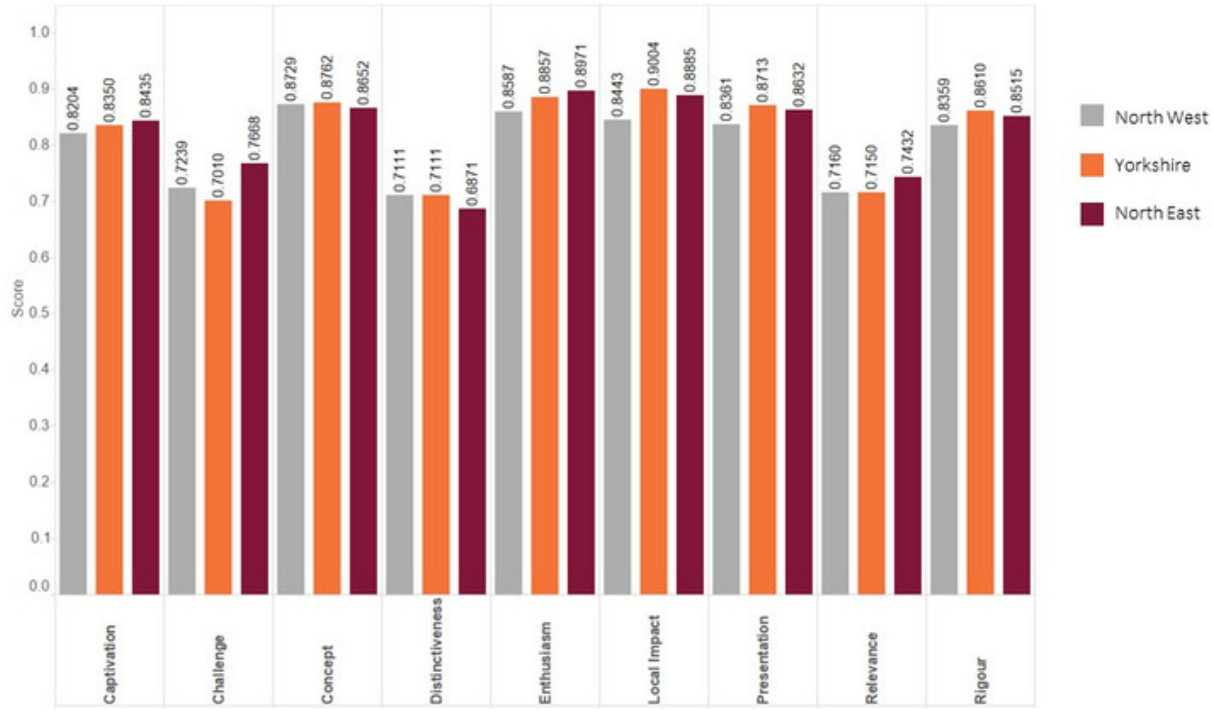


Figure 23 - Aggregated public response dimensions across the Midlands

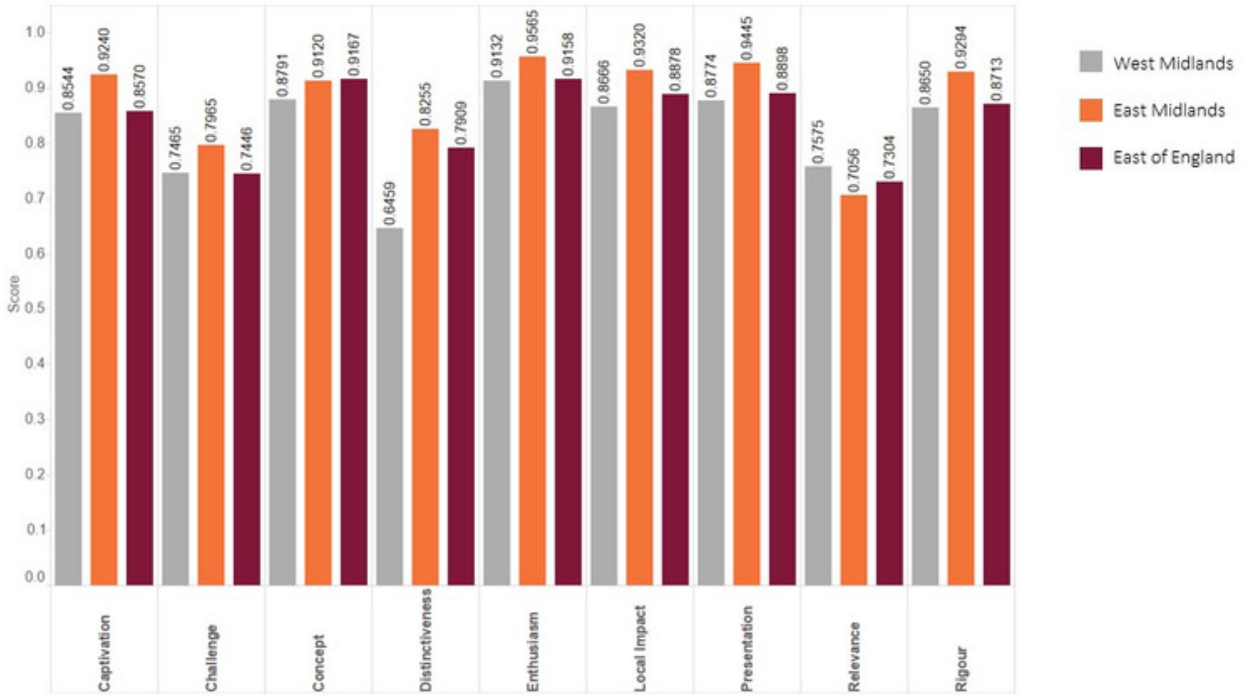
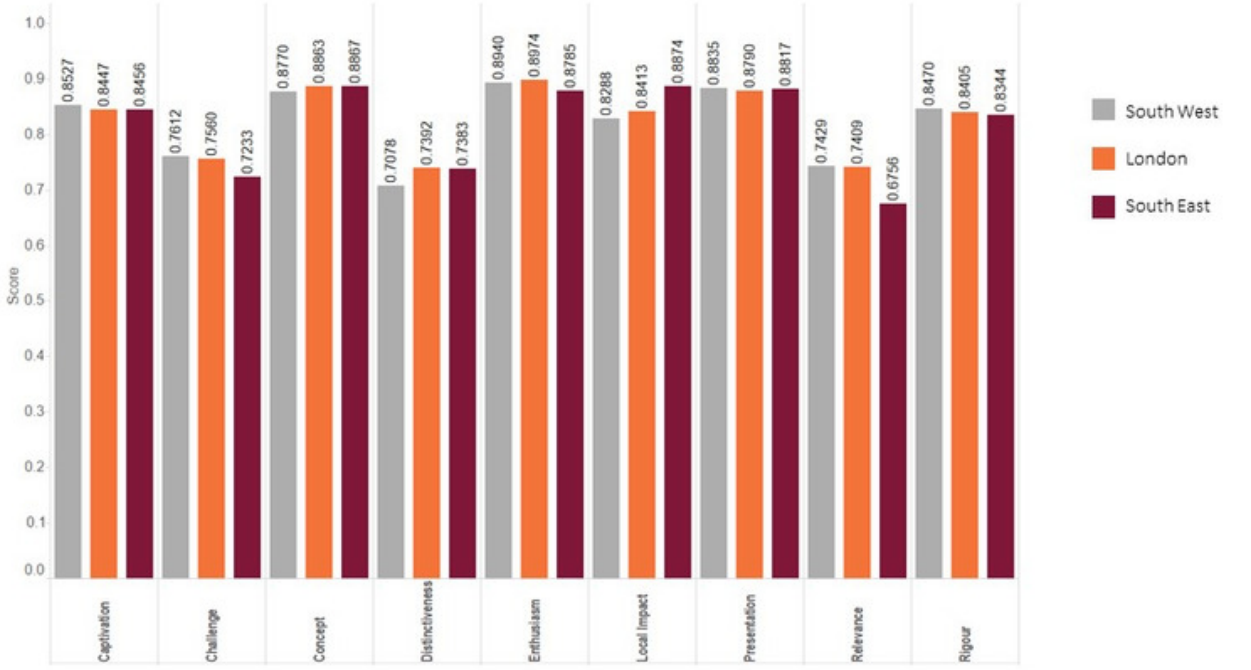


Figure 24 - Aggregated public response dimensions across the South of England



To what extent did evaluations in rural and urban areas attract different profiles in terms of dimension scores? Figure 25 summaries the aggregate public responses by rural / urban areas across seven of the dimensions (concept, presentation, distinctiveness, rigour, relevance, challenge, and captivation). As the chart demonstrates there are no consistent variations in the profile of public dimension scores as you move from the most rural areas (rural 1) to the most urban (urban 6). In other words, distinctiveness is not being rated much higher in urban as opposed to rural areas.

Figure 25: Aggregated public response for 7 dimensions across rural / urban areas

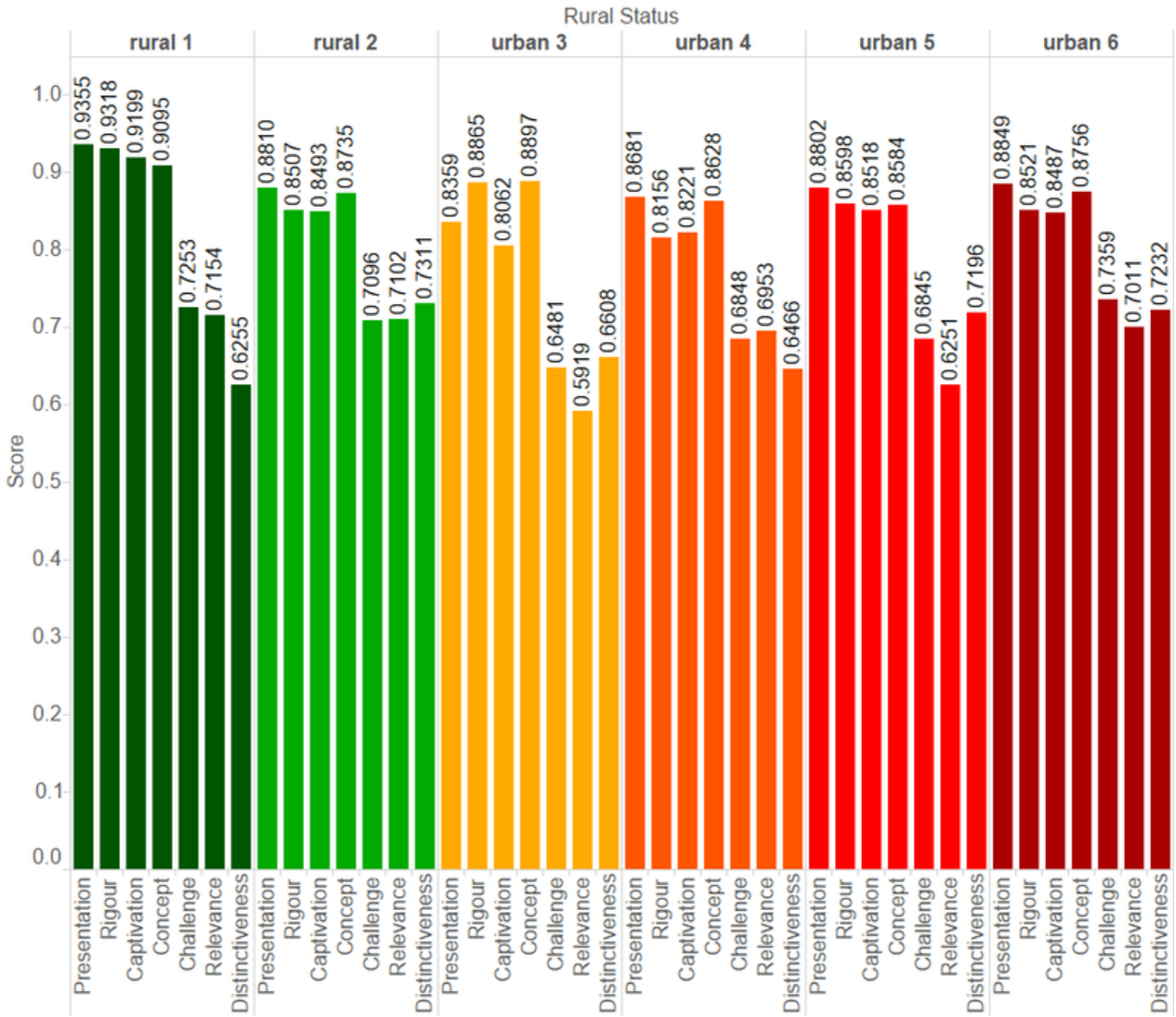
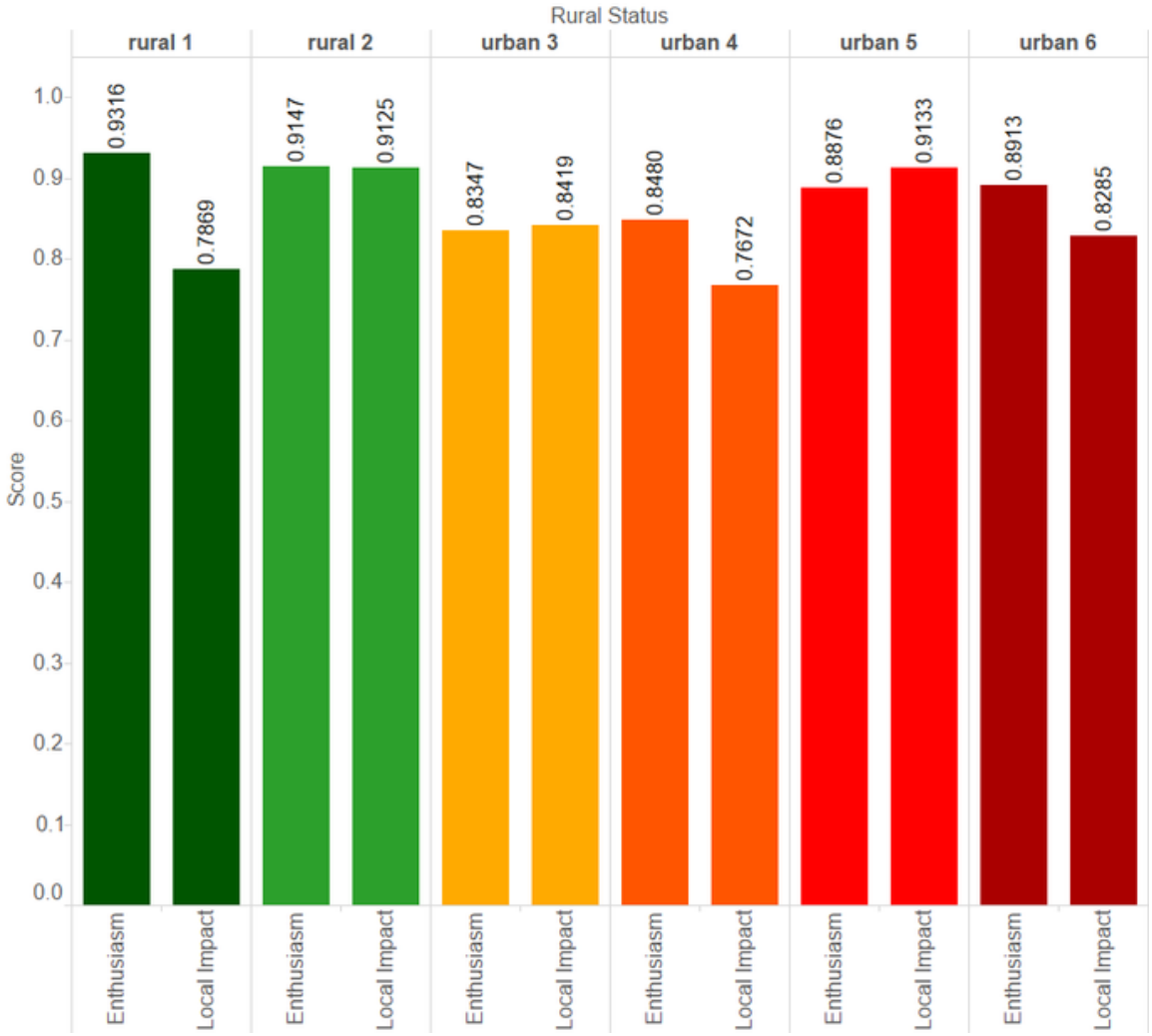


Figure 26 below presents aggregated public response across the rural and urban areas for the other 2 dimensions, enthusiasm ('I would come to something like this again') and local impact ('it is important that it's happening here'). Given the differential access to cultural provision in rural as opposed to urban areas, one might expect public ratings for 'enthusiasm' in more rural areas to be high, and they were (higher than in other urban areas). Similarly, one would naturally hypothesise that local impact scores would also attract high public ratings in rural areas. In this cut of the data, the rural status alone does not seem to have a strong influence on local impact scores.

However, it is also true to say that these findings show no evidence that well served audiences in strongly urban areas (urban 5 and 6) are 'jaded', or significantly less enthusiastic about the work they are viewing, than those in rural areas, or relatively less urban areas (urban 3 and 4). In contrast, enthusiasm and local impact scores were impressively high, particulalry in areas urban 5 and 6.

Figure 26: Aggregate public response for enthusiasm and local impact by rural / urban area

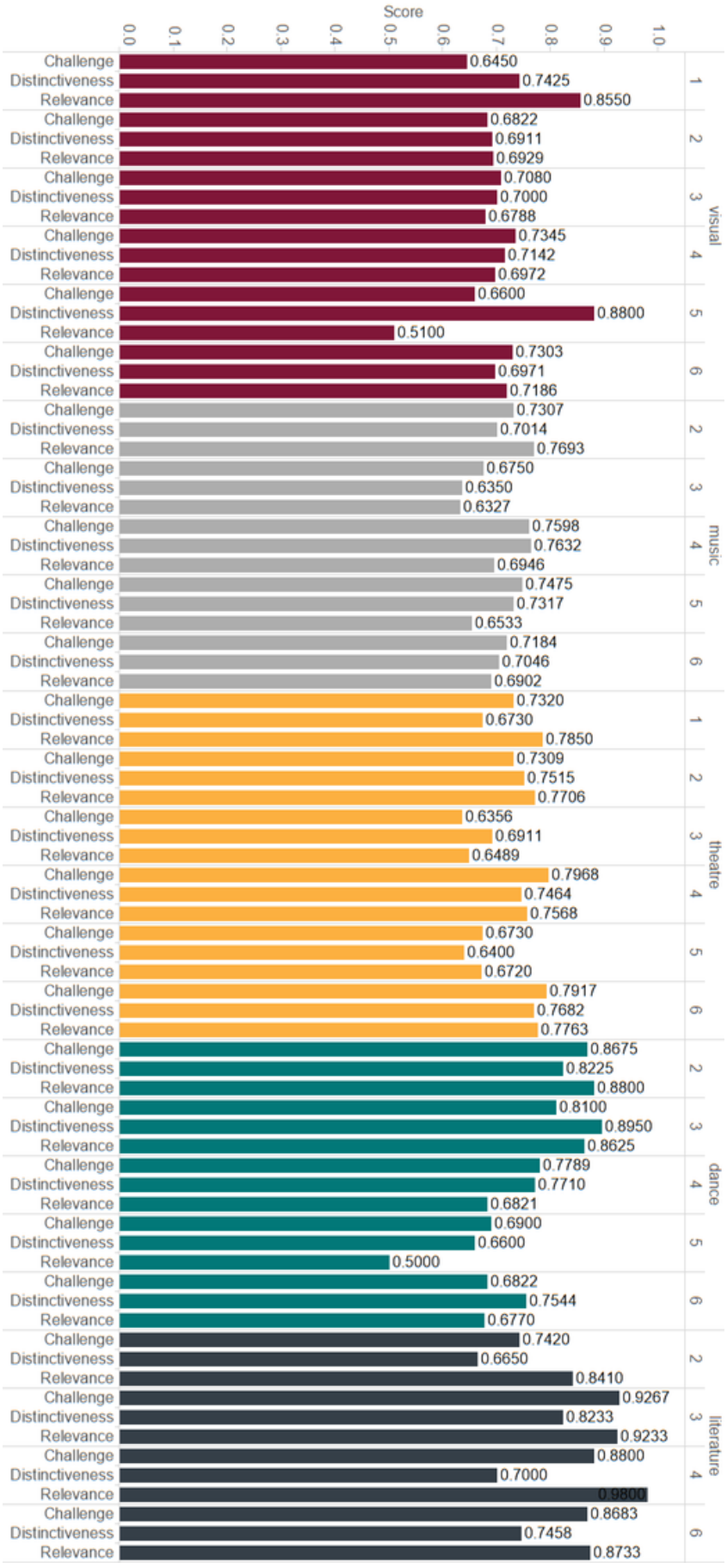


The regional analysis charts presented thus far include the evaluations conducted by the Royal Shakespeare Company (RSC), whose work takes place in a location defined as rural (Rural 1). They are therefore an anomaly in the funded portfolio, being a large scale recognised 'national' company, with an international reputation, but located and presenting work in an area defined as rural. In the Rural 1 area – 'mainly rural' – there were 7 evaluations in the study overall. The small sample of 7, which includes data from one of the largest organisations in the cohort, is unlikely to be representative of the full scope of work produced and/ or performed in the most rural communities in England.

Are there any noticeable regional differences by artform? Figure 27 presents aggregated public dimension scores across 3 dimensions (challenge, distinctiveness and relevance), cut by broad artform (music, theatre, dance, literature, and film), and then our rural / urban classification areas (numbered 1 to 6 for each art form with a matching set of public responses to the 3 dimension questions). It is interesting to look visually at the profile of the dimension scores (for the 3 metrics) by the rural / urban area for each artform. The chart shows quite clearly that music and theatre have quite similar profiles across the rural and urban regions in this study, whereas visual arts, dance, literature and film have more pronounced differences in regions and dimension profiles.

We would be cautious about this data, as of course we have much less data in some areas, but if the quality metrics are used widely across the cultural sector it illustrates the potential to start to map audience response by geographical location of the work and explore how far this may impact on the sentiments of public respondents to the quality of the work. This is particularly so when the audience profiles of the audiences are included within the analysis (Culture Counts users can view their results in the Culture Counts dashboard cut by age, gender, and download their results and matching postcode data into a CSV file for further analysis).

Figure 27: Aggregated public scores across 3 dimensions cut by artform and rural / urban area



3.8.2. Does touring work get different receptions in different places?

One obvious way to explore regional differences is to examine whether touring work gets a different reaction from the public in different places. Given the limited number of examples available in the test, it isn't sensible to look at patterns in the wider data however a case study from English Touring Theatre does provide some interesting insight to potential factors influencing dimension scores.

CASE STUDY: ENGLISH TOURING THEATRE

Does having a collaborative relationship with a host venue have an impact on dimension scores for touring companies?

With very special thanks to English Touring Theatre for sharing their results and insight for this case study.

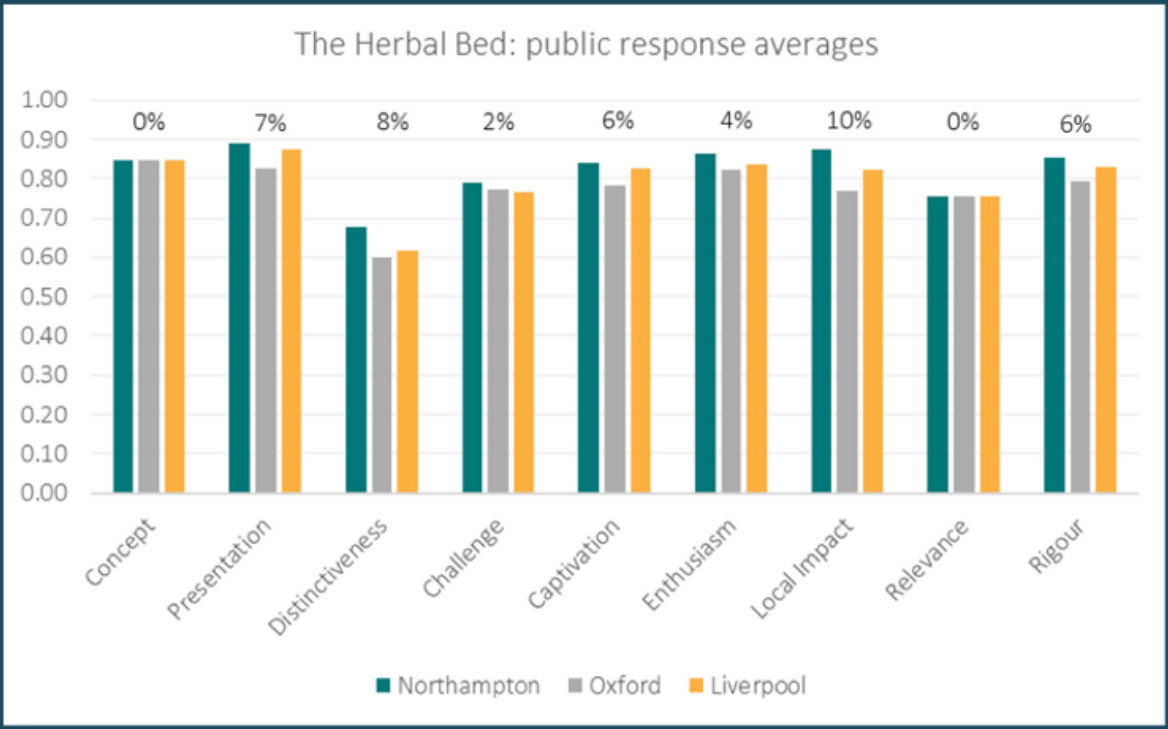
As a touring company, English Touring Theatre (ETT) had a different experience to many of the other users throughout the Quality Metrics National Test. They were using the metrics and the Culture Counts platform to evaluate one specific work, The Herbal Bed, in different locations across the country. Evaluating one piece of work in different locations enables the organization to compare the reception of the work across locations. And it wasn't just the audience's experience; ETT also measured the experience of peer and self assessors in each location.

As the staff were not able to travel to each venue, it was decided that the most practical method to collect public responses would be to use the 'Online' method, whereby an email would be sent to attendees after the performance to ask for feedback. In order for this to be successful, ETT needed to rely on the cooperation of the host venues. Offering to share their results with the host venue, this was generally not a problem. However, it certainly proved a challenge in one case, where one venue chose to not assist in the collection of public responses. This resulted in no public responses being gathered for that particular venue, which was naturally frustrating for ETT. That said, they were still able to gather peer and self responses which have contributed to their overall evaluation.

When comparing the public responses across the three locations, the shape of the graph remains constant. However, responses in Northampton were consistently higher than in Oxford and Liverpool, and generally the lowest scoring was present in Oxford. When comparing the results between Oxford and Northampton, the dimension with the largest difference in scoring was Local Impact, with a 10% difference between the two locations. Whilst this may not seem a large difference, when comparing with the average differences at 4.8%, its significance can be appreciated. Something worth highlighting here is that The Herbal Bed was coproduced with the host venue in Northampton.

This collaboration might have resonated with the audience in a particular way, causing them to feel a stronger attachment to the work, as opposed to other locations and venues. There could of course be other things playing a part – this is where perhaps looking at demographics and other cultural activities in the various locations could also be of interest.

The chart below compares the different scorings across the locations and the percentage in difference between the highest scoring location and the lowest for public respondents is marked above the bars:



Despite the smaller sample sizes of peer and self assessors, the results they present are interesting and highlight the importance of variety within the different respondent groups.

The self assessors were from the creative team at ETT, and yet they scored the production very differently. Is this due to the fact that each self assessor focuses on a different element of the production, and therefore takes a different approach to assessment? Or is it because their individual backgrounds within the cultural sector have caused them to receive the production differently? The different perspectives revealed in the self assessment highlight the value of using multiple assessors. Results such as these broaden the discussion surrounding creative intention.

The peer assessment presents a similar reception to that of the public assessment, where the Northampton production generally scored highest. The responses from peers can be subject to how well they know the work of the organisation and their previous experience as a reviewer.

The results largely mirror with what could be expected in comparison with broader trends emerging from the Quality Metrics National Test: for example, the difference between the public and peer scores for the Distinctiveness dimension is large. In addition, the peer assessors tend to score lower than the self assessors.

Upon greater reflection, it seems that the significance of the coproduction between the host venue in Northampton and English Touring Theatre must not be underestimated, as it seems from the scores that the experience of that production was more positively felt by all the respondents. The value of the insight being that ETT could explore this notable difference as a point of reflection when reviewing the production and planning future work.

If more data is collected on regional touring work, allied to a more detailed analysis of audience breakdowns and other relevant metadata, some enduring patterns may emerge in particular places and / or presenting venues. For example, it is possible that the reputation of a hosting company / presenting venue may have an inflating affect on some aspects of public dimension ratings. Or may be that a strong local impact effect is visible that relates directly to rural / urban classifiers. We would need more data to start to confidently explore these types of questions.

3.9 For the same evaluations are there any marked differences between online and interviewer modes of data collection?

We had 14 evaluations which used both interviewer and online methods of data collection where at least ten responses were collected by each method.

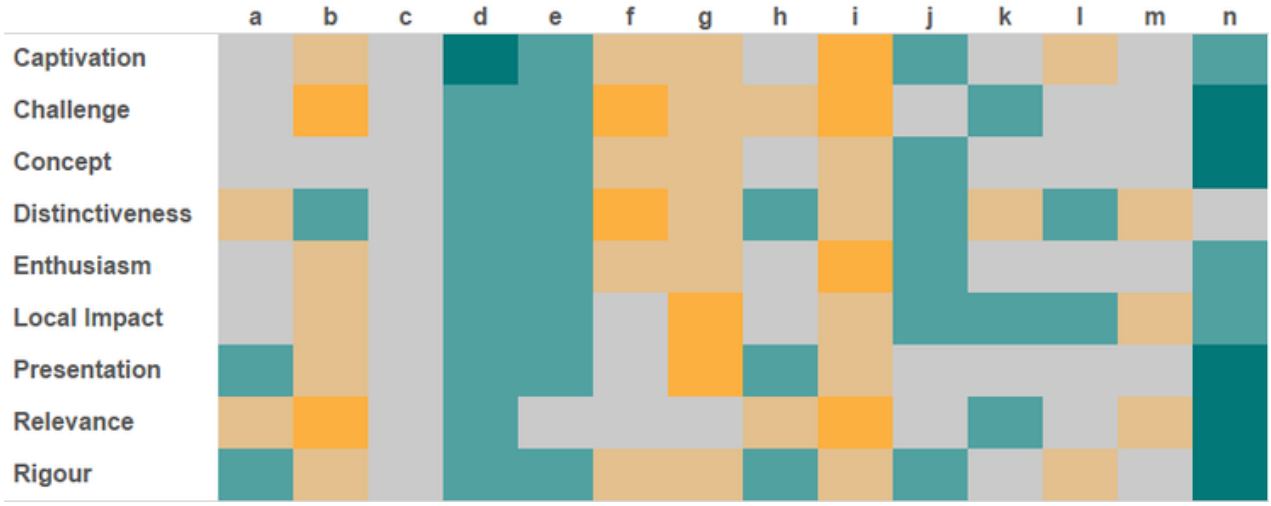
In 4 evaluations, there was a consistent pattern towards higher dimension scores for online surveys. In a further 4 there was a consistent pattern towards higher dimension scores collected via interviews. In the remaining 6 there were no patterns between the two delivery methods. The difference between scores were picked up if there was greater than a 5% difference between delivery methods. In the 8 evaluations with clear patterns, the differences were present for at least 3 different dimensions. In 2 evaluations, a difference of more than 10% was picked up for at least 3 dimensions (1 had higher scores by interview, the other via online delivery). None of the dimensions were more likely to have a discrepancy between delivery method than another.

This sample suggests that there is no direct or consistent effect of the delivery method on dimension scores.

Given the small sample size – both in total evaluations and the very low threshold of a minimum of ten responses via each method, it is difficult to confirm whether this would be observed at a larger scale.

Figure 28 visually demonstrates the inconclusive results of the delivery method in the trial. Squares that are teal coloured are where a dimension scored higher by online delivery method by at least 5%. The darkest teal squares are where there was at least 10% higher scores captured online. Pale orange squares are where dimensions scored higher by interview method by at least 5%, with the brighter orange on dimensions in specific evaluations scoring higher by at least 10%. Grey squares are where there was less than 5% difference either way.

Figure 28: Differences in public ratings by mode of data collection



The results are also suggestive that there may be an indirect relationship between the delivery method and the way in which dimensions are scored as individual differences are observed on some evaluations for both methods equally. This is may be a related factor to interviews which is different to the related factor to online delivery methods, potentially explaining the differences in both directions.

4. CHAPTER FOUR : Digging Deeper into Data

4.1 Introduction

In designing the data collection and analysis strategy for this Quality Metrics National Test, Culture Counts were clear from the outset that a suitable artform categorisation, representing the breadth and diversity of the work being evaluated, was needed to meet our aspirations to bring real granularity and subtlety to our aggregated analysis.

In creating that artform categorisation there was a clear opportunity to obtain terms used by the sector and represent them in a semantic data model. Not only does this enable broad artform categorization and aggregate group analysis but it also means that individual attributes of work, such as whether it is contemporary, immersive or performed by community artists can also be grouped and interpreted in the aggregate results.

This chapter explores some early analyses using data cuts exploiting those artform attributes. There are many more possibilities for exploration enabled by this dataset and the flexibility of this categorization enables combinations of attributes and artforms that for example accurately describe work that is multi-disciplinary. Moreover, it enables a growing way of categorizing work produced by Major Partner Museums (MPMs), particularly around subject matter.

The terms referred to throughout this paper are used in a minimum of 5 evaluations. With a larger dataset for robust analysis of the influence of artform attributes, this number would be higher, but for the purpose of testing the potential of these methods, creating a starting set to demonstrate some of the questions that can be asked of the data, this was deemed the smallest figure that would retain anonymity of the contributing underlying data. It should also be noted that attributes were not added to the metadata where it wasn't specified, so not all performances in the cohort will have been labelled with 'performance'. This was because it could not be accurately inferred computationally with the data available.

