Thesis title:

‘How sports events shape host cities’ - Development of a comparative framework for assessing the impact of sporting events on the host location

For admission to the degree of

Doctor of Philosophy

Research Institute

Institute for Health and Sport, Victoria University Melbourne, Australia

Submission by

Michael Linley, December 2021

Doctoral Candidate: s4358981

Abstract

As the costs associated with bidding and hosting mega-events continues to escalate, the need to establish the benefits of these undertakings rises in step with governments coming under increasing scrutiny over the investment in sports events. Despite the billions spent each year in bidding for and hosting major sports events; the assessment of claimed benefits lacks a recognised or comparable method for assessing the hosting of major sporting events, with each federation, organiser or agency applying its own methodology.

Despite an increasing body of research on event outcomes and potential classification frameworks, the mega-event dominant research focus neglects other events and their potential impacts; hampering current and prospective hosts from critically reassessing their event portfolio, bidding more selectively, and grounding events within their longer-term development plans and financial resources.

This research challenges how terminology for the associated and myriad of outcomes the hosting of events has morphed to one of 'legacy' without any agreed definition of either the term itself or its constituent components. It seeks to explore whether a conceptual framework for the comparative assessment of event impact across events of different scales and types, including recurrent editions of events for the same host, could be established.

Adopting a focus on event impacts rather than legacies represents not a semantic choice but a determination to create a more constrained framework that allows smaller, higher-frequency events to be considered alongside quadrennial large-scale events on a comparable and consistent basis.

The development of the conceptual framework was grounded in an initial literature review of event impact and legacy assessment from which six core areas (Pillars) of event impact were identified; Economic, Social, Sport, Media, Brand, and Environment with 30 sub-areas (Drivers). From 350+ potential measures identified, 200+ metrics were included in the study with panellists rating measures on both their importance and reliability in assessing event impact. Using a Delphi method, the study sought to establish if consensus views on the relative importance of each of the pillars, drivers, and measures could be established across three survey rounds.

Despite the rhetoric that event legacy is unable to be defined, the findings show that there exists a solid underlying consensus on event impacts and how different areas contribute to the overall impact. Three tiers of impact emerged with Economic and Sport outcomes forming the top tier, with Media, Social and Brand outcomes grouped in a second tier, and the final pillar of Environment forming the lowest tier.

Evidence of significant bias within groups was found that reflects a lack of cross-discipline awareness and the need for greater collaboration between the event sector and academia. Uncovering bias enabled their influence on the consensus scores to be explicitly addressed and placed the importance of each dimension in the context of overall event impact.

The framework established in this study was shown to provide a strong basis for a consensus on the attribution of event impact. Using a multi-level structure allows for the core areas of impact to be consistently assessed in a standardised framework allowing for greater comparability across events, and for hosts to be more informed in building an event portfolio.

Further work on the technical development of the framework and its application by host cities is discussed.

Student Declaration

"I, Michael Linley, declare that the PhD thesis entitled "'How sports events shape cities' - Development of a comparative framework for assessing the impact of sporting events on the host location" is no more than 80,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references and footnotes. This thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work".

"I have conducted my research in alignment with the Australian Code for the Responsible Conduct of Research and Victoria University's Higher Degree by Research Policy and Procedures."

"All research procedures reported in the thesis were approved by the Victoria University Human Research Ethics Committee (VUHREC) under project ID: HRE21-126."

Signature

Date

9th December 2021

TABLE OF CONTENTS

1 Introduction 9
1.1 Background 9
1.2 Nature of the Problem 10
1.3 Significance of the study 11
1.4 Purpose and research questions 12
1.5 Theoretical Framework 13
1.6 Assumptions and Limitations 13
1.7 Definitions and Abbreviations 14
1.8 Summary 15
1.9 Thesis structure 15

2 Literature Review 16
2.1 The Legacy Demand 16
2.2 Unmasking the legacy assumption 18
2.3 Legacy and event scale bias 19
2.4 Re-grounding the concept of legacy 20
2.5 Legacy as an 'Essentially Contested Concept' 23
2.6 Reweighting or reassessing the value of events? 30
2.7 Agency relationship and influence on legacy incentives 32
2.8 Events, stakeholders and legacy 34
2.9 Model for all stakeholders - Zero-base legacy 42
2.10 Classification of existing legacy approaches 43
2.11 Framework structure for event assessment 47

3 Development of the Conceptual Framework 49
3.1 Development of the Event Impact Framework (EIF) 49
3.2 Conceptual framework dimensions 49
3.3 The emergence of key dimensions of a conceptual framework 54
3.4 Legacy limitations 58

4 Foundations of Framework Content 62
4.1 Development of core content 62
4.2 Proposed framework model 63
4.3 Review of Pillar and Driver foundations as related to event impact 65
4.4 Pillar - Economic impact 65
4.5 Pillar - Sport 78
4.6 Pillar - Media 85
4.7 Pillar - Social 90
4.8 Pillar - Brand 97
4.9 Pillar - Environment 105
4.10 Political 109
4.11 Assessing content importance 110

5 Research methodology 111
5.1 Research project structure 111
5.2 Research method 113
5.3 Research design 116
5.4 Survey instrument 120
5.5 Delphi content by Rounds 121
5.6 Research fieldwork 128

6 Results 130
6.1 Approach to data analysis 130

6.2 Delphi Panel Respondent Profile 131
6.3 The Weighting of Pillars – Round 1 135
6.4 Pillar weightings – Round 1 and Round 2 139
6.5 Driver Weightings – Round 2 and Round 3 144
6.6 Summary table of scores for all Pillars and Drivers 157
6.7 Indicator ratings 158
6.8 Framework internal reliability 173
6.9 Stakeholder priority outcomes 174
6.10 Pillars weights by Event Typology 174
7 Discussion of findings 176
7.1 Study base and panel profile 176
7.2 Weightings of event impact 177
7.3 Sources of bias within pillar weightings 181
7.4 Drivers of event impact 184
7.5 Event outcome measurements 189
7.6 Event scale and Pillar weightings 192
7.7 Stakeholders and assessing event impact 193
7.8 Key characteristics of the framework and its benefits and application 196
7.9 Limitations and further developments 199
7.10 Planning for future uncertainty 200
8 Conclusion 202
9 Reference List 205
Appendices 228
Appendix 1 - Pillar definitions provided to respondents 228
Appendix 2 - Delphi study surveys : Rounds 1 – 3. 229

Reference list of Tables

Table 1 - Gallie's ECC conditions and legacy 24
Table 2 - Gallie's ECC conditions: Legacy vs. Impact 28
Table 3 - Classification of stakeholders 38
Table 4 - Event stakeholder typology 40
Table 5 - Conceptual framework dimensions 49
Table 6 - Sheffield Hallam SIRC Event Typology 50
Table 7 - Comparison of research focus and legacy dimensions 54
Table 8 - Resolving legacy typology limitations 60
Table 9 - Emerging themes of impact areas 63
Table 10 - Economic Drivers and descriptions 66
Table 11 - List of potential economic impact measures 76
Table 12 - Sport Drivers and descriptions 78
Table 13 - Three Types of Fair Play 81
Table 14 - List of potential sport impact measures 83
Table 15 - Media Drivers and descriptions 85
Table 16 - List of potential media impact measures 90
Table 17 - Social Drivers and descriptions 91
Table 18 – Government commitment to volunteering leading up to London 2012 94
Table 19 - List of potential social impact measures 96
Table 20 - Brand Drivers and descriptions 97
Table 21 - List of potential brand impact measures 103
Table 22 - Environment Drivers and descriptions 105
Table 23 - List of potential environment impact measures 108
Table 24 – Levels of anonymity as applied to Delphi studies 114
Table 25 - Number of potential Indictors per Pillar 117
Table 26 - Dimensions of research by Pillar 120
Table 27 - Overview of Round 1-3 content 121
Table 28 – Best/Worst Scaling (BWS) element combinations 125
Table 29 - Internal reliability tests 127
Table 30 - Respondent communications by round 128
Table 31 - Survey response rates by round 128
Table 32 - Expert panel profile by Age 131
Table 33 - Role description 131
Table 34 - Tenure in field of expertise 132
Table 35 - Tenure by Age Group 132
Table 36 - Location base for role 133
Table 37 - Breadth of responsibility in role 133
Table 38 - Nominated areas of expertise 134
Table 39 - Cross-incidence of expertise selections 134
Table 40 - Initial weightings of Pillars from Round 1 135
Table 41 - Differences in mean pillar weightings by workstream 136
Table 42 – Differences in mean pillar weightings by workstream, Role = Academic 137
Table 43 - Differences in mean pillar weightings by workstream, Role = [Not] Academic 137
Table 44 – Differences in Academic vs. [Not] Academic Mean Pillar Weightings 138
Table 45 - Differences in Tenure (grouped) Mean Pillar Weightings 139
Table 46 – Pillars– Round 1, Initial weightings 140
Table 47 - Round 1 Pillar Consensus Range presented in Round 2 140
Table 48 – Pillars – Round 2, Revised weightings 140
Table 49 - Pillars consensus significance (All respondents) 141

Table 50 - Pillars consensus range (excluding outliers)..... 142
Table 51 - Pillars consensus significance (excluding outliers)..... 143
Table 52 – Balanced Overall Pillar Weights..... 144
Table 53 - Economic Drivers – Round 2, Initial weightings..... 144
Table 54 - Round 2 Economic Drivers, Consensus Range presented in Round 3..... 145
Table 55 – Economic Drivers – Round 3, Revised weightings..... 145
Table 56 - Driver Consensus testing - Economic..... 146
Table 57 - Sport Drivers – Round 2, Initial weightings..... 147
Table 58 - Round 2 Sport Drivers, Consensus Range presented in Round 3..... 147
Table 59 – Sport Drivers – Round 3, Revised weightings..... 148
Table 60 - Driver Consensus testing - Sport..... 148
Table 61 - Media Drivers – Round 2, Initial weightings..... 149
Table 62 - Round 2 Media Drivers, Consensus Range presented in Round 3..... 149
Table 63 – Media Drivers – Round 3, Revised weightings..... 150
Table 64 - Driver Consensus testing - Media..... 150
Table 65 - Social Drivers – Round 2, Initial weightings..... 151
Table 66 - Round 2 Social Drivers, Consensus Range presented in Round 3..... 152
Table 67 – Social Drivers – Round 3, Revised weightings..... 152
Table 68 - Driver Consensus testing - Social..... 153
Table 69 - Brand Drivers – Round 2, Initial weightings..... 153
Table 70 - Round 2 Brand Drivers, Consensus Range presented in Round 3..... 154
Table 71 – Brand Drivers – Round 3, Revised weightings..... 154
Table 72 - Driver Consensus testing - Brand..... 155
Table 73 - Environment Drivers – Round 2, Initial weightings..... 156
Table 74 - Round 2 Environment Drivers, Consensus Range presented in Round 3..... 156
Table 75 – Environment Drivers – Round 3, Revised weightings..... 156
Table 76 - Driver Consensus testing - Environment..... 157
Table 77 - Summary table of scores for all Pillars and Drivers..... 158
Table 78 - Measure ratings: Economic..... 159
Table 79 – Rating disparities: Economic..... 160
Table 80 - Measure ratings: Media..... 161
Table 81 - Rating disparities: Media..... 163
Table 82 - Measure ratings: Sport..... 163
Table 83 - Rating disparities: Sport..... 165
Table 84 - Measure ratings: Social..... 165
Table 85 - Rating disparities: Social..... 167
Table 86 - Measure ratings: Brand..... 168
Table 87 - Rating disparities: Brand..... 169
Table 88 - Measure ratings: Environment..... 170
Table 89 - Rating disparities: Environment..... 172
Table 90 - Cross workstream BWS measure ratings..... 172
Table 91 - Tests of Framework Reliability..... 173
Table 92 - Event Host vs Event Owner priorities..... 174
Table 93 - Mean Pillar weights by event scale..... 175
Table 94 - Differences in Pillar mean weights from Type A events..... 175
Table 95 – Exemplar Delphi study typology..... 177
Table 96 - Summary of Delphi consensus outcomes..... 188
Table 97 - SIRC Typology for Pillar Weightings by scale..... 192
Table 98 - Assessing Gallie's conditions from the Delphi study..... 195

Table of Figures

Figure 1 - Qualitative Classes of Stakeholders 37
Figure 2 - Subsets of Qualitative Stakeholder classes 37
Figure 3 - Stakeholder Typology: One, Two or Three Attributes Present 38
Figure 4 - Preuss' Legacy Cube 39
Figure 5 - Classification of existing legacy approaches 44
Figure 6 – Framing measure dimensions 52
Figure 7 – Dimensions in structural approaches 55
Figure 8 – Framework structure = Single Level 55
Figure 9 - Framework structure = Nested (Measures) 56
Figure 10 - Framework structure = Nested (Levels) 56
Figure 11 - Framework structure = Tiered 57
Figure 12 - Target EIF structural choice 57
Figure 13 – EIF Tiered model terminology 58
Figure 14 - Process of primary model determination 63
Figure 15 - Balanced Scorecard in evaluating events 64
Figure 16 - Proportion of Private vs. Public funding - Olympic Games 1984-2012 70
Figure 17 - VNV Employment Growth Effect 71
Figure 18 - TOP: The Olympic Partners - Revenues vs. Partner Numbers 74
Figure 19 - EIF Framework reliability model 127
Figure 20 - EIF and Triple Bottom Line 180

1 Introduction

The chapter introduces the challenges that cities face as prospective hosts of sporting events and the need for an approach that enables events of different scales to be considered in their development plans.