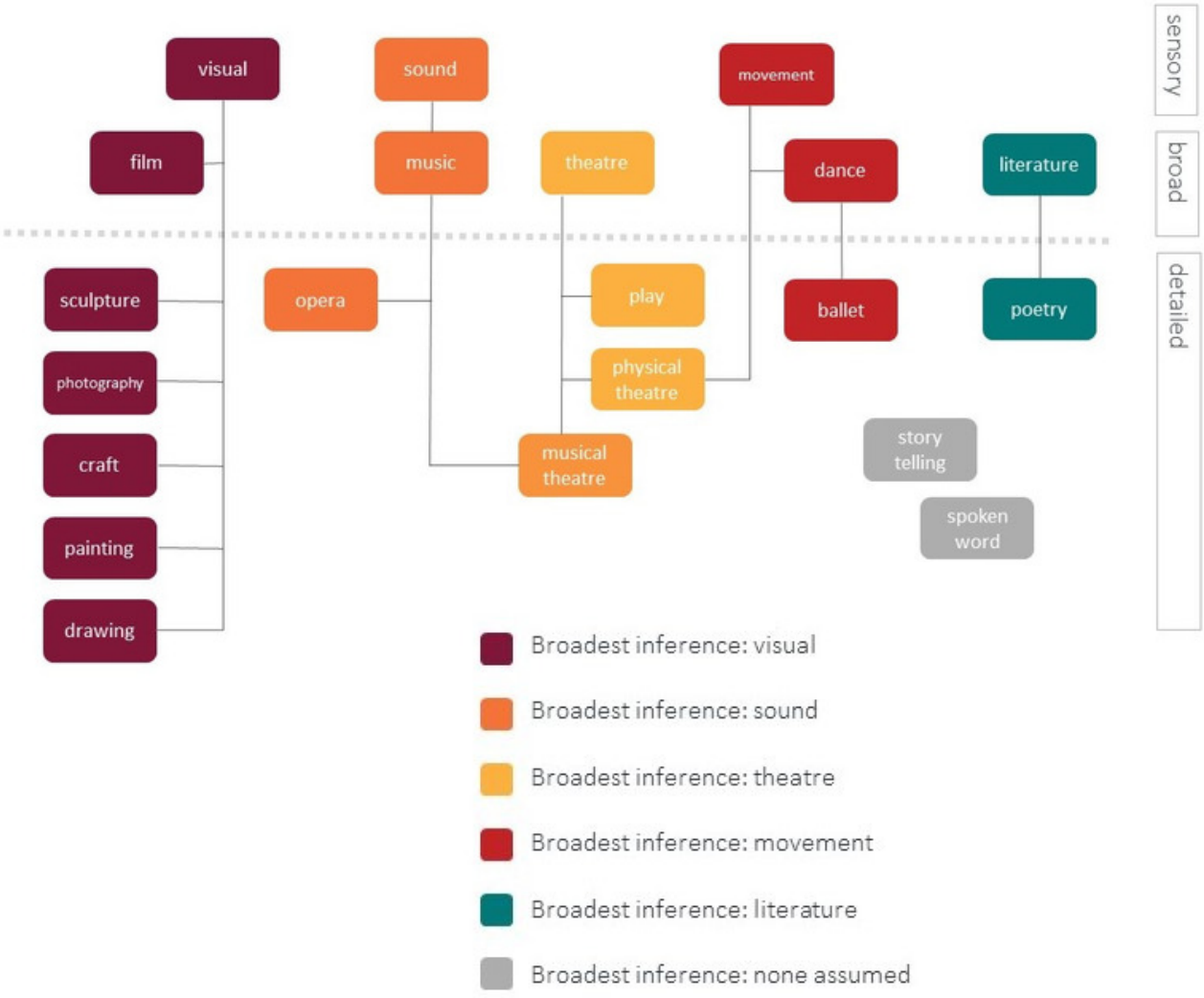
Based on the terms provided to us by the participating organisations, and using the approach outlined in Chapter 2, artforms were defined with logic developed by the Culture Counts team. Examples of this logic is as follows:

- Sensory artforms: sound, visual, movement
- Broad artforms: theatre, music, dance, literature, film
- Artforms are defined in broad and narrow terms. e.g. all music is sound, but not all sound is music.
- Visual is a broader term for film
- Sound is a broader term for music.

- Movement is a broader term for dance
- Detailed artforms must be associated to a broad and/ or an appropriate sensory artform:
- Visual is a broader term for sculpture.
- Movement is a broader term for dance, which is a broader term for ballet.

Figure 29 illustrates the artform categories that were used in at least 5 evaluations in the national test. Any number of artforms can be assigned to a piece of work, and in some cases inferred terms were used. Inferred terms were only assigned where a narrower term was used without a corresponding broader term as suggested above. It is noted that these terms may require grammatical consistency in future iterations of this data model but it was decided here that the language provided by organisations should be retained.

Figure 29: Inferred Artform Terms



This diagram is purely for illustrating the lines of inference. It does not cover the full spectrum of artform terms that can be accommodated within this model, nor does it indicate the cross-artform possibilities of the work evaluated in this study. For museum-curated work or combined art presentations in our analysis, artform attributes have been assigned and are not covered in this diagram.

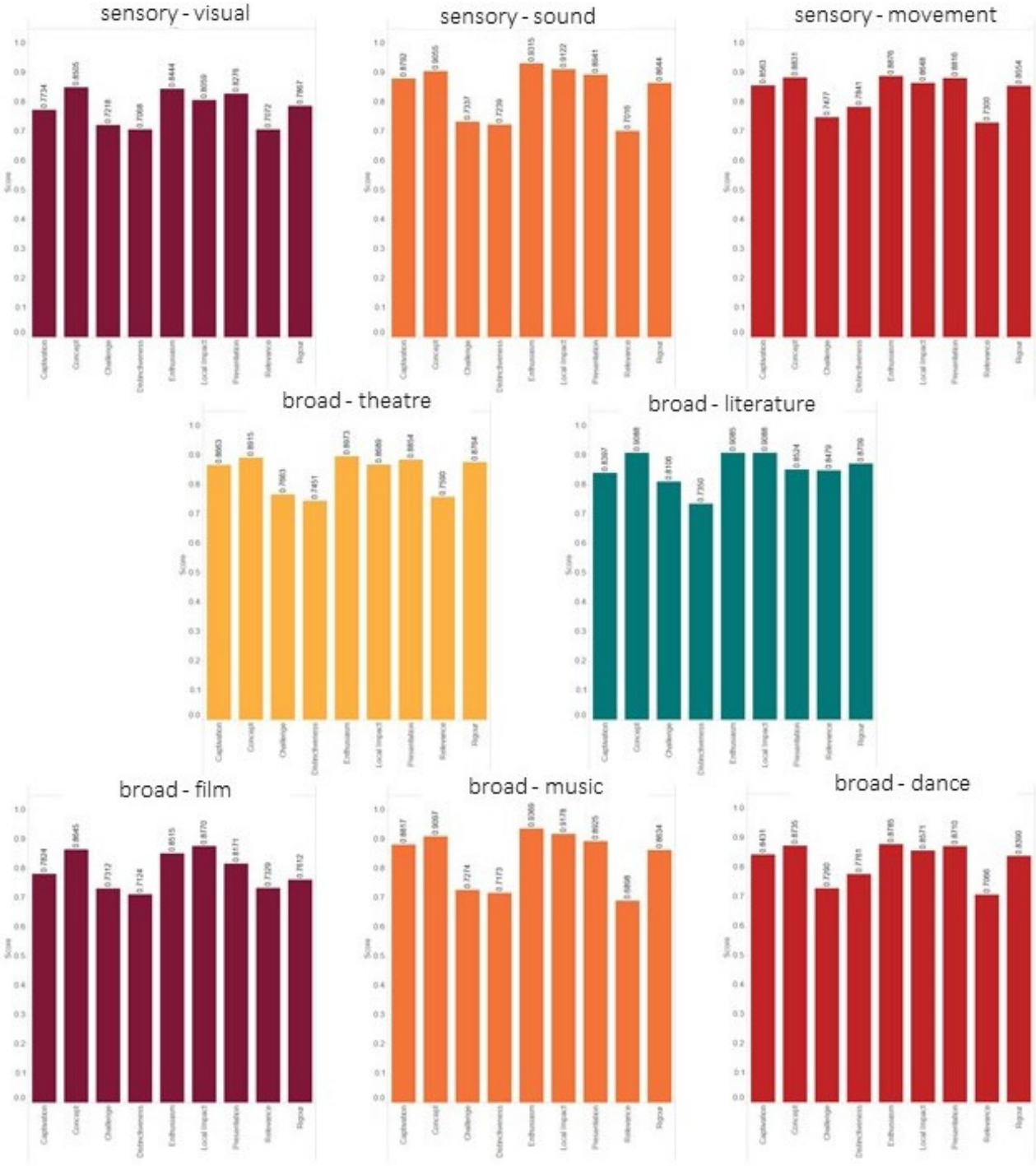
It is also noted that the key is a guide and in some cases an artform crosses into two broader categories, for example physical theatre is both a type of theatre and has a sensory form of movement. It can also be argued that some artforms, e.g. theatre, is a sensory amalgamation, hence the use of it as a distinct broad term.

A further use case is that there are many more dance works covered in the evaluations in addition to ballet. In other cases, the dance had attributes rather than detailed artforms, such as contemporary or hip hop. More detail is provided on artform attributes later in the chapter.

An initial aggregate analysis looked at these broad and sensory categorisations to view any overall dimension profiles (see Figure 30). As this chart clearly shows, the (public) profiles look very similar.

Broad artforms and sensory artform categorisations do not, generally, have distinctive dimension profiles, notable exceptions are for literature, and the public scores for challenge and relevance. Looking generally at sensory artforms – visual sits below the average and sound sits above it for 'high scoring' dimensions (concept, captivation, enthusiasm, local impact, presentation, rigour) and movement sits above the average for distinctiveness – a 'lower scoring' dimension.

Figure 30: Sensory and Broad Artform Profiles – Public Respondents[19]



19 See Appendix 2 for self and peer comparison versions of these charts

4.2 Do multidisciplinary pieces of work produce dimension profiles that are distinctive as compared to single art presentations?

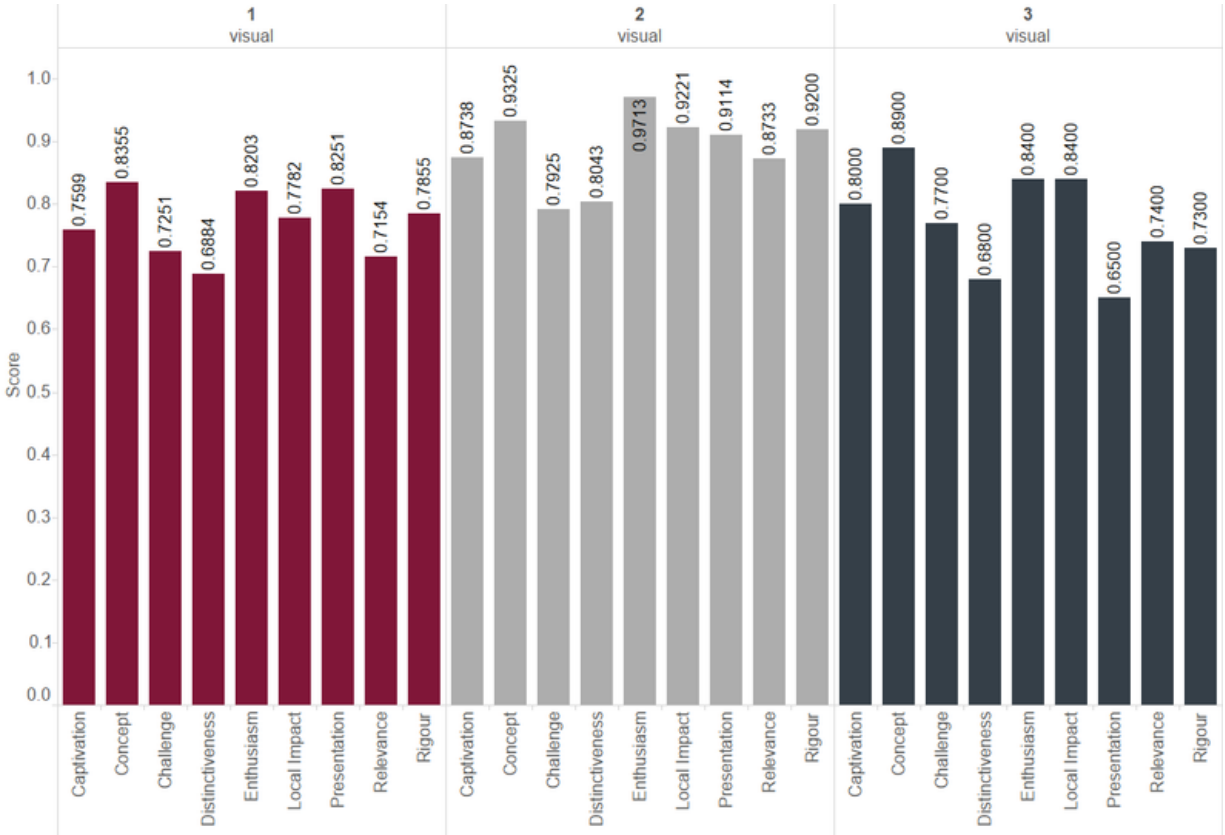
As the following figures show, multidisciplinary work does seem to produce different dimension profiles than work described with purely one artform.

The charts can be read with the coloured section to the left as a single artform; light grey as that artform with one other artform defined; and dark grey with two other artforms defined.

The categories chosen to be explored cover a large sample of the work in the study. Film as a broad term falls under visual work and was tested for differences[20].

The differences observed across dimensions do not fall in clear patterns in this cut of the data. This may be due to the mixture of artforms in groupings for two or more artforms which could have any number of combinations, or may indeed have differences to do with the nature of multi-disciplinary and combined arts presented in different ways.

Figure 31: Multidisciplinary Profiles - Public Respondents: visual



20 See Appendix 2, Chart A1

Figure 32: Multidisciplinary Profiles - Public Respondents: music

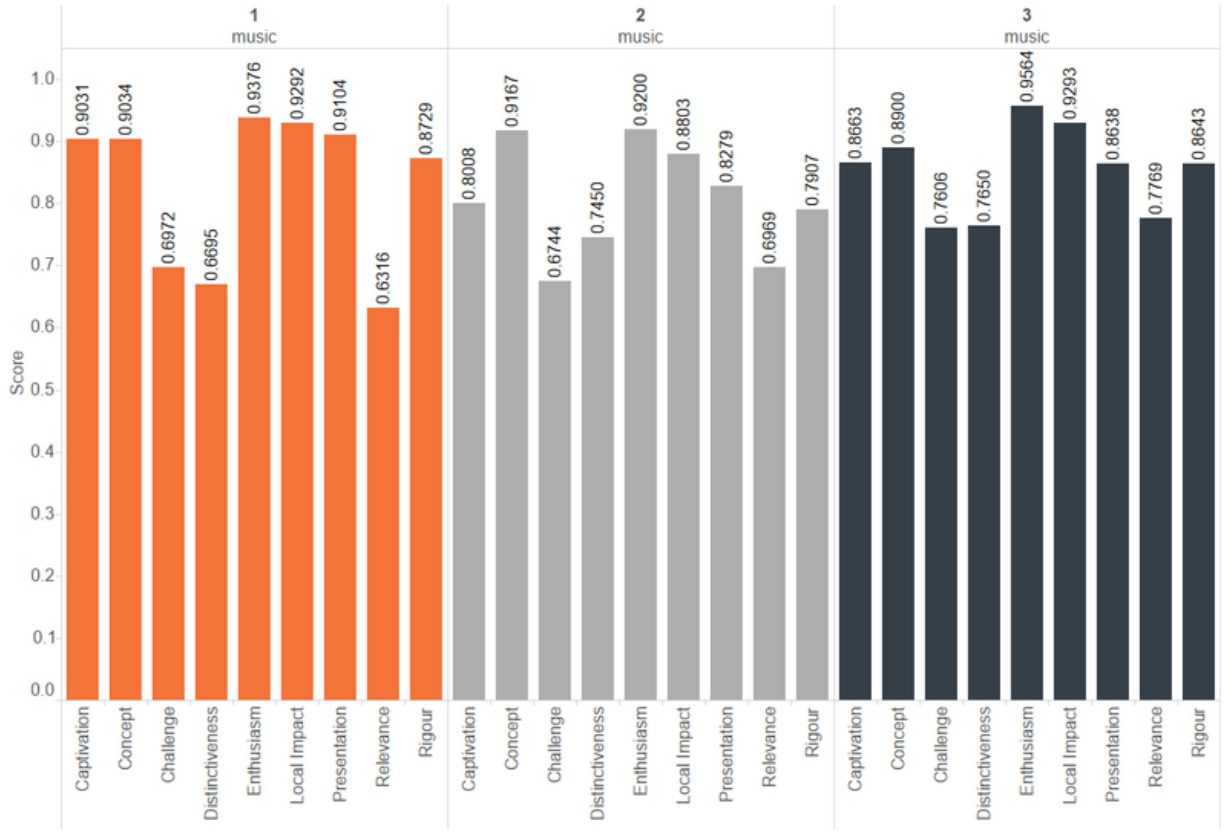


Figure 33: Multidisciplinary Profiles - Public Respondents: dance

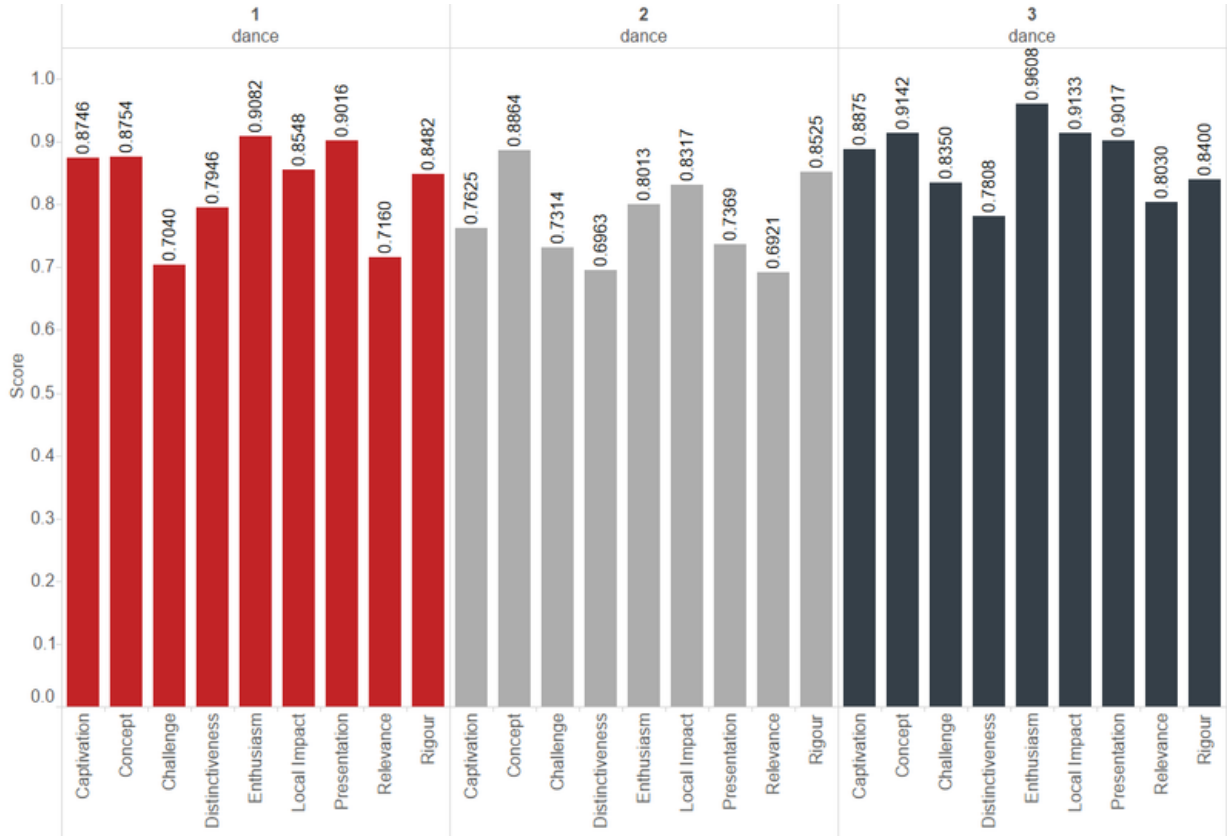


Figure 34: Multidisciplinary Profiles - Public Respondents: theatre

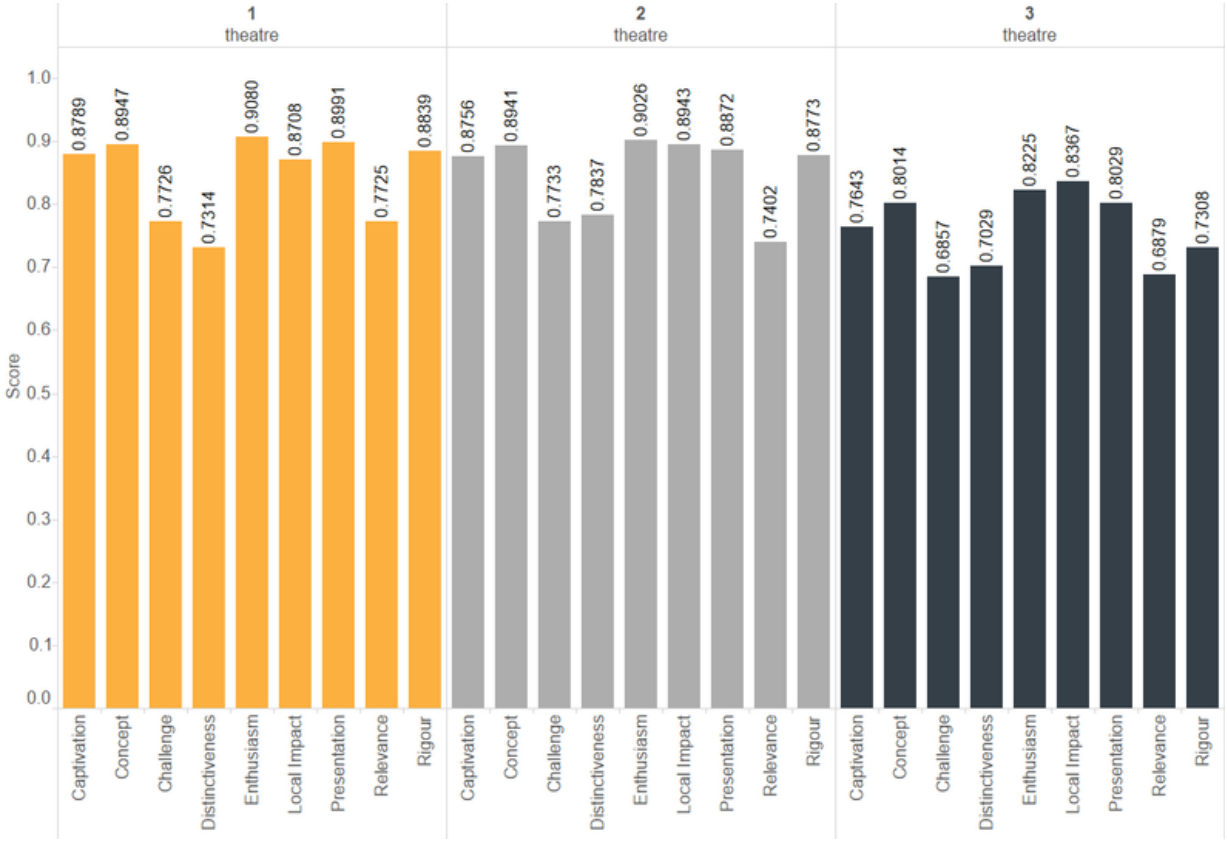
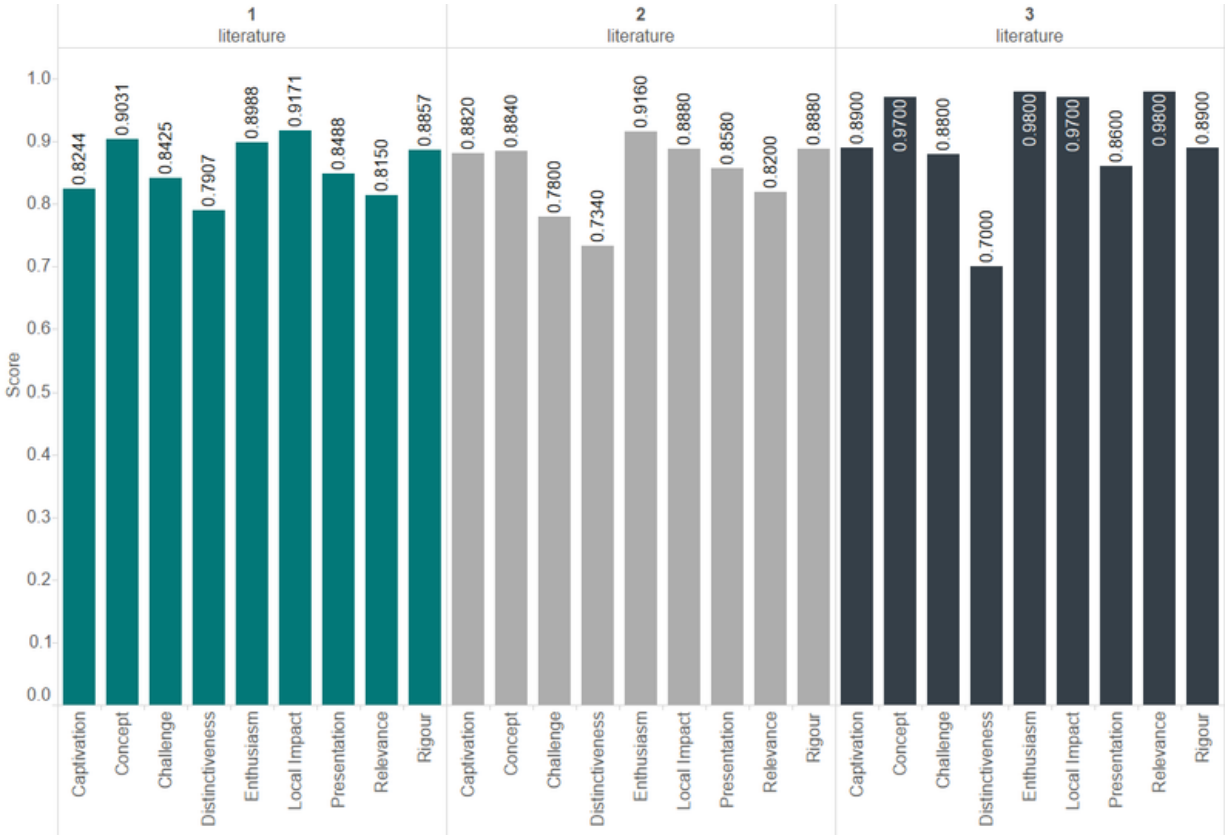


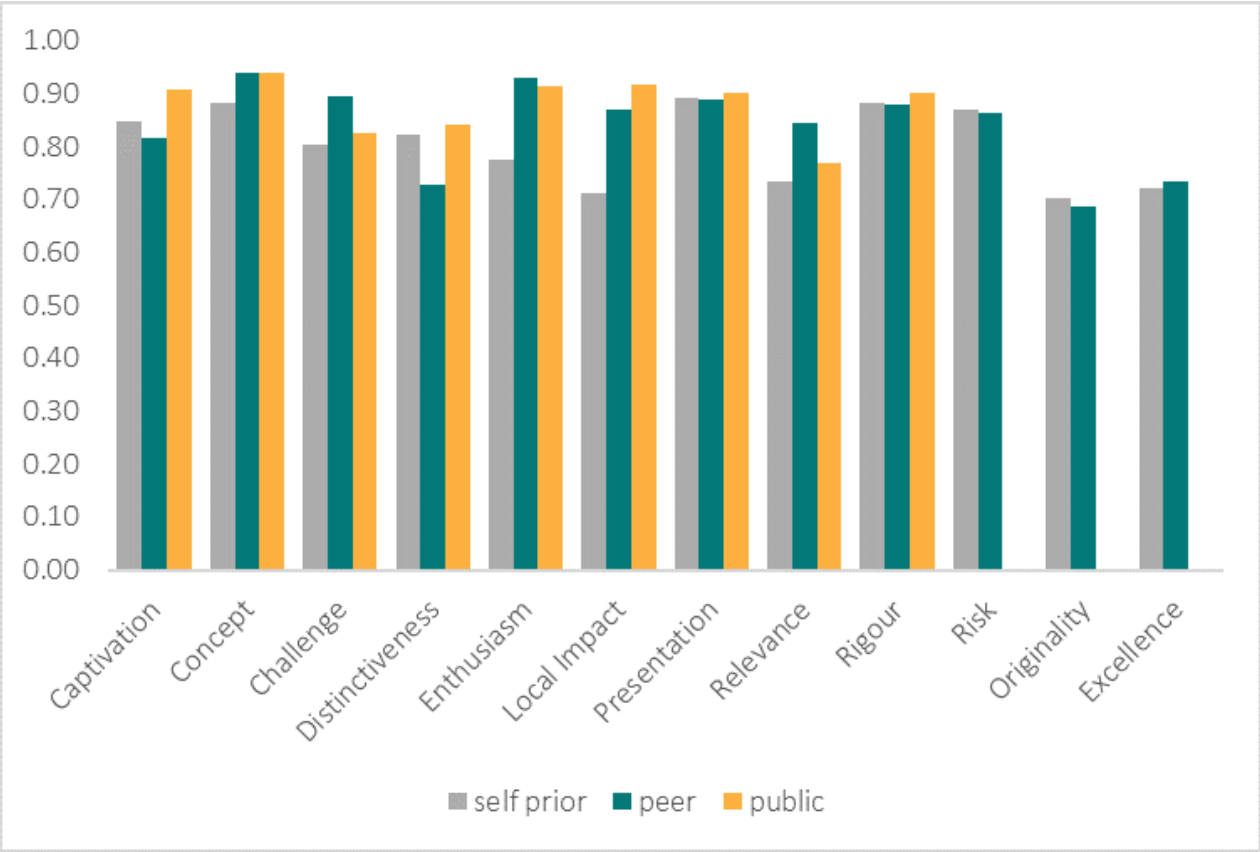
Figure 35: Multidisciplinary Profiles - Public Respondents: literature



4.3 Detailed Artform Analysis

What of detailed artform comparisons? Perhaps as one might intuitively expect, different artforms do have distinctive dimension profiles – but this only becomes clear when detailed artforms are considered in their own right. Figure 36 is a respondent comparison chart looking at the detailed artform term of opera.

Figure 36: Detailed Artform Respondent Comparison: Opera



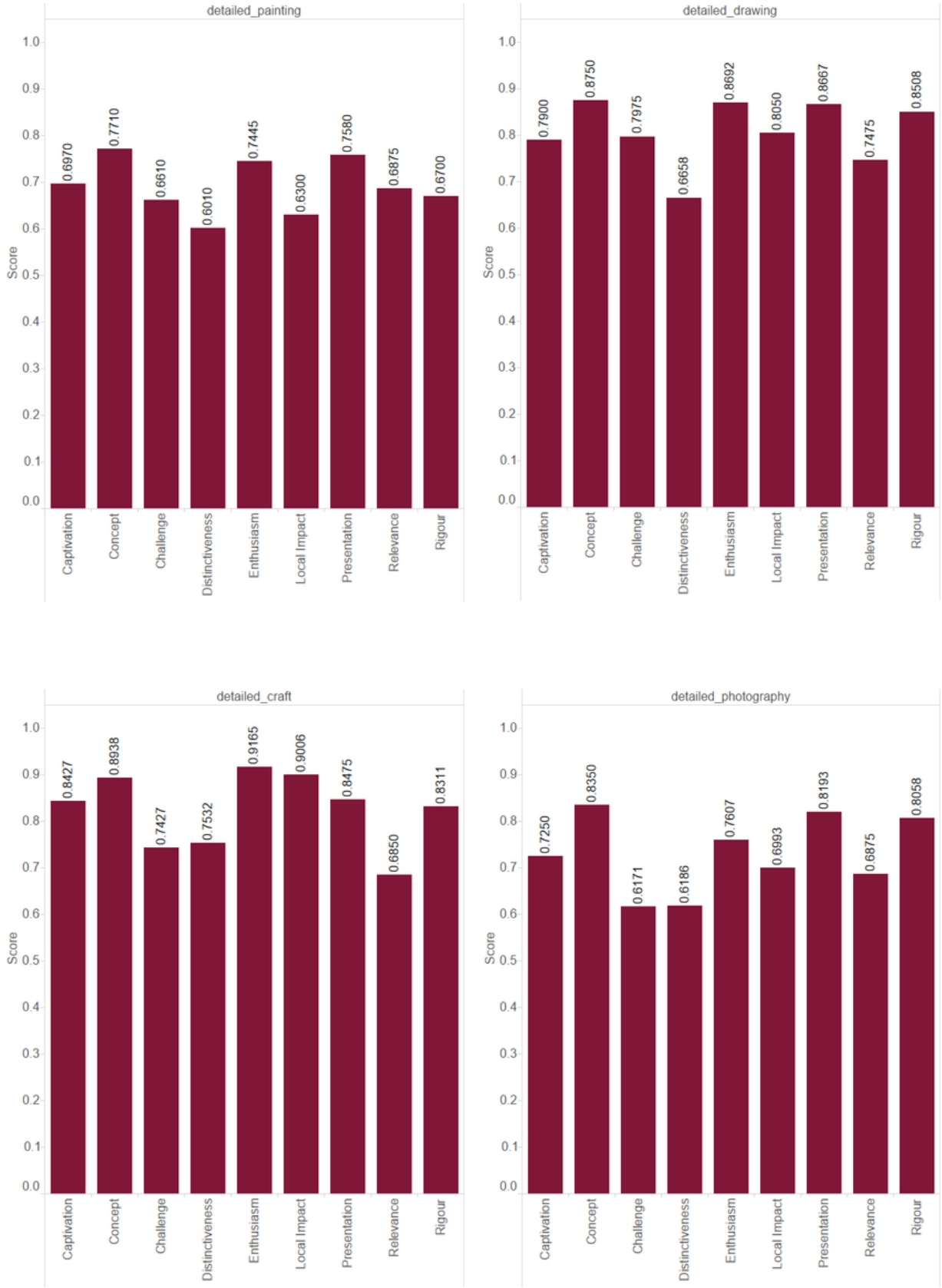
Interestingly here, peer and public scores sit above the overall aggregate baseline average scores. Moreover, the three lower scoring dimensions – distinctiveness, challenge and relevance – do not follow the aggregate pattern for music or sound either. This starts to show the influence of opera as an inflator on these dimensions in this particular study. It may be of note here that the opera work included in the trial may not be representative of opera 'at large'.

Further data cuts (with more data available) could consider these opera works with respective attributes, such as presentation (broadcasting and live performance), chronological interpretations (contemporary, new, revival), or key performers or directors (potentially cross-referencing with other datasets such as operabase[21]).

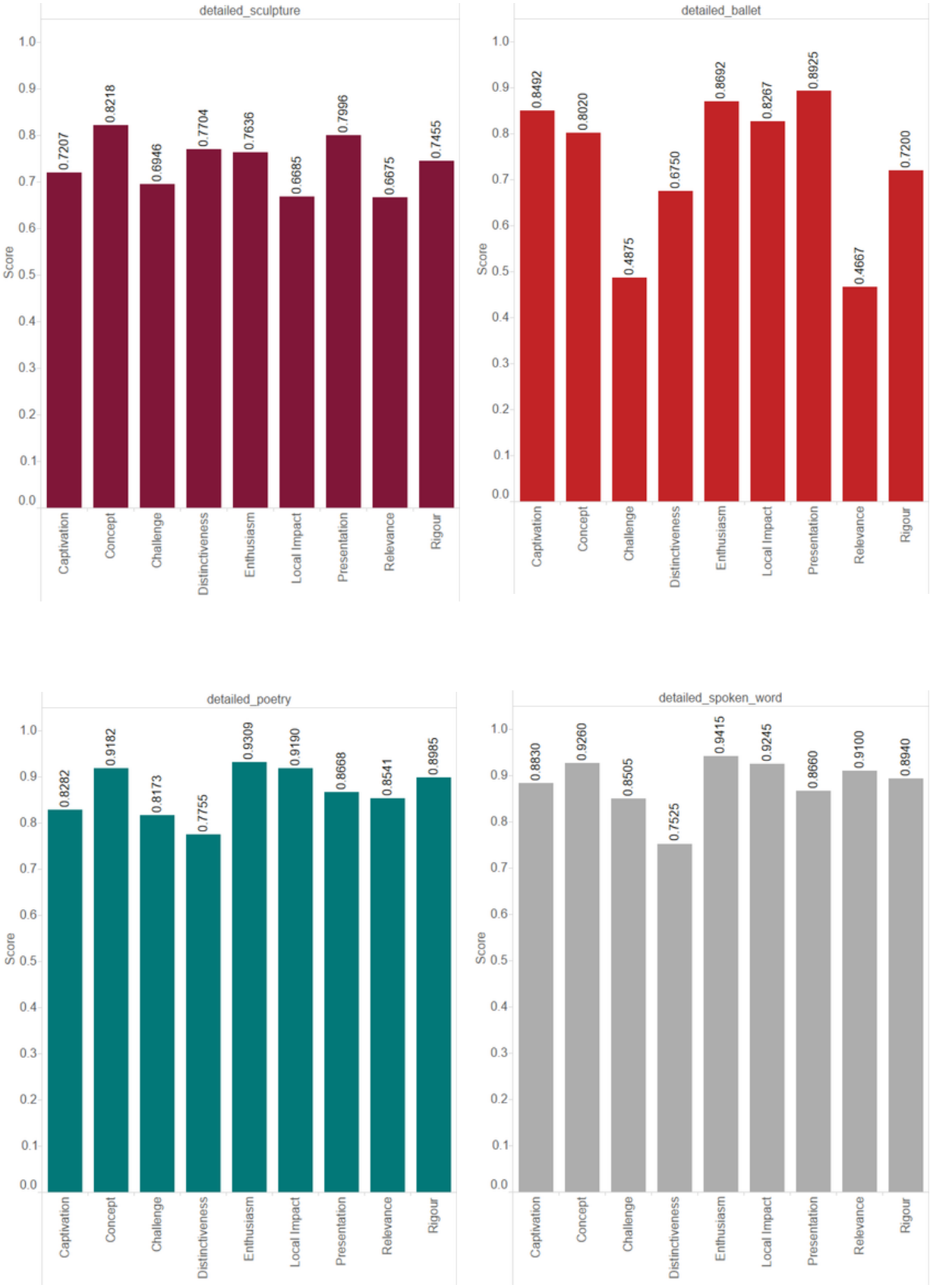
Figures 37a-l shows the dimension profiles for public responses across a variety of other detailed artforms. Notable differences start to emerge in basic patterns.

21 http://operabase.com/

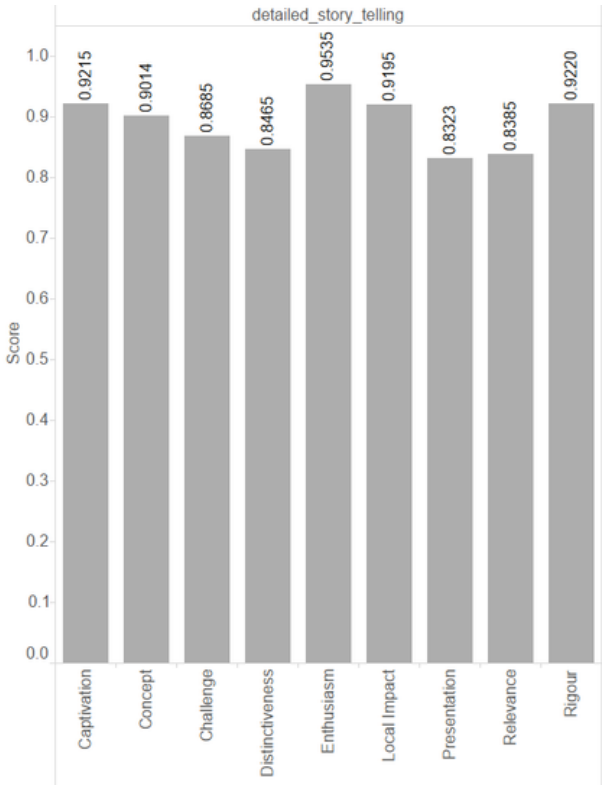
Figures 37a, b, c & d: Detailed artform profiles – public respondents

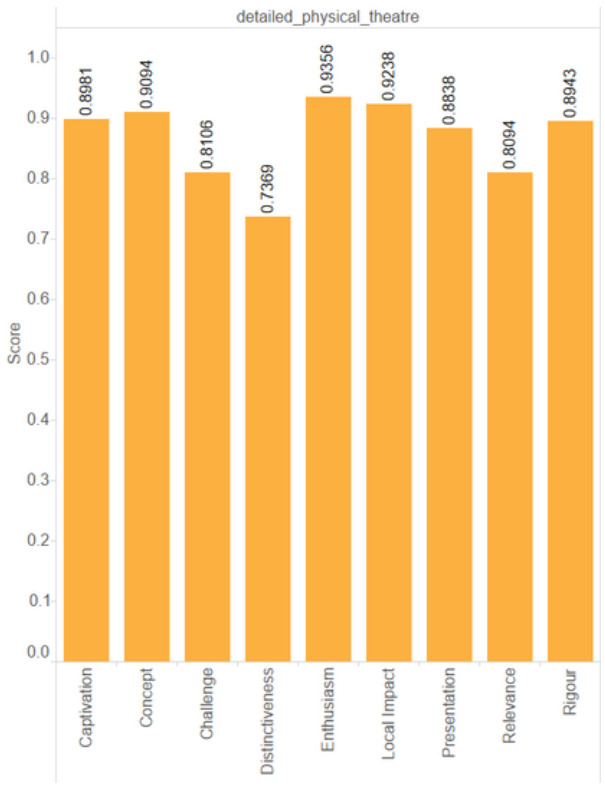


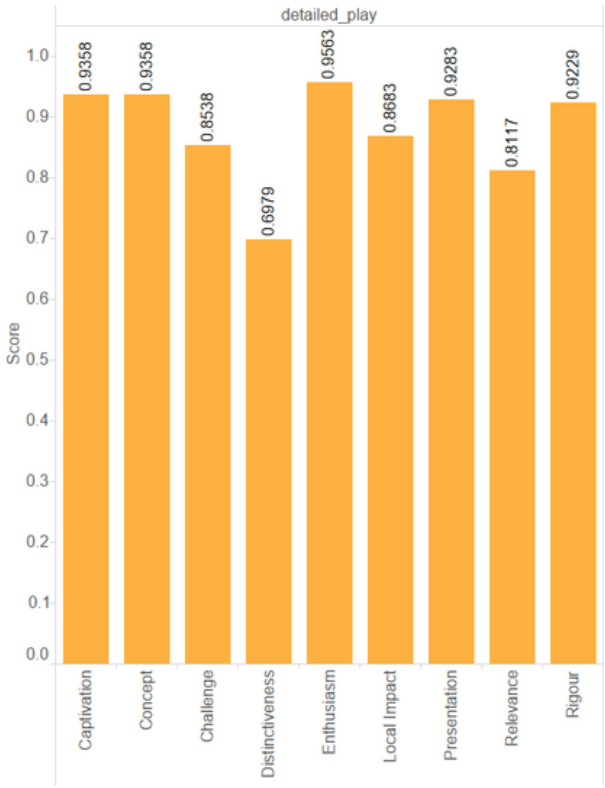
Figures 37e, f, g & h: Detailed artform profiles – public respondents

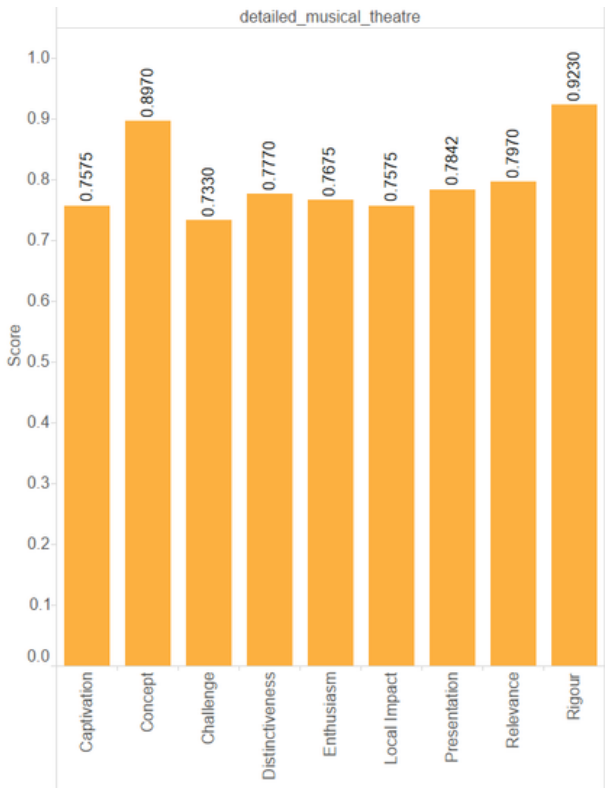


Figures 37i, j, k & l: Detailed artform profiles – public respondents









For comparisons with self prior and peer responses for these detailed artforms, please refer to the Supplementary Data Charts, in Appendix 2. The variations existing for each artform could be for a variety of reasons, not least the particularities of the work evaluated in this QMNT. Exploring these differences requires dialogue and debate within and across artforms; a process that is certainly preferable to speculation here.

4.4 Artform Attributes

Artform attributes were assigned to the majority of the evaluations in this study. It emerged very quickly when gathering artform data from organisations that simply an artform categorisation did not accurately portray the essential context for those artforms. Moreover, for some work which does not fit in to artform categorisations, such as work produced by MPMs, a flexible data model was required to capture suitable metadata beyond artform. Attributes were kept as they were provided by the participating organisations i.e. no inference was applied. This work also provides a basis upon which to better understand the minimum metadata fields to start to delve in to artform analysis. Table 4 lists the emerging artform attribute fields, with common examples from this study. In the analysis charts that follow, only terms used at least five times have been included.

Any number and combination of attributes can be accommodated by the data structure.

Table 4: Artform Attribute Fields

ATTRIBUTE FIELD	EXAMPLES PROVIDED BY THE PARTICIPATING CULTURAL ORGANISATIONS
Artist Attribute	young, disability
Audience	children, community
Chronology	contemporary, traditional
Genre	documentary, comedy
Locality	school, outdoor
Medium	digital, percussion
Person	Shakespeare, Mozart
Place (culture)	indian, welsh
Presentation	festival, installation
Process	creative, writing
Purpose	participation, conceptual
Subculture	jazz, punk
Subject	christmas, social
Transaction	immersive, interactive

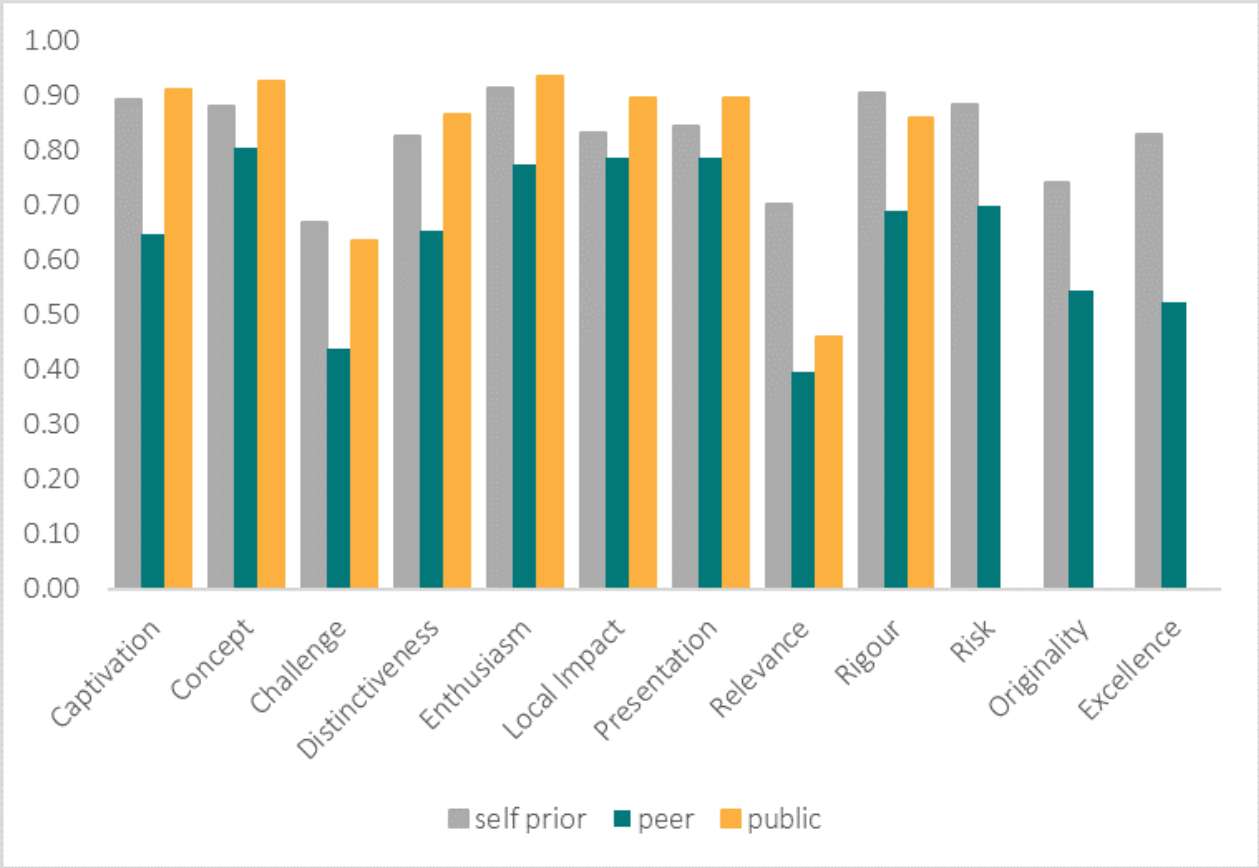
It is recognised at this stage that these categories are first attempts at categorising attributes linked to artforms. This will undoubtedly expand as more data is collected, both in terms of extending the fields, and the terms within fields.

4.5 Presentation Analysis

The presentation of art can be a significant component of how that art, or combination of arts is defined. Therefore, assigning a presentation method can be a useful way of segmenting the data. For example, that is very much the case with circus, cabaret, or mulit-arts festivals which are by nature combined in form. If we sought to assign detailed artform attributes to say cabaret (singing, dance, theatre) it would not aid the analysis process in this case where the distinctiveness comes from their combined nature and the mode of presentation.

Figure 38 compares self, peer and public ratings for circus work evaluated within this study.

Figure 38: Presentation Respondent Comparison: Circus



The above chart really highlights the public popularity of circus, particularly on the dimensions which generally score highly across the portfolio – captivation, concept, local impact, presentation, and rigour, with an inflated score on a 'lower' scoring dimension - distinctiveness. Lower scores on challenge and relevance are perhaps to be more typically anticipated with this presentation method, although of course variation between specific productions would be expected.

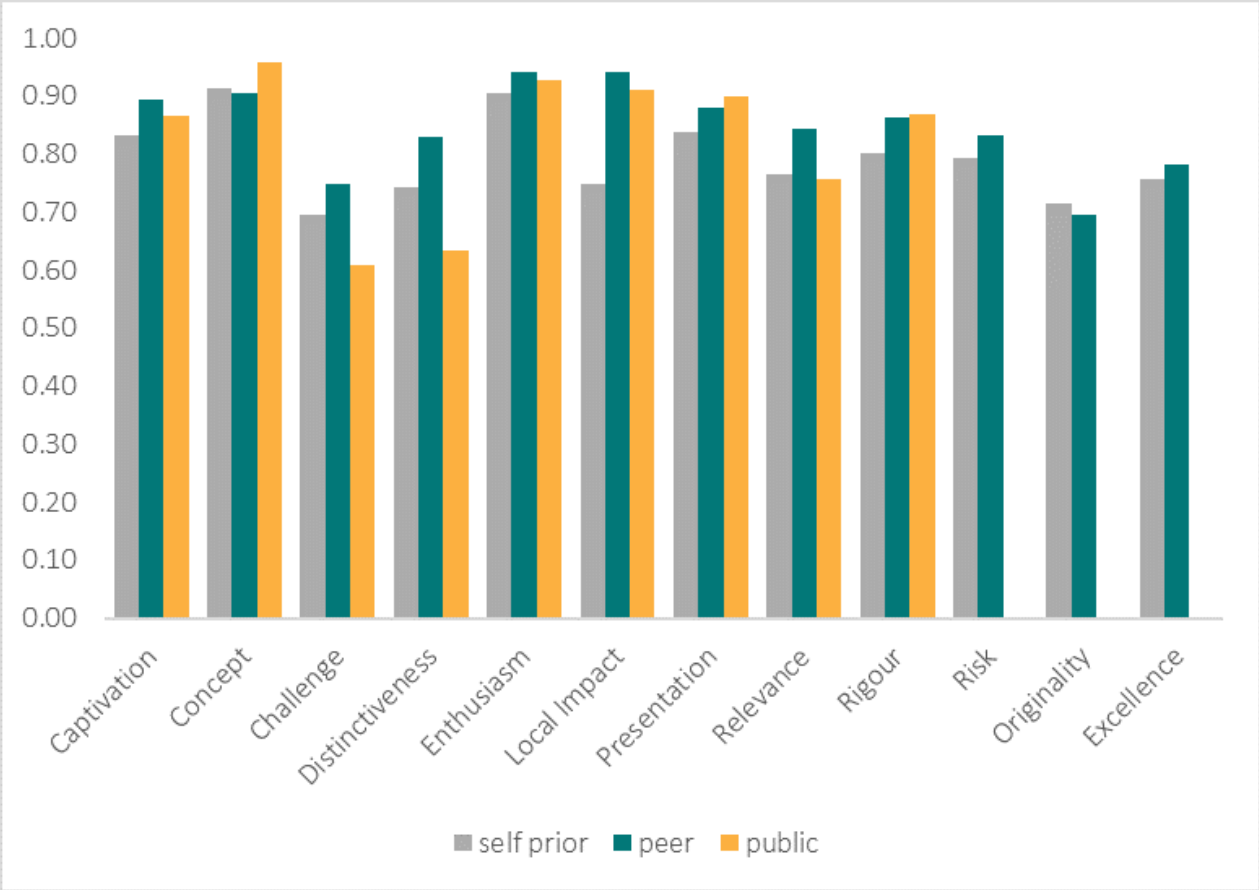
For more respondent comparisons on different presentations of work, please refer to Appendix 2: Supplementary Data Charts.

Presentations of work are different to mediums. A medium here is defined as a physical tool that is used to express the art, whereas presentation refers to the overall production of the art. For comparative charts of different mediums, refer to Appendix 2.

4.6 Genre and Subculture Analysis

Genre and subculture can also be grouped for analysis and presentation purposes. For an authoritative version of these types of categories, further work would need to take place to cross reference with established ontologies, or otherwise professional uses of the terms in relation to the cultural sector. By way of example, Figure 39 below looks specifically at punk (for clarity, this term was only used in the context of music in this study, but this field could refer to other punk-art).

Figure 39: Subculture Respondent Comparison: Punk



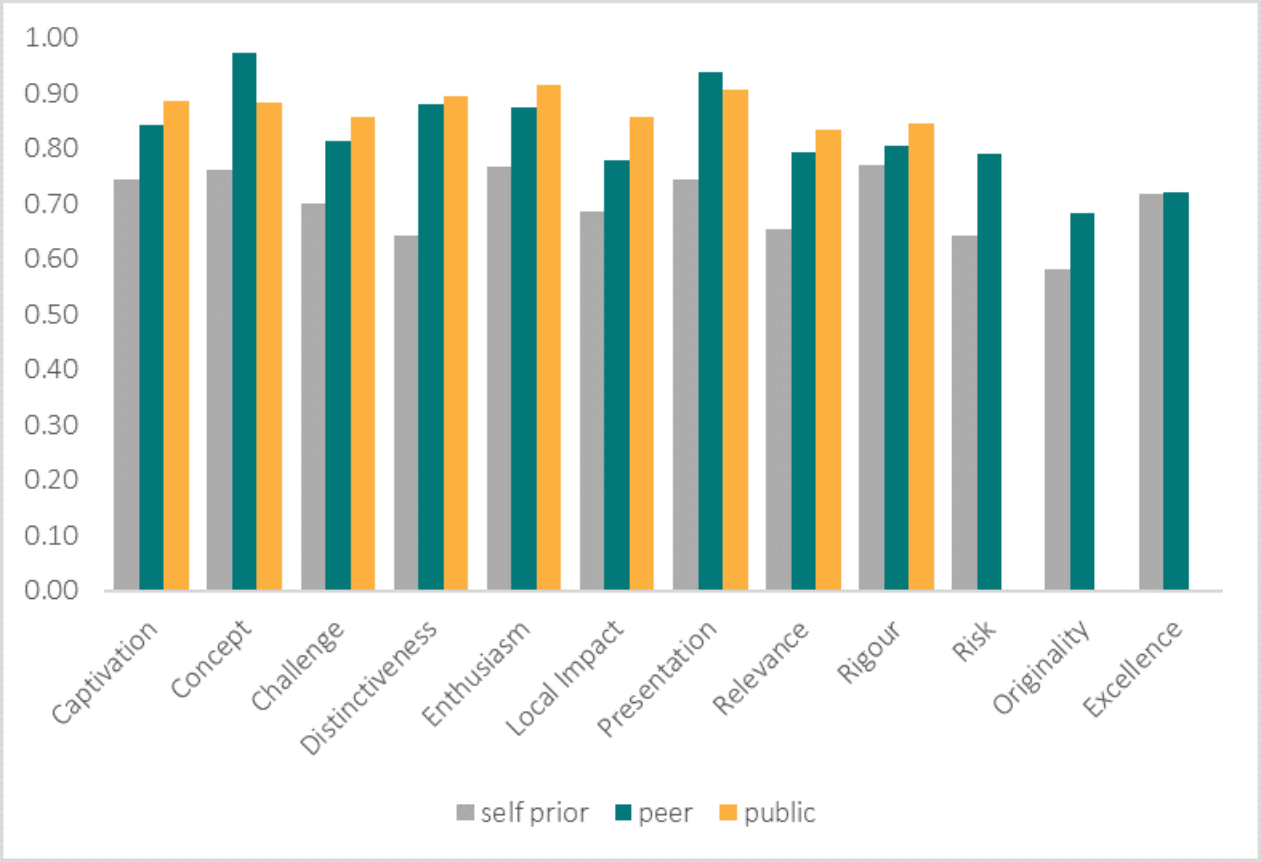
This example was chosen as the peer scores are (overall) the highest scoring respondent group. This is a more unusual pattern. In addition, scores for self and peer distinctiveness (it was different to things I've experienced before) and originality (it was ground-breaking) clearly contrast with the lower score for distinctiveness attributed by the public respondents. In addition, the peer and self prior scores are much more aligned than the public scores, particularly in relative pattern. This may indicate that the professional perspective of this group of evaluations differs from the public perspective.

4.7 Examples of other Attribute Dimension Profiles

Immersive

Figure 40 summaries a transaction respondent comparison – for immersive work. The transaction category is somewhat a misnomer as it implies something more concrete than the terms it encompasses here. Transaction is the mode or description of a human experience and importantly, the interaction with surroundings in some way e.g. experiential, interactive, immersive. These terms in some cases may be considered as synonyms, but at this stage of the attribute development process, each word is held in its own right as there are conceptual differences as well as overlap.

Figure 40: Transaction Respondent Comparison: Immersive

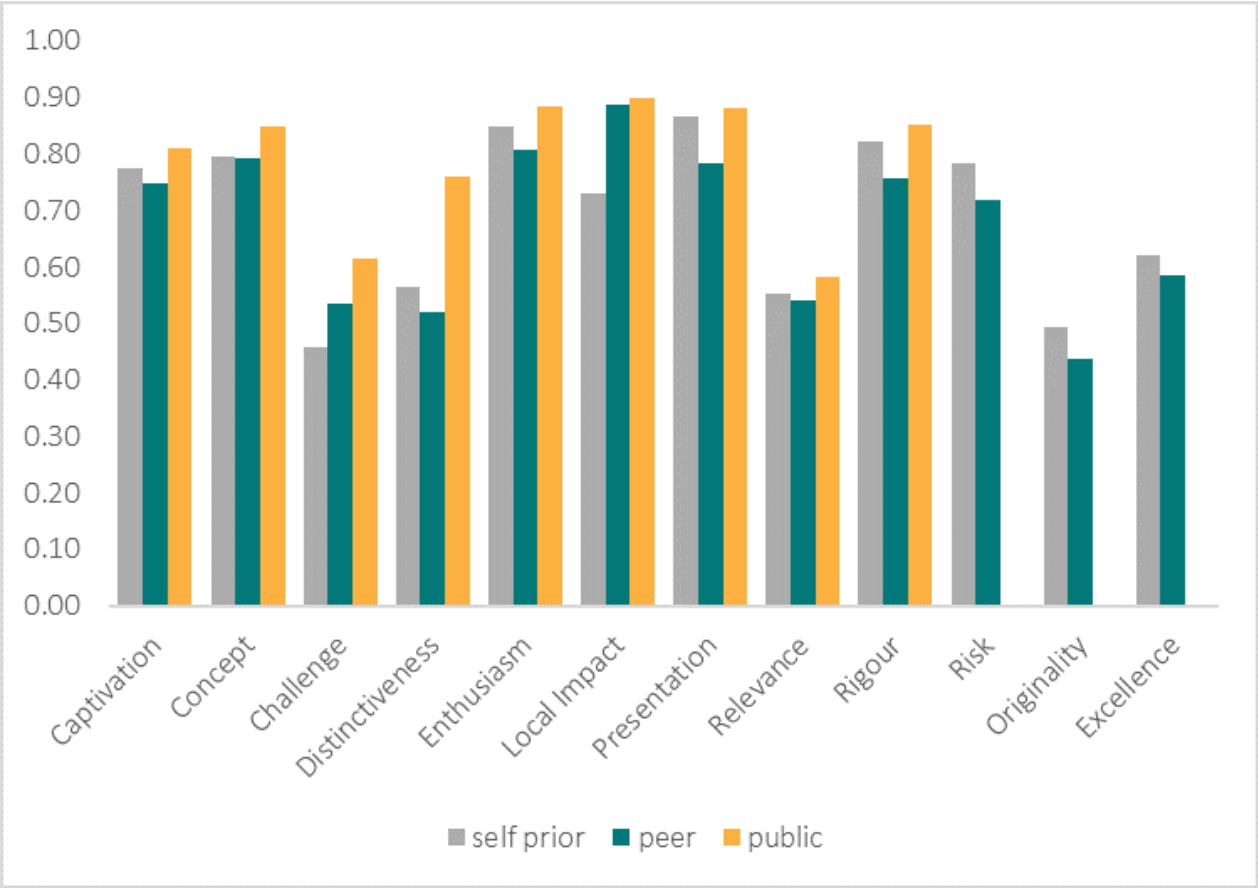


Interestingly for work defined as 'immersive' in this study, the public and peer ratings are much higher than the aggregate average for challenge, distinctiveness and relevance (the 'lower' scoring aggregate dimensions across all evaluations). With more data it would be interesting to explore whether 'immersive' work is a consistent inflator in peer and public ratings across particular dimensions.

Christmas

Figure 41 presents a self, peer and public dimension comparison for 'Christmas' work.

Figure 41: Subject Respondent Comparison: Christmas

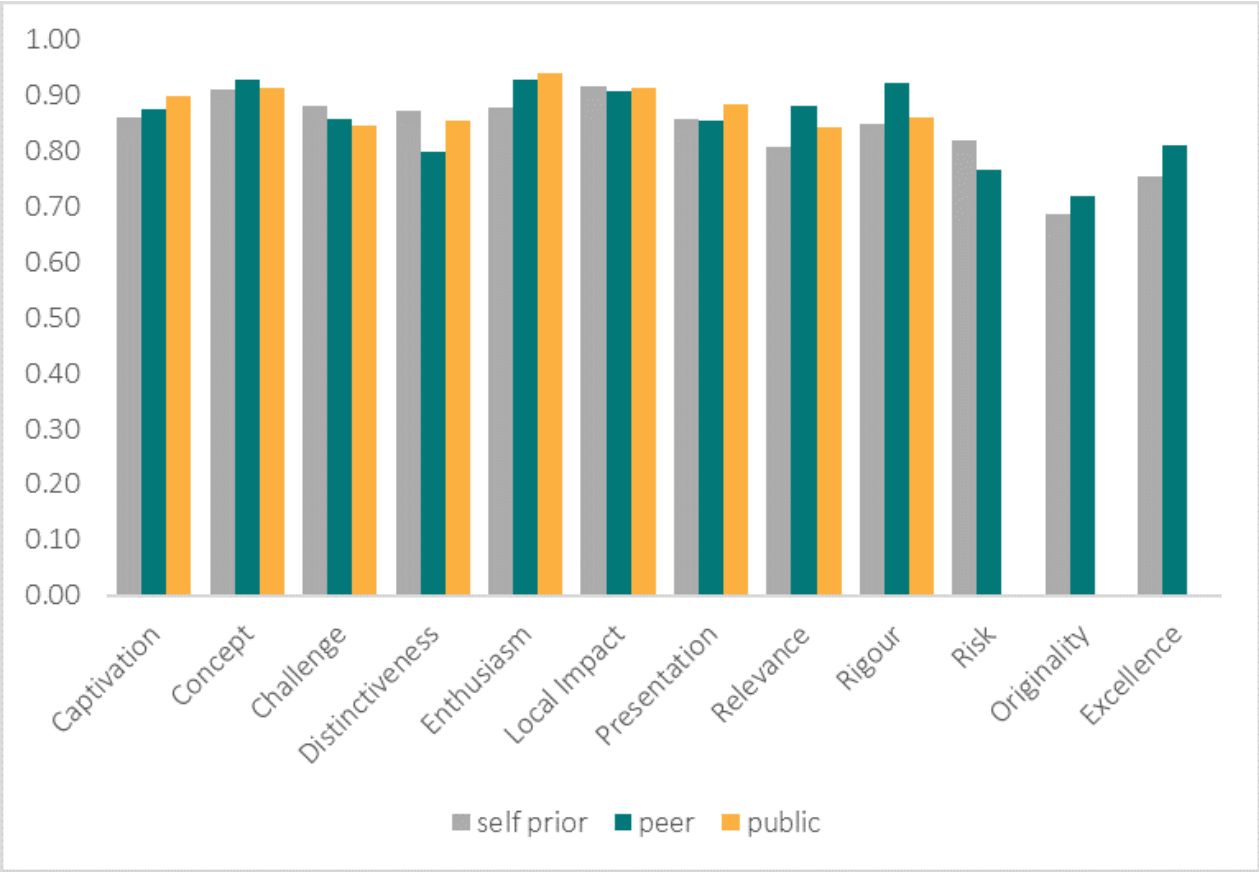


As an attribute, Christmas is an important one as it represents a popular season for programming, particularly across theatres. Challenging, original work might be assumed to be unusual focuses for 'Christmas' shows, however many of the organisations in the cohort put on 'alternative' festive productions. This is reflected in a public score for distinctiveness showing a relatively higher score than anticipated even by the self assessors. This is also reflected by the relatively high scores for risk.

Participatory

Figure 42 shows self, peer and public responses for work classed as participatory work that was evaluated within the Quality Metrics National Test. The cultural organisations who presented and evaluated this work were not participants in the separate participatory metrics strand, and therefore used the quality metrics to evaluate their events as opposed to any of the participatory metrics, or combination of the two. Therefore, their results are reported here rather than in the participatory metrics report.

Figure 42: Purpose Respondent Comparison: Participation



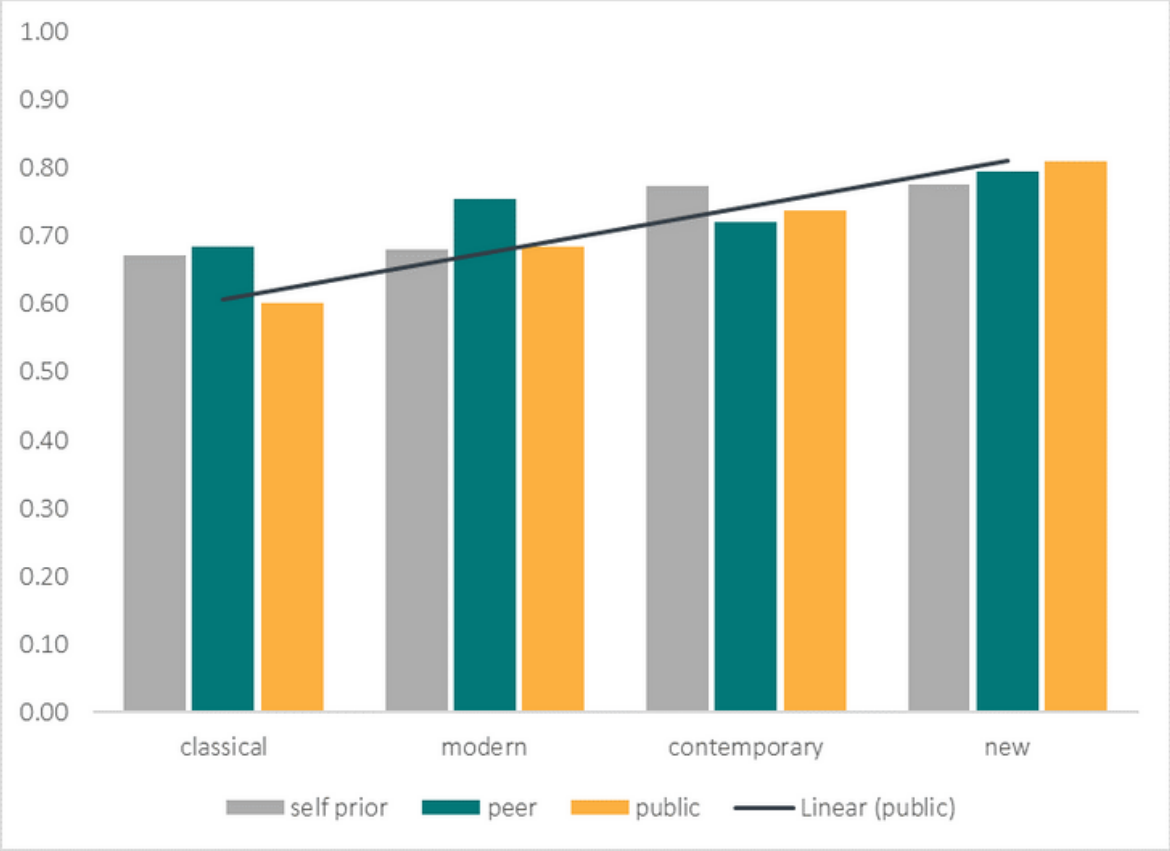
Interestingly, it seems that participatory work generally acts as an inflator on scores, particularly on the generally 'lower' scoring dimensions – challenge, relevance and distinctiveness. Assuming that this work is providing relatively new experiences for respondents which are more likely to challenge and be relevant to them, this highlights the importance of the development and use of the participatory metrics.