1.1 Background

The Summer Olympic Games dominate the global event calendar in terms of international audience reach. However, the roster of quadrennial 'mega-events' (including FIFA World Cup, the Winter Olympic Games, Rugby World Cup, Commonwealth Games, each sports' own international championships and the global F1 Grand Prix circuit) all compete for audience attention, sponsorship, and host bids. Globally, the sporting industry was USD $388 billion in 2020 down from USD $458 billion in 2019, an outcome of global pandemic restrictions, but is expected to regain momentum be worth nearly USD $600 billion by 2025 (Business Research, 2021).

Events are seen as a key pillar in development strategies with "more competition than ever from emerging cities and countries that see major sporting and entertainment events as a fast track to global recognition and influence" (Pellegrino & Hancock, 2010, p. 4). The pinnacle of these 'showcase' events are the mega-events of the Olympic Games and FIFA World Cup. Those events, while able to deliver on their 'global recognition', come at a price.

As the costs of events have continued to rise in the period since the debt-laden Montreal Olympics, prospective and past hosts have sought to both expand the claimed gains from direct, and induced economic benefits, and balance the cost equation through expanded soft 'benefits' (e.g. social, health, cultural, community pride) – all together now commonly described under the catch-all term 'legacy'. To date, the measurement of sporting event outcomes has predominantly focussed on measuring their economic value, with ascriptions of extended legacy often selectively applied to justify any gap between economic benefits calculated and taxpayer funds invested.

In response to fears that the 1976 Olympic Games might incur a debt for the city as the 1967 World's Fair had, then Montreal Mayor Jean Drapeau responded claiming "The Montreal Olympics can no more have a deficit than a man can have a baby" (Borders, 1974). Cruelly, the result was a cost blowout in the hosting and a resultant CAD $1.5bn debt its residents only finally paid off in 2006, 30 years after the Games. That impost of a significant public obligation for debt service from cost overruns means that "budgetary gaps must be filled either by cutting other government services or increasing taxes; either will slow down the local economy" (Siegfried & Zimbalist, 2006, p. 423). The likelihood of drawing on public funding is exacerbated when "less than reliable economic benefit calculations cast a dark shadow on the return on investment claims made by both sport event organizers and government backers" (Lee & Taylor, 2005, p. 596). The shadow of those claims is revealed from an analysis that identifies mega-events not just as risky projects with poor returns, but with the worst history of cost overruns for projects of this scale. Indeed the exposure placed on citizens from the use of public funds is not the exception, as "cost overrun is found in all Games, without exception" (Flyvbjerg et al., 2016, p. 2).

Despite the history of cost overruns by successive hosts of mega-events, most have claimed theirs to be the exception, referring to 'wider legacies' from events to counter increasing gaps between projected costs and budgets. Indeed, "public policy planners and event organizers are increasingly promoting potential economic, tourism, urban, social, and/or environmental legacies to justify significant public investments required to host special

events" (Thomson et al., 2013, p. 111). So dependent have hosts and event owners become on legacy in justifying their case for special events that "for some events, particularly large scale public events, the issue of legacy has become central to the decision to host or create them"(Allen et al., 2008, p. 115). Unsurprisingly, defining the legacy has become an embedded requirement for mega-event bid plans submitted by prospective hosts.

While not universal, a widely used definition states that legacies are "all planned and unplanned, positive and negative, tangible and intangible structures created for and by a sport event that remains longer than the event itself" (Preuss, 2007, p. 211). Leaving aside the catch-all nature of the definition to include all outcomes as legacies, the embedded requirement to evidence a lasting structural change is ill-suited to most events. This is especially true for hosts that are seeking to be economically prudent and target hosting events within their existing infrastructure assets and/or avoid those requiring special conditions. Events with those outcomes do not leave a lasting 'legacy' of structures, but despite the pejorative labelling of the term, they do have an impact.

The production of event legacy has also become a central focus by event organisers due to the increased scrutiny and accountability on the social, environmental, and economic impacts left following the hosting of an event. Leopkey and Parent (2012) argued event legacy has become institutionalised within the Olympic Movement as a way to help justify the spending of exorbitant sums of money on hosting. This ascendency of legacy is despite the absence of "longitudinal studies that are able to trace the long-term benefits of mega-events" (Allen et al., 2010, p. 115). An examination of the extended timeframe over which legacies outcomes are claimed hampers the identification of definitive outcomes, as "over time, the influence of other mitigating factors obscures them, making such benefits difficult to trace" (p. 115).

Another notable consequence of the legacy focus is that the requirement under such a focus to only include items of long-term consequence also excludes significant annual events, where each edition's impact horizon is only 12 months by nature. Intentionally or not, the exclusion of these events is to exclude iconic sporting occasions that are inextricably linked with their host, including Wimbledon, the Monaco Grand Prix, the Melbourne Cup, and the US Masters. As noted above, the legacy focus also excludes smaller events, which are likely to be challenged to provide or evidence outcomes over a lengthy timeframe.

1.2 Nature of the Problem

Due to the predominant focus on legacy and mega-events, there is currently no consistent and comprehensive framework for measuring the outcomes of events of differing scales and types. Despite recognition of sporting events as complex and multifaceted enterprises, the challenge in understanding the results of hosting these events is that "most legacy studies focus on one or two legacy outcomes" (Karadakis & Kaplanidou, 2012, p. 246) rather than seeking to research the breadth of outcomes claimed. More specifically, "much of the growing body of literature on legacy focuses on either the economic effects or the infrastructural changes" (Preuss, 2015, p. 3). This has two immediate consequences. Firstly, researchers are unable to offer a perspective of their findings in context to other areas of event outcomes, and secondly, hosts are not provided with an integrated and weighted view from the research that they can use to determine the outcomes of a specific event. A review of research priorities notes that "academics had largely focused on economic and tourism impacts, while little attention has been given to other dimensions, such as environmental, socio-cultural, and psychological impacts" (Kassens-Noor et al., 2015, p. 667).

Even within the most established and researched component of legacy frameworks - the economic value of events - the assessment of economic impact from direct visitor spending shows concerns over methods of assessment and "the methods used to research the economic impact of major events remain contested" (Davies et al., 2013, p. 31). This lack of agreement on the most substantively researched area of legacy exemplifies the similar lack of consensus over the measurement and contribution of the other impacts or legacies of events observed by Dickson et al. (2011). This lack of consensus has complicated the measurement of events' legacies (Horne & Manzenreiter, 2006).

While there exists general agreement on the categories of outcomes to be included in event assessments (even if much of the research is constrained in scope), the issue of assessing events consistently is further compounded by a scarcity of specific metrics used to underpin those categories. The lack of measures applies not only to events overall but even to the areas that are being studied, where "only a few researchers have suggested actual indicators to measure legacies, and those who did, were restricted to a specific subcategory of their selected field of study." (Bob & Kassens-Noor, 2012, p. 13).

This study seeks to address these deficiencies by identifying the key dimensions of event outcomes that might apply to events of differing scales and types. Through drawing on the expertise of academic and industry practitioners, the research will seek to explore a consensus on both the dimensions of event outcomes relevant to a wider range of events and their relative weightings and identify specific measures that underpin those dimensions. In seeking a more inclusive approach than existing legacy models, such a framework ought to be able to be consistently applied across events of different types rather than tailored to each edition, enabling comparability of events.

This research will seek to provide a new perspective for current and prospective event hosts, cities seeking to bid for events, and key stakeholders in the events industry including rights holders, the media, and partners. By establishing a consistent and comparative basis for the wide range of events hosted annually or below the mega-event scale, the framework would allow for a more objective and structured assessment of potential event outcomes and the expected return from the investment of public or private funds in bidding for and hosting events. Both event hosts and event rights holders may be able to uncover where properties are presently undervalued by the absence of consistent and specific measurement.

1.3 Significance of the study

With the rising complexity of hosting major events, even the cost of the bidding process for the Olympics rose to peak at USD $40-60m (ASOIF, 2014), promoting changes in IOC Candidate City process to reduce costs (IOC, 2021b). Despite that, the costs to bid for the FIFA World Cup in 2021 still bid commitments excess of USD $30m (Stensholt & Korporaal, 2021). Hence in considering bidding, city and national governments need to give increasing attention to the question of what value is realised from hosting such large-scale events and if hosting smaller events might offer a better return for their taxpayers and their strategic goals.

Being able to choose both the right event to host and to be clear on the outcomes sought from that event means cities need to be able to compare different events to find those that fit with and support their development plans. The reality is that very few cities will ever host a mega-event, with many cities now seeking to build a portfolio of events of different types and scales. While smaller events may pale in comparison to other large government investments, the impact and consequences of multiple poor event decisions may endure in

terms of opportunity cost, loss of resident engagement, or fading support for city development plans.

If event owners and governments require 'legacies' from the hosting of events (for proof of their value and prudent expenditure of taxpayer funds respectively), then the application of a comprehensive and consistent assessment framework grounded in specific measures would appear to be a minimum requirement.

New knowledge in the research domain

To date, academic models have proposed and assessed areas of impact or legacy, and some have proposed detailed measures that could be used to evaluate event legacy. However, their overarching limitation has been a lack of applicability across different types and scales of events.

It would appear few have considered weighting different areas of event impact. Of those that have such as the quantitative approach by Mair and Whitford (2013b) who scaled different event research themes for importance, the analysis was designed to guide where future event research may be directed rather than to underpin a model of event measurement.

This research seeks to evaluate and extend the prior academic and practitioner models, identify the key areas of impact and relevant measures, and combine them into a comprehensive framework that can be applied across a wide range of sporting events. Such a framework aims to empirically determine the relative importance of different areas of event impact and identify the measures that would underpin those weightings.

Contribution and Significance

As discussed previously, the requirement to leave a 'legacy' is biased against smaller events. By nature, their scale means they do not impose a structural change on the host city, hence cannot be assessed on an equivalent basis to mega-events. Without a standardised approach, governments cannot consistently assess the value derived from events on an equivalent basis and cannot critically assess the value and benefits derived from taxpayer investment across their event portfolio. Unlike the stock or bond market, there is no visible 'market price' for events that governments can use to assess the likely value to be derived from hosting any particular event compared to another. Therefore, the ability to compare and contrast the impact of events remains inconsistently assessed and subject to the limited information provided by various event owners/promoters. The establishment of a consistent framework would underpin a more efficient market (Fama, 1970) for the promotion of and bidding for events. Further, reducing information asymmetry (Stiglitz & Weiss, 1981) on the potential event impact would contribute significantly to ensuring public funds used for bidding and hosting of events might be more rationally and efficiently allocated.