Chronological Attributes

The final artform attribute comparison in this section looks at chronology, focussing in on one dimension – relevance ('it had something to say about the world we live in'). Chronological attribute is assigned when a term implies something about the era in which the work is made.

An intuitive hypothesis would be that generally audiences are likely to experience the newest work as the most relevant, and the oldest work as the least. Whilst of course this would not be a hard and fast rule for all work, the public responses do seem to reflect this, as can be seen in Figure 43. Self prior scores also numerically follow this pattern albeit with less noticeable difference between some groups. Interestingly, peer scores do not follow this pattern.

Figure 43: Chronological Attribute Respondent Comparison: Relevance



For more charts looking at chronological attributes, please refer to the Appendix 2.

4.8 Summary

The aim of this chapter (supported by the supplementary data charts in Appendix 2) was to demonstrate the wider potential of combining self, peer and public ratings on the standardised quality metrics with cross-cutting analysis rooted in artform categorisations generated from the terms provided by the cultural sector (with these meta data tags attaching to individual responses within each survey and evaluation across the Quality Metrics National Test).

As the analysis shows there is enormous potential in this approach, allowing for larger scale aggregation of the data whilst maintaining real granular detail in the results. It also underlines that part of the wider value of the standardised quality metrics, and a platform like Culture Counts, is to capture both survey data and metadata at scale such that as the user base of the quality metrics grows, and we move from small scale to genuine big data, all kinds of patterns, subtleties, and added value analysis becomes possible, with the interpretation discussed and driven by the creative professionals that make the work.

5. CHAPTER FIVE: Cohort Engagement, Insights, and Data Culture

5.1 Introduction

At its simplest this project has had some very clear output targets, namely:

- Recruit 150 NPOs and MPMs to take part in the Quality Metrics National Test in accordance with the published Expression of Interest terms
- Provide those 150 organisations with logins to the Culture Counts system and support them to conduct their own evaluations
- Analyse and interpret the resulting data
- Produce an aggregated data set for all of the evaluations that took place.

Whilst these headline outputs are clearly very important, and have been fully detailed in this report and the separate participatory metrics report, both Arts Council England and Culture Counts were also very interested in gaining insights into the data culture of the participating cultural organisations, and the challenges and opportunities they identified, or demonstrated, in their evaluative practice.

Nordicity, an independent consultancy, were commissioned by Arts Council England to carry out an independent evaluation of the Quality Metrics National Test examining in some detail the views and perspectives of the participating organisations. In this chapter we share some of the key insights gained by the Culture Counts team as to the data culture and evaluative practices of the participating cultural organisations. Some of the participating organisations also took part in the separate participatory metrics strand, supported within the Quality Metrics National Test.

5.2 Cohort engagement: 'supporting' not 'pushing'

Arts Council England made it very clear to Culture Counts that whilst the aspiration was for every one of the 150 participating NPOs to complete 3 evaluations, with a target of 1 self assessment, 5 peer assessments, and 30 public responses for each evaluation, and that this headline target was important, it should not be treated as the priority KPI of the project. Arts Council England were more interested in examining what proportion of participating NPOs would stay engaged during the lifetime of the project and what challenges and barriers they faced in delivering on these aspirations.

This had a very practical implication for Culture Counts in terms of how we sought to engage with the cohort. The Culture Counts delivery team were encouraged by ACE not to 'push' the organisations towards the target number of evaluations. This was because this national test was designed to examine whether the participating organisations had the resources and desire to engage and carry out their own evaluations using the quality metrics and the Culture Counts system.

Therefore, we did not directly call or email the cohort of organisations to 'chase' them to complete more evaluations or undertake more activity.

Rather our approach was to support the use of the Culture Counts platform by responding quickly to any support requests initiated by the participating organisations. We also supported evaluation activity through the development of a wide range of resource materials on the Quality Metrics National Test site[22]; and through discussions at the Learning and Insight sessions (which that took place in December, January, February, April and May at various locations in England (see Figure 43)). These sessions built understanding and interest in the project by encouraging the participating organisations to share their evaluation experiences and their data stories. As a consequence of this engagement approach, we gained a wide range of insights into the evaluation practices and data culture of the participating organisations.

Figure 43: Learning and Insight sessions with the participating organisations

Participatory Learning and Insight sessions took place on:

15th December – London
14th January – Manchester
28th April – London
29th April – Manchester
10th May – Manchester
11th May – London

Quality Metrics Learning and Insight sessions took place on:

26th January – London
4th February – Manchester
5th February – Newcastle
8th February – Birmingham
9th February – Bristol
28th April – London
29th April – Manchester
10th May – Manchester
11th May – London

In terms of attendance Culture Counts estimates that approximately three quarters of the active participating organisations attended one or more than one of the Learning and Insight sessions.

22 http://www.qualitymetricsnationaltest.co.uk/resources/

5.3 The evaluative and data culture of the participating organisations as judged by the levels of support they required

From the outset of the project Culture Counts expected the cohort of participating organisations to need very different levels of support. 'Support' in this context encompasses a number of elements:

- Support in setting up self, peer and public evaluations in the Culture Counts system
- Support in using the criteria for peer selection and in how best to invite peers to take part
- Support with an organisation's overall evaluation / data collection strategy and how best to integrate the quality metrics
- Support with using the Culture Counts system to segment their audiences and integrate with their surveying strategy through the Culture Counts dashboard
- Support with negotiating partnerships with venues not taking part in the national test
- Support with sharing their data with other organisations / partners
- Support with data manipulation and analysis

The amount of support a participating organisation sought from us was influenced by their overall capacity and resources; their general familiarity with software system and tools; their level of evaluation expertise; and their data analysis expertise.

Culture Counts offered support in a number of ways. A member of the Culture Counts team could be reached over the phone or skype to talk through any support request; via email; or through the support portal of the Culture Counts platform

In addition, Culture Counts launched a dedicated Quality Metrics National Test website in January 2016 and began directing the cohort to a wide range of instructional videos and other resources to help participating organisations carry out their evaluations and interpret their data. Therefore, in assessing the support offered to the cohort of organisations during the lifetime of the project we have reviewed both our direct contact with the participating organisations and all our web analytics for the Quality Metrics National Test site.

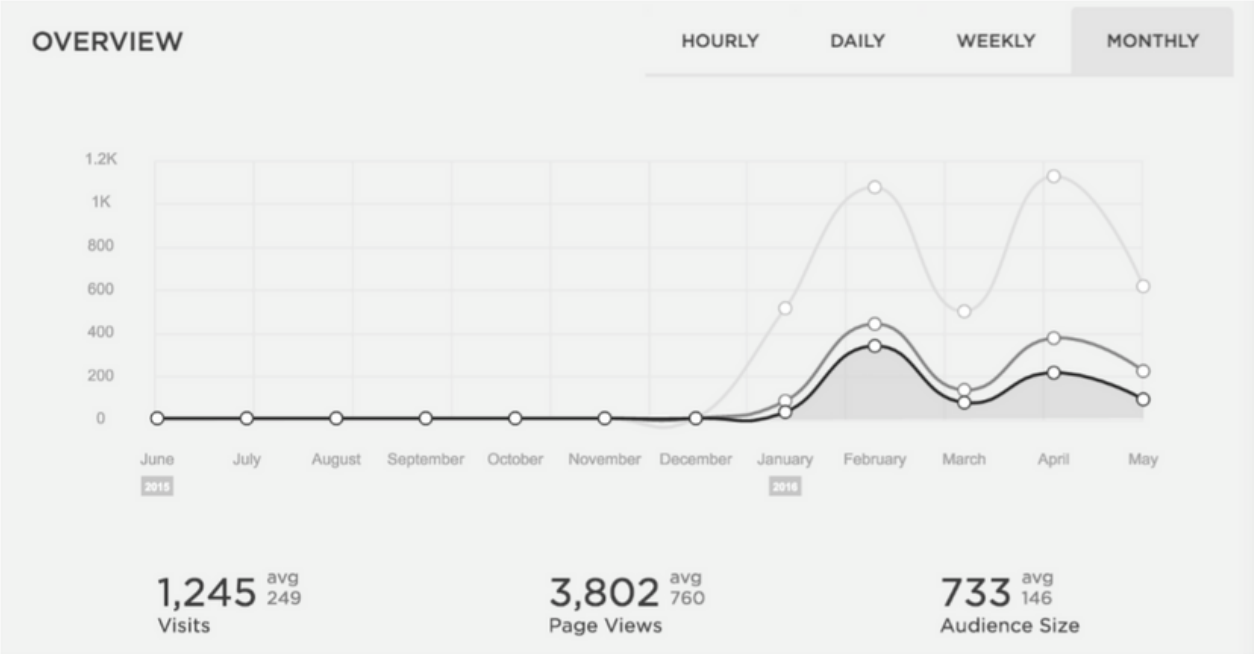
In broad terms, just under half the cohort required what we would label 'high' or 'significant' levels of direct support (in other words the cultural organisations in this group initiating multiple calls and emails across the evaluation period). Around a third of the cohort required moderate levels of support (initiating some calls / emails). Just under a fifth required low levels of support (initiating a few calls / emails).

There are of course some subtleties here, in that some organisations made lots of contact with Culture Counts because they were trying to do imaginative things with their evaluations, as opposed to needing 'hand holding' support to carry out a basic evaluation. Nonetheless, overall it is true to say that the majority of organisations seeking the most help were those with lower capacity and / or evaluative expertise.

When cross-referenced with well-configured and executed evaluations, these findings are consistent with our previous depiction of the cultural sector in England as being 80% data shy; 15% data ready; and 5% data driven.

What do our web analytics tell us about cohort behaviour and engagement levels? After the launch of the Quality Metrics National Test site in January 2016, Culture Counts regularly uploaded videos, blogs and other materials to the website, and video material to the Culture Counts UK YouTube channel. Figure 44 summarises our website visit statistics. The lightest grey line shows page visits, the medium grey shows site visits, and the darkest grey shows unique visitors. After the home page of the site, the most popular pages were the resource and blog segments. The site has clearly been visited regularly, although we cannot be definitive about the proportion of visits that were from the cohort as opposed to other interested users.

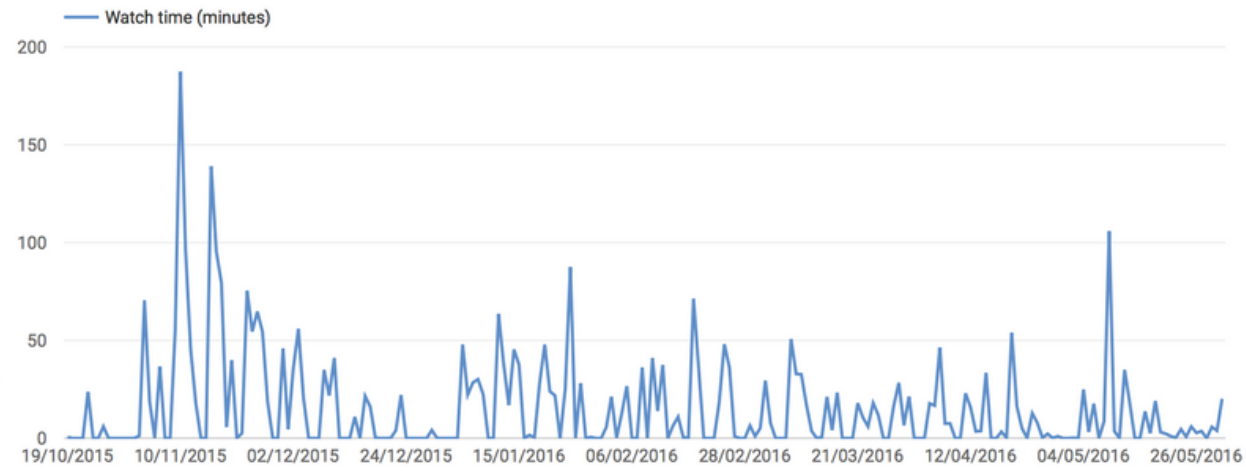
Figure 44: Quality Metrics National Test website analytics



The YouTube analytics for the Cultural Counts Quality Metrics National Test channel are rather more insightful. During the lifetime of the project, our YouTube channel has attracted 3,527 minutes of watching time, from a total of 1,364 views. Again, whilst not all user activity may have been from the cohort, if we divide the number of total minutes spent by the number of active cultural organisations taking part in the national test (137), it emerges that on average each organisation spent 25 minutes and 45 seconds watching our videos. This highlights how popular the channel has been amongst the cohort as a form of evaluative support.

This is confirmed by the fact that views of the YouTube channel strongly spiked during periods when we increased direct face to face engagement with the cohort. Figure 45 shows when the channel was most popular during the national test. The largest spikes are on 10/11/15 and 16/11/15, just after Culture Counts had provided all organisations with a login to the Culture Counts system along with their induction and opening support materials which featured the link to the instructional videos on our YouTube channel. However, after this initial viewing immediately after enrolment and induction, we can see that the predominant spikes in watch time are around the time of the Culture Counts Learning and Insight sessions on 25th January and 9th May. This confirms that face to face contact had a powerful impact on online activity suggesting that these sessions with the cohort, at which they shared experiences and results, encouraged greater use of the support materials and increased the engagement of participating organisations in the Quality Metrics National Test as a whole.

Figure 45: Watch Time Graph for the Culture Counts YouTube Channel during the Quality Metrics National Test .



We were also able to track which of our 'support' videos were watched the most, in terms of total minutes watched and number of views. The most watched videos were:

- Configuring the public survey
- Interviewing briefing on the Culture Counts Survey
- Copy surveys
- Brief run through of the Culture Counts system
- Inviting self and peer assessors
- Logging in and setting up your first evaluation
- Participatory Evaluation run-through

All of which confirms that that these materials acted as an important source of support and guidance to the participating organisations.

5.3.1 Engagement led to improved evaluation practice and outcomes

What also became very clear during the short span of this study (the majority of evaluations took place in a concentrated period between January and May 2016) was that as the participating organisations grew more familiar with the quality metrics; and the self, peer, public triangulation approach embedded within and facilitated by the Culture Counts platform; this in turn led to more accomplished evaluative practice and better outcomes, as evidenced by:

- Culture Counts observing more accurately configured evaluations (i.e. no mistakes in configuration or in URL attribution for self, peer and public responses) as organisations moved through the project
- A declining number of support calls on how to set up evaluations in the Culture Counts dashboard
- More interest from the cohort in 'value adding' activity, such as creating separate URLs for audience sub groups, and collaborating / sharing data with other organisations
- Improving data collection outcomes in terms of the total number of public responses being achieved by the participating organisations

5.3.2 Engagement led to strong adherence to triangulation / creative intention measurement

The Culture Counts team would also note that the adherence to carrying out self prior and self post evaluations is also a good indicator of cohort engagement in the project. With producing this outcome in mind, at the Learning and Insight sessions, Culture Counts shared with the participating organisations the experience of evaluating the Cultural Programme of the Commonwealth Games in 2014[23] and the importance of the participating organisations building a clear sense of their creative intentions in order to inform their self assessments (the cultural organisations in the Commonwealth Games Cultural Programme evaluation prepared their own creative intentions statements for each piece of work evaluated using the quality metrics and self, peer, and public responses).

We also stressed to the participating organisations the importance of building a reflective community inside their organisations, so that they would be able to get more out of their results and interpretation processes. So for example, we suggested that organisations could think about creating broad self assessor groups from across their organisation (e.g. not made up of just artistic, and curatorial staff – but also including education and marketing teams) in order to enrich both the data and subsequent dialogue about results. At the Learning and Insight days we shared Figure 46 as a way or prompting and guiding the participating organisations through this process.

23 http://www.creativescotland.com/__data/assets/pdf_file/0017/31652/Evaluating-the-Quality-of-Artist,-Peer-and-Audience-Experience-.pdf

Figure 46: Building a reflective evaluative community inside your organisation

We have already reported that self prior and post ratings were well aligned to public ratings across the evaluations in this study, showing that the self assessor communities did indeed think through their creative intentions clearly and with considerable predictive accuracy as judged by alignment with peer and public response. Culture Counts also saw interesting examples of how results were shared and discussed within organisations (see the following case study from Derby Museums).

CASE STUDY: Derby Museums

With very special thanks to Derby Museums for their results and insight for this case study.

Derby Museums completed three evaluations using Culture Counts over the course of the trial period, all of which were focusing on very different events.

The results were shared at a full staff briefing, where a cross section of the whole team was interested to hear the results. Using a more qualitative style of data presented via the quality metrics enabled the team to view a more rounded perspective, as opposed to only looking at hard figures, which doesn't necessarily do justice to the work or its outcomes. Enabling the team to see how satisfied their customers were provided a real morale boost.

The results revealed some surprises, showing high levels of satisfaction where perhaps the team hadn't felt things had been as successful or relevant. Seeing the nuances between the different events and the scores received can be interesting, as there may be elements of the event that particularly resonated with the attendees, which the organisation may not have previously realised with such clarity. This highlighted the usefulness of the self-assessor completing both the prior and post surveys. Through completing the self-assessments, they also realised that their intentions could be more focused. Having clearer intentions from the offset will allow the focus of the work to really come through. By mapping expectations and experiences, a richer frame for interpretation is formed.

Combining results from the Quality Metrics survey and demographics from other survey providers, Derby Museums have been able to further develop their applications for funding. Many funding bodies have specific goals to fulfil or demographics to reach and therefore their funds must be allocated appropriately; for example, this could be the elderly, rural communities, ethnic minorities etc. Incorporating the results from the Quality Metrics survey, Derby Museums have presented a positive case in funding applications and funding reports. In particular, this has been useful when asked questions about the audience's thoughts and experiences, as they were able to refer directly to the positive feedback gathered from the attendees.

There is also the potential to collect more detailed demographics and to gather further data focusing on Generic Learning Outcomes. This could be achieved using the system as it currently stands; however, adding further questions was not actively encouraged during the Quality Metrics National Test.

5.4 Challenges, Issues and Opportunities identified by the cohort

Culture Counts received feedback on the challenges being encountered by participating organisations both through their direct support request and calls; and through their contributions at the the Learning and Insight sessions. During the course of the study the organisations also detailed their responses to some of these challenges and identified opportunities around the quality metrics and this type of evaluation approach. The key challenges identified by the organisations did not relate to usability issues with the Culture Counts platform and dashboard (on which participating organisations gave positive feedback to the Culture Counts team), or the core quality metrics themselves. Rather the key points raised were:

5.4.1 Peer Management, Engagement and Building a Peer Community

Planning and managing the peer process was a new experience for all of the participating organisations and represented the greatest challenge in successfully completing their evaluations. Not only do organisations have to select their peers, but they then have to invite them and secure their participation, and follow through with peers checking that they have attended their event and completed their evaluation. Some organisations also mentioned that depending on the location of the particular event the distance required to travel for some productions is a barrier to obtaining peer assessors[24]. This peer engagement and management process was the most demanding part of the process in terms of the time investment of participating organisations. The original target aspiration of engaging 5 peers per evaluation proved a very tough target indeed (with the cohort engaging 777 peers, completing 921 assessments - 41% of the 2,250 target).

The paradox around the peer review element was that whilst it was the most demanding element of the evaluation process, it was also seen as a very positive aspect of the Quality Metrics National Test. Participating organisations agreed that, irrespective of what happens around the future use and development of the quality metrics, one of the vital legacies of this project is the bank of peers who have been engaged for the first time in this formal way to give feedback on work across the cultural ecology. This was regarded as a hugely positive thing. We also saw innovation, in terms of organisations seeking peers from other artforms to review their work as they wanted a broader interpretative context to be reflected in their peer ratings. At the Learning and Insight sessions there was considerable discussion around how to secure ongoing benefits from this legacy of a newly engaged community of peers.

Participating organisations welcomed the opportunity to invite their own peers. Strong support was also expressed for the idea that as a result of this national test a peer database is formalised across the arts and cultural sector – in other words support is given to create an open searchable data base of peers for the sector to draw on – in which each peer could list their artform expertise and interests.

24 For example, see http://www.qualitymetricsnationaltest.co.uk/new-blog/2016/5/3/royal-shakespeare-company

Some of the participating organisations also showed interest in the idea of creating stable cohorts of peers for their work, who they would use to track peer sentiment over time, with that cohort becoming well placed to assess changes and inflections in their approach and work over time.

5.4.2 Enhancing Peer Continuing Professional Development

One clear potential deficit in the current process identified by some of the participating organisations was that having secured the engagement and participation of their peer evaluators, they had received feedback from peers that they had found the process 'too short'. Peers would have happily answered more questions and would have welcomed more discussion around the results.

Clearly, the length of the quality metrics question schedules has no bearing on how the peer community is engaged around the results. Even with the current peer dimensions, and additional open questions, organisations could choose to bring their peer evaluators together on a conference call to discuss their opinions and evaluations, and their reactions to the triangulated self, peer, and public ratings.

By way of international comparison here, the Department of Culture and the Arts in Western Australia, which is now using the quality metrics and the Culture Counts platform with all of their funded organisations, believes that the baseline quality metrics data that is being collected by the organisations in their funded portfolio is giving them the ability to open up ongoing, consistent long term relationships with peers, with the baseline data becoming the starting point for dialogue and discussion about the quality of work being produced.

These observations notwithstanding, the feedback from the participating organisations suggests that the current evaluation process may be under utilising the insights that could come from the peer evaluation process, both for the organisations, but also in terms of critical reflection and continuing professional development for the peers.

The participating organisations made a number of suggestions on possible additional questions that could be asked of peers:

- *'Should we be asking more open reflective questions?'*
- *'Many peers are active artists/curators: should we be asking them about whether the piece of work they saw made them reflect on their practice or work?'*
- *'Should we be including, or encouraging organisations to include open questions about the artform contributions of the piece they saw?'*
- *'Should we be asking an open question about how might / could the piece of work been improved? Are we missing a learning loop here?'*
- Organisations also discussed how the peer's interpretative context is vital. The current ACE AQA form asks the assessor – *'Please give details of the context you bring to this assessment etc.'* Organisations noted: '*Should we be including a similar open text question for peers?'*

What is encouraging about these suggestions from the participating organisations is that the ease and quickness of giving feedback via the Culture Counts platform actually left some peers feeling they would like to have to been engaged more. Meeting this demand needs to involve the right mixture of post event discussion and engagement around their feedback and overall results, and the use of some additional peer questions.

Engaging with peers so that they get something out of the process in terms of their continuing professional development is clearly an important outcome. Given how important the participating organisations thought the creation of a peer community was as a legacy outcome of this Quality Metris National Test, Culture Counts will seek to work with users to explore how to meet these requests.

The self, peer and public triangulation at the heart of the quality metrics process requires building dialogue and understanding across an interconnected community in the arts and cultural sector. Figure 47 ('sine wave') and Figure 48 ('Jackson Pollock') visualise how deeply inter-connected the community in the Quality Metrics National Test were.

In both visualisations, the key elements are:

- Orange circles = self assessors
- Teal circles = peer assessors
- Grey circles outlines = evaluations
- Grey lines = individual assessor connected to an individual evaluation

So for example, with regard to Figure 47, this depicts a sine wave type shape made up of orange circles (self assessors); teal circles (peer assessors) and grey circles (evaluations), with the grey lines linking the self and peer assessor to an individual evaluation. This chart, and the 'Jackson Pollock' chart, very clearly demonstrate the interconnectedness of the national test evaluations and peer assessors.

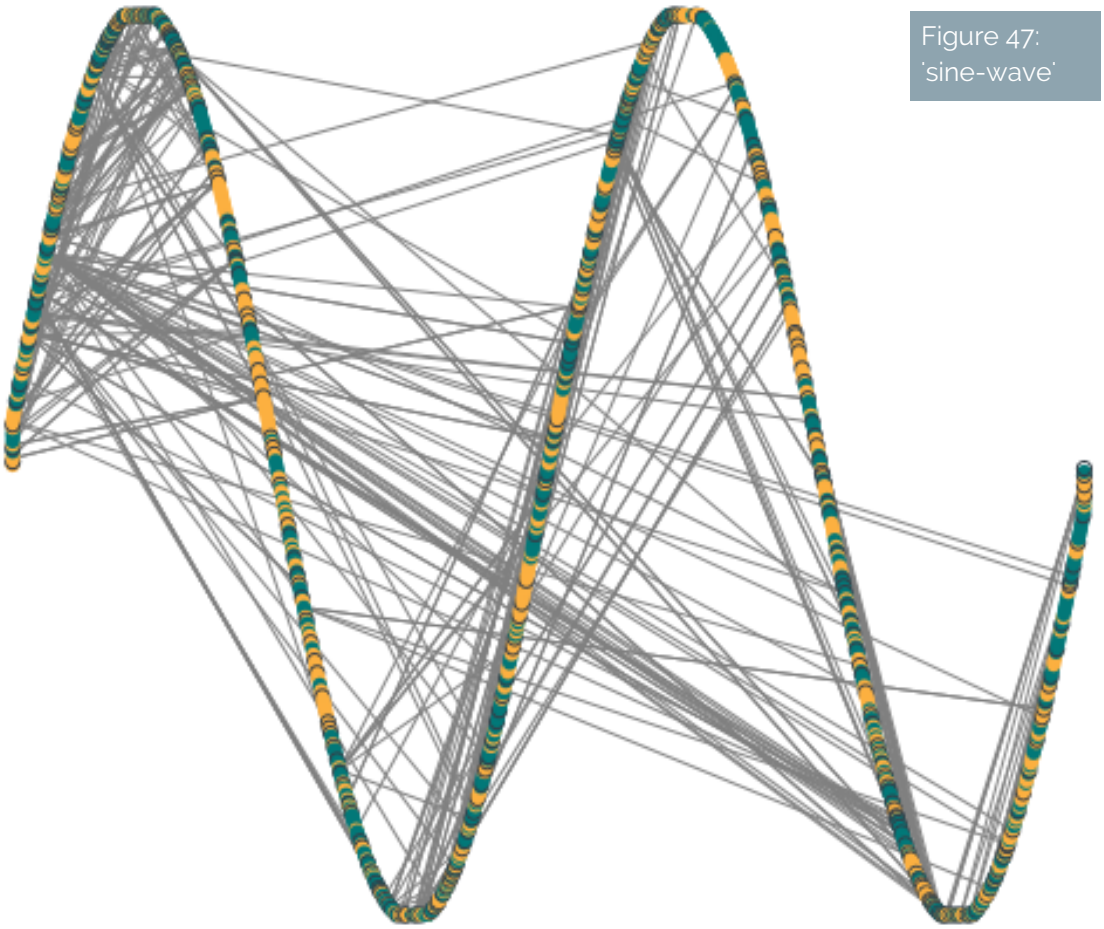


Figure 47:
'sine-wave'

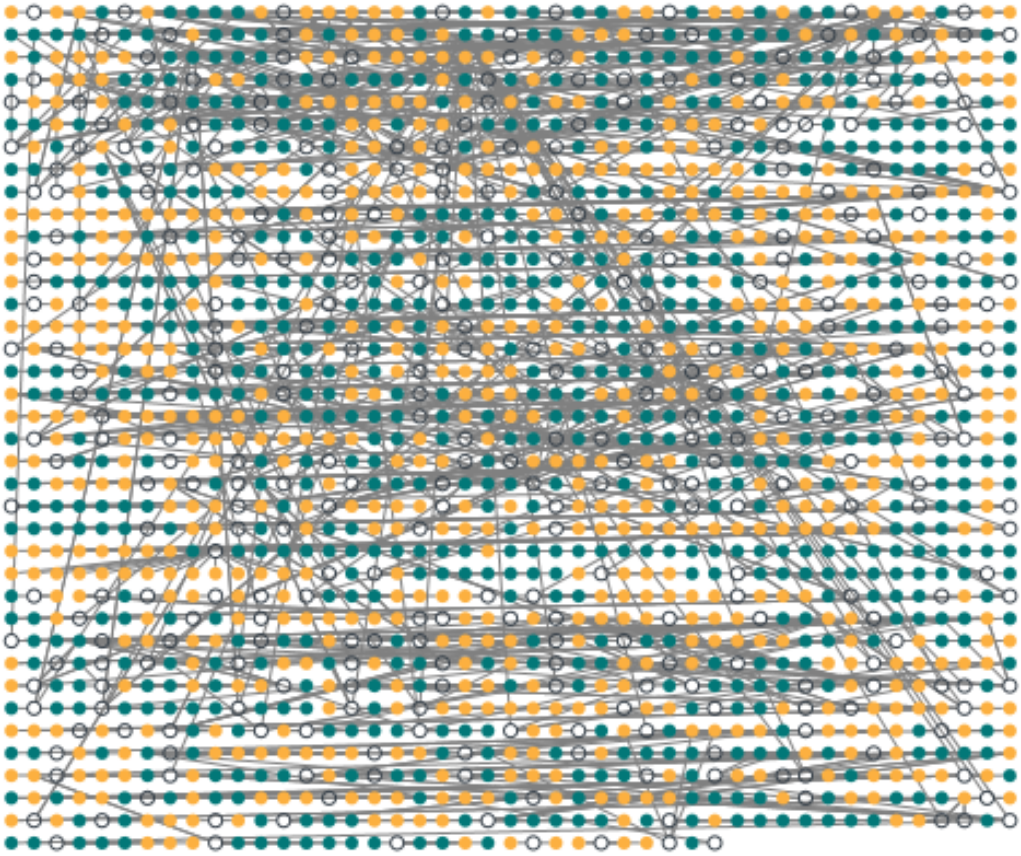


Figure 48:
'Jackson Pollock'

5.4.3 Integration with other databases; online databases; and ticketing / CRM systems

At the Learning and Insight sessions the participating organisations asked a range of questions about how a system like Culture Counts could integrate with the other databases, tools and CRMs they already use. The organisations were both interested in the future possibilities for integration with existing systems, and informed by the desire to ensure that their evaluation activity in the round is as efficient and as effective as possible both now, and in the future. Broadly speaking their key questions, and Culture Counts responses, were as follows:

- **Culture Counts and future interoperability:** Culture Counts as a software system has been conceived and built from inception to ensure ease of inter-operability with both open source and commercially provided data base tools and CRM systems. Culture Counts has a fully documented API, and as a native software company Culture Counts is planning a range of API integrations with other providers. Given the likely proliferation of data tools it is important that providers can demonstrate an openness and facility for such integration.

- **Using the quality metrics with CRM based audience segmentations:** We were asked by a number of participating organisations to support this type of activity, and we saw organisations innovate around segmenting and reporting their data. This normally involved creating different respondent groups – for example different groups of public respondents (e.g. regular attendees; TV audiences vs live audiences), and then using the Culture Counts system to create distinctive surveys (often with additional customer questions) and URLs for however many groups they wanted to focus on in their analysis. The most common question (see 5.4.4) being added in to the surveys related to organisations segmenting their respondents by previous attendance.

- **Interpreting results by demographic profile of respondents:** The Culture Counts platform collects basic demographic data (age, gender, postcode) and can be configured to collect more. In response to requests from some of the participating organisations we provided materials at the Learning and Insight sessions on how they could upload their postcode data from their quality metrics evaluations to open source demographic profiling tools, such as GeoConvert[21], which is based on ONS data.

5.4.4 Integrating with existing evaluation practices – adaptation and innovation

In both this quality metrics strand, and the participatory metrics strand, those organisations with well developed evaluation frameworks and practices had to think carefully about how best to integrate their quality metrics evaluation work alongside other evaluation activity they already had planned or were committed to.

21 http://geoconvert.ukdataservice.ac.uk/

Sometimes these integration challenges concerned using the metrics within multi-stranded evaluation approaches; or focused on how to design surveys and manage their distribution through a range of URLs targeting different audience segments in ways that added value to existing evaluation activity.

In practical terms these issues of integration and complementarity need to be explored by users in real evaluation examples. However, a few general comments can be made at this stage. Culture Counts as a data collection platform can be used to ask any 'closed' or 'open' question and generate data from self, peer and public respondents, with multiple URLs for each survey directed at sub-sets of self, peer and public respondents. Therefore, it is a straightforward data gathering task to combine the quality or participatory metrics with other bespoke evaluation efforts (and any question schedule).

In both the quality metrics and participatory metrics strand, one response to this integration challenge saw participating organisations innovating in their survey designs, adding in bespoke questions or picking additional questions from the Culture Counts interface. It is important to remember here that as part of the EOI terms all participating organisations had agreed to use the quality metrics in their standardised format, and Culture Counts did not actively encourage the participating organisations to create new bespoke questions. However, it was of course made clear to users that the Culture Counts system allowed them to use and generate other questions as this is an intuitive feature of the user interface.

So how much bespoke innovation around metric choice and creation did we see in this quality metrics strand of the Quality Metrics National Test?[22] In total, 485 custom questions were added to surveys in spite of no recommendation from Culture Counts encouraging organisations to do so. The appetite to innovate (but also ask audiences lots of questions) is clearly present. Most questions were used between one and three times, indicating the tailoring of questions to individual pieces of work or organisations. The top 5 custom questions used by a variety of organisations were as follows, and represent:

a) A desire to segment: 'Have you visited [org name] before?' (the most popular additional question)

b) Insights into marketing impact: 'How did you hear about the event?'

c) Connections with audience demographics: 'How would you describe your ethnicity?'

d) Gauging the accessibility of work: 'Do you consider yourself to have a disability?'

e) Encouraging additional open ended feedback: 'Do you have additional comments?'

22 Culture Counts has produced a separate participatory metrics report from the Quality Metrics National Test which contains a detailed account of the equivalent set of metric choices and design innovations carried out by the participating organisations.

The Culture Counts system captures which questions are being asked most frequently, both from the established quality and participatory metrics sets, but also new bespoke questions with an art-form or audience focus. As users endorse the established metrics, and new bespoke questions, through frequent use, this offers up the opportunity to consolidate them by theme (e.g. artform; cultural consumption; marketing; net promoter score etc.) in the Culture Counts dashboard and offer them to users. This is vital in building knowledge and expertise around evaluating cultural value across the sector globally; encouraging the exchange of views and data; and ensuring a lively, open conversation about what quality is and how we interpret the data arising from these types of metrics and evaluation approaches.

In terms of additional metrics being chosen by the participating organisations, interestingly, alternative dimensions available in the Culture Counts dashboard were selected 163 times across 22 organisations. The top alternative dimensions were:

- Diversity: 'It could engage people from different backgrounds'
- Meaning: 'It moved and inspired me'
- Growth: 'It could appeal to new audiences'
- Currency: 'It made me reflect on the world we live in today'
- Platform: 'It has the potential to inspire other artists and artforms
- Like: 'Finally, what were your overall feelings about the work'
- Inquisitiveness: 'It made me want to find out more about the work'
- Atmosphere: 'I enjoy the atmosphere here'
- Innovation: 'It was introduced to the audience in a new way'
- Connection: 'It helped me to feel connected to people in the community'

A particular interesting example of how organisations approached the integration of the participatory metrics through combining them with new metrics, or by using variations of the existing metrics, was Ludus Dance.

CASE STUDY: Ludus Dance

How the identity of the work can shape metric choice

With very special thanks to Ludus Dance for sharing their data and insight for this case study.

Ludus Dance is an organisation that had signed up to be a part of the Quality Metrics National Test, whereby the focus was on testing the quality metrics. That said, Ludus made the choice to experiment further with the metrics, choosing their metrics according to respondent type. They also added a further respondent group by including the participants, and testing the participatory metrics, within one particular evaluation. This took place over their event The Lancashire Youth Dance Festival. The Festival was a combination of two days of dance workshops and classes, followed by a showcase which was open to the general public, as well as friends and family of the participants.

Through choosing their own metrics in what could be considered a more organic approach to metric selection, a further layer is understood in the organisation's intentions or focus of the work. Interestingly, although different metrics were used to across the respondent groups, peer, self, public and participants, we can see similarities in the themes.

The selection of the below metrics indicates the importance of inclusivity and connection between those from different backgrounds, both amongst the audience and the participants. Interestingly, Growth is the only metric Ludus chose for the three respondent groups: public, peer and self.

Audience: Growth: It could appeal to new audiences
Self: Collaboration: It connected other artists
Self: Growth: It appeals to a large community of interest
Participants: New people: I got to know people who are different to me
Peer: Growth: It could appeal to new audiences
Peer: Diversity: It could engage people from different backgrounds

When looking at the survey for the participants, Ludus' choice of participatory metrics presents a well-rounded selection. Aligned to the participatory metrics clusters, the emphasis for this piece of work seems to be on the development of the participants.

Generally, the metrics chosen were also frequently used by other organisations; however, it is worth noting that the above themes of inclusivity & connection are quite specific to this piece of work, which contrasts with other works evaluated. In addition, the selection of the metric New People indicates that the opportunity to meet new people was again quite specific to this piece of work whereas in the other test events across all the participatory organisations in the trial, this metric was much less likely to be chosen. Similarly, with other work evaluated with participatory metrics, the metrics selected demonstrate the unique position the participatory activity has in contribution to the broad range of participatory work produced by organisations.

As well as the core quality metrics, Ludus has also used a metric in their audience survey from the Place category, Accessibility: I find it easy to get to and from here. Whilst this is a place metric, it also relates to the importance of inclusivity, as shown overall in the chosen metrics.

When looking at the custom questions, the use of the word 'highlight' features in both the questions for the self assessors and the participants, revealing the importance of the experience of those that have been involved as part of both the creative and participating teams. Ludus also included open text questions regarding the experience for the peer and public assessors.

The importance of inclusivity and opportunity for connection with those from different backgrounds amongst the participants and audience is clear when looking at the metric choice. Not only can the metric selection enable one to see how Ludus' Festival sits amongst other participatory works, but it also enables one to see what makes it unique. This understanding contributes to the overall picture of participatory work. Ludus' specific focus on inclusivity and connection is strongly featured, with the emphasis on the participant experience. When looking at the selected metrics collectively, across all respondent groups, we get a sense of the specific intentions of the Festival.

A number or participating NPOs were involved in strategic touring work. This led to some interesting innovations around sharing their results in the Culture Counts dashboard. In other words they invited other participating NPOs in the study to share their evaluations directly through the Culture Counts dashboard (the platform allows a user to share surveys or results with any other registered user of the Culture Counts platform). The project saw ten examples of this data sharing through the dashboard. One interesting variant of this was where organisations toured the same piece to different locations, but shared all the results in one dashboard. Those organisations that started to share data in the national test through their Culture Counts dashboard tended to do so for each of their evaluations.

Finally, by way of adaptation and integration, because some NPOs were taking part in this quality metrics strand and the participatory metrics strand, some of their work had overlapping objectives and therefore those NPOs needed to gain insight into the quality of a participatory process and the the quality of any event or performance arising from that participatory process (as measured through self, peer and public opinion). As a result, we saw examples of organisations combining the quality and participatory metrics, as described in the Ludus Dance case study.

5.4.5 Staff Turnover and Resource Challenges

As advised by Culture Counts, the majority of participating cultural organisations designated one member of staff to be a super-user of the Culture Counts system. In other words, on behalf of a participating organisation that super-user familiarised themselves with the Culture Counts system; configured the self, peer and public surveys; and interfaced with their colleagues around securing the requisite levels of self, peer and public feedback, including where necessary working with volunteers or front of house teams on coordinating intercept interviewing. From the outset of the project (in evaluation terms) on November 1st 2015 to the close of the project on May 31st 2016, 14% of the the originally designated super-users of the system either left their job role, or that role disappeared for resourcing issues inside the participating organisation.

Understandably, this was very challenging for the participating organisations in terms of the continuity of their engagement in the trial which definitely impacted on the ability of some organisations to complete the target of three evaluations. This turnover of roughly one seventh of the initially inducted users presented continuity challenges within the organisations evaluating, subsequently impacting on delivery of the overall project.

5.4.6 Accessibility Issues

The evaluation processes highlighted a range of accessibility challenges that need ongoing attention, and the participating organisations also innovated in trying to overcome some of these issues. The specific accessibility issues identified by the cohort were as follows:

i. Those with visual impairment would struggle to complete the survey alone with the Culture Counts interface as it currently stands

ii. Working with children and adults where English is a second language can in some cases pose difficulties in accurately understanding the questions (e.g. 'hard to decipher between some specific words e.g. 'produced' and 'presented')

iii. Specific groups, such as those with dementia, pose very specific challenges. (from issues of informed consent to the appropriateness of a survey-based format)

iv. The survey response scales are unlikely to be clear enough for participants with 'complex individual needs'

v. Elderly respondents (e.g. are more likely to be unfamiliar with touch-screen technology and have a higher chance of conditions such as Parkinson's)

iv. For 'early years' participants (0-8) the text base interface is not appropriate

In response to these challenges the organisations innovated in a number of ways. For example, Arnolfini adapted the survey interface into a scale of unhappy (strongly disagree) to happy (strongly agree) faces to capture the response of the children and young people they were working with.

It is clear from both this strand, and the wider Quality Metrics National Test work, that like other digital platforms, and text based survey interfaces, Culture Counts will need to work with the cultural sector to facilitate access as much as possible. Culture Counts is already working with partners on exploring different screen based response modes and other interface innovations alongside efforts to accommodate within the dashboard the full range of languages in which the metrics can be expressed.

5.4.7 The language of assessment versus evaluation

In a strong mirror of the Quality Metrics National Test work on the participatory metrics, the participating organisations discussed their attitudes to evaluating their work and sharing their findings with peers and other organisations.

Organisations acknowledged that the use of standardised metrics could create anxiety around particular pieces of work being 'judged' in particular ways. Clearly, these types of evaluation approaches will only thrive if the data proves insightful to cultural organisations, and they are encouraged and supported to explore the resulting data in ways that put the emphasis on critical reflection and improvement, as opposed to a narrow emphasis on 'audit' and 'performance reporting.'

The participating organisations talked about the importance of a number of enabling factors that will help build openness across the cultural sector to these forms of evaluations:

i. To use the language of evaluation, and improvement, as opposed to the language of 'audit' and 'assessment'

ii. To build understanding that the value in these types of evaluation is when they are a collaborative exercise, between the self, peer and public (and participants), in which dialogue and reflection are vital to interpreting and gaining insights from the results. Organisations talked about the how the metrics 'lead to useful evaluative exchanges'

iii. That insightful evaluation is about asking good questions; being open to the answers, and working through with others what they might mean

5.4.8 Resource and context specific challenges

Some organisations indicated to the Culture Counts team that they would have carried out some, or more, intercept interviewing at their events if they had access to more tablets and supporting technology. In some instances, organisations were in a 'work-around' situation from the outset of the national test. For example, a few evaluations involving rural touring work in presenting locations without WIFI or 3G/4G coverage, necessitated the use of paper versions of the quality metrics survey, with the data uploaded electronically after the event.

5.5 Summary

This chapter has detailed the contours of cohort engagement in the Quality Metrics National Test, and explored in more detail some of the challenges, issues and opportunities arising out of this project. As the Culture Counts team expected, we witnessed a wide range of evaluation practices across the cohort, including interesting examples of adaptation and innovation. Overall, given the demanding nature of the original EOI terms and conditions, which meant completing evaluations in a short time frame, and balancing other evaluation commitments, we think the levels of engagement shown by the participating NPOs and MPMs was impressive, resulting in real insight and of course a very substantive data set on the quality of the work being presented.

6. CHAPTER SIX: Conclusion

What are the key conclusions that can be drawn from this Quality Metrics National Test? They cluster into three themes:

6.1. Self-driven scalability

It can be tempting when a project has met its aims to forget that at the outset of this project there was genuine uncertainty on the part of ACE and Culture Counts, given the demands on funded organisations, as to whether they would sign up in large numbers to take part in this Quality Metrics National Test, and then be able to self-drive large scale evaluation activity using the quality metrics and the Culture Counts platform, all within what was effectively a six-month time frame for the evaluation activity. As we noted in Chapter 1, it is therefore pleasing that 91% of the 150 participating NPOs and MPMs became 'active' participants in this national test.

The project has therefore resoundingly confirmed that funded arts and cultural organisations, if offered the right tools and support, can self-drive large scale evaluation activity, engaging in new ways with peers and audiences about the quality of their work. This would suggest that the quality metrics and the sector's evident interest in being able to measure their creative intentions, allied to tools that help the arts and cultural sector to collect and analyse data easily and at scale, offers up the prospect of a much richer conversation about cultural value in the future, informed by big data.

6.2. The quality metrics and the aggregate analysis

Taken together the aggregate scores in this Quality Metrics National Test suggest:

- The work presented and analysed in this study received a broadly positive response from peer and public respondents, and largely met the (quite high) prior creative expectations of the creative teams involved in its production (self assessors)
- When it comes to measuring the quality of a cultural experience three dimensions in particular - challenge, distinctiveness and relevance - generally score lower across all respondent categories than the other six dimensions
- The clustering of self, peer and public responses in relation to these metrics suggests that audiences are adept at assessing them, with their judgements showing broad alignment with self and peer responses.
- The participating cultural organisations largely met their creative intentions, as measured by the degree of alignment between their self prior scores for each dimension and the corresponding aggregate scores for peer and public respondents
- Peer responses (as we have seen in all previous evaluations) are consistently lower across all dimensions than self and peer responses

The strong adherence of the participating NPOs and MPMs to the self prior and post rating process within their evaluations indicated their understanding of how the self, peer and public triangulation process provides them with real insight into how far they are meeting their creative intentions for a particular piece of work. The aggregate results confirm that self prior and post ratings were well aligned to public ratings across the evaluations in this study, showing that the self assessor communities did indeed think through their creative intentions clearly and with considerable predictive accuracy as judged by alignment with peer and public response.

The analysis also revealed very interesting differences by artform. As we noted in Chapter 3, there is a greater degree of misalignment between some sets of self prior and peer ratings on the 'risk' and 'originality' dimensions when examined by artform, with the peer ratings being much lower than self ratings for both dimensions in some artform categories. Culture Counts thinks these artform differences are a particularly interesting area of future research, as are the other artform attribute differences presented in Chapter 4.

If the metrics become widely used across the cultural sector, and as we move from mid-scale to big data across these types of dimensions and artform categorisations, any persistent and marked variations in say risk ratings by artform; or between self and peer ratings on risk and originality by artform; could then by explored through detailed dialogue across the sector. One outcome might be that 'risk' as a dimension measure for self and peers is thickened up with additional metric components. Helpfully, the data will inevitably drive consideration of what is notable and discussion worthy for the sector, and where insight can be gained by further development and detailed enquiry.

The aggregated data set from this Quality Metrics National Test is available from Arts Council England.

6.3. The future potential of this approach

The overall evaluation approach facilitated by the features of the Culture Counts system, mirrored in the design and analysis of the aggregate data set from this Quality Metrics National Test, allows the arts and cultural sector:

- To present a very clear story on quality which does not over simplify the findings

- To use flexible co-produced metadata frameworks, for example relating to artform descriptions, to demonstrate both the variety and plurality of work being produced by the funded portfolio; and to allow a rich analysis of quality by artform and artform attribute.

The analyses in Chapters 3 and 4 demonstrates that there is enormous potential in this approach, allowing for larger scale aggregation of the data whilst maintaining real granular detail in the results. The approach effectively unites data across the standardised quality metrics, artform, artform attributes, and open data into a powerful prism through which to better understand quality. As we have seen this will deepen understanding of how artform and certain attributes of work influence quality, and offers up the potential to produce a very wide range of analytical and reflective insights.

More broadly, the approach underlines that the wider value of the standardised quality metrics, and a platform like Culture Counts, is the capacity they offer to capture both survey and metadata at such scale so that as the user base of the quality metrics grows, and we move from small scale to genuine big data, all kinds of added value analyses becomes possible revealing new patterns and subtleties around cultural value.

Crucially, those patterns and the interpretation of that data will be driven and widely discussed by the creative professionals that make the work, ushering in an era of co-produced quality metrics, co-produced analytical frameworks, and a co-produced conversation about cultural value, informed by big data.

Richard Long, TIME AND SPACE, 2015

© Stuart Whipps
Courtesy: Arnolfini

Acknowledgements

All of the cultural organisations who took part in this strand of work volunteered to give up their time to carry out their evaluation activities, and attend the Learning and Insight sessions across the country. As this report has demonstrated, they all carried out a huge amount of work in a comparatively small period of time.

Everyone at Culture Counts would like to thank all the individuals and organisations for taking part, for giving up their precious time, and being so candid and generous with their insights.

We would also like to thank Carl Stevens, our project manager at Arts Council England, who offered constructive support throughout the project.

All errors and omissions remain ours alone.

Culture Counts Team, June 30th, 2016.

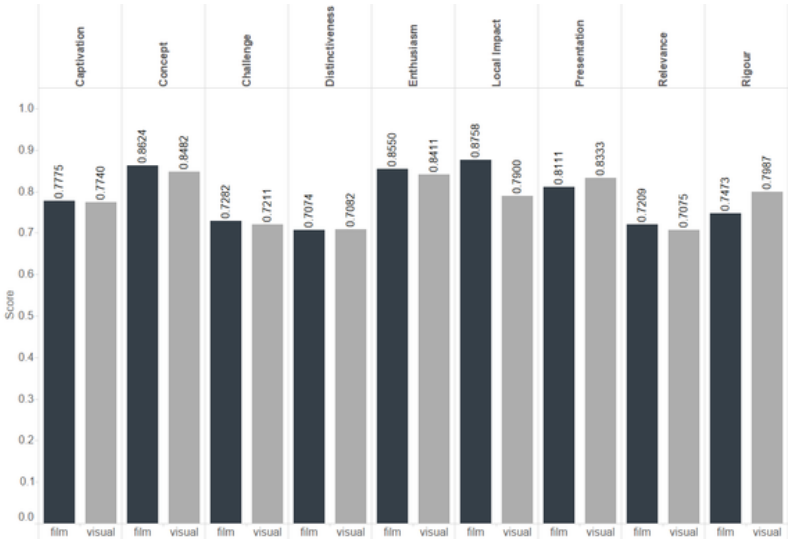
Appendix 1: Participating Organisations

This list contains all organisations who signed up including those who focused solely on the Participatory Metrics Strand.

&Co.
20 Stories High
20-21 Visual Arts Centre
Abandon Normal Devices
Action Transport Theatre
Akram Khan Company
Albert and Friends Instant Circus
Anvil Arts
Arc
Arnolfini
Artsadmin
Aspex
Baltic Mill
Barbican
Battersea Arts Centre
Beaford Arts
Bike Shed
Birmingham Repertory Theatre
Birmingham Royal Ballet
Birmingham Symphony Orchestra
Black Country Living Museum
Book Trust
Bournemouth Symphony Orchestra
Brass Bands England
Brewery Arts Centre
Brighter Sound
Brighton Dome
Bristol Museums Galleries and Archives
Bristol Old Vic
Bush Theatre
Cambridge Junction
Candoco
Carousel
Centre for Chinese Contemporary Art
Cheshire Dance
Children's Discovery Centre
Clod Ensemble
Collective Encounters
Colston Hall
Coney
Contact
Corn Exchange Newbury
Creative Arts East
Crying Out Loud
Curve
Customs House
Dada Fest
Dance East
Dance Umbrella
DARTS (Doncaster Community Arts Limited)
Dash Deda
Derby Museums
Devon Guild of Craftsmen
Durham County Council
Eastern Angles
Eden Arts
Emergency Exit Arts
Engage
English National Ballet
English Touring Theatre
Entelechy Arts Limited
Extant
FACT
Farnham Maltings
Fevered Sleep
Fitzwilliam Museum
Forced Entertainment
Free Word Centre
Fuel Theatre
Fun Palaces
Future Everything
GemArts
Gulbenkian Theatre
Halfmoon Theatre
Halle
Helix Arts
HOME
Improbable
Ironbridge Gorge Museum Trust
Jacksons Lane
Jazz North
J-Night
Junction
Lakeland Arts
Lawnmowers Independent Theatre Company
Lawrence Batley Theatre
Leeds Art Gallery
Lift Festival
Live Theatre
Liverpool Philharmonic
Ludus Dance
MAC Birmingham
Mahogany Opera
Manchester Art Gallery
Manchester Camerata
Manchester Museum
Meadow Arts
Mercury Theatre
Merseyside Dance Initiative
Mind the Gap
More Music
Museum of London
Music in the Round
National Centre for Craft and Design
National Youth Jazz Orchestra
New Vic Theatre
New Writing South
Norfolk & Norwich Museums
Norfolk Museums Service
Northern Stage
Nottingham Playhouse
Oldham Coliseum
Oxford Playhouse
Pavilion Dance South West
Peckham Platform
Pentabus
People's Palace
Pilot Theatre
Prism Arts
Queens Theatre
Rambert
Rifco Arts
Roundhouse
Royal Exchange Theatre
Royal Shakespeare Company
Rural Arts
Sadlers Wells
Sage Gateshead
Salisbury Arts Centre
Salisbury Playhouse
Serpentine Galleries
Seven Stories
Shape Arts
Sheffield Theatres Trust
Siobhan Davies Dance
Situations
Sound and Music
South Asian Arts UK
South East Dance
Southbank Centre
Spike Island
Spot On Lancashire
Spread the Word
Streetwise Opera
Tangle International
The Albany
The Lowry
The Met
Theatre Royal Stratford East
Titchfield Festival Theatre
Tobacco Factory Theatres
Tullie House
Turner Contemporary
Tyne and Wear Archives and Museums
Vane
Watermans
Watershed
Whitworth Art Gallery
Wolsey Theatre
Wolverhampton Art Gallery
Wordsworth Trust
Writers Centre Norwich
Writing Squad
York Museums Trust

Appendix 2: Supplementary Data Charts

Figure A1: Visual Arts and Film
As a sensory artform, visual contains the broad artform film and many detailed artforms and medium attributes. Differences between visual arts (not film) and film were measured and the only difference observed was for Local Impact with a small difference for Rigour.



Figures A2-A50: Isolated artforms and attributes by respondent category
The following 49 respondent comparison charts present data based on isolated artforms or artform attributes. Self prior scores are in light grey, peer scores in teal and public scores in orange.