1.4 Purpose and research questions

This research aims to deliver an empirical assessment of sporting event impact; establishing the foundational underpinning to assess the potential impact of a wide range of events.

The research seeks to redress the core limitations of event measurement that exist in the global sector, especially as governments seeking to bid for and host events become more accountable for the funds invested. As such, the goal is to integrate event impact assessment into a comprehensive framework, instead of one dominated by the single dimension of 'economic impact' due to a lack of models that consistently capture other dimensions of impact. Through the provision of the framework, it is desired to reduce the

potential for events to be underestimated due to not capturing the value from other areas, or overestimated by extending the boundaries of the 'economic impact' longitudinally to justify taxpayer funds invested. Lastly, the purpose is to enable sporting event outcomes to be compared across differing event types and relative to the scale of events.

Arising from that purpose is the following major research question and sub-questions.

Major research question

- *How might sporting events of different types be assessed for their impact on their host location within a standardized comparative framework?*

Within that key question, additional sub-questions that the research seeks to answer are:

- *What are the key areas of impact that should be addressed in an event impact framework, such that they are relevant across a wide range of sporting events?*
- *What weightings should be applied to the different areas of event impact?*
- *What measures might be included under each area of impact to maximise the relevance and reliability of the assessment of event outcomes?*
- *What might be the benefits to host cities of applying a standardised comparative framework?*

1.5 Theoretical Framework

To establish the key dimensions, weightings, and measures of a standardised framework, the research will draw on a global panel of academics and industry practitioners through a Delphi study process.

Lack of consistency in the definition of legacy creates challenges to establishing a standardised framework. The study will need to be able to establish if a degree of 'consensus' exists on the importance of legacy elements within a contested space, allow for clear definitions in terminology to establish a consistent understanding, and be able to distinguish between responses from different stakeholder perspectives, without compromising anonymity or biasing responses.

A 'Delphi study' approach was identified and selected as a recognised "group facilitation technique…designed to transform opinion into group consensus" (Hasson et al., 2000, p. 1008). Critically it is not just the transformation of opinion, but acceptance of the consensus that is powerful from this approach. The application of the Delphi approach has shown "the method produces useful results which are accepted and supported by the majority of the expert community" (Czinkota & Ronkainen, 2005, p. 122).

Hence the proposed approach was to use a Delphi method towards consensus forming on the key dimensions of event assessment within a prospective framework. It is important to note that the selection of a quantitative Delphi study in the form of a 'judgement function' as the research method may exclude some subjective variables or elements that are not compliant with that model. Any exclusion does not connote any lesser value to those as event outcomes, but rather reflects the constraints within the model selected for this research.

1.6 Assumptions and Limitations

The assumption in this study is that sporting events are discussed in the context of cities as hosts. Some large-scale events (such as the respective football, rugby and cricket world

cups) are hosted across multiple cities. However, the impact of the event can be considered in the context of the matches held in each host city. There are potentially additional outcomes at the national level, but these are not explicitly differentiated in terms of outcomes.

The limitations in the study are that the research has not considered non-sporting events, festivals, meetings, and conferences as events, although there will be strong overlaps with those events on several dimensions. Extending the application of the framework to address these is outside the scope of this research.

While noted in the methods, the provision of anonymity for respondents under the Delphi study conditions meant the respondent base of the panel was not controlled and could only be assessed post-hoc for the profile of those who had responded.

The research views sought from a global panel of experts were only available in English. Mitigating that choice are the considerations that for academics to be invited, they must have published their research and presented and discussed their findings in conferences in English. Likewise, the newsletter content provided by industry collaborator Sportcal is only published in English and subscribers actively engage on that basis. Hence the provision of English only was considered consistent with the established expertise of participants from those sources.

1.7 Definitions and Abbreviations

The following abbreviations are used in the study:

Organisations

AFL: Australian Football League

CGF: Commonwealth Games Federation

EPL: English Premier League

FIFA: International Federation of Association Football

IAAF: International Association of Athletics Federations

IOC: International Olympic Committee

IPL: Indian Premier League

ISF: International Sports Federation

LOCOG: London Organizing Committee of the Olympic and Paralympic Games

MLB: Major League Baseball

NFL: National Football League

OOC: Olympic Organising Committee

Other terms

CSR: Corporate Social Responsibility

EIF: Event Impact Framework

TBL: Triple Bottom Line

Within the Event Impact Framework, the following definitions were used:

Pillars

Pillars are the top-level major areas or disciplines of event impact and include Economic, Sport, Social, Environment, Brand, and Media.

Drivers

Drivers are sub-elements or second-order elements contained under each of the individual Pillars. These are not direct measures of impact, but grouped sub-areas within the Pillars that are considered critical dimensions of an event's impact.

Indicators or Measures

Indicators are the specific measures used in assessing event outcomes. The terms Indicators and Measures can be considered synonymous and are used interchangeably in the literature.

1.8 Summary

The research aims to address the gap in the literature on the assessment of sporting events underdeveloped through the focus on legacy and mega-events. In seeking to develop a framework that might compare events of different types and scales, the research will identify the core dimensions needed, their weighting across the breadth of event outcomes, and the specific measures most relevant to capturing their performance.

1.9 Thesis structure

The following outlines the structure and organisation of this thesis. Expanding on the challenges identified, the Chapter 2 – Literature review establishes the development of legacy as the dominant narrative for event outcomes, its limitations in application, status as a contested concept and the role of event stakeholders in that narrative. Chapter 3 considers the structural elements required of a framework and the extent to which it may address issues identified in the literature, before moving to Chapter 4 which establishes the basis for the framework content across the identified areas of impact. Chapters 5 outlines the methods used in the Delphi study to test the framework with the global panel, with results of the panel's responses detailed in Chapter 6 – Results. The final section Chapter 7 – Discussion reviews the results considering the research questions before a final summation in Chapter 8 – Conclusion.

2 Literature Review

This chapter explores the emergence of legacy as the dominant narrative for events outcomes, and how the 'catch-all' definition of legacy limits its application to a wider range of events. Legacy is contrasted with the term impact, and their status contested concepts is explored, as is the role of event stakeholders in the event outcome narrative.

2.1 The Legacy Demand

With the rising complexity of hosting mega-events the bidding process itself had also gained in complexity with 'peak bidding' arising in the case of Tokyo's failed bid for the 2016 Games which alone exceeded USD $100m (ASOIF, 2014). While some measures have been taken to lower the cost of exploring hosting mega-events it remains a considerable undertaking in both financial and human resources. Hence in considering bidding, city and national governments need to give increasing attention to the question of what actual value is derived from hosting events, and whether a portfolio of smaller events might offer a better return for their taxpayers and their strategic goals.

In the face of rising hosting costs since the debt-laden Montreal Olympics in 1976, prospective and past hosts have sought to both expand the claimed gains from direct, indirect, and induced economic benefits, and balance the cost equation through expanded soft 'benefits' (for example social, health, cultural, community pride) – all together now commonly described under the catch-all term 'legacy'.

Indeed, the decision on whether to bid or invest in an event has meant that "for some events, particularly largescale public events, the issue of legacy has become central to the decision to host or create them"(Allen et al., 2008, p. 115). Despite the demand "the justification of legacies from events remains complicated due to inconsistent conceptualizations of legacy across academic and industry practice" (Thomson et al., 2013, p. 111).

Given the rhetoric surrounding the critical role of legacy and its reported importance in decisions on the scale of investment in events, attention from researchers on the categorisation of legacy and a strong foundation of research has been established. That research has been both in developing conceptual frameworks of event legacy (Dickson et al., 2011; Getz, 2003; Gratton & Preuss, 2008; Preuss, 2015) and through reviews legacy literature (Bob & Kassens-Noor, 2012; Mair & Whitford, 2013a; Thomson et al., 2019; Thomson et al., 2013).

Components of legacy

Despite the challenges in defining the term, legacy is believed "to encompass tangible and intangible outcomes which are also classified as hard and soft legacy aspects respectively" (Kaplanidou & Karadakis, 2010, p. 111). The 'hard' or tangible legacies such as sport infrastructure and telecommunication and transportation networks are those considered to be easily identified and measured, whereas with 'soft' or intangible legacies, such as enhanced destination image and renewed community spirit, the converse is true. The literature on sport events presents several classifications of legacies. Cashman (2006), for example, proposed that there are six main categories: sport; infrastructure; economic; information and education; public life, politics, and culture; and symbols, memory, and history. In contrast, Chappelet (2006) argued that there are five categories: sporting, economic, infrastructural, urban, and social. Due to the lack of agreement on the classification of legacy, it is often simplified to include three major categories: economic, social, and environmental (Kaplanidou & Karadakis, 2010), the same dimensions ascribed to the "Triple Bottom

Line". In addition, practitioner-based models have been established to measure event impact by associations and federations such as the Olympic Games Impact (OGI) study (IOC, 2015). Elsewhere, national authorities have developed localised event assessment models such as eventIMPACTS (UKSport, 2015) and NZ Major Events (MBIE, 2013) to assess an individual event's impact within a consistent framework.

Inconsistent or developing definitions?

If there is a common aspect of the definition of legacy, it is the lack of agreement of both its content and form.

The notion of a legacy as something 'bestowed' upon a host as described by Getz (1991) appears somewhat idealistic as cities bid and pay for the rights to host events (often with tax-payer funds) and those costs have continued to soar in the intervening decades. The broad sweep of these endowments included "the physical, financial, psychological, or social benefits that are permanently bestowed on a community or region by virtue of hosting an event" Getz (1991, p. 340). Ameliorating the positivist language in that definition is also a recognition that legacy "can also be used to describe negative impact, such as debt, displacement of people, pollution, and so on." (p. 340)

Other researchers looked to specific dimensions of legacy as the core outcome in their definition. Sport is more central other event legacy narratives looking towards "new opportunities for participation, and stirring examples of human achievement, inspiring wider and wider circles of men, women and children to train, clubs to be formed, and public and private sporting investments to be made" (Kidd, 2003, p. 135). A sentiment still reflected in London 2012's bid to "inspire a generation" (LOCOG, 2004, p. 7). The built environment is perhaps the default version of legacy that people associate with an event hosting. It is tangible, and for the most part, enduring to the extent they are a 'permanent improvement' to the built environment. (Hiller, 2000).

More recent definitions have sought to expand the conceptual dimensions – but this has been at the expense of specifics; as reflected in the open-ended definition of "all planned and unplanned, positive and negative, tangible and intangible structures created for and by a sport event that remain longer than the event itself" (Preuss, 2007, p. 211). The use of 'planned and unplanned' enables event hosts (and owners) significant latitude in their claim of legacies from the event. It at least retains an acknowledgment of a requirement to show causation from the event itself and offers a position that improves on the earlier definition from Essex and Chalkley (2003a) where outcomes may be claimed for an event "even if there is evidence that the development may have emerged in the fullness of time irrespective of the event" (p. 95).

Amidst the confusion and overlapping definitions proffered over time, there emerges some commonality and consistent themes on the assessment of event outcomes and legacy.

Confusion in defining and assessing event legacy

Noting the confusion in the definition of legacy across frameworks, researchers have sought to identify shared elements. Thomson et al. (2013) identified five elements that were common across the researchers in their framing of the definition of event legacy. They identified them as follows:

1. Terminology: use of "legacy" as opposed to another term.

2. Legacy as automatically bestowed or needing to be planned.

3. Temporal nature of legacy: permanent or long term.

4. Legacy as positive and/or negative.

5. Legacy as a local and global concept.

Of the definitions cited by the authors, the emergence of legacy as the default and common terminology for describing event outcomes is not perhaps a sign of agreement per se. Rather it reflects an increasing adoption of, and acquiescence to, the IOC's preferred terminology; given that most of the authors and work cited refer singularly to the Olympic Games in their definition instead of considering sporting events of different scales or types.