Figure A2: Sensory Artform – Visual

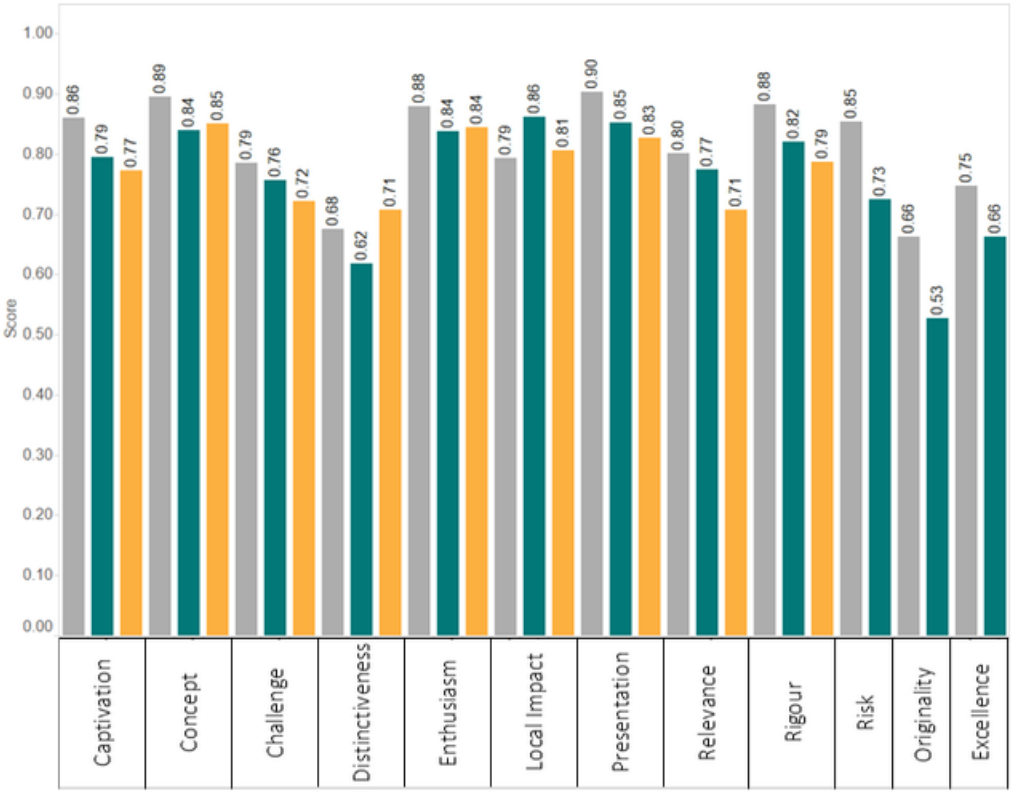


Figure A3: Sensory Artform – Sound

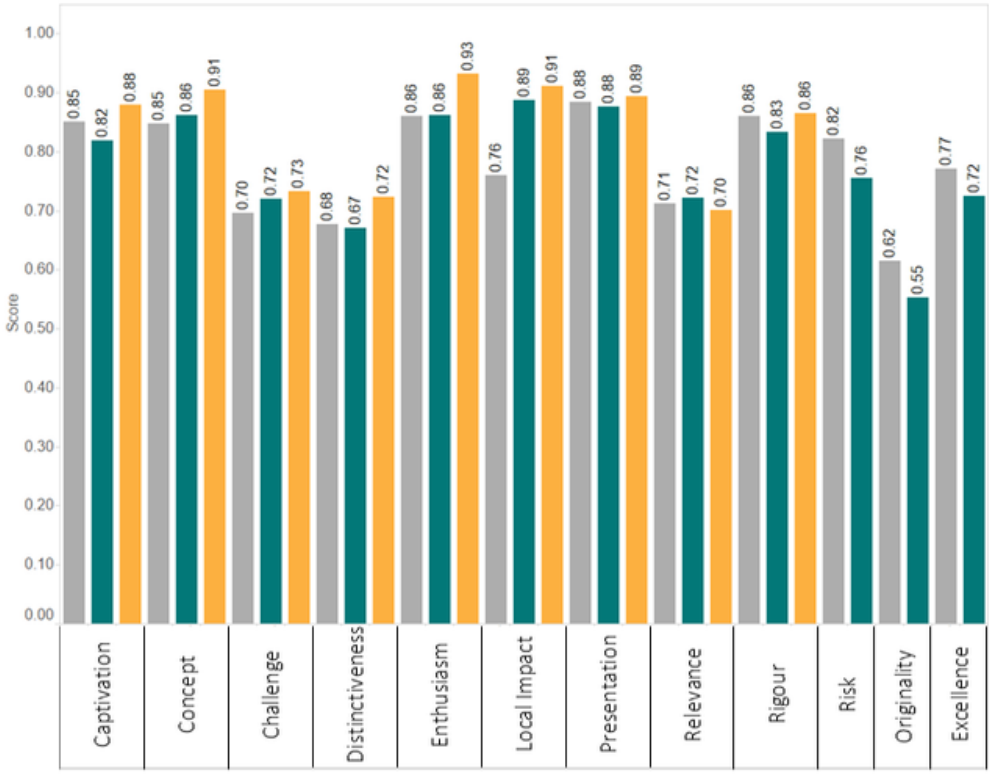


Figure A4: Sensory Artform – Movemennt

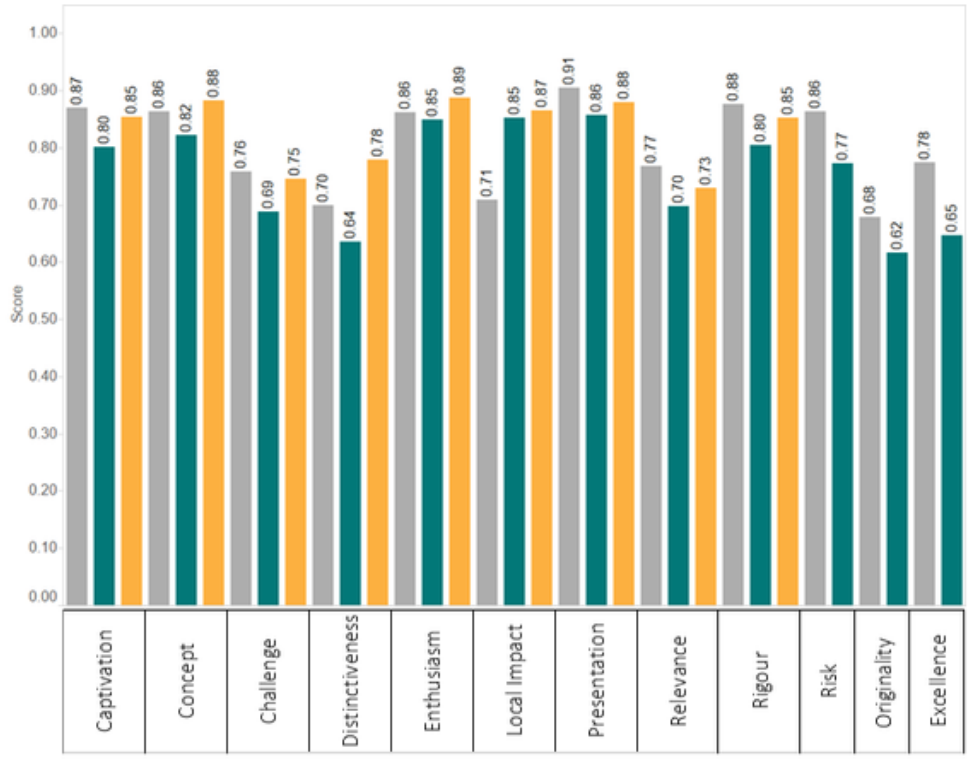


Figure A5: Broad Artform – Dance

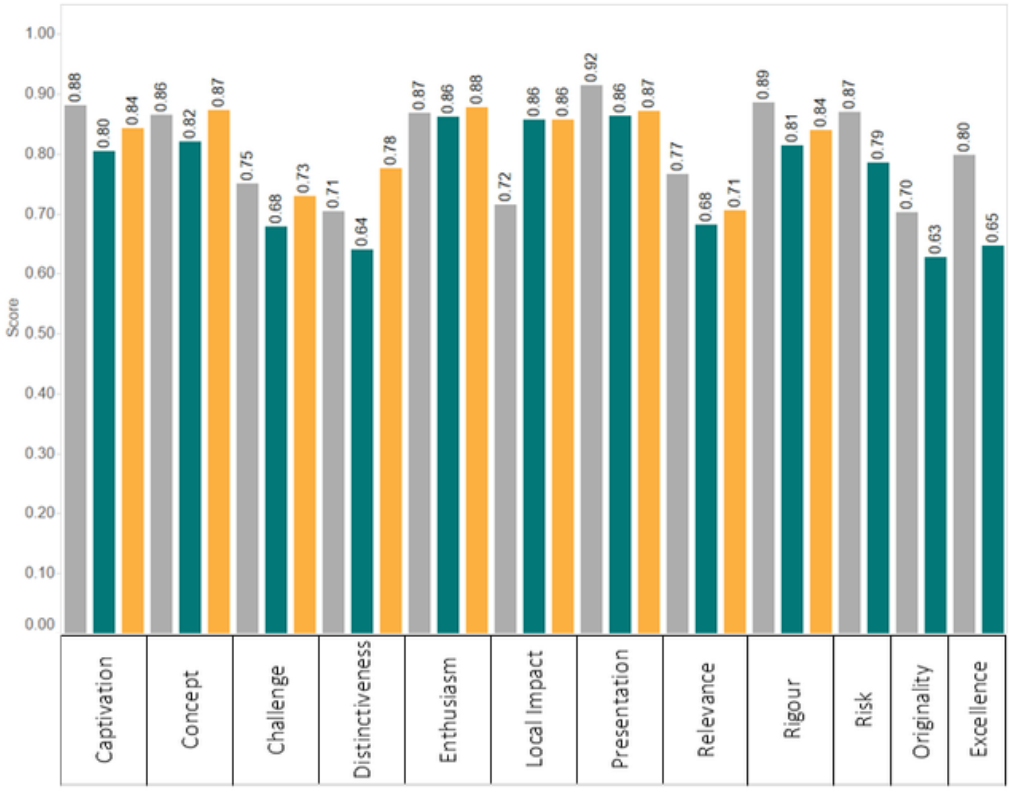


Figure A6: Broad Artform – Film

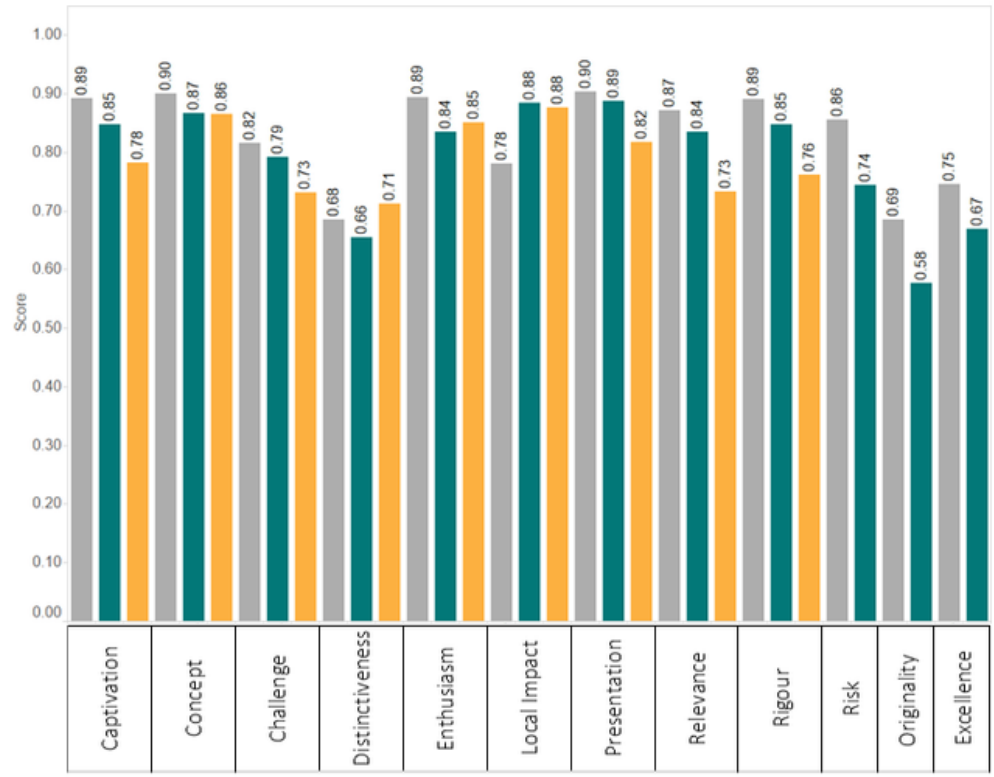


Figure A7: Broad Artform – Literature

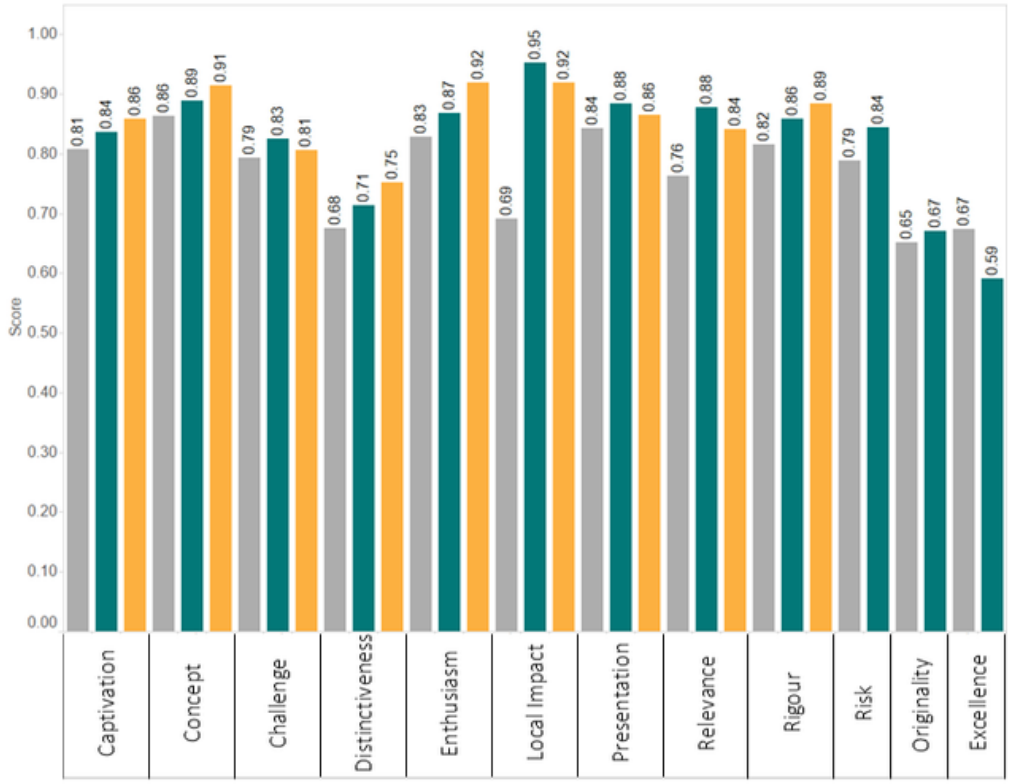


Figure A8: Broad Artform – Music

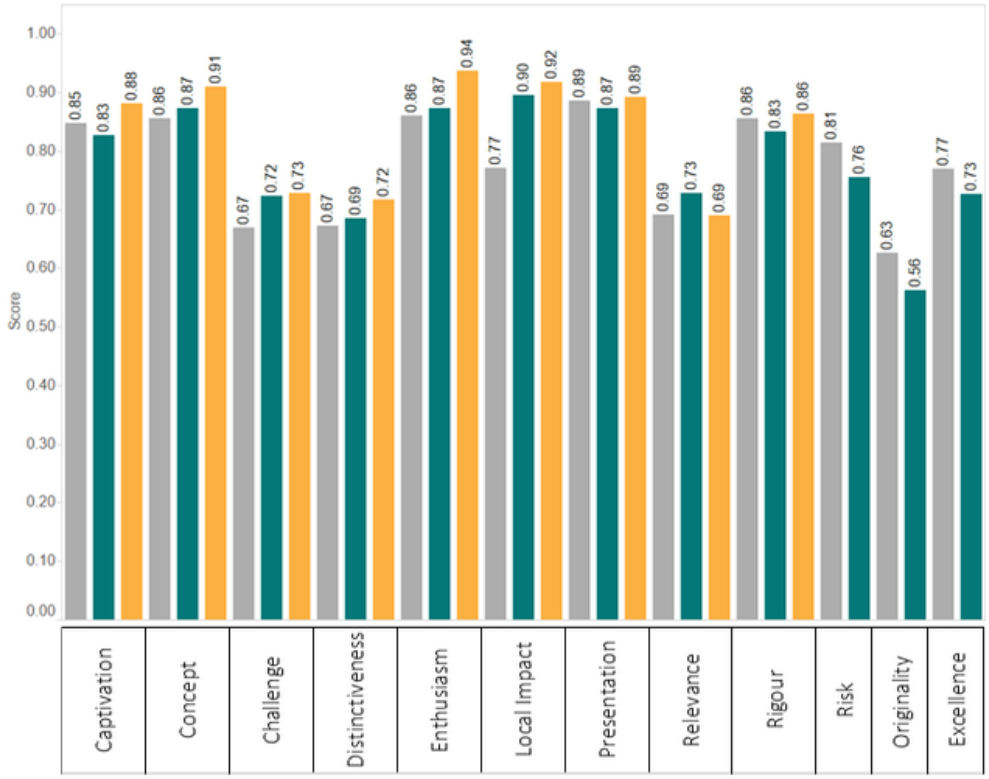


Figure A9: Broad Artform – Theatre

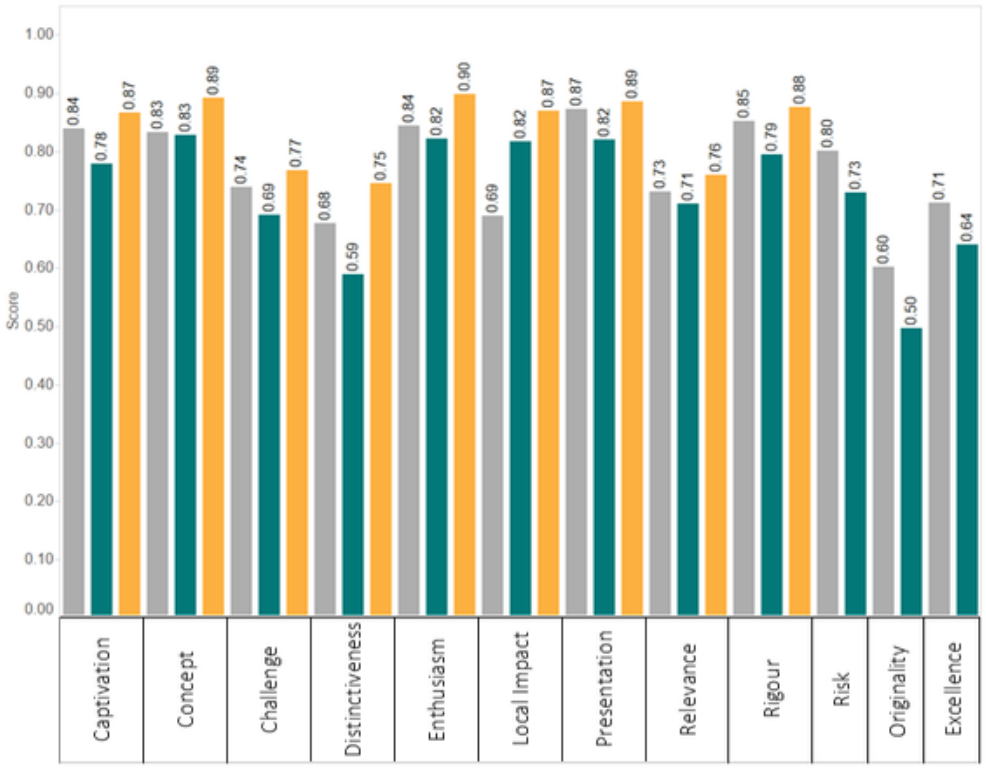


Figure A10: Detailed Artform – Ballet

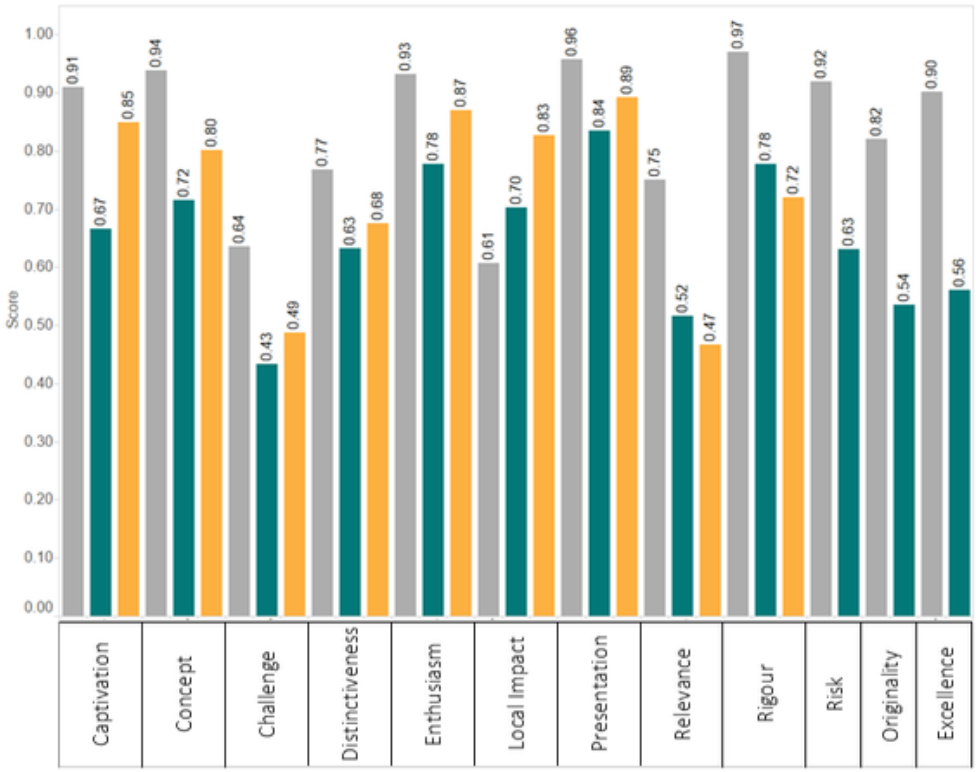


Figure A11: Detailed Artform – Craft

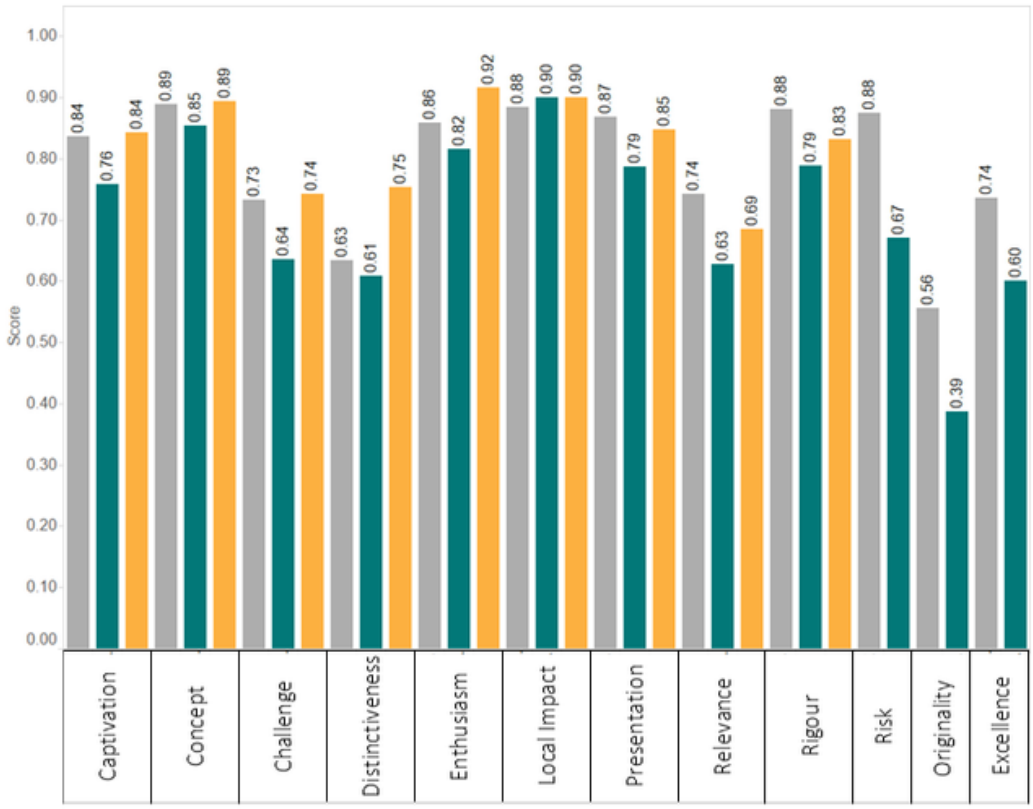


Figure A12: Detailed Artform – Drawing

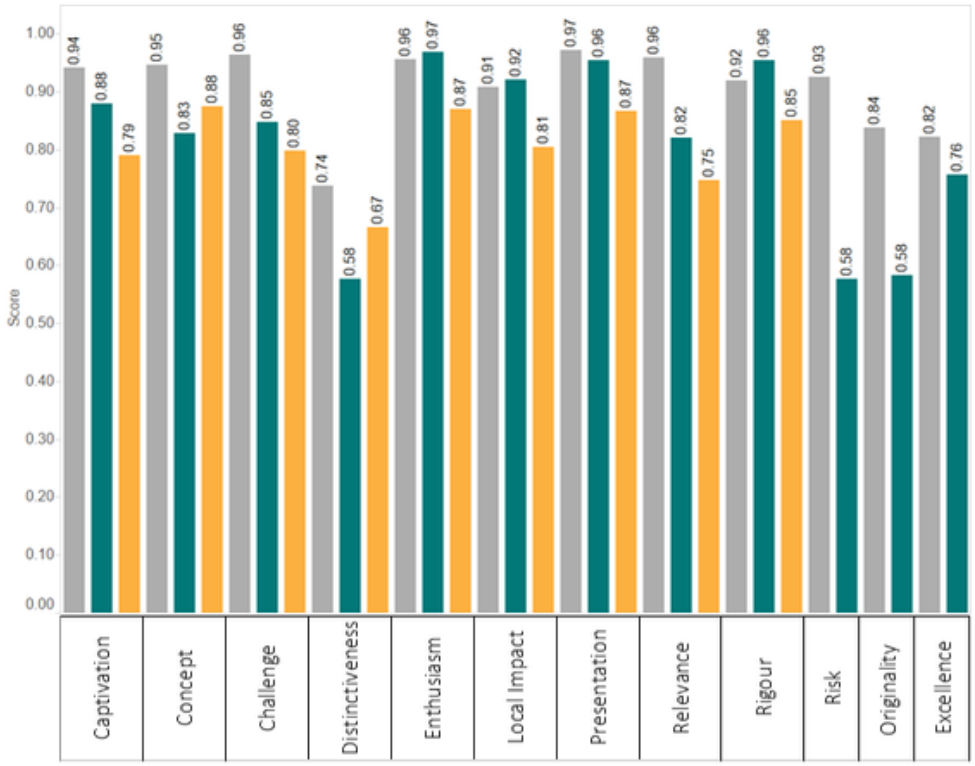


Figure A13: Detailed Artform – Musical Theatre

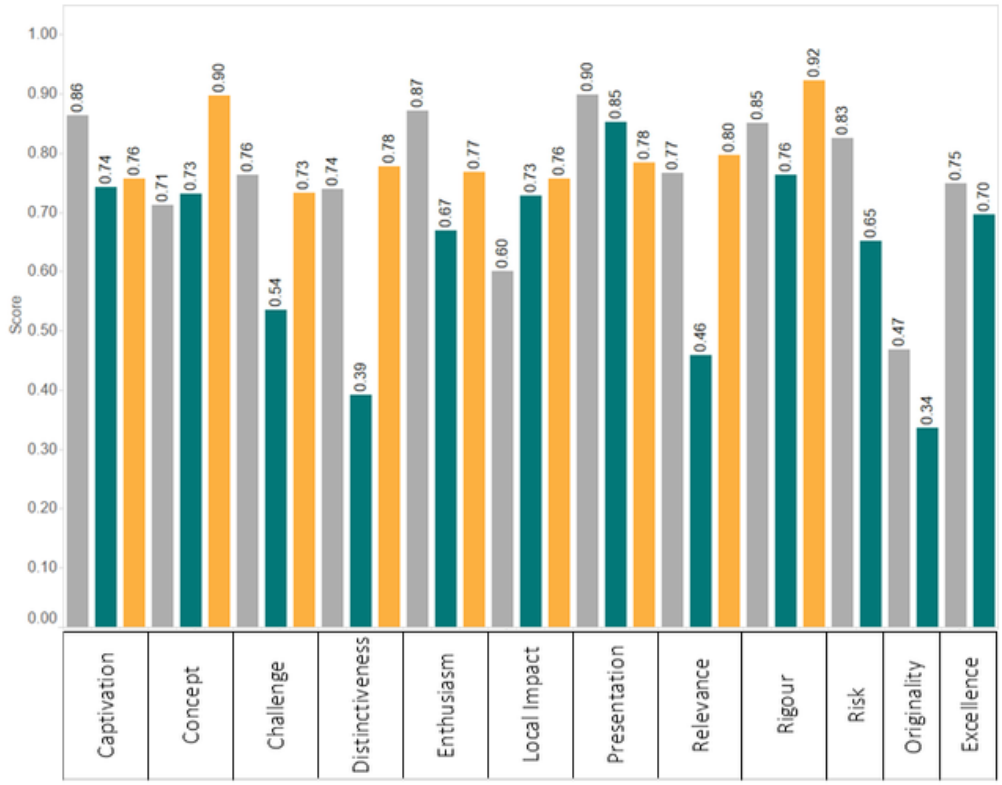


Figure A14: Detailed Artform – Painting

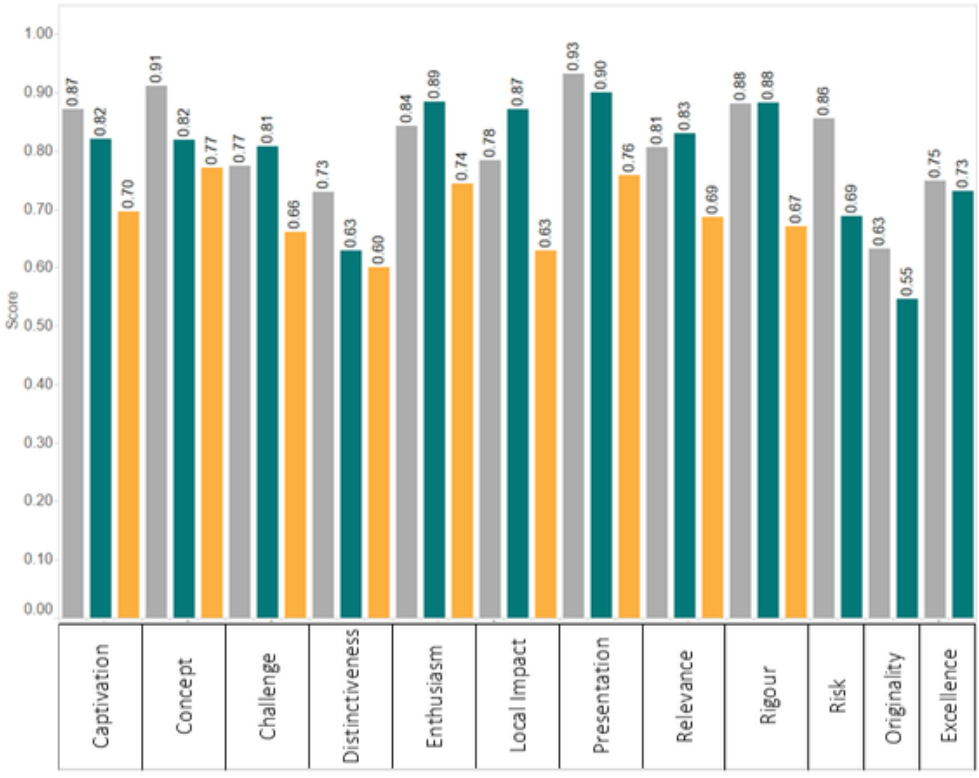


Figure A15: Detailed Artform – Photography

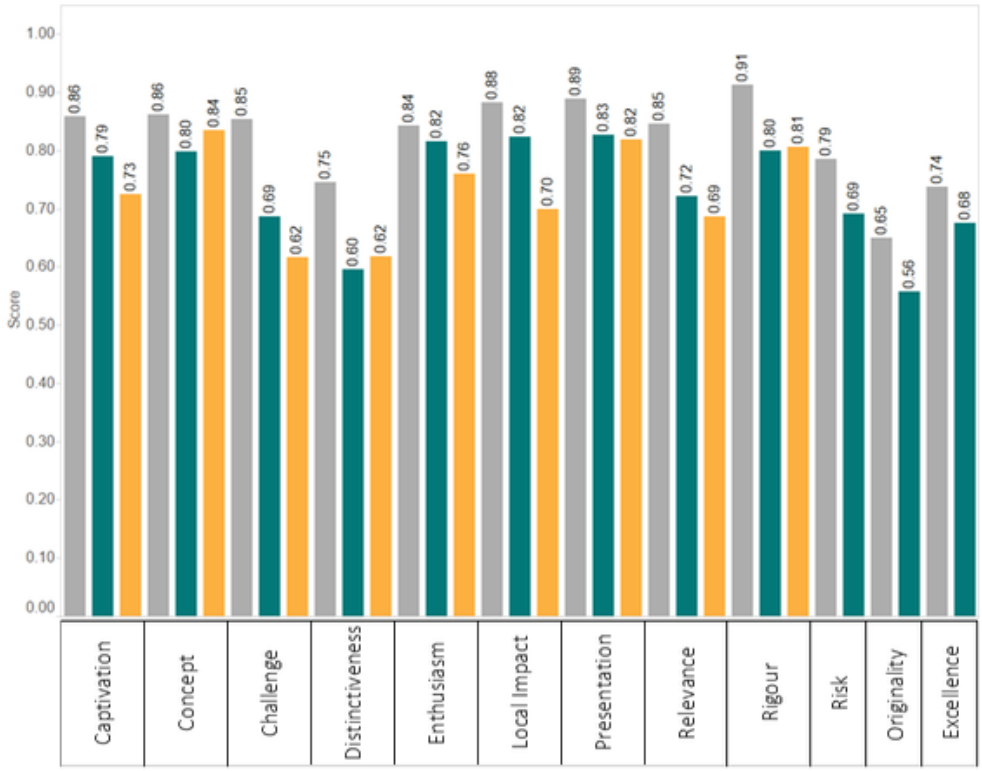


Figure A16: Detailed Artform – Physical Theatre

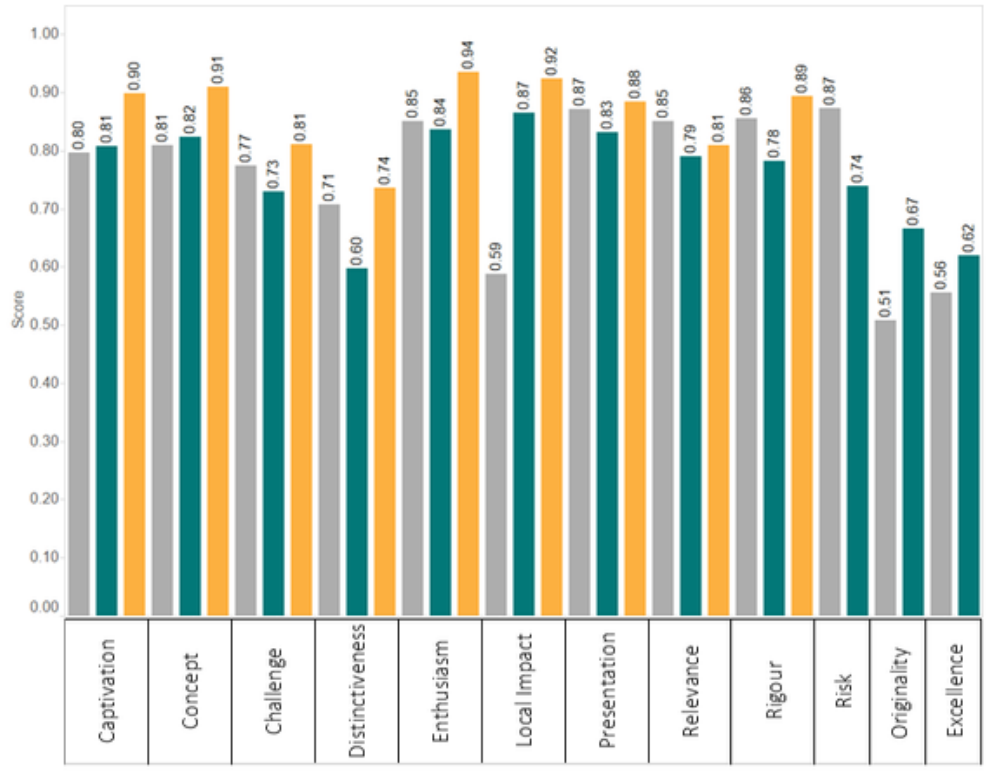


Figure A17: Detailed Artform – Play

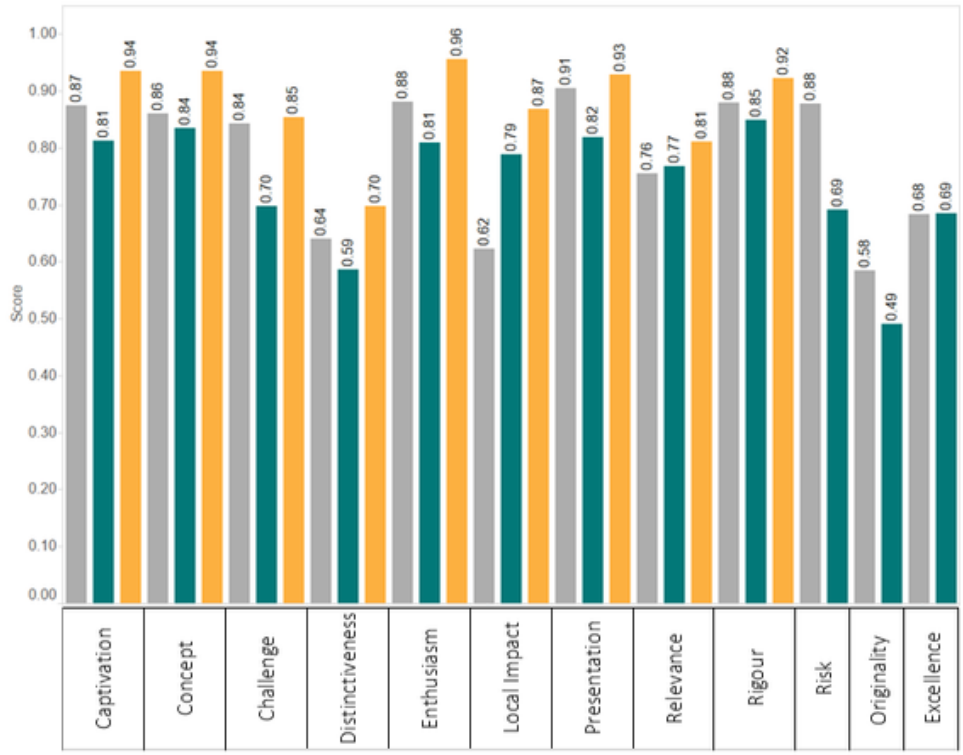