This focus on a specific mega-event as the lens through which legacy frameworks should be conceptualised and described means the authors are not starting from a fundamentally neutral viewpoint. The challenge in seeking to apply legacy frameworks to a wider range of events is not simply one of resourcing and attention alone – but that the foundational assumptions of what legacy might mean are framed by the largest and most costly event on the planet. There is an assumption that a legacy must exist, and if not, it must be created. It is embedded in the requirement to provide a detailed legacy plan to even bid for the Olympic Games, and hence becomes an implicit element built into conceptual frameworks presented.

Alternate approaches have been developed by some researchers (academic), event owners (Commonwealth Games Federation), event funders (UK Sport), and professional services (PwC, Deloitte). However, rising cynicism from citizens on the value of hosting mega-events has derailed bids from prospective hosts, with the promise of legacy benefits too vague to garner their support. The confused and vague definitions of legacy have fuelled the mistrust and hampered the development of a constructive dialogue of what legacy should mean, and more critically is it, or should a legacy, always be required for every type of event?

2.2 Unmasking the legacy assumption

The dominant characteristic of the research on the outcomes of hosting events (whether supporting or challenging their value) is reliant on the assumption that the concept legacy is central to the value of events.

In exposing the mythology around the rise of 'leadership' as a panacea to all organisational issues, Wilson (2013) captured pithily "We have come to live in an age where leadership is the solution, regardless of the problem" (p. i). Sporting events have become afflicted by a similar irrationality. The application of legacy has become so central to the narrative around major events that it negates exploration of the implicit and underlying assumption of a legacy being a necessary and requisite condition of hosting an event.

Rather than accept the implicit assumption that legacy is an inevitable outcome of event hosting, Byers et al. (2021) draw on contradictory evidence found in the legacy literature to question the "assumption that legacy is a solution to economic, social, cultural, or political challenges"(p. 172).

This viewpoint is further reinforced by the IOC, CGF and FIFA formalising the requirement that bidding cities and countries make explicit the plans, form, and scale of the legacy that will arise should the candidate be awarded the right to host the event. In the case of the Olympic Games, prospective candidate cities need to develop and detail their Legacy Governance plans including "the governmental and non-governmental organisations responsible for the planning (pre-Games) and delivery (pre-Games and post-Games) of

legacy programmes [how funding will] be secured for the implementation of the legacy programmes and their continuation in the post-Games period''(IOC, 2021a, p. 25). Despite moving to accept a greater use of existing and temporary venues the IOC has not relinquished its intention on the Games to leave an enduring imprint on the host city.

2.3 Legacy and event scale bias

That an explicit 'legacy' is included in the bidding requirements for all major events premises that major events implicitly create a legacy. Compounded by the IOC's positioning that the term 'legacy' implies long-term or enduring, this implicit assumption that major events are associated with legacy creates a bias towards researching major events and their outcomes. Further, the presumption that events have to leave a 'legacy' has extended implications for host city strategies. It creates a bias towards large-scale events as the mechanism for legacy creation and diminishes both the interest in, and resourcing towards, researching the effect and assessment of smaller events for prospective host cities; leaving them less informed about the potential contributions of smaller events on a comparative basis.

Much of the research on the hosting of sporting events have focussed on mega-events (Summer and Winter Olympics), FIFA World Cups, and to a lesser extent Commonwealth Games and Rugby World Cup (Thomson et al., 2019). This skew on research towards global events, a pattern of 'research and requirement' (research on the legacy from mega-events), establishes the precedent for IOC to make legacy a requirement for future bids and sets up a seemingly virtuous cycle of event bids, awarding, hosting and research. Underlying this is that all frameworks, irrespective of their definitions and dimensions, are predicated on the assumption that there is a legacy from sporting events because under an 'availability bias' (or more correctly 'availability heuristic' (Tversky & Kahneman, 1973)) of only researching mega-events, this has always been true.

However, more sporting events occur globally every month than could ever be researched using the current approaches to legacy as applied to major events. No single week in the world's highest attended sports leagues (including NFL, IPL, AFL, Bundesliga and EPL (Gough, 2021)) produce a legacy that would be consistent with the models under which the current operating understanding of legacy is being framed and researched. Perhaps in aggregate, multiple years of competition may lead to longer-term or enduring 'legacies', but outside the economic effect, the link to 'legacy' as it is positioned for events is tenuous at best. Despite this, those four named competitions consistently deliver the highest average game attendance of any competition and total attendance at any single round of their respective competitions would surpass the attendance at all except the largest of the major events. In even starker contrast to the relatively short aforementioned competitions, Major League Baseball (MLB) teams in North America play 162 individual games each regular season for a total of 2,340 games across the 30 teams of the competition. In this regard, a fixture of any of these competitions might be considered to be the closest pure 'event without legacy' – and certainly a challenge to the concept that scale and legacy are synonymous or inextricably linked.

It is not just large-scale recurrent events that are missing from within legacy frameworks, but also events that have a large scale relative to the host location. The failure of legacy frameworks to account for the lack of fit with recurrent events is highlighted with globally obscure but locally impactful events such as the "Lorne Pier to Pub" – a swim event that typifies the need for a critical review of legacy frameworks and their disproportionate focus on major events. Held in the seaside town of Lorne, two-hours drive from Melbourne,

Australia the swim has a self-imposed limit of 4,000 competitors and occurs in a single afternoon annually – in a town with a population of just 1,114 permanent residents (ABS, 2016). Despite its scale compared to the town's population, the influx of competitors and visitors use only existing infrastructure and temporary marshalling fences that within 24 hours leave nothing but the spike in economic activity for local business and the revenues that underpin its lifesaving club as the 'legacy' of another edition. Hence despite its popularity over 35 years, few of the dimensions of traditional legacy frameworks could be applied at events of this scale and certainly not the 'enduring' dimensions - most often applied to post-event infrastructure - required for major events.

2.4 Re-grounding the concept of legacy

In assessing the resultant outcomes of event hosting, the work by Burns et al. (1986) on the 1985 Australian Grand Prix on its then host city (Adelaide) addressed the effects on a breadth of areas including tourism, accommodation, local business and local residents. This pioneering study as noted by Davies et al. (2013) is considered to be the foundation of the event legacy field. However, it is also noteworthy that in this seminal work the term 'impact' rather than legacy adorned their title i.e. "The Adelaide Grand Prix: The Impact of a Special Event".

Understanding how and when the concept of 'legacy' from sporting events emerged and came to dominate the narrative is critical to determining why the field of research has developed without a consensus viewpoint on what legacy is.

Based on the literature there are three strong potential pathways as to how the concept of legacy emerged around events. "These aim to describe the 'mostly likely' pathways by which the concept of legacy developed and how the concept is shaped in the literature by that heritage. Although presented as distinct they not mutually exclusive or exhaustive of all possible pathways. The pathways relate to the concept of an event legacy, not the legacy of any specific event."

Legacy pathway #1

The legacy concept emerged from inductive reasoning based on observation studies

Here the assumed pathway is that the legacy concept emerged as an observed phenomenon from the hosting of events and that under repeated observations researchers sought to codify the outcome into frameworks. This would be consistent with the foundations of Grounded Theory (Glaser & Strauss, 1967) where "theory is generated by showing the dynamic relationships among the emerging concepts" (Gehman et al., 2017, p. 286). That would established the legacy concept as arising from the observation of data-to-theory connections (Gioia & Chittipeddi, 1991) in the context of the impact of events, rather than the imposition of a theory onto the target.

The absence of an original exemplar case on which a consensus viewpoint on the definition of legacy could be applied (Thomson et al., 2013) indicates this is a less likely pathway from which legacy emerged; as if it had, then later work should have built on or expanded a previously agreed definition from the original case – of which there is none.

Legacy pathway #2

Legacy is the use of an event as a mechanism to realise existing long-term host plans

An alternative pathway is that 'legacy' is a term applied to the realisation of outcomes extant in long-term development plans of host cities: that is, the pre-existence of "plans for

future city development represents the 'without case', the city development that will take place without the event" (Preuss, 2007, p. 217). Hence rather than being unique to the event itself, these plans are brought forward or fulfilled by the hosting of events. Therefore, legacy under this viewpoint is an acceleration of, or agreement to, the execution of existing plans for example "urban renewal legacy (Istanbul), improved transportation legacy (Osaka), environmental legacy (Beijing), or improved sporting facility legacy (Toronto). In each case, the term legacy appears to be used as a synonym for the realisation of long-term planning" rather than isolated to the event itself (Agha et al., 2012, p. 131).

In testing the veracity of this proposition then the lack of agreement on both the definition of legacy and its constituent elements is indeed reflective of the diversity and unique nature of each host city's long-term plans. However, it would also follow then that a consensus definition is unlikely to emerge, and any specific method of assessment (and therefore a comparative assessment) would not possible as the objectives vary from one edition to other editions of even the same event, let alone across events.

Legacy pathway #3

Legacy = A concept devised to satisfy stakeholders to offset significant costs of hosting.

The third pathway is that the identification of and attribution of a legacy from an event grew from a need to close the gap between the cost of hosting events and the satisfaction of stakeholders. As economic impact reports were becoming more critically assessed for their subjectivity (Davies et al., 2013) the gap between the cost of event bidding and the credible economic impacts widened. The response was to look for additional areas of event impact to be included under the umbrella term 'legacy' to justify the expenditure of public funds to satisfy the needs of stakeholders. The increased academic and public sector attention on event impact arose "due to the needs of stakeholders, including event organisers, sponsors and government to justify the investment of both private and public funds to support major events" (Davies et al., 2013, p. 31). Despite the noted lack of definition on what legacy entails "public policy planners and event organizers are increasingly promoting the legacies of sport events to justify significant investments required to host them" (Thomson et al., 2013, p. 111).

Hence not only has the development of legacy served to protect stakeholder interests it has become an industry in itself as "the Olympic legacy framework turned the idea of sustainable sports development into an enterprise rationalizing and legitimizing its major stakeholders, organizations concerned with monitoring and measuring the legacy and myriad of delivery partners"(Girginov, 2012, p. 547). That role of legacy as a rationalising element in event bids has been a deliberate strategy for the development of mega-events. The primacy of legacy in mega-event thinking is such that researchers have sought to investigate how "the emergence of legacy and the process through which it becomes a taken-for-granted institutional rule that has impacted how organizations plan and implement the Games" (Leopkey & Parent, 2012a, p. 437). The degree to which legacy has been embedded within the IOC rhetoric specifically and more broadly in other global events has given rise to the need to provide a framework for measuring this institutionalized requirement. The absence of a common comparative framework seeking to work across events has seen the "emergence of a double myth where the myths associated with event legacy lead to the creation of institutional rules that also function as myth." (Leopkey & Parent, 2012a, p. 448). So where does that leave us?

Current state

There has been substantial confusion in determining how to measure legacy as there has not been a clear single origin of the development of event legacy, a definition for it, its dimensions, or the time over which benefits can be accrued. Worse still is that without a consistent framework for the assessment of events, their contribution to the local host is open to highly subjective choices within impact models "because the motivation undergirding them usually is to prove the legitimacy of the sponsor's economic case, the temptation to engage in mischievous practices is substantial" (Crompton, 2006, p. 67).

The fact that there has not been any agreement on the definition of legacy challenges the development of the concept from a 'grounded theory' approach (Pathway 1), where the initial observation should have established an original definition – even if that definition is subject to later revision or refinement. The current state of confusion is more consistent with the development of Pathway 3, where legacy has been a deliberately abstract and artificial concept developed as a necessary function to address the issues of cost vs. outcomes and one or more stakeholder groups – which by nature are inconsistent across events and unique to each event. In addition, since the third pathway of needing to satisfy stakeholders predates the emergence of the second pathway in the literature - where legacy is the convenient rebranding of existing long-term plans through event hosting - then pathway three appears the most likely route for the development of the legacy field.