Figure A18: Detailed Artform – Poetry

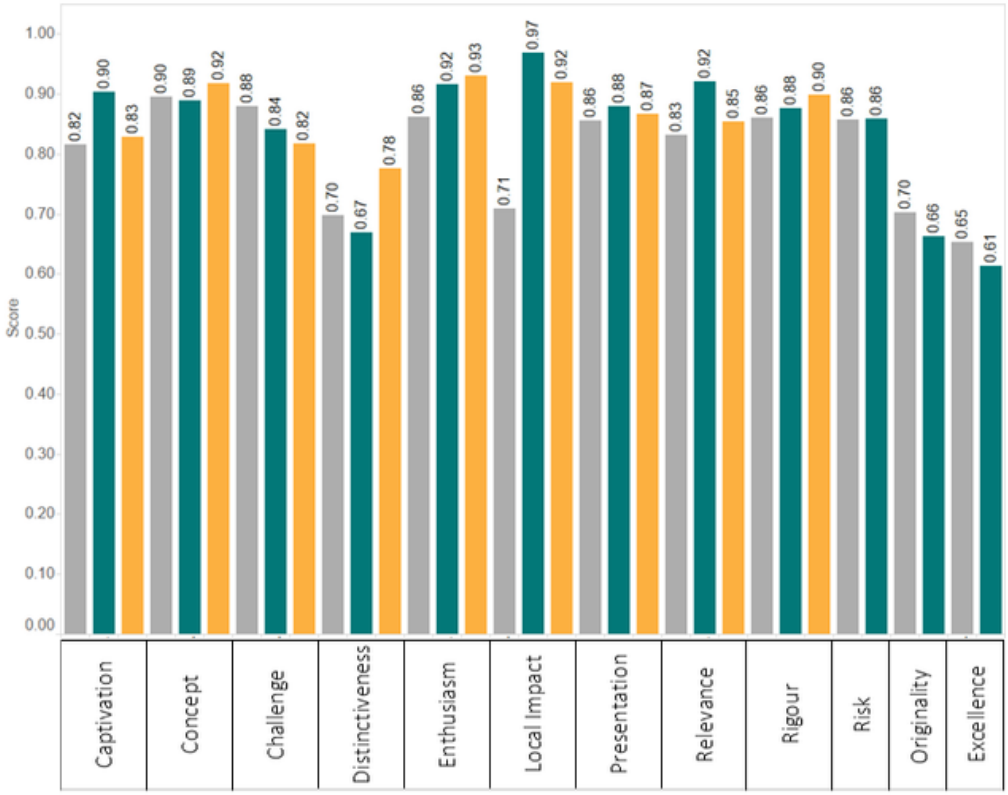


Figure A19: Detailed Artform – Sculpture

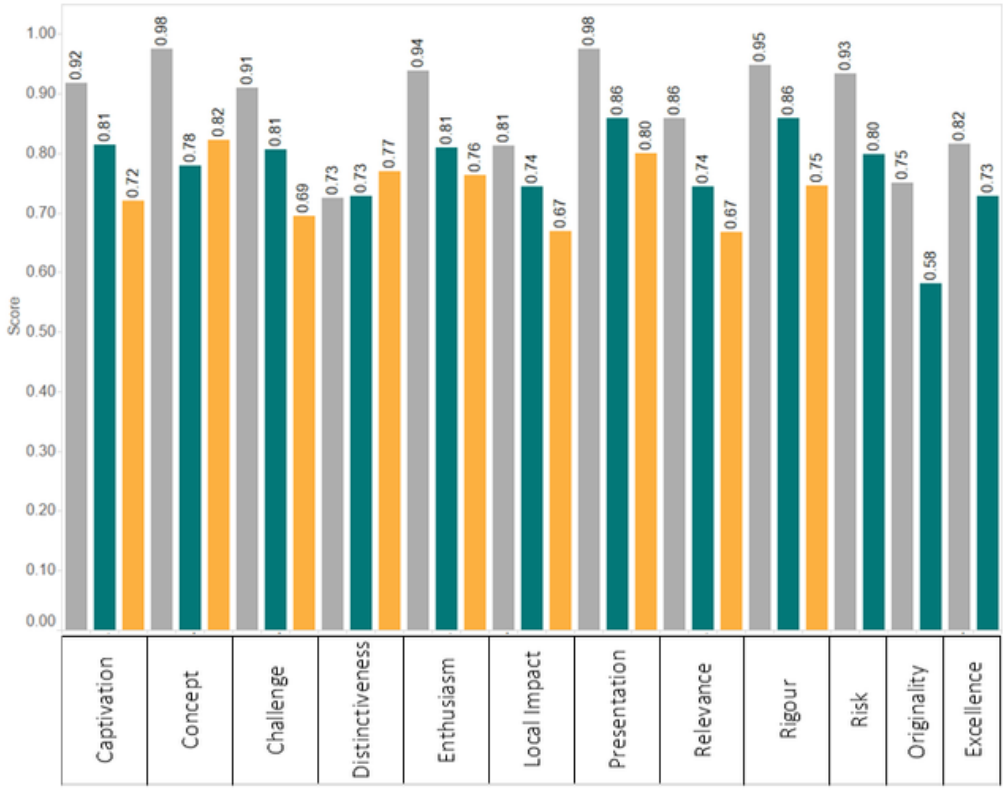


Figure A20: Detailed Artform – Spoken Word

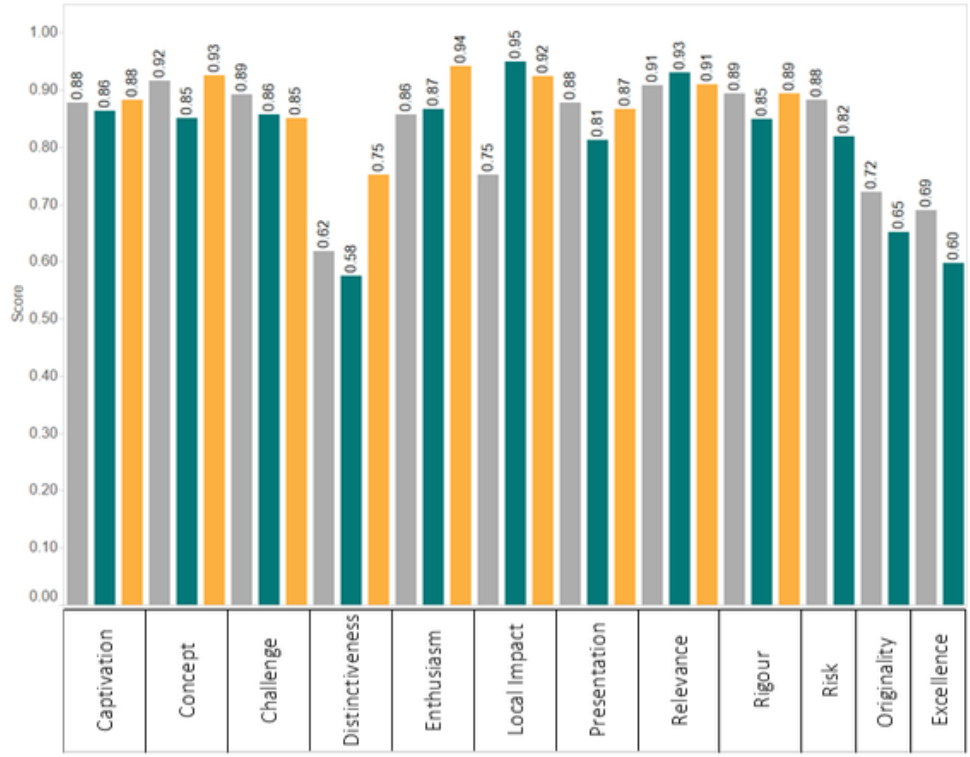


Figure A21: Detailed Artform – Storytelling

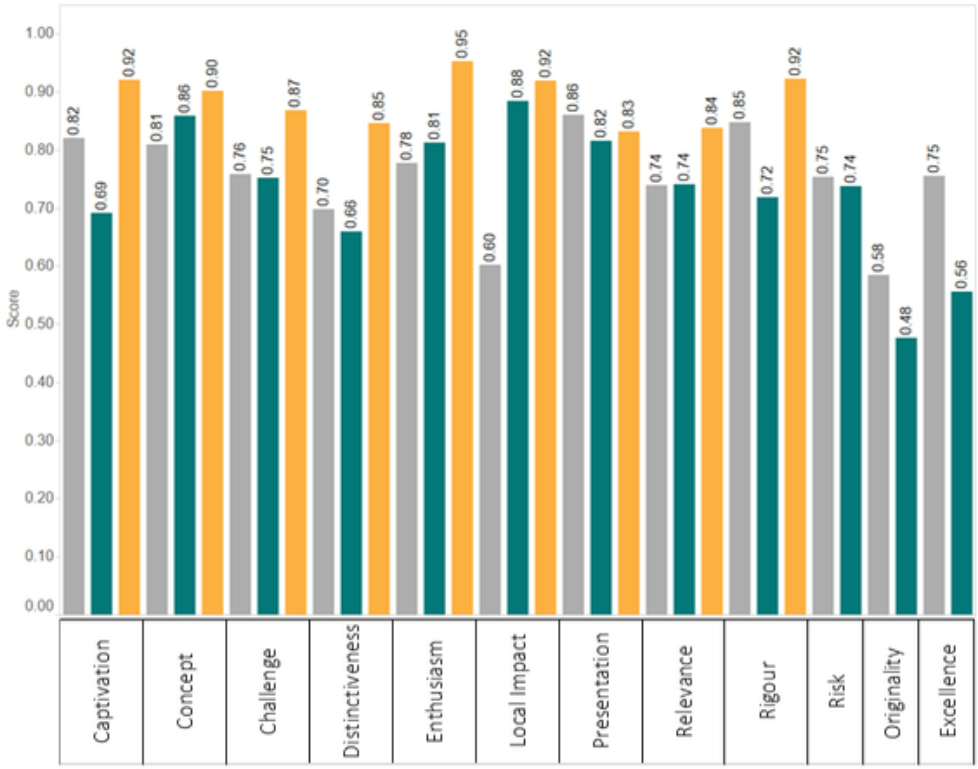


Figure A22: Chronological Attribute – Classical

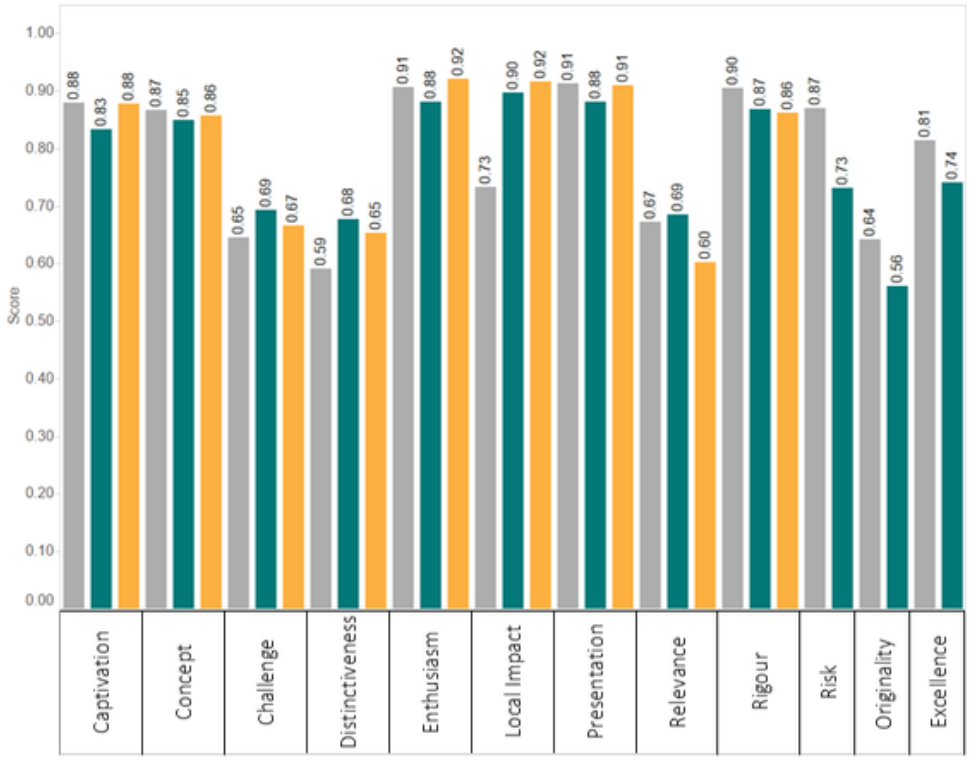


Figure A23: Chronological Attribute – Modern

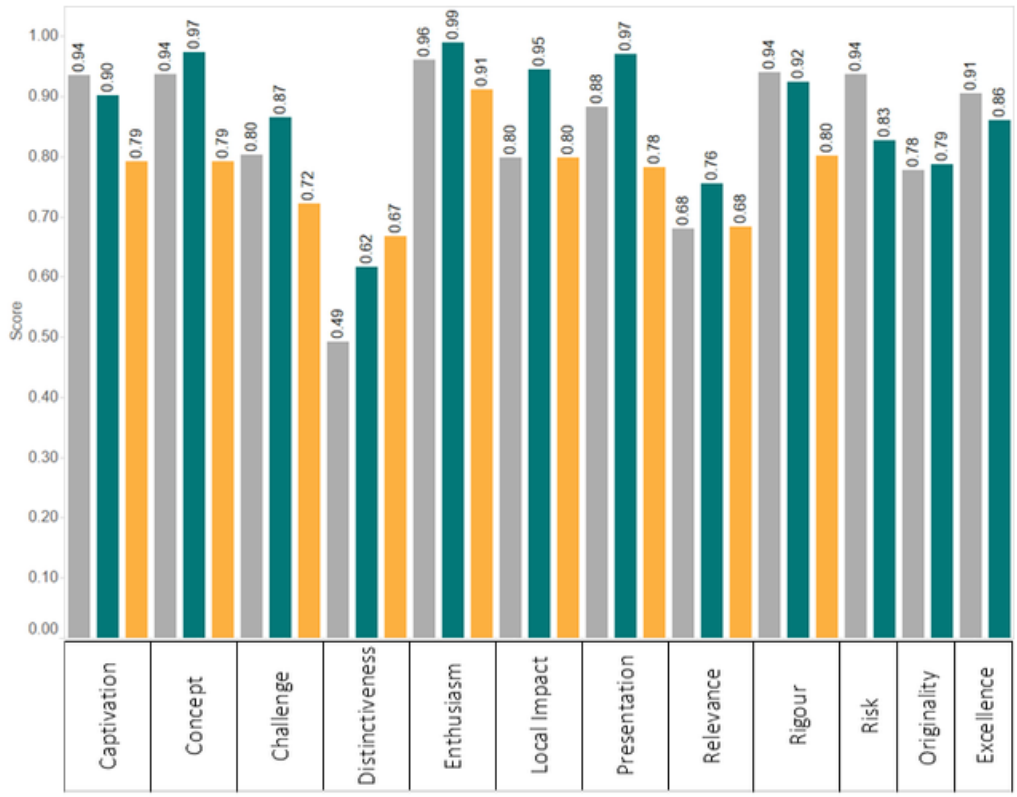


Figure A24: Chronological Attribute – Contemporary

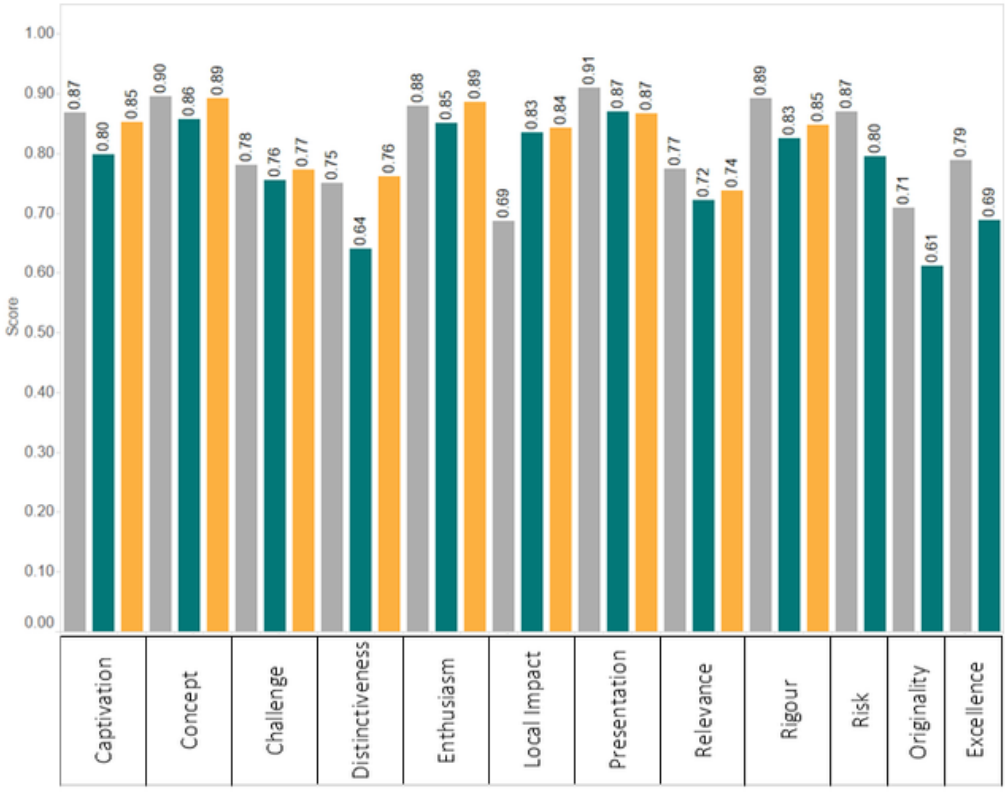


Figure A25: Chronological Attribute – New

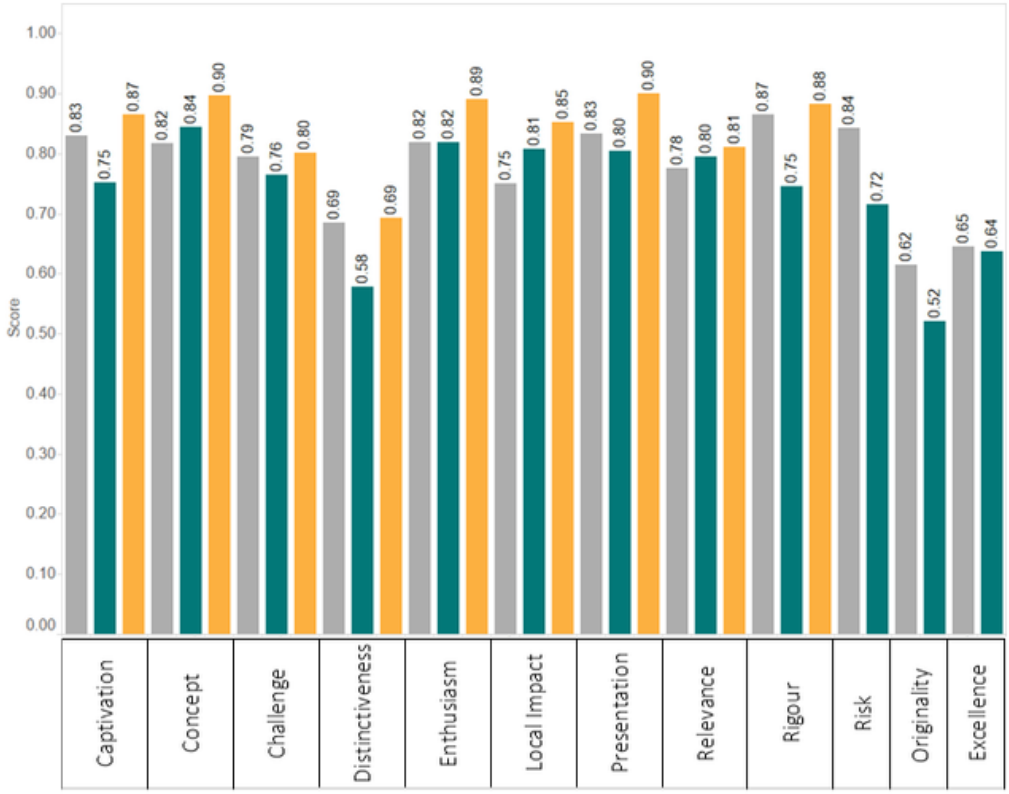


Figure A26: Genre – Comedy

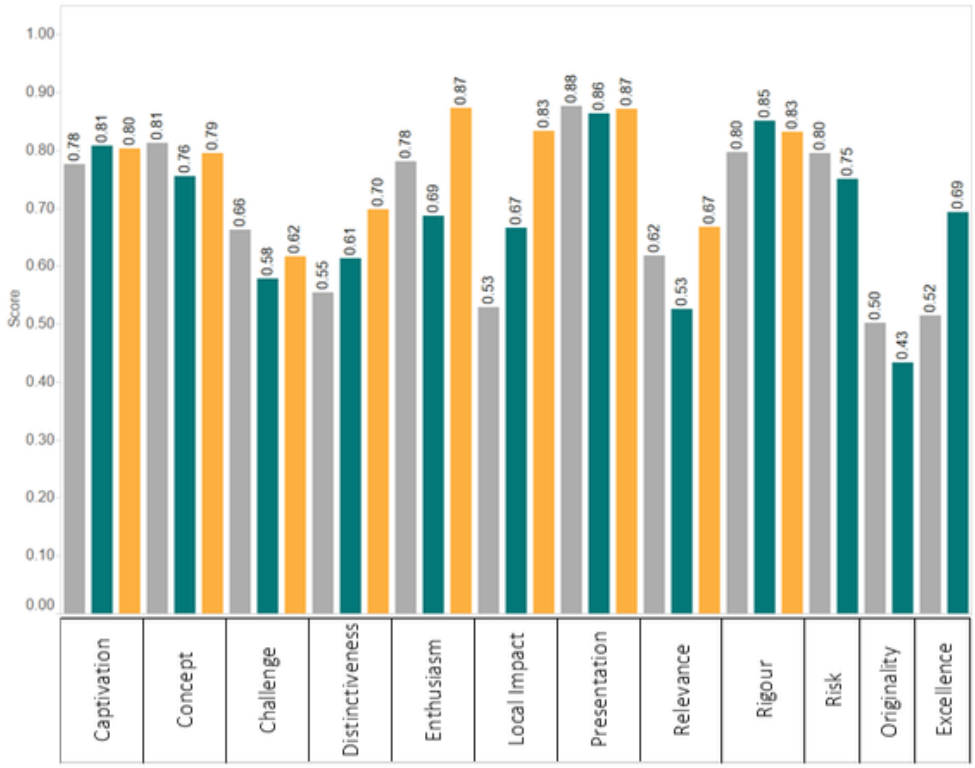


Figure A27: Genre - Documentary

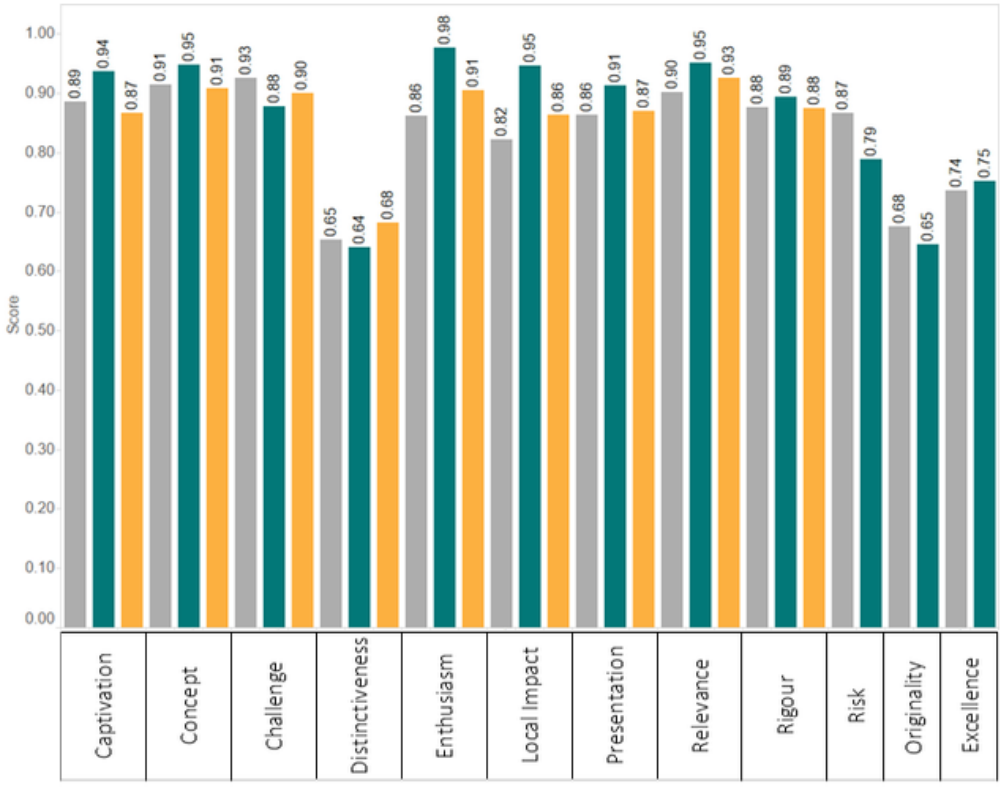


Figure A28: Genre – Drama

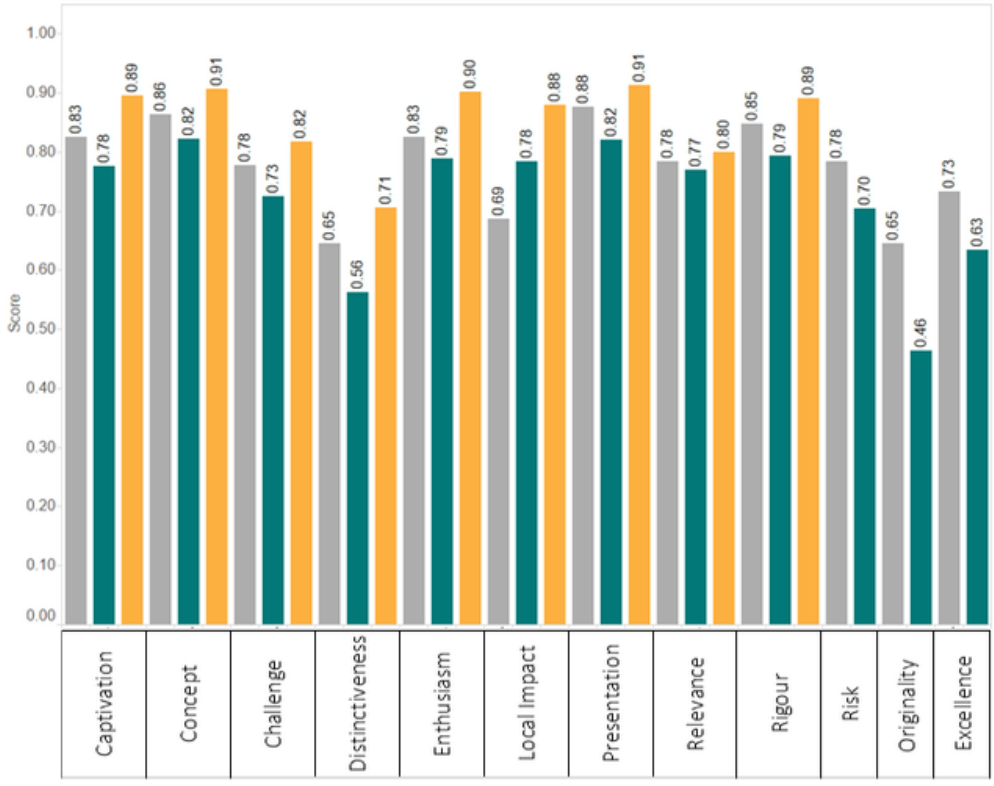


Figure A29: Medium – Ceramic

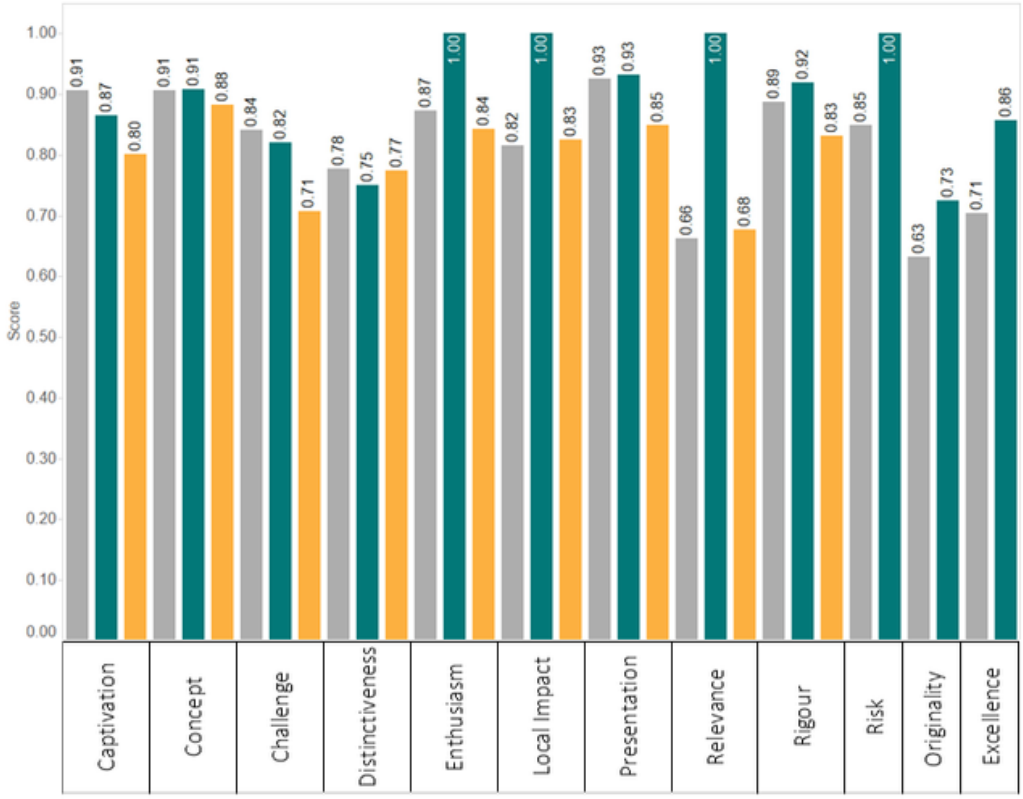


Figure A30: Medium – Digital

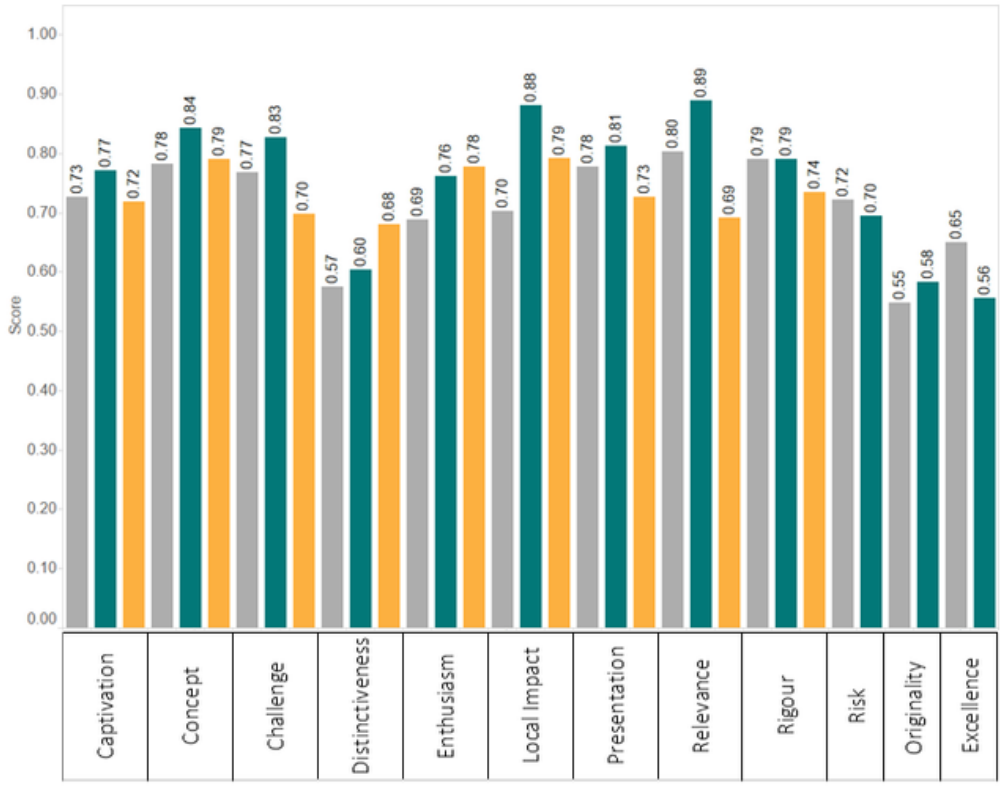


Figure A31: Medium – Electronic

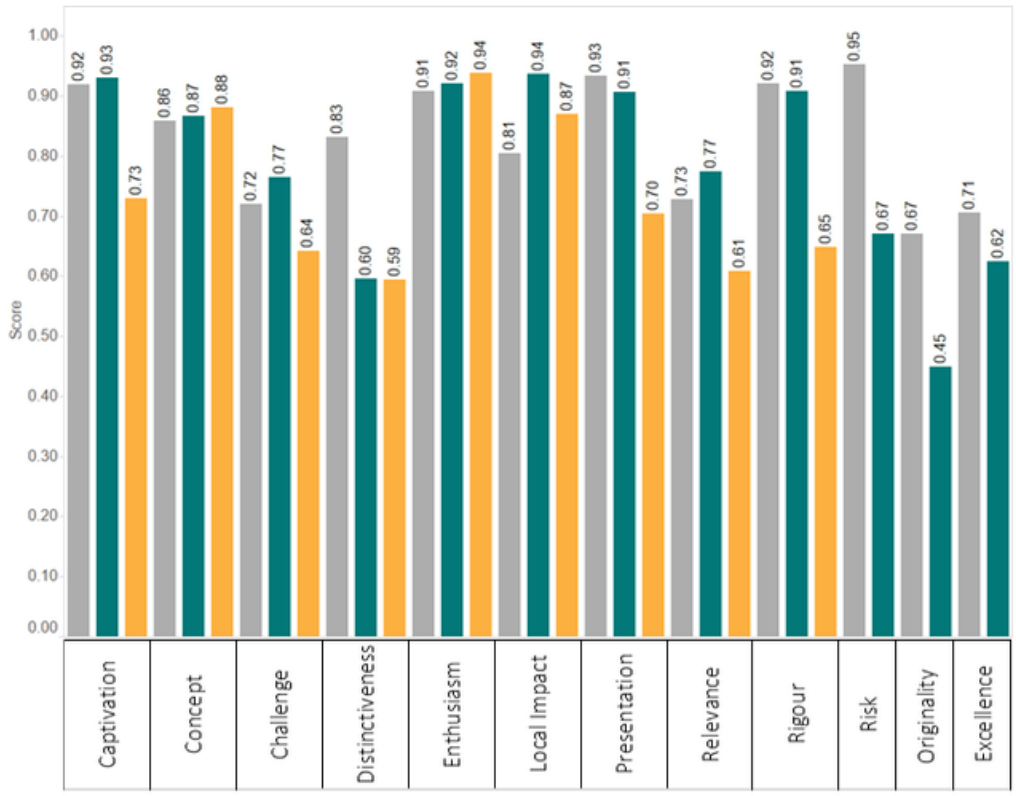


Figure A32: Medium – Mixed

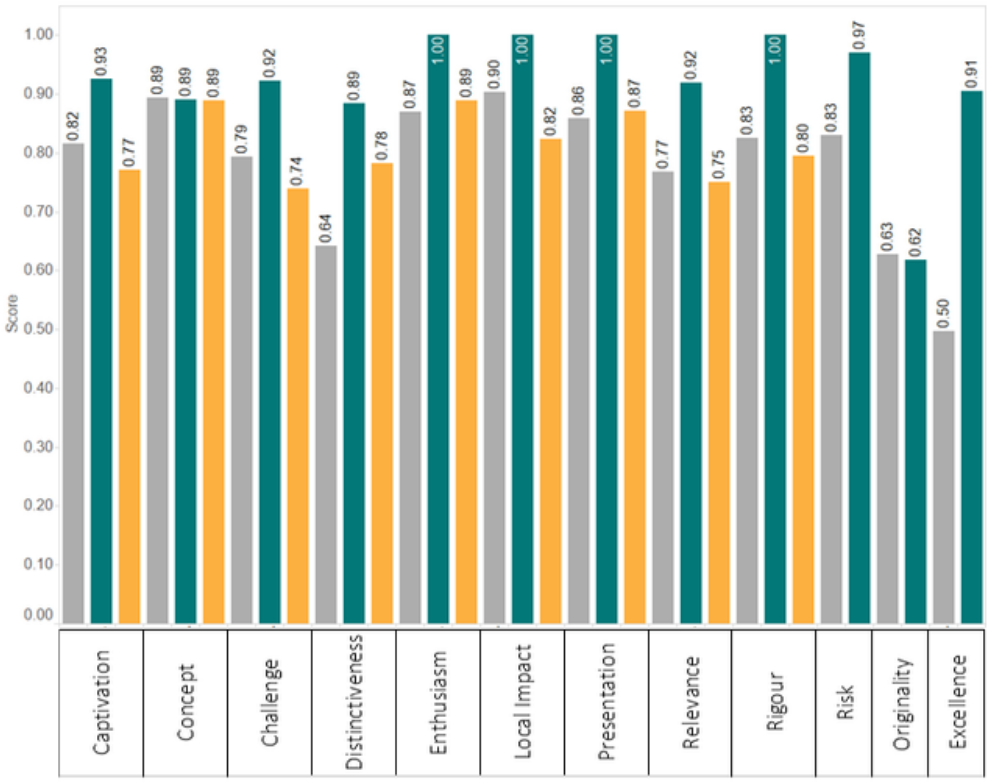


Figure A33: Medium – Orchestral

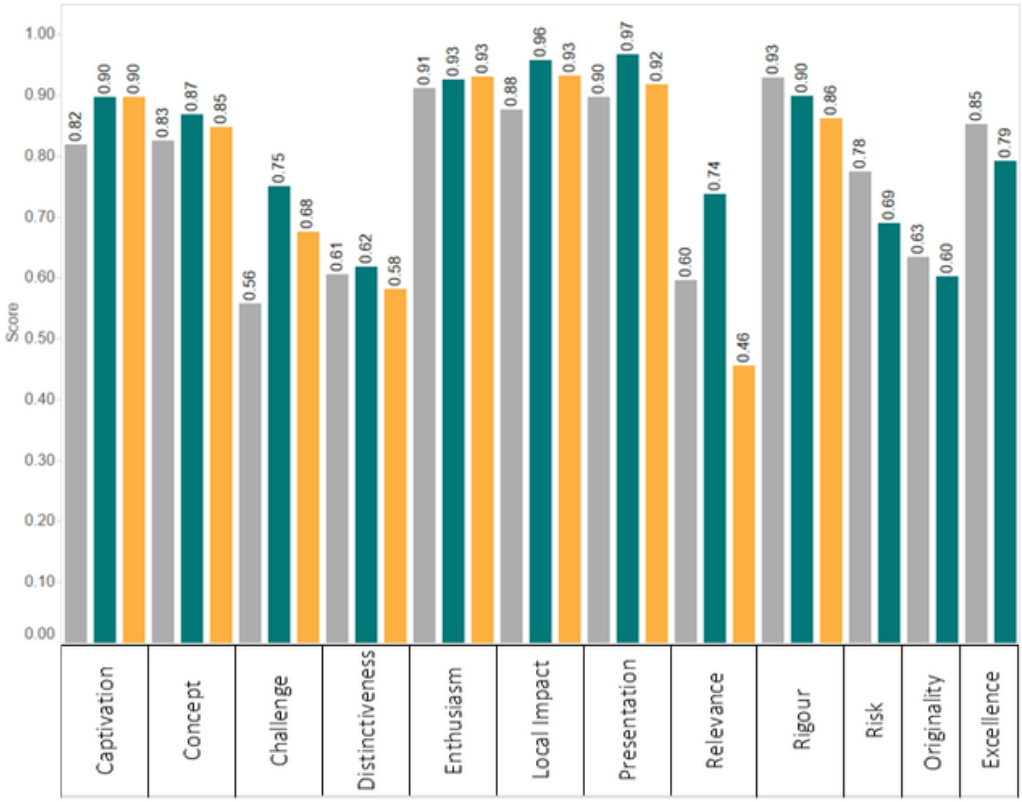