Reinforcing the supposition that Pathway 3 is the most likely development pathway is the criticism on the emergence of legacy as the superior form of event outcomes from Gold and Gold (2013). As a direct recommendation from the "The Legacy of the Olympic Games 1984–2000: Conclusions and Recommendations" conference in 2002, the IOC actively sought to ensure the concept of legacy "was sufficiently flexible to provide an all-inclusive framework that was able to embrace, with equal facility, outcomes that could be tangible and intangible, planned and unplanned, direct and indirect, short- and long-term, and positive and negative." (p. 3530). A 'catch-all' definition that cut across all possible outcomes at the expense of specificity.

If Pathway 3 is correct (or at least the dominant pathway to which researchers, owners and practitioners have contributed) then the source of emerging confusion and lack of agreement on the definition is not as much surprising as designed.

For the promoters of the 'flexible' legacy concept, the unintended consequence of seeking legacy to be weakly defined and subjectively assessed has over time undermined its credibility; and with it, the credibility of its promoters. The owners of mega-events now face empowered and hostile citizens urging and voting for their city governments to not bid for these events (Gold & Gold, 2010). The depth of cynicism around the outcomes of these events is such that citizens are in effect resisting being recipients (perhaps burdened would be more appropriate) of the legacies these events claim to deliver.

Irrespective of the pathway by which the concept of legacy developed, buried within the narrative is the assumption it is a requisite element of hosting, and how that influences the framing of event outcomes. In a systematic review of legacy literature mega-events dominate the domain with over half of the research articles using the Olympic Games in their title, with the next dominant event being the FIFA World Cup. (Thomson et al., 2019). It is difficult to disentangle whether the bias towards researching legacy on mega-events has developed because the scale of the event means that both costs and intended (under this condition) legacies are most significant, or whether funded centres of research (such as the global network of over 50 Academic Olympic Studies and Research Centres (OSRC))

dominate the narrative and prevent smaller events from matching, replicating, or contributing to an alternative legacy description. Given the prominence of the mega-events in framing the legacy narrative, the extent to which smaller events and their smaller impacts were considered in the development of existing legacy frameworks is not explicit.

2.5 Legacy as an 'Essentially Contested Concept'

Is legacy, as it exists in the literature, a contested concept or just poorly defined? Whilst some literature on legacy has continued to develop in its critical assessment of legacy and its characteristics. For example, the recognition of legacies as potentially negative and not just positive (Preuss, 2007) and reviews of the extent to which intended outcomes are even realised (Brittain et al., 2017), a critical review shows "there continues to be a trend in the research of attempting to categorise legacy purely as positive, mixed or negative"(Thomson et al., 2019, p. 309). Within these remains a consistent theme within the research on legacy – that is the presumption of the concept of itself.

Researchers make an implicit assumption that for any event, a legacy must exist and be planned for. Rather than challenging the assumption of legacy, the quality or absence of legacy outcomes are given as a function of the hosts' execution ability rather than the appropriateness of the legacy to the event. This promulgation of legacy is consistent because hosts are required to articulate and plan for 'a legacy' in bidding for the event, but researchers are not challenging the premise of legacy even though legacy is without an agreed definition or comparable framework for assessment.

The characteristic of a concept being simultaneously readily used but still lacking an agreed definition has been identified previously and categorised as an Essentially Contested Concept (ECC) (Gallie, 1956).

Is the concept of legacy actually contested or just confused?

The term essentially contested concept is "used to indicate situations where there is widespread acceptance of a concept but disagreement on the best instantiation of it" (Miles, 2012, p. 286). In reflecting the dispute surrounding the components and definition of event legacy, the concept of 'legacy' appears at prima facie to qualify as an essentially contested concept. That is, the term legacy is accepted in its use and existence to the extent that it is both required and included in the bidding for major events, but without the grounding of an agreed exemplar of its meaning.

The application of Gallie's criteria for parsing the differences between failure to grasp differences between definitions and the selective use of one definition over others means that it remains actively relevant in social sciences today and "has attracted wide attention over the intervening decades." (Collier et al., 2006, p. 212). More recent applications of ECC have addressed Corporate Social Responsibility (CSR) (Okoye, 2009), social entrepreneurship (Choi & Majumdar, 2014) and the 'sharing economy' (Acquier et al., 2017).

Noteworthy amongst recent papers is the application of essentially contested concepts within the wide field of sustainability – a term (like that of "Legacy") that is readily used but with widely varying definitions. Applying a similar approach as Thomson et al. (2013) took to legacy, Korhonen et al. (2018) gathered and analysed conflicting definitions of the concept of a 'Circular Economy' (CE) from the literature and applied Gallie's EEC criteria to the CE concept "to initiate a more scientific, research-orientated or scholarly discussion on the newly popularized concept of the circular economy." (p. 551). Amongst their findings, the authors noted the concept's definitions are "still fragmented and…have been

developed and discussed in the literature without in-depth and critical discussions on the theoretical foundations, system boundary limitations, and frameworks for methodological inquiries" (p. 551); of which the latter two reflect closely the current state of legacy for events.

The rationale for moving from a wasteful linear (extract, produce, use, dispose of) to a 'closed loop' production approach is compelling in both economic and sustainability terms but that doesn't ensure the concept remains uncontested by differing stakeholders. Likewise, the concept of "legacy" at prima facie as another case of an essentially contested concept appears rational. Fortunately, Gallie (1956) also provided the specific and testable conditions under which, if applicable, a concept could be considered contested.

The application of Gallie's conditions, in addition to substantiating the status of legacy as an essentially contested concept or not, reveals the specific sources of dispute on event legacy as it is currently used.

2.5.1 Defining the conditions for an essentially contested concept (ECC).

In assessing the conditions under which concepts may be considered 'contested' rather than just misunderstood or wilfully misused Gallie (1956) proposed five key operating conditions:

I. ''It must be appraisive in the sense that it signifies or accredits some kind of valued achievement''

II. ''This achievement must be of an internally complex character, for all that its worth is attributed to it as a whole''

III. "An explanation of its worth must therefore include reference to the respective contributions of its various parts or features...the accredited achievement is initially variously describable''

IV. ''The accredited achievement must be of a kind that admits of considerable modification in the light of changing circumstances; and such modification cannot be prescribed or predicated in advance".

V. ''That each party recognizes the fact that its own use of it is contested by those of other parties, and that each party must have at least some appreciation of the different criteria in the light of which the other parties claim to be applying the concept in question." (pp. 171-172).

Contrasting the state of 'legacy' assessment against the five core conditions defined by Gallie confirms the 'contested' state of the event legacy and its associated research.

Table 1 - Gallie's ECC conditions and legacy

Condition	Gallie's description of Conditions for an Essentially Contested Concept (ECC)	Legacy perspective
I	It must be appraisive in the sense that it signifies or accredits some kind of valued achievement.	The inclusion of 'legacy' and the requirement by event owners to specify and invest in 'legacy' from their event, signals its perception as a valued and valuable outcome.

II	This achievement must be of an internally complex character, for all that its worth is attributed to it as a whole.	The complex nature of events of bringing infrastructure, sport, competitors, technology media, volunteers, officials, transport, spectators etc. together is the driver of 'legacy' rather than any specific element isolated from the event
III	Any explanation of its worth must include reference to the respective contributions of its various parts or features	Each event's edition has specific requirements and challenges based on the event and the host's extant infrastructure, capabilities and needs. Legacy contribution and worth are thus inseparable from the context of its specific hosting.
IV	Achievement must admit considerable modification in the light of changing circumstances; and such modification cannot be prescribed or predicted in advance.	Legacy outcomes are highly fluid with 'planned and unplanned', and 'positive and negative' outcomes within the event results
V	Each party recognizes that its own use of it is contested by those of other parties, and that each party must have at least some appreciation of the different criteria in the light of which the other parties claim to be applying the concept in question.	Those assessing the 'legacy' of events recognise there are multiple different definitions and criteria, but in reporting, stakeholders' selective use of definitions are maintained.

In reviewing the preliminary case for legacy as an essentially contested concept, it is in Gallie's fifth condition that the central contention in the application of legacy is found.

The promulgation of the concept of 'legacy' has been used aggressively by the International Olympic Committee (IOC, 2015) to expand and reinforce the claimed territory for the range of benefits a host city might accrue from bidding, let alone hosting, an edition of the event. Indeed whilst initially acknowledging that researchers often use the terms of 'legacy' and 'impact' interchangeably, the IOC's pejorative framing of 'impacts' as "more often regarded as implying an adverse effect or a damaging or destructive result" (p. 4) seeks to actively redirect attention towards a more favourably and positively framed legacy by representing the latter as "more often used when presenting positive effects. 'Legacy' also tends to be used in association with those effects that are of longer duration" (p. 4) .

That self-interested framing of the concept of legacy and the resistance to other forms of the concept was anticipated by Gallie in the fifth of the ECC conditions. Those positioning legacy to align with a mega-event agenda use it both "aggressively and defensively''(p. 172) to further their position. That is, 'aggressively' as in portraying short-term impacts as a negative, and 'defensively' in supporting 'catch-all' definitions such that all outcomes can be claimed as a legacy. Hampered by the absence of an agreed definition of 'legacy' within academic and practitioner communities, the IOC's pejorative positioning of 'legacy' and 'impact' seeks to frame researcher and host language about their events. An effect compounded by the hosts of each edition (Summer and Winter Olympic Games) is the use of a tailored version of 'legacy assessment' without direct comparison to prior host results. This lack of comparability between editions of the event hosting does not preclude claims

being made of an overall 'Olympic legacy' without a common set of outcomes or the concept being clearly defined.

Use of such subjective terminology is perhaps matched only by the bullish defence in the cost and usage of scarce public resources by bidding host cities who seek to justify the costs imposed on their citizens under the named concept of legacy (Sant & Mason, 2015). Hence despite the confused state of the concept of legacy, it is being used as predicted by Gallie aggressively and defensively by different 'activist' stakeholders who also seek to limit others' usage of the same concept in assessing event outcomes.

The extent to which the perspectives of others is appreciated or considered is a key dimension this research will explore. It will seek to test whether a revised framework for considering event outcomes may deliver a more universal appreciation of how others view 'legacy', and contribute to at least reducing (rather than eliminating) the 'contested' nature of the concept. This reduction in contestedness is vital. As any progress made towards legacy becoming less contested and better defined will contribute to it being better understood and more likely to be consistently applied across sporting events. That consistent application to a range of events will also contribute to a better grounding of expectations by potential hosts regarding the outcomes that might be achieved, and hopefully a more rigorous assessment if the outcomes sought are commensurate to the level of tax-payer funding.

2.5.2 Terminology of Impact vs. Legacy – is it semantics?

In agreement with the IOC, Preuss argues that 'impacts' are too short in time to consider the contribution of events to the city because the event 'legacy' is a longer-term shift in the city's outcome. Positing that "long-term economic growth requires a constant influx of autonomous money" and mega-events can only achieve their legacy if tied to "structural changes that improve the host cities' location factors" (2007, p. 213). That viewpoint accords with Ritchie (2000) in asserting that legacy comes from embedding the event into the host city processes – that is, Pathway #2 - that events are simply a mechanism by which the longer-term city development plans are brought to fruition. If these are the foundational case for event legacy, then the notion of 'a legacy' as a distinctive and separable outcome from an event is flawed. It is not more than a catalyst for existing long-term development plans.

Despite years of claimed legacies from mega-events they remain inconsistent and elusive at best (Brittain et al., 2017) and are the worst class of projects for cost overruns (Flyvbjerg et al., 2016) a city can undertake. To gamble on a single mega-event as the catalyst for an enduring city transformation has a poor risk-adjusted return. By contrast, the shorter time frame of 'impacts' fits well with the vast majority of events outside the mega-event level. These can still be major events but also include events that are run annually and fit within the city's capabilities and longer-term strategy. In contrast to seeking a 'legacy' from a single event, such an approach may be more fiscally prudent as it allows the city to integrate and promote a recurrent event to build a longer-term narrative for both its residents and prospective tourists. It may not require the 'structural changes' Preuss advances as being required to qualify for 'legacy' status, but it needs to make a contribution appropriate to its scale and delivers recurrent value to its host.

Hence while impacts may be seen as less ambitious and shorter-term, it does allow for a wider range of events to be considered without the hubris of an enduring 'legacy' to be claimed or required. Under an 'impact' view, each event is assessed on its contribution in the shorter term and benchmarked against expectations and the business case, however, if

applied consistently it should enable hosts to learn from other editions of the same or similar events.

2.5.3 The difference between Contested vs. Confused concepts

In establishing the five essential conditions for a contested concept, Gallie also identified the need to distinguish between genuinely 'contested concepts' and those that are just "radically confused" (p. 180). To do so, two further conditions were outlined.

VI. ''The derivative of any such concept from an original exemplar whose authority is acknowledged by all contestant users of the concept''

VII. ''The probability or plausibility, in appropriate senses of these terms, of the claim that the continuous competition for acknowledgement as between contestant users of the concept, enables the original exemplar's achievement to be sustained and/or developed in optimum fashion''.

As already noted, there being no authoritative exemplar of legacy on which to draw raises a challenge as to whether legacy is 'contested' or 'confused'. In exploring Gallie's semantic origins of an exemplar Evnine (2014) notes that in "essentially contested terms, the exemplar is something like a stage of a tradition. The exemplar will therefore consist in anything that might be an element of a tradition" (p. 127). Whilst lacking a single agreed exemplar there is a strong tradition of the concept of legacy being applied to events. The 1992 Barcelona Games is presented as the positively appraisative model of events that can deliver sustained legacy outcomes. However, that event is predated by other editions of the same event that could also be said to have delivered distinctive legacies (or at least elements thereof) in their own right - including the 'commercial success of 1984 Games'; the politicisation and negative legacy of the 1936 Berlin Games, the sporting records and political statements of the 1968 Mexico City, the international terrorism of the 1972 Munich Games and the cost burden legacy of 1976 Montreal Games (Gold & Gold, 2010).

Hence while the identification of a single exemplar would be ideal, the strong tradition and consistent presence of the concept of legacy being associated with these events is sufficient. That is, there is less dissension on the existence of event outcomes than there is on the method by which those might be acknowledged and the boundaries of those outcomes being claimed.

That a concept does not as neatly conform to the extended conditions is neither unexpected nor a critical flaw in the application of contested concepts to the issue of legacy. Indeed, as if in anticipation, Gallie himself observed that not all concepts fitted neatly into the justifying conditions noting "I must admit that my first justifying condition…(in this case a single tradition of art) cannot be simply or directly applied" (p. 182). Hence while legacy lacks a singular exemplar legacy, Gallie's primary five conditions for contestedness remain intact and are still applicable. The lack of a singular exemplar extends into the seventh condition but has not precluded the development of the field of legacy research. The contest between versions of legacy promulgated is not on the presence or absence of legacies but rather the method by which they might be classified, claimed or the extent to their effect. That does not reduce the contestedness of the concept but does to a large degree ameliorate the absence of the singular exemplar in this instance.

Hence the concept of 'legacy' meets all five primary conditions for an 'essentially contested concept'. The goal in reducing contestedness would be to enable hosts with an objective and comparative framework from which to assess event options. Reduced contestedness would shift the focus from 'aggressively and defensively' applied definitions of legacy with

flexible usage and 'catch-all' outcomes and instead work towards defining the measurable substance of legacy. In the development of a new framework, any progress to reduce or resolve at least one of Gallie's five conditions for the assessment of event outcomes would make a substantive contribution to the literature in this critical field of 'legacy' research.

2.5.4 Does using 'impact' resolve any 'legacy' contestedness?

As noted previously, the term 'impact' has been pejoratively characterised as a negative and criticised as focussed on shorter-term outcomes. An alternate view is that rather than being a limitation, it is the shorter-term horizon that is central to achieving a clearer definition and more widely applicable framework for the assessment of event outcomes.

A critical rather than subjective characterisation of 'impact' recognises that a shortened 'time-frame' brings with it three simultaneous benefits to the framing of event outcomes. The first is that a shorter horizon allows for a greater level of certainty in the outcome and that the outcome is directly attributable to the event and not related to other confounding factors. The second is that the overwhelming majority of sports events held in any single year are not mega-events. The shorter timeframe allows for individual editions of annually recurrent events to be included in standardised assessments, as opposed to using an extended timeline of mega-events as the 'base case' for framing assessable event outcomes. The third is that the shorter time allows for greater frequency and number of events being assessed and hence more rapidly developing, relevant, and shared learnings and practices to be applied to future events. Therefore, a shift to framing outcomes in terms of impact offers greater clarity (by avoidance of confounded effects), certainty (directly measurable from the event), objectivity, and consistency over the 'catch-all' use of legacy.

Applying an event assessment approach grounded on an impact perspective rather than legacy perspective to Gallie's conditions for essentially contested concepts considers how the more constrained 'impact' might be able to moderate or move closer to resolution on a greater number of conditions over a legacy perspective of event assessment.

Table 2 - Gallie's ECC conditions: Legacy vs. Impact

	Gallie's ECC conditions	Legacy perspective	Impact perspective
I	It must be appraisive in the sense that it signifies or accredits some kind of valued achievement.	The inclusion of 'legacy' and the requirement by event owners to specify and invest in 'legacy' from their event, signals its perception as a valued and valuable outcome.	Acknowledges outcomes and activity from the event, and it is valued for its contribution relative to scale.
II	This achievement must be of an internally complex character, for all that its worth is attributed to it as a whole.	The complex nature of events of bringing infrastructure, sport, competitors, technology media, volunteers, officials, transport, spectators etc. together is the driver of 'legacy' rather than any specific element isolated from the event	The complexity of event hosting is the same as legacy, but outcomes are focussed on shorter-term outcomes and operational drivers rather than subjective narratives.

III	Any explanation of its worth must include reference to the respective contributions of its various parts or features	Each event's edition has specific requirements and challenges based on the event and the host's extant infrastructure, capabilities and needs. Legacy contribution and worth are thus inseparable from the context of its specific hosting.	Specific and measurable dimensions are common to all events and differ only in scale or relevance.
IV	Achievement must admit considerable modification in the light of changing circumstances; and such modification cannot be prescribed or predicted in advance.	Legacy outcomes are highly fluid with 'planned and unplanned', and 'positive and negative' outcomes within the event results	The scale of impact effect is subject to the same conditions of legacy but removes the need for positive and negative perspectives from assessment.
V	Each party recognizes that its own use of it is contested by those of other parties, and that each party must have at least some appreciation of the different criteria in the light of which the other parties claim to be applying the concept in question.	Those assessing 'legacy' of events recognise there are multiple definitions and criteria different but in reporting for their stakeholders' selective use of definitions are maintained.	No ascribed value to performance. Provides a single 'truth' in the results to all stakeholders with explicit identification of stakeholder viewpoints.

2.5.5 Implications of contestedness on the development of sporting events

How does the lack of agreement on the definition of legacy and the lack of a consistent assessment of events limit the development of sporting events?

Inconsistent perspectives mean differing expectations, and differing responses to information about the event both in the lead-up, during, and after its hosting. This is not unique to events with capital markets also seeing this effect with investors displaying "'motivated sensitivity' as a response to information, suggesting that investors will analyse and reject information that is inconsistent with their expectation" (Cahill et al., 2017, p. 162). That 'sensitivity' effect is moderated in the capital markets by high volumes of trades and pricing signals, which events do not share

The lack of an agreed understanding of differing perspectives on legacy – even if an agreed definition is not achievable – comes at a cost for both the hosting cities and the development of the sporting events themselves. The ways that can be seen is as follows:

1. *Stakeholder perspectives – depending on stakeholders' viewpoints, the concept of legacy is different, but often not expressed explicitly.*
   - *Hence stakeholders may all agree on legacy's importance but concurrently hold very differing views of what that means, their expectations of outcomes, and resourcing needed to realise them.*

2. *Sporting events are complex and require multiple stakeholders to execute, but how can we engage and get buy-in if they cannot work from a single common understanding and framework?*
   - *This may work for a single event (Olympic Games) that has a specific legacy framework designed for its event but is then not applicable to other events.*
3. *Agency theory identifies that a lack of clarity in the principal-agent model of stakeholder relations for events which understanding of who is the real principal and who are the agents?*
   - *That lack of clarity means there is no reconciliation of stakeholders to a single agreed view and by consequence no ability to adapt that viewpoint in light of changes to the bid/hosting plan.*

The above points reflect some of the practical implications arising from a contested view of legacy. If a more limited impact perspective is able to moderate conditions for an essentially contested concept, the consequences are more than academic clarity but avoid the effects and implications noted above.

2.6 Reweighting or reassessing the value of events?

There is more at stake in the critical assessment of legacy frameworks and ascription of event outcomes than simply the determination of the appropriate use of 'impact' and 'legacy' in terminology, or even the broader inclusion of the range of events. There is a fundamental shift also in the potential for even mega-events to create 'legacy' as understood in the past.

Underpinning that shifting narrative on the nature of event outcomes – or those that might be expected - is the use of the terms 'tangible outcomes' and 'intangible outcomes'. The terms are frequently found in proposed definitions of legacy but lack clarity on the specific usage being applied or the constituent components of each. Further confusion exists in that the terms are also used interchangeably with hard (tangible) and soft (intangible) outcomes. Given the increasingly important role intangible outcomes are expected to have in underpinning events assessments, a framework would need to capture multiple dimensions that might be classed as intangible outcomes. The breadth of intangible outcomes are not limited to social outcomes, but should address elements of media, brand, sport, environmental and economic dimensions. Likewise, the dimensions captured should recognise the beneficiary of developed intangible assets may include either or both the event host and the event owner.

Taken as a common viewpoint, the "early impacts of the Games are typically associated with sporting or local infrastructure" (Leopkey & Parent, 2012b, p. 931) but if, as the report "Cities Alive – Rethinking Legacy for Host Cities" contends mega-events are moving to a design that will deliver 'less legacy' in terms of infrastructure (Arup, 2017). This direction means a reduction in the tangible dimensions of legacy will accelerate in the years ahead, moving towards a model where "legacy may be more about emotion and memories and environmental legacies than enduring sporting infrastructure"(Dickson et al., 2011, p. 297).

Given the existing dependence on physical infrastructure to underpin the tangible dimensions of legacies within current models (Dickson et al., 2011; Gratton & Preuss, 2008; Preuss, 2007; Taks et al., 2014), this could result in a lower perceived 'overall legacy' in real terms. Alternatively, there may be a significant rebalancing in the contribution to the value of events ascribed to tangible and intangible aspects of event legacy (Dwyer et al., 2000). Indeed, that shift is already occurring, and earlier in the event lifecycle than perhaps

widely recognised. Reflecting on the discourse around bidding for the 2010 Winter Olympic bid of Vancouver, Sant and Mason (2015) analysis of the shifting legacy narrative from those championing the bid showed that across the bidding process "supporters moved away from discussions of new infrastructure development and economic impacts toward intangible event benefits" (p. 42). Longer term a shift in weight of value from tangible to intangible presupposes a robust valuation framework for intangible legacy and the results can supplement any decline in tangible legacy value.

Given the seeming shift in the balance of hard and soft outcomes from events, the ascription of an increasing value of future events to the non-infrastructure based elements is predicated upon the ability to:

- *Agree on the components of legacy*
- *Measure them consistently across events*
- *Ascribe a value to each of the components*

An assumption that there may an overall reduction in the value of events because the number of potential visitors hosted may be reduced, reflects both the current 'visitor economics' dominated viewpoint on legacy research and the perceived difficulties in ascribing value to the intangible elements of legacy. A position is not helped by the lack of agreement in the definition of legacy or the timeframe over which the outcomes should be assessed.