Figure A34: Medium – Screen

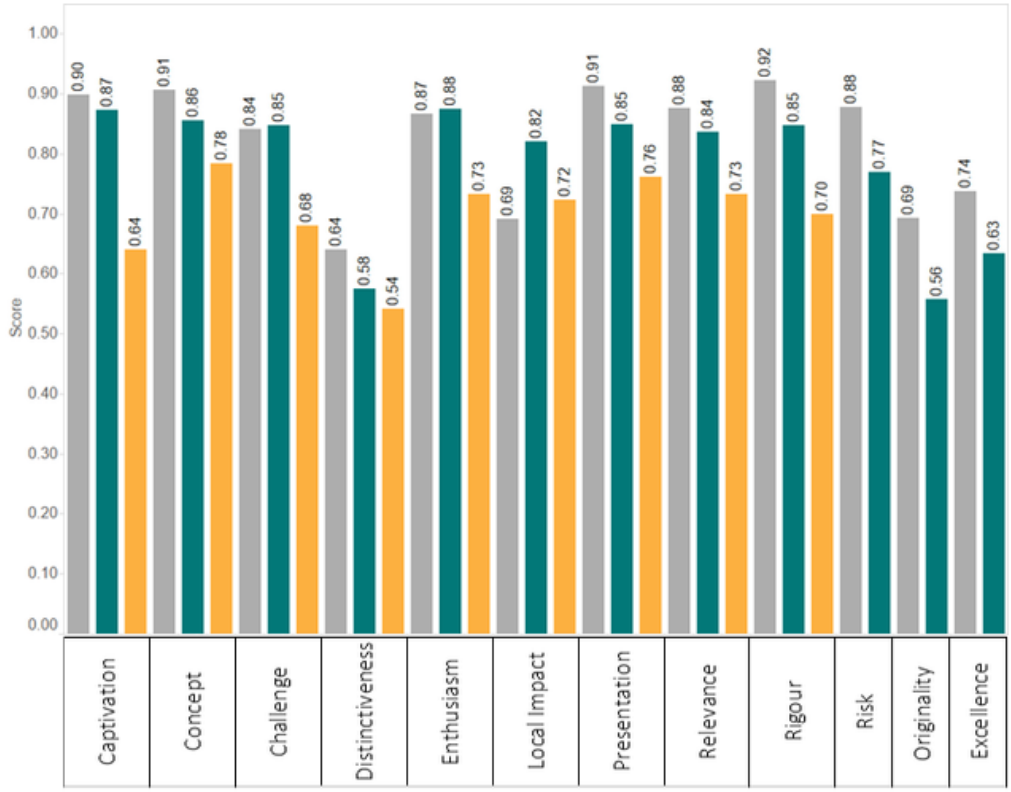


Figure A35: Medium – Textile

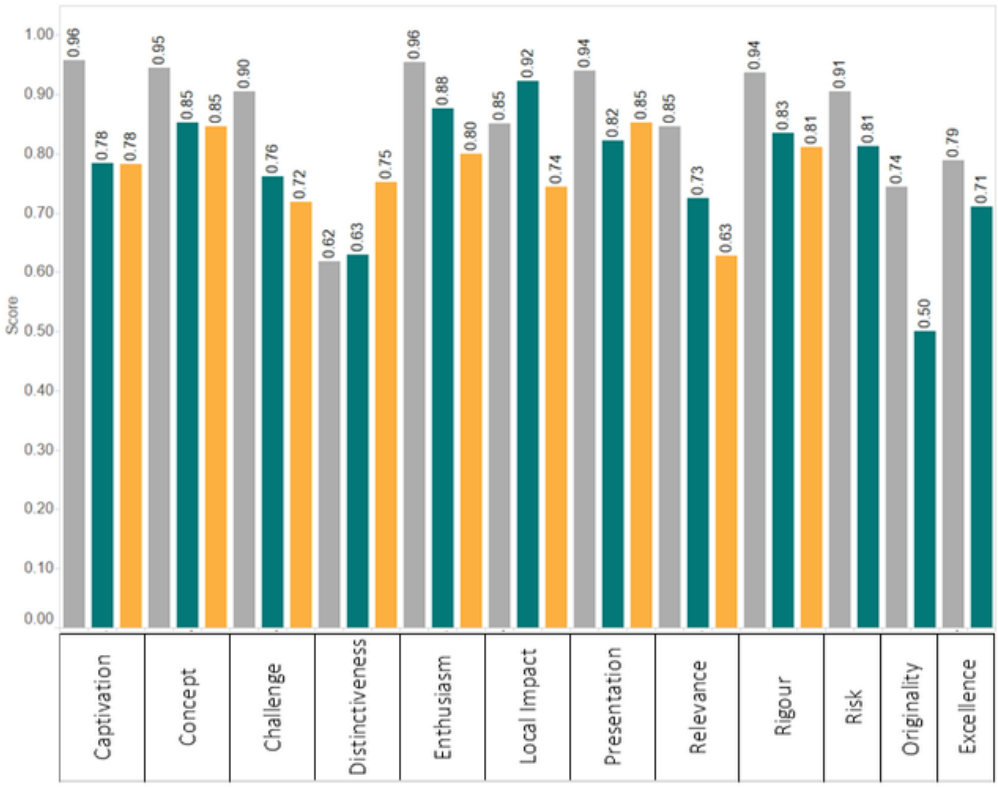


Figure A36: Presentation – Activity

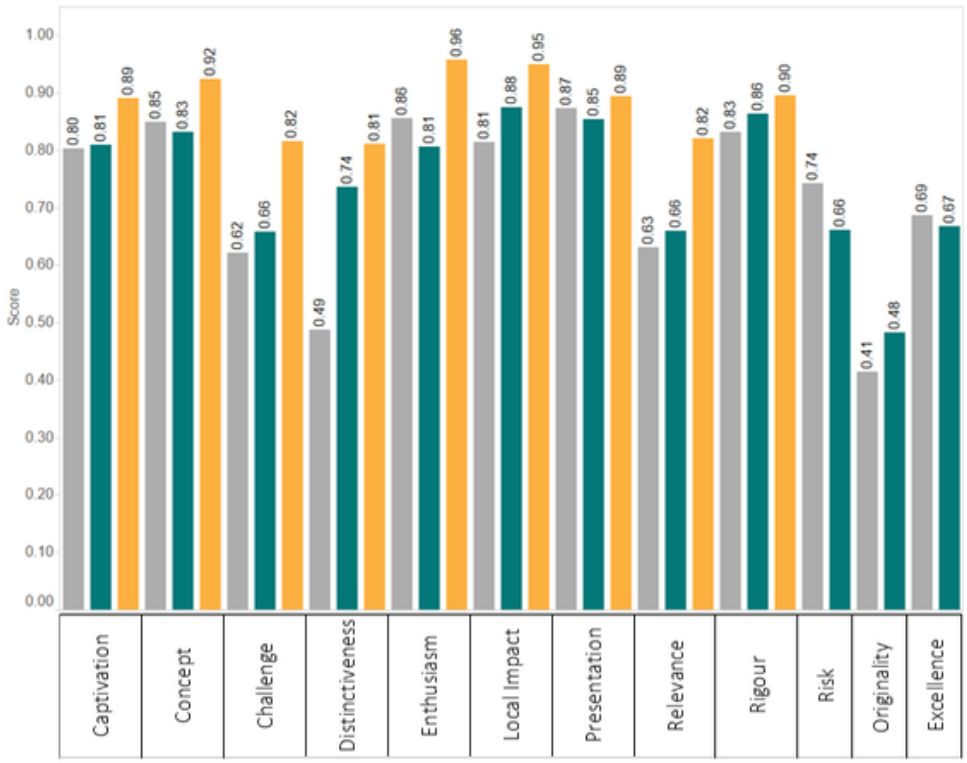


Figure A37: Presentation – Concert

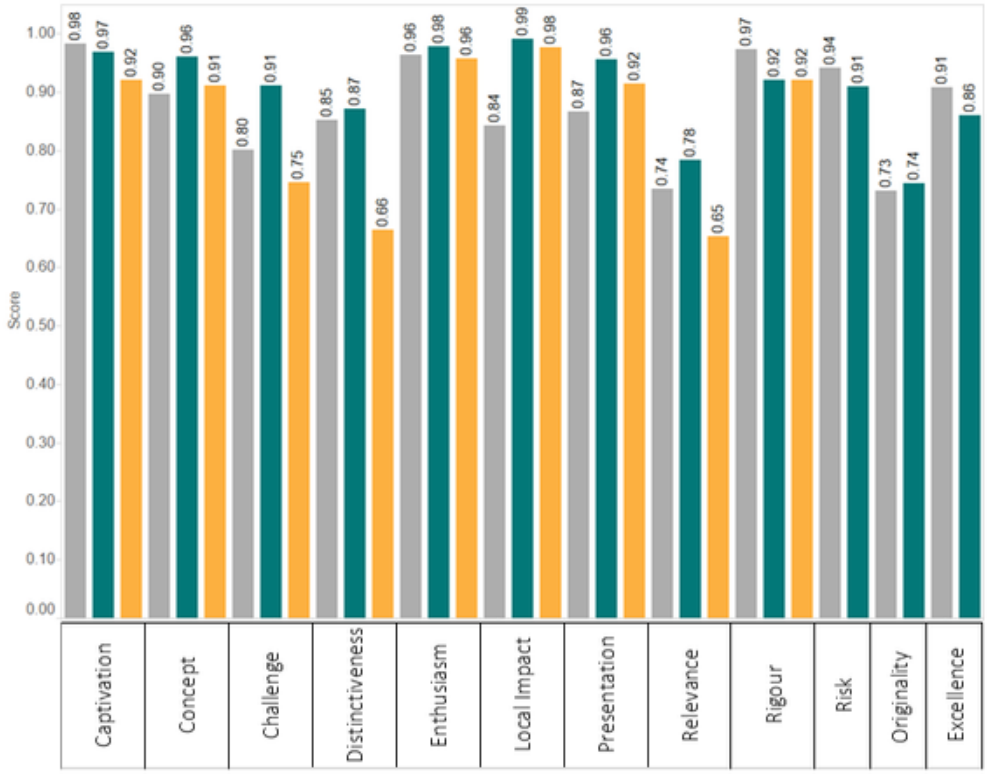


Figure A38: Presentation – Conversation

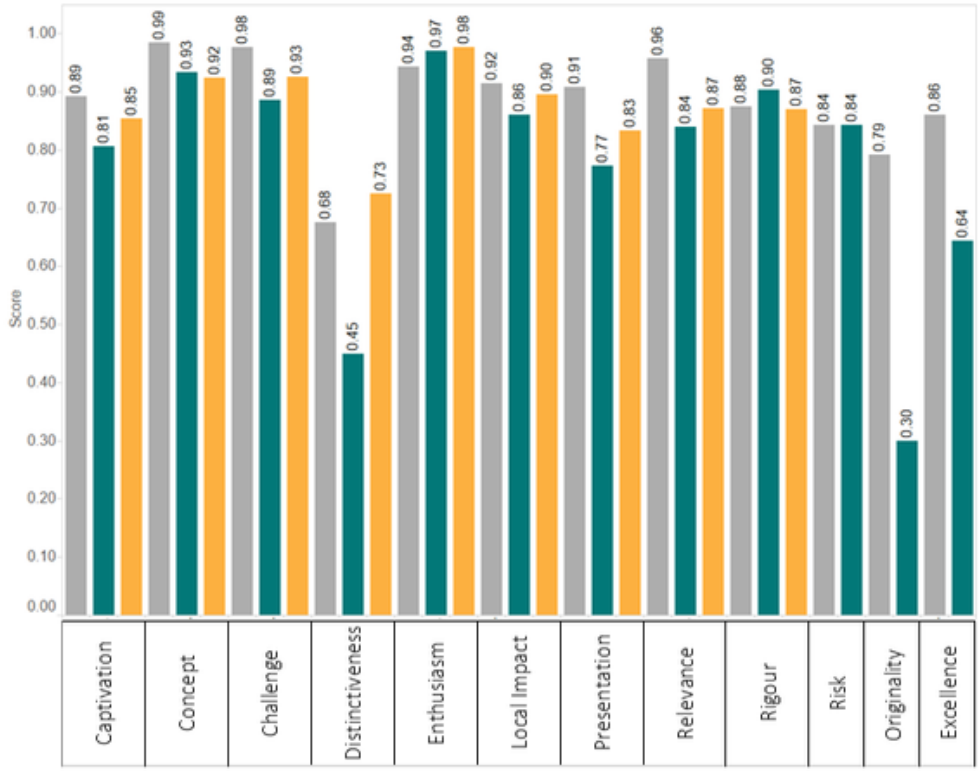


Figure A39: Presentation – Event

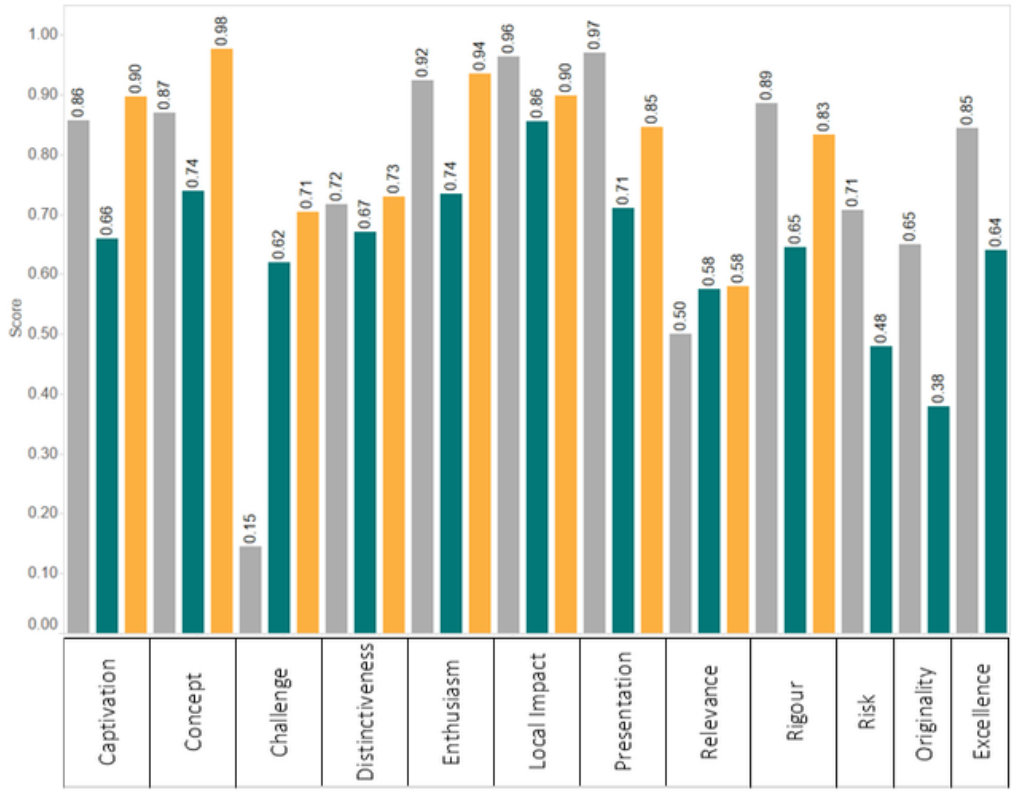


Figure A40: Presentation – Exhibition

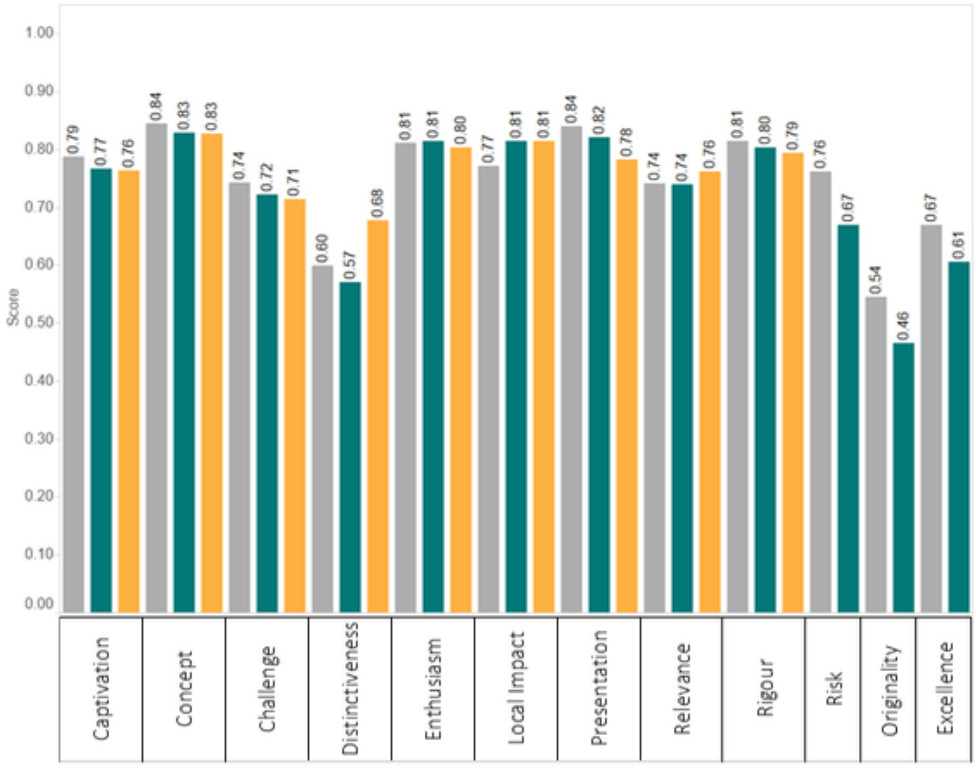


Figure A41: Presentation – Festival

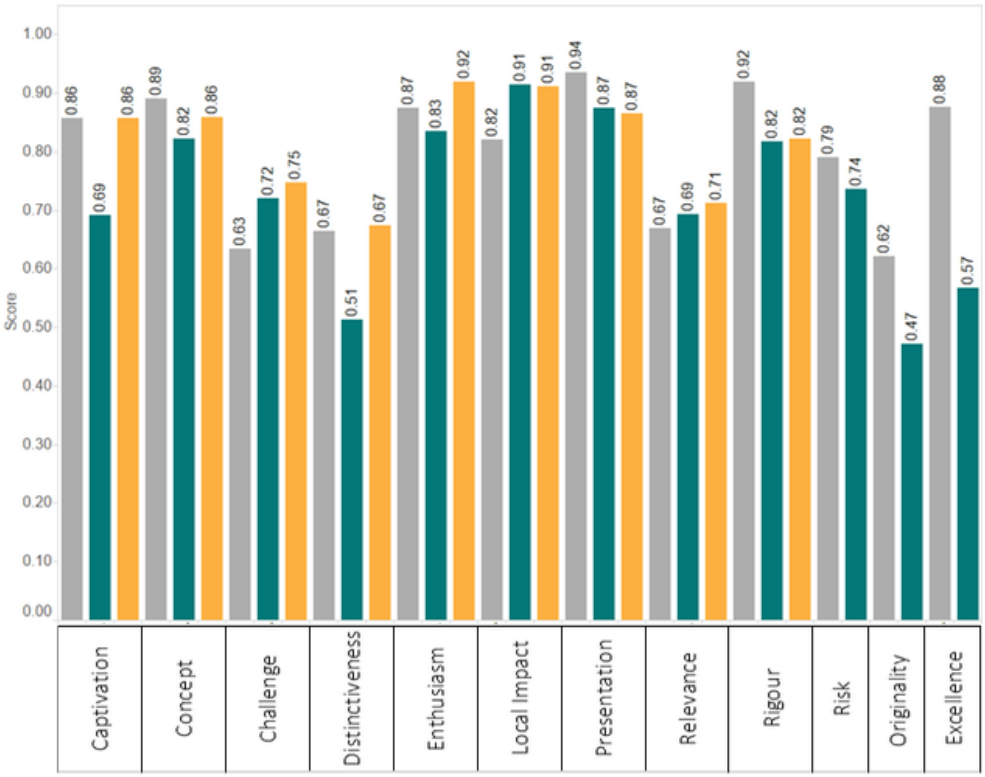


Figure A42: Presentation – Installation

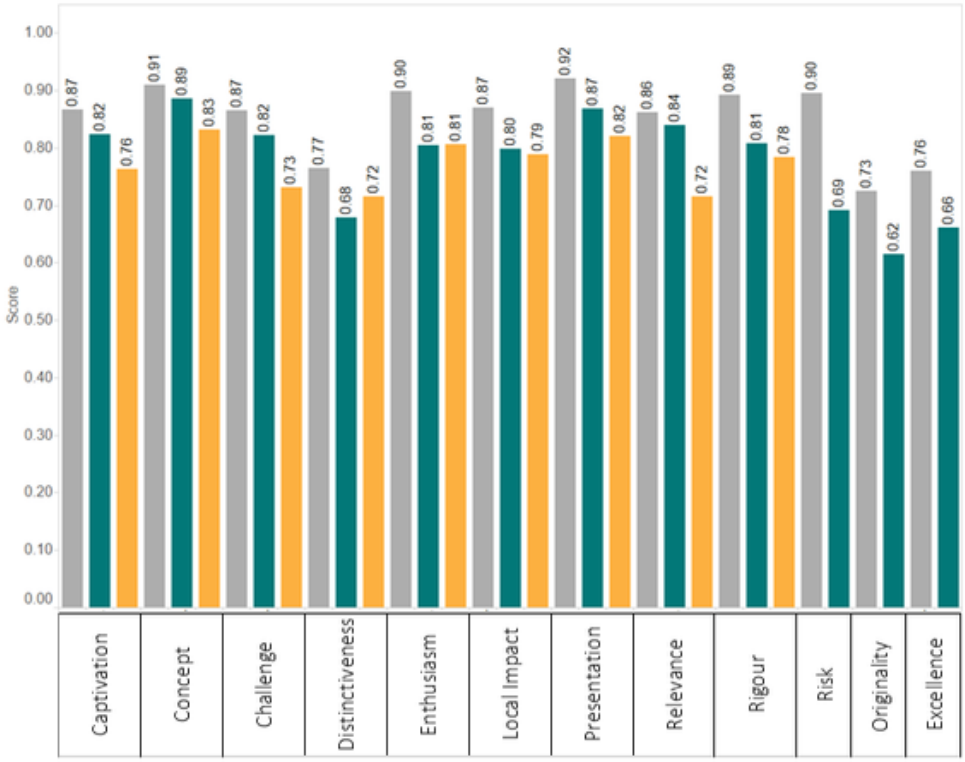


Figure A43: Presentation – Live

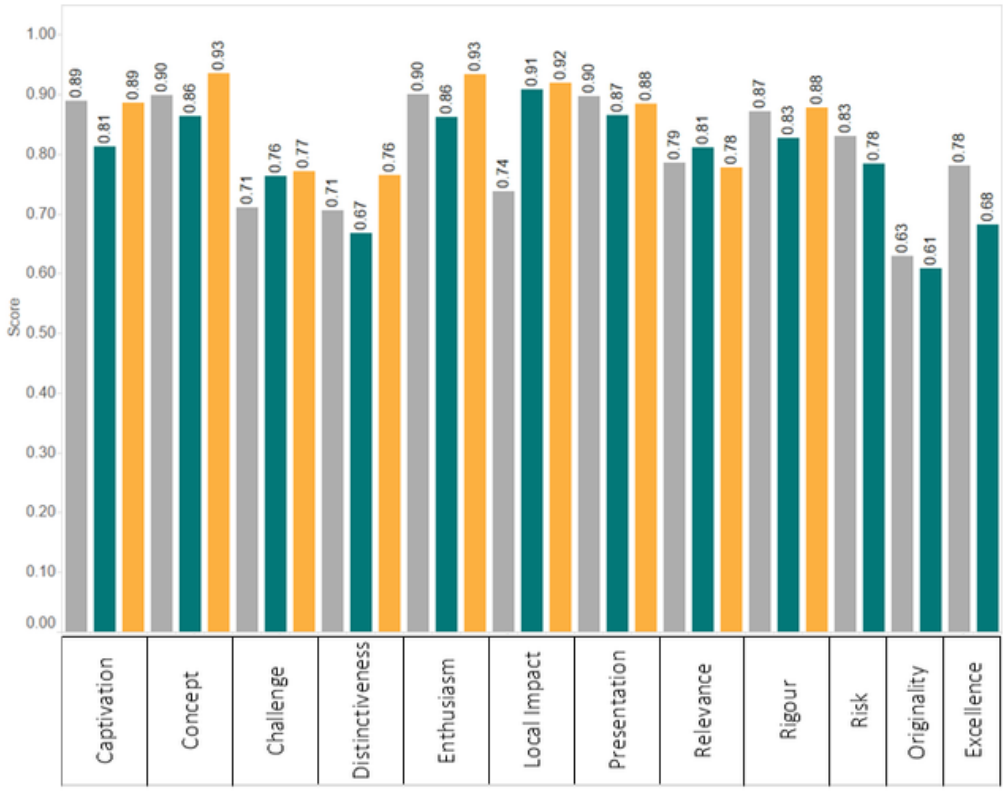


Figure A44: Presentation – Performance

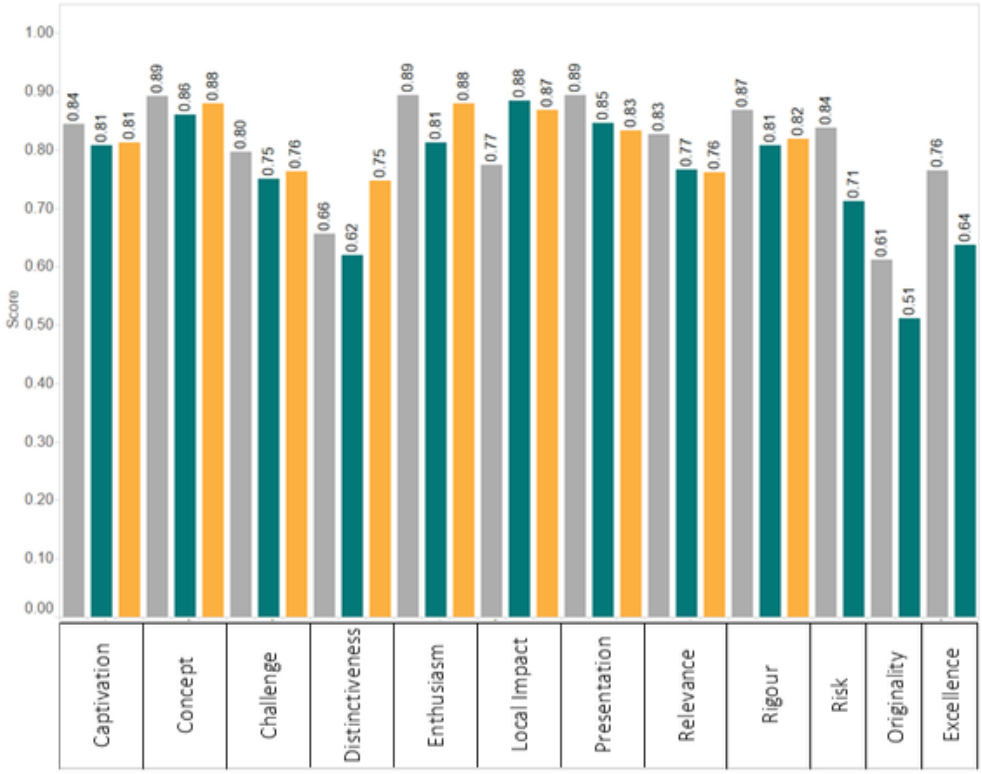


Figure A45: Presentation – Show

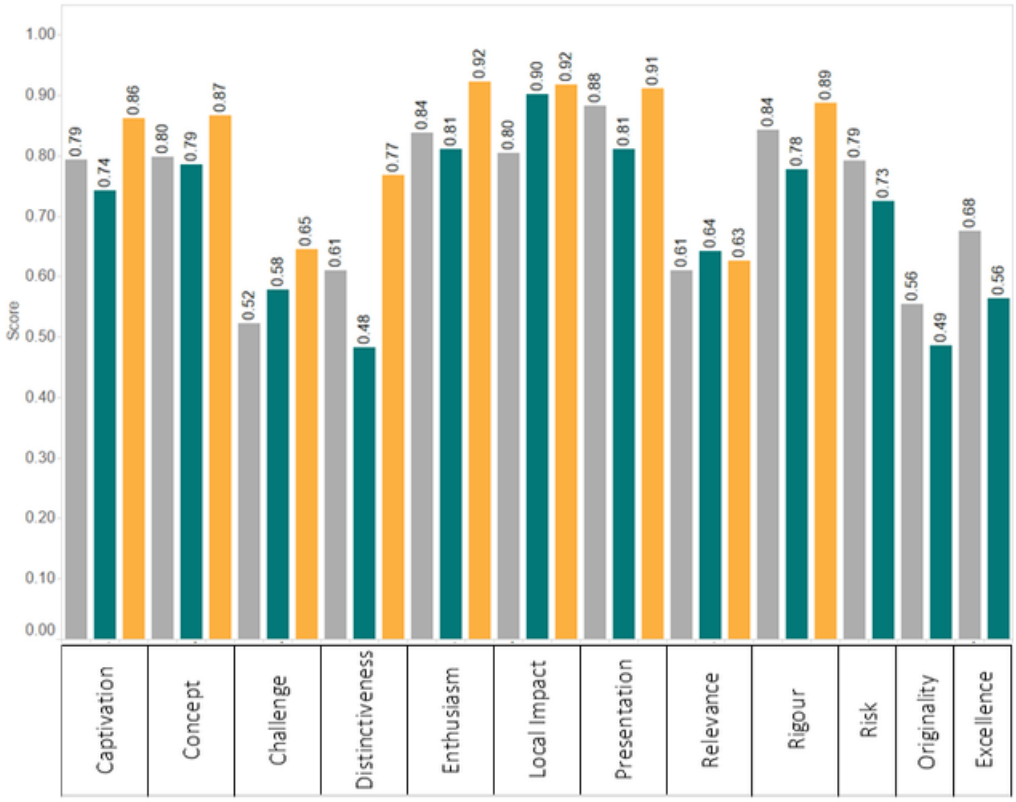


Figure A46: Presentation – Talk

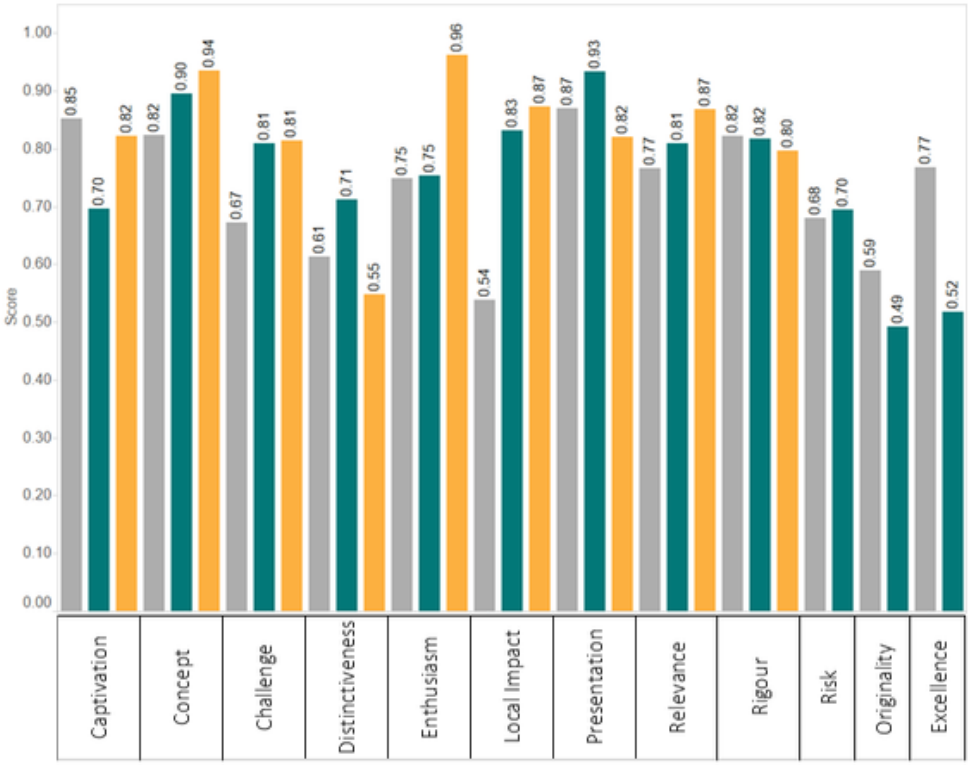


Figure A47: Presentation – Workshop

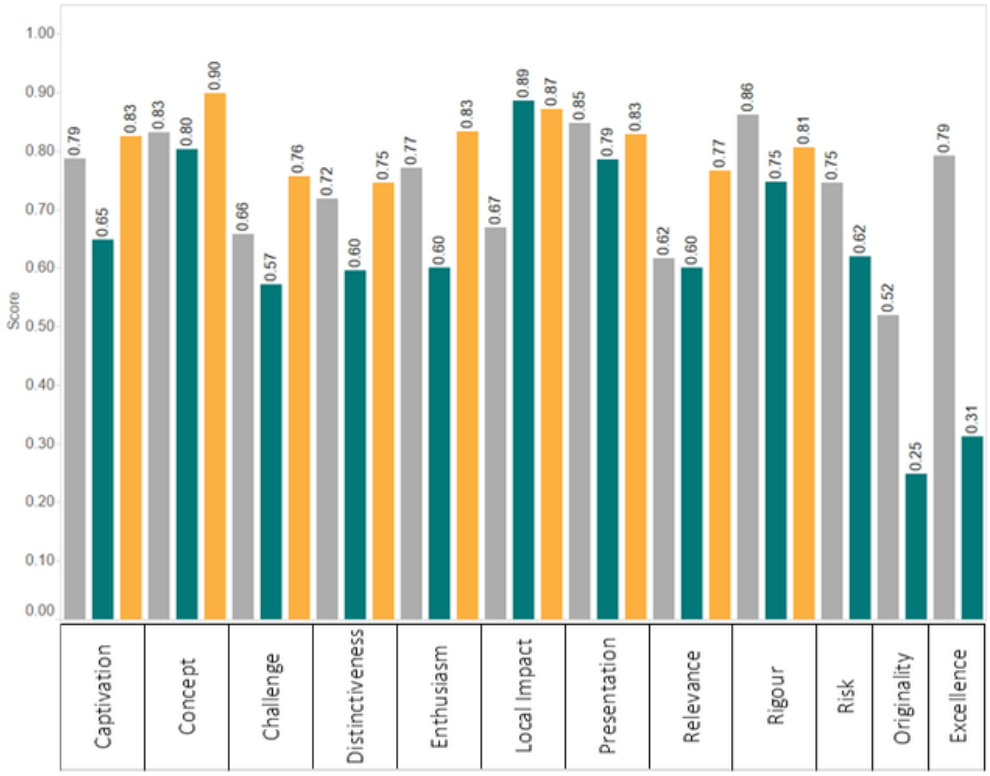


Figure A48: Subculture – Jazz

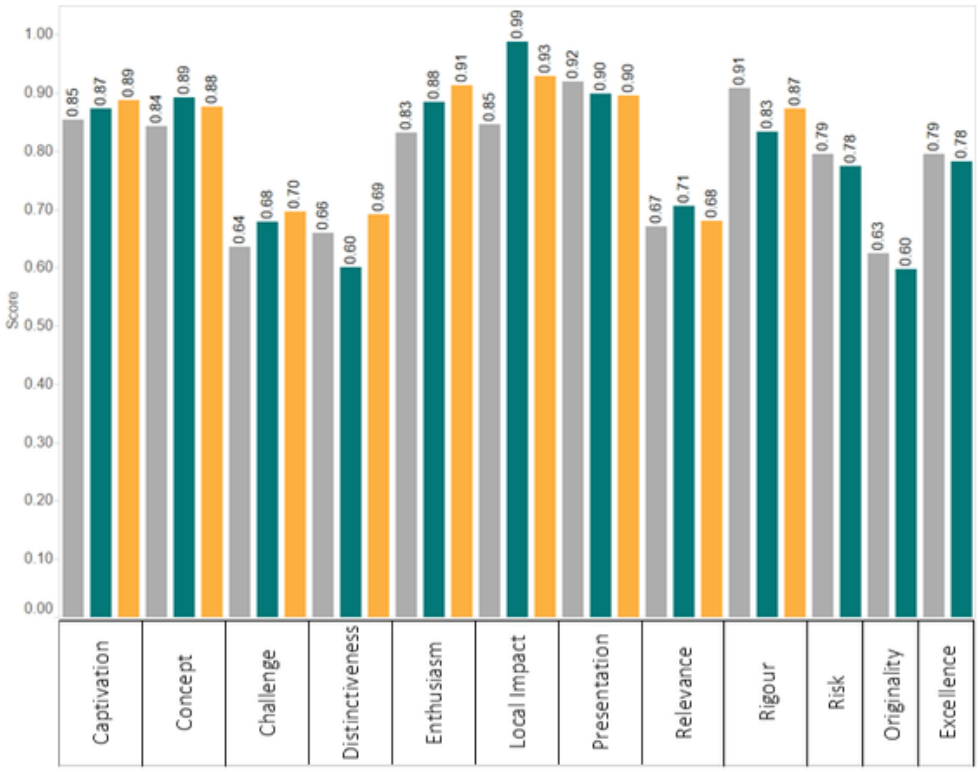


Figure A49: Subculture – Pop

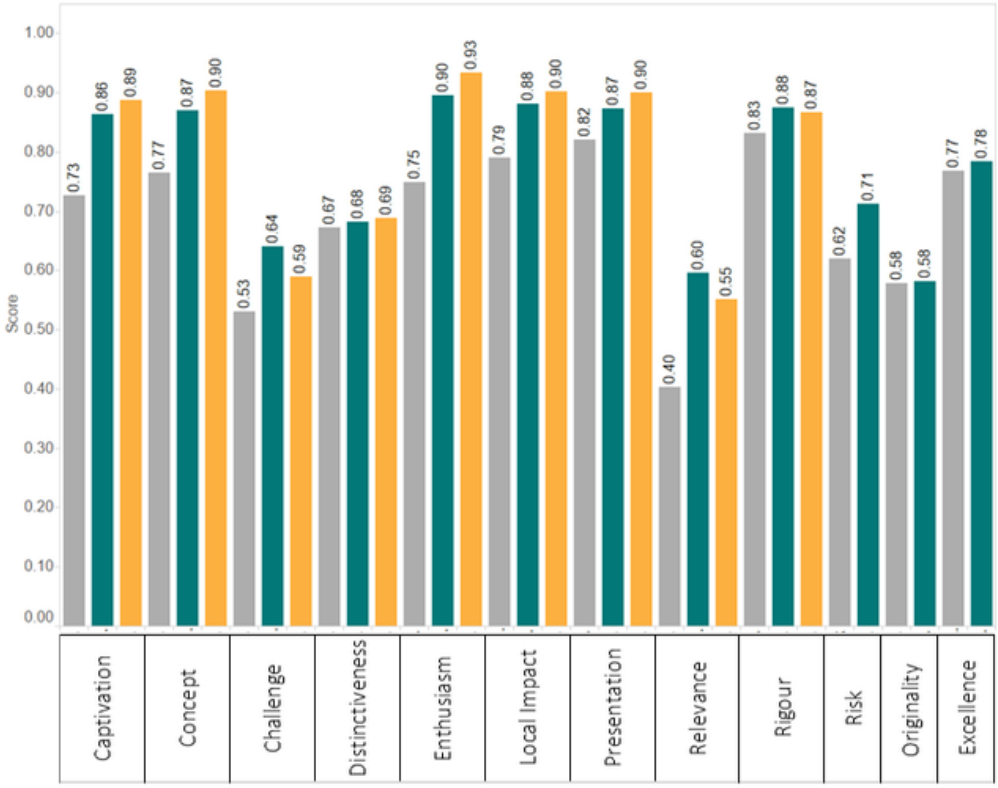


Figure A50: Subject – History