If the outcome of achieving a more consistent method for the assessment of events were to see an overall erosion in event value ascribed to legacy, then the rational market response should be either a fall in the number of bids being made for specific events and/or the pricing of bids reduced to meet the lower expected return achieved from the investment of public funds in their hosting. Alternately, if the intangible aspects increase in value to mitigate the loss of value from the tangible components, and perhaps even exceed them, then the market should see the value and price of events could continue to rise – along with the cost of attending them.

At present these three elements – agreed, valued and consistent components – of legacy are not found within the research literature, but a key outcome of this research is the development of a conceptual framework to test if a consistent assessment of events across types and scale is achievable in principle.

2.6.1 Shifting value of event assets

According to the Australian Accounting Standards (AASB) the definition of "assets" are: "A resource: (a) controlled by an entity as a result of past events; and (b) from which future economic benefits are expected to flow to the entity." (AASB, 2018). There has been a shift from 'asset heavy' industries dominating the S&P 500 most valued companies to service and IP based companies –including Alphabet (Google's parent company), Apple, Amazon, Facebook, and Microsoft (Edwards & Lazzara, 2019). The basis of the value of S&P 500 companies has dramatically inverted from 1975, when tangible assets comprised 83% of its total value, to 2015, when it now accounts for just 13% of the total value (Ocean Tomo, 2015). This reflects a shift from tangible to intangible value for businesses over the past four decades. This global trend has occurred at both the individual firm and the national economy level, as markets move from a large manufacturing base towards an economy based around commercial services (banking, finance) and consumer experiences (tourism, events, culture).

That change has also been reflected in the event value perspective when considering the very competitive market for younger audience attention, which is a key consideration for prospective host cities. In response to the question 'Why host the Youth Olympic Games?', the CEO of the Lausanne YOG Committee articulated "Today, cities compete globally for youth attention, trying to brand themselves as an attractive place for young people to visit, study in and work in, in a healthy, sport environment. The Youth Olympic Games…position a host city as the leading destination for future generations of tourists, students and workers." (Logan, 2018). This critical perspective on the potential for the Youth Olympic Games to create intangible value for its host city is critical because it is an event that in its bidding requires that the host city build no new infrastructure to host the event.

That focus on intangible or soft elements of legacy is an exemplar of the prospective shift in thinking about the value of the events. When even those who benefit most from heavy infrastructure investment have recognised the future tilting towards a 'less infrastructure' event future, as posited by global engineering firm Arup (2017), which in turn means greater attention needs to be given to intangible value within a legacy assessment.

2.7 Agency relationship and influence on legacy incentives

In their seminal work on the 'Theory of the Firm', Jensen and Meckling define agency relationship "as a contract under which one or more persons (the principal(s)) engage another person (the agent) to perform some service on their behalf which involves delegating some decision making authority to the agent" (Jensen & Meckling, 1976, p. 308).

Recent work in the application of agency theory has sought to extend research into the 'principal-agent problem' within the domain of sports (Mason & Slack, 2005; Schubert, 2014) rather than solely focus on organisational relationships between stockholders and professional managers as their appointed agents. In the context of mega-events in which much of legacy research is grounded, the application of 'theory of the firm' agency relationships develops new perspectives on the incentives that have driven the determination on legacy and uncovers some of the dynamics that have contributed to the 'contestedness' of its definition.

A limitation of the application of agency theory within event management is that whilst a major sporting event is a significant commercial venture, is it not one in which event management can be self-maximising for their outcomes. Those restrictive conditions reflect the original authors warning that "if both parties are utility there is good reason to believe that the agent will not always act in the best interests of the principal" (Jensen & Meckling, 1976, p. 308). In seeking to maximise their utility, event owners place highly restrictive conditions under which their events may be bid for and hosted. The prescriptive terms under which events are awarded and managed are designed for the protection of the intellectual property and intangible assets of the event's owner, and not to provide the event host with the opportunity for unfettered self-maximising design.

Despite that limitation, there are indeed characteristics of mega-event hosting that are aligned with the issues of agency relationships and the associated 'agency costs' needed to overcome the incentive for both parties in the relationship to be self-maximising. As noted, "it is generally impossible for the principal or agent at zero cost to ensure that the agent will make optimal from the principal's viewpoint. In most agency relationships, the principal agent will incur positive monitoring and bonding costs and in addition there will be some divergence between the agent's decisions and those decisions which would maximize the welfare of the principal." (1976, p. 308).

Rather than the agent having the opportunity to be self-maximising, the 'bonding costs' are inverted and not paid by the principal to align the agent's interest with their own but rather by the agent within the price and conditions agreed to in winning the right the host the event. The restrictive finances, the aggressive and contentious brand and intellectual property policing, and accommodation and travel demands to host an edition of a mega-event have become so punitive that it is more comparable to a franchisor model than principal shareholders working with independent management as agents.

The extent to which agency relationships do not follow the classical model may be due in part to the project-based nature of the venture. Rather than being able to operate on a 'continuing' basis, the hosting of many major events is a 'one-off'. Hence the enterprise framework within which the principal and agent are operating is temporary. The contract in which the stewardship of the property lies is only momentarily with the agent and then reverts to the principal. "Events, unlike manufacturing or service operations, are by definition unique to the location in which they are held, and strictly temporary" (Rose, 2002, p. 726). Indeed the exclusive rights to act as the agent for the defined period are not even unique to the event itself, as even during the period of hosting the next event edition has already been awarded to the subsequent host, and is frequently referenced, most often culminating in a ceremonial handover at the conclusion of the event.

The model of events operating on a recurrent basis is closer to a 'continuing' enterprise, even if only for a contracted period. International tour events like Formula 1 Grand Prix are held annually but the venues are re-contracted periodically, creating a balance of stability and familiar venues, with new circuits added from time to time. The owners of the tour events remain independent principals with each venue owner acting as an agent for that edition of the event.

Permanent recurrent events like the tennis 'Grand Slam' venues are locked into venue and timing annually and able to invest over the long term and represent the closest to the continuing model seen in business. However, the agency relationship shifts in that the principal and agent relationship is overlapped, unlike that of shareholder and manager on which classical stakeholder theory is based.

Complicating the agency relationships for major events further is the potential for multiple 'principals' to be seen as embedded with the principal-agent model of events. For example, defining the primary 'agent' at the Organising Committee and then the multiple principals in the city/event binary are:

Principal-Agent relationship #1: Principal = Event owner*; Agent = Organising committee^{+}

Principal-Agent relationship #2: Principal = Host city^; Agent = Organising committee^{+}

*Event owner	^Host city	+Organising committee
International Federations (IOC, FIFA, CGF, World Rugby, IAAF, UCI etc.) and/or private owners (IMG, ASO, Wanda group)	The primary host, which may be a city, region or a series of cities within a national event	The Local Organising Committee is responsible for the delivery and management of the event on behalf of the bidder.

The implications from viewing events and their requirement for a legacy to be an outcome of event hosting aligns and highlights the incentive for stakeholders to identify a 'legacy' as proposed under 'Pathway 3'… but extends that to acknowledge there is equally an incentive for the event owners (principals) to require an explicit 'legacy' to be included as part bidding for the event by the principal. The effect of making legacy a requirement is to make

effective the ability to limit divergences from the rights owner's interest whilst delegating limited authority to the agent. This is not just to avoid paying 'bonding costs' but to pass the costs on to the agent while reducing or removing the incentives to be solely self-maximising and instead maximise the welfare of the principal. Indeed, the requirement for identifiable legacies enhances the principal's property and raises its market value for future bidders and editions. Hence, through the demand for a 'legacy', they avoid both bonding costs and dilution of their 'welfare' whilst simultaneously raising the value of their property.

Therefore in bidding for large-scale events under highly restrictive conditions based on "less than reliable economic benefit calculations cast a dark shadow on the return on investment claims made by both sport event organizers and government backers" (Lee & Taylor, 2005, p. 596). Where those events require significant public funding and/or debt for delivery, then both the principal (owner) and the agents (bidding stakeholders) have an incentive to claim success in legacy – and avoid deeper scrutiny, especially if legacy remains elusive in both its definition and measurement.

2.8 Events, stakeholders and legacy

The investigation on the application of 'agency theory' and the principal-agent problem uncovered a key dynamic in event legacy assessment, namely the interdependent incentives to claim for event legacies from both principal and agent. However, it also prompts a deeper assessment across a wider range of stakeholders involved in the planning, hosting and assessment of events. Freeman's definition of stakeholders as "any group or individual who can affect or is affected by the achievement of the organization's objectives" (Freeman & McVea, 2001, p. 2) provides significant scope for the inclusion of interested parties when considering the hosting of global sporting events as the organisation's objectives.

Grounded in corporate strategy the origins of "stakeholder theory has been intimately connected to the idea of strategy from the earliest days [and] clearly aimed at making business policy and strategy more effective." (Freeman et al., 2020, p. 214). Here is where Freeman's definition of a 'stakeholder' and making 'strategy more effective' cuts to the core of the legacy issue faced in assessing events. The legacy promises made by major events such as those presented in bidding for the London 2012 Games to 'inspire a generation' promise to imbue a generation nationally with both a passion for sport and the Olympic ideals. That generation would under Freeman's definition be part of "those affected by the achievement of the organization's objectives" (Freeman & McVea, 2001, p. 2). That is, the achievement of the objectives of the London Organising Committee would prospectively have as stakeholders, every person in a single generation.

So how can effective boundaries be placed around stakeholders such that their role in events and the assessment of event legacy might be determined?

2.8.1 Decoding stakeholders and sporting events

To establish effective boundaries an understanding of the critical differences between a 'stakeholder approach' and 'stakeholder theory' is vital. The "stakeholder approach to understanding the firm [is] intended to broaden management's vision of its roles and responsibilities beyond the profit maximization function" Mitchell et al. (1997, p. 855). By in contrast to stakeholder theory which "attempts to articulate...in a systematic way: which groups are stakeholders deserving or requiring management attention, and which are not?" (p. 855). This establishes a useful functional approach to refining the decoding of stakeholder involvement in events. Critically within the above definition, the potential breadth of stakeholders that could be included in the hosting of major sporting events is

almost limitless, hence the first challenge of (event) management is to prioritise their attention in working with stakeholders.

The use of the term 'deserving' in the language of Mitchell et al. (1997) could be seen as disrespectful towards many potential stakeholders in the context of a time-bound project like the hosting of a major sports event. The term may be more reflective of management's need to stretch 'allocation of limited resources' rather than it accruing an unnecessarily pejorative meaning. Whilst there is merit in the application of 'stakeholder approach' in seeking to broaden management's vision, major sporting events do not suffer under the restrictive singular dimension of 'profit maximisation'. To the contrary, with the existing range of outcomes already attributed to event legacies making 'breadth of vision' not the primary limiting factor. Hence, 'stakeholder theory' rather than 'stakeholder approach' provides a more aligned frame of reference in seeking to identify the influential and critical stakeholders across different events in a systematic way.

2.8.2 Defining stakeholder groups

The application of stakeholder theory requires consideration of how to define the boundaries between groups and which approach fits with the context of events and legacy. In the process of identifying stakeholder groups and their interactions, researchers have used three broad classes of approaches: namely a 'descriptive/empirical approach', an 'instrumental approach', and/or a 'normative approach' (Donaldson & Preston, 1995).

The descriptive/empirical approach can be seen in descriptions of the nature of the organization, of the way managers think about managing, of how board members think about the interests of stakeholders, and of how some organizations are managed in reality. The instrumental approach is used to identify the connections (or lack thereof) between stakeholder management and the desired (traditional) objectives of the focal organization. The normative approach provides moral/philosophical guidelines for the operation and management of an organization and has been used to analyse the functions of an organization (Donaldson & Preston, 1995).

The descriptive and instrumental approaches stem from social science-based research while the normative approach stems from ethics-based theory (i.e., moral obligation concerning social issues) (Jones & Wicks, 1999). Stakeholder research can also focus on one of three aspects: the focal organization, the focal organization's stakeholders, or the relationship between the focal organization and its stakeholders. These descriptions suggest that the normative approach would not be appropriate, since we are not seeking to answer an ethics-based question on how event organisations should be run but rather a description of their current relationships and power dynamics and how that influences classification and definitions of outcomes and legacy.

While the descriptive approach looks on the surface to be useful, it would require a deeper view of how managers think about managing across events and qualitative interviews would be required. Hence the instrumental approach seems the best approach for framing assessment of stakeholders for event hosting, as it focuses on the objectives of the organisation and the connections between stakeholders needed to successfully deliver the event.

2.8.3 Stakeholder salience and legacy narrative

Within stakeholder theory, the development of stakeholder salience as proposed by Mitchell et al. (1997) provides an effective basis for delineation between active and passive

stakeholders. Its application gives a structured method for understanding both the breadth of stakeholders involved and the interactions of the complex relationships within the hosting of events.

It is that complexity in relationships and differing perspectives that established Gallie's fifth condition of an essentially contested concept and its legacy application. That is, understanding whether their different viewpoints are considered and recognised by others as being equally valued and valuable, and whether the value of event outcomes are assigned or weighted differently according to their various stakeholder perspectives. In the corporate application of stakeholder theory, the notion of different stakeholder perspectives as being 'equally valued and valuable' is at odds with Jensen's critique that "it is logically impossible to maximize in more than one dimension" (Jensen, 2002, p. 238) and leads to the conclusion that purposeful behaviour requires a single valued objective function. Event hosting lacks an equivalent singular objective like shareholder value in the field of legacy. Indeed, given the open-ended outcomes that stakeholders are meant to maximize under a 'catch-all definition' of legacy', the confusion and contestedness on the concept of legacy are more expected than surprising. There is however a key counterpoint noted by Wood (2008) that Jensen's logical argument is framed within an underlying assumption that maximising a single dimension "necessarily involves limited resources that cannot be deployed in multiple directions without loss" (p. 160).

Can the complexity and differing agenda of actors in the stakeholder network be aligned in the co-production of an event such that outcomes are enhanced (if not maximised) across multiple elements simultaneously? This viewpoint is closer to the notion of the 'interlocking cubes' than the 'singular valued objective function' seen by Jensen (2002). That inherent disconnect between a value-added economic model of value creation and the co-creation across stakeholders echoes Ramirez's reflection on co-production in that "value co-produced by two or more actors… invites us to rethink organizational structures and managerial arrangements for value creation inherited from the industrial era (Ramirez, 1999, p. 49). Hence in considering the differential perspectives and how those influence views of legacy, the application of stakeholder theory and its ability to describe and understand the network of actors in the delivery of a complex project as a sporting event are vital.

So who are the stakeholders in events? A key dimension in the determination of drawing a boundary on the conditions for stakeholders is where the risks are placed. That is, in understanding dynamics in bidding for and hosting events, who decides on the events and who bears the risk for their legacy? That delineation based on voluntary vs. involuntary assumption of risks echoes Clarkson's (1995) perspective on the undertaking of risk in the private sector. Recognising "voluntary stakeholders bear some form of risk as a result of having invested some form of capital, human or financial, something of value, in a firm. Involuntary stakeholders are placed at risk as a result of a firm's activities. But without the element of risk, there is no stake" (Mitchell et al., 1997, p. 856). Within that definition, residents would be 'involuntary stakeholders' as taxpayers do not choose to directly 'invest' in events, their representatives decide how to invest their contributions but are exposed to the risks of events in their host location. So, whilst involuntary, it does not mean residents are passive stakeholders as noted by the recent shift in residents seeking to influence their governments from bidding for overly costly (and hence higher risk) events.

The dynamics of stakeholder relationships is, therefore, more graded than in risk alone. In seeking to codify understanding the relationships between stakeholders. Mitchell et al. (1997) established three core dimensions on which the relationships operate as being (1) the

stakeholder's power to influence the firm, (2) the legitimacy of the stakeholder's relationship with the firm, and (3) the urgency of the stakeholder's claim on the firm (pp. 865-867).

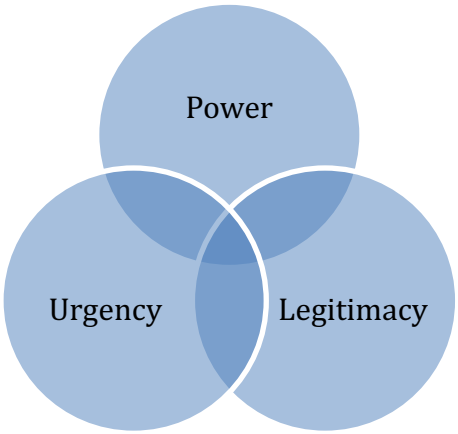

Figure 1 - Qualitative Classes of Stakeholders

The intersection of the dimensions of power, legitimacy and urgency as shown in Figure 1 further refines the classifications of stakeholder typology in which Mitchell et al. (1997) defined groups of stakeholders by their ability to address the dimension.

The authors further clarify the dimensions of the typology noting that urgency "exists only when two conditions are met: (1) when a relationship or claim is of a time-sensitive nature and (2) when that relationship or claim is important or critical to the stakeholder." (p. 867)

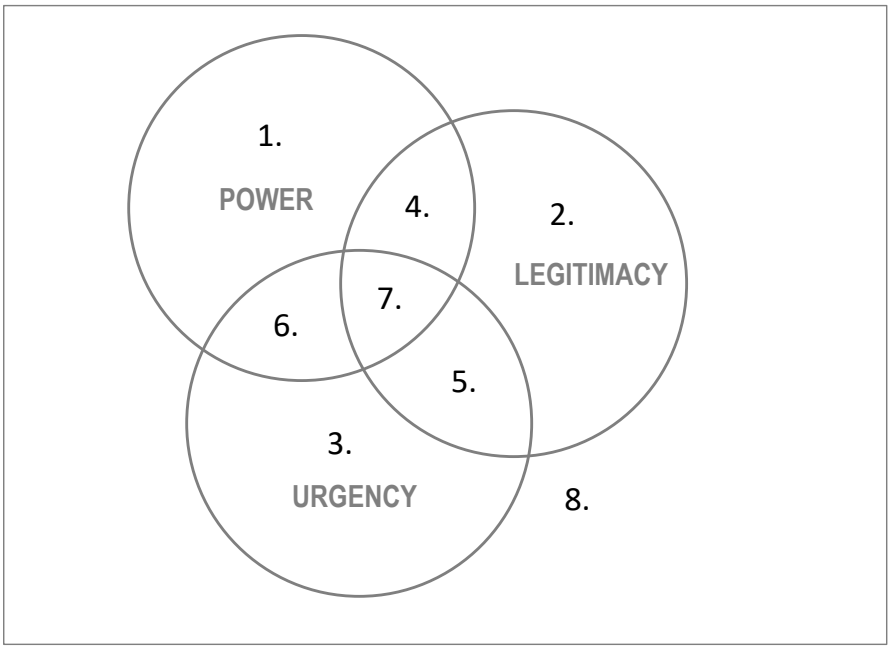

Figure 2 - Subsets of Qualitative Stakeholder classes

Working through the intersections of the typology, (Mitchell et al., 1997) identified the qualitative classes into a typology of up to seven sub-categories of stakeholder classes, with an eighth category of non-stakeholder as shown in Figure 2, and all being dependent on the stakeholder dimensions applicable to the entity being studied.

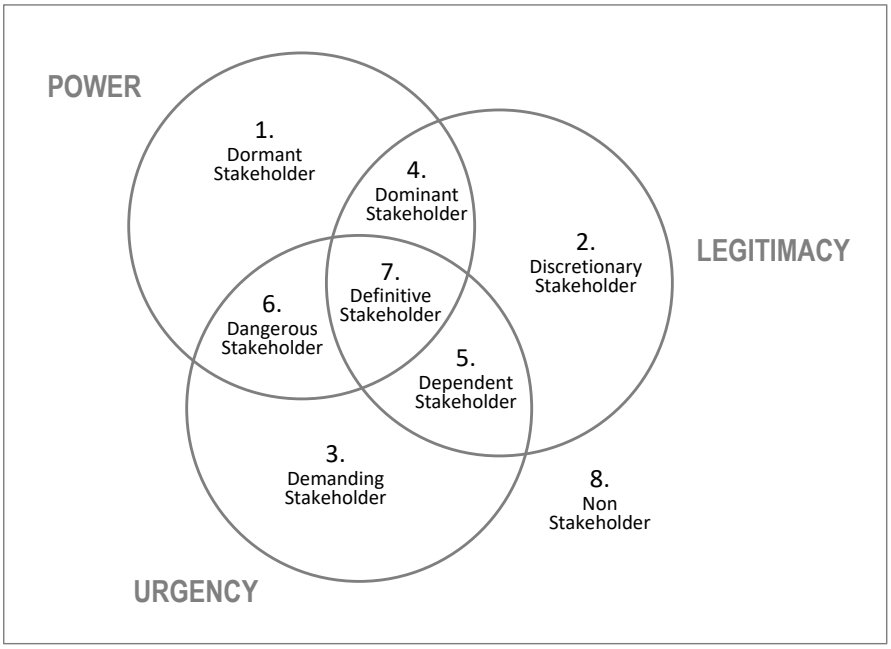

Figure 3 - Stakeholder Typology: One, Two or Three Attributes Present

Noting the intersections in which the stakeholder groups can occupy one, two or all three of the core dimensions of power, legitimacy, and urgency the authors represented their influence in the number of dimensions they hold.

The descriptions for each are as follows:

- *Latent stakeholders: occupy only one of the three attributes*
- *Expectant stakeholders: occupy any two of three attributes*
- *Definitive stakeholders: occupy all three attributes*

Assigning descriptive names to each of the individual groups rather than relying on a numbering identity the names are as shown in Figure 3.

Table 3 - Classification of stakeholders

ID	No. of Attributes	Qualitative Class	Attribute dimensions	Stakeholder group
1	One	Latent	Power	Dormant stakeholders
2	One	Latent	Legitimacy	Discretionary stakeholders
3	One	Latent	Urgency	Demanding stakeholders
4	Two	Expectant	Power + Legitimacy	Dominant stakeholders
5	Two	Expectant	Legitimacy + Urgency	Dependent stakeholders
6	Two	Expectant	Power + Urgency	Dangerous stakeholders
7	Three	Definitive	Power + Legitimacy + Urgency	Definitive stakeholders
8	None	-	None	Non-stakeholder

Table 3 summarises all dimensions of Mitchell et al. (1997) into a single view to identify the relationships between stakeholder perspectives around event bidding, hosting and critically in defining the narrative on event legacy.

Key stakeholders determine the legacy narrative

The controlling determination of legacy's assessment is subject to a narrative determined by the strongest stakeholders in the event eco-system, that is, those with the highest stakeholder salience. The determining factor on stakeholder salience is the number of stakeholder attributes of power, legitimacy, and urgency they hold. Hence the Dominant and Definitive stakeholders within the stakeholder typology are the strongest candidates in the formulation and control of the legacy narrative as they encompass the dimensions of Power + Legitimacy, and Power + Legitimacy + Urgency respectively.

The disparity in power of the dominant stakeholders has already been noted in the restrictive hosting agreements rights event owners impose on prospective hosts. Indeed the terms of those "agreements underscore the power and control that the international event rights owner possesses". (Kelly et al., 2019, p. 52)

Conceding that the narrative will always be influenced by the dominant stakeholders, therefore an event assessment framework able to deliver a less contested viewpoint on event legacy should be able to at least make explicit any biases within the interpretation of the event outcomes introduced by stakeholder groups.

Interlocking 'legacy cubes'

The conceptual 'legacy cube' framework developed by Preuss (2007) introduces three axes on which event outcomes are classified based on their characteristics, but not actually measured on these dimensions.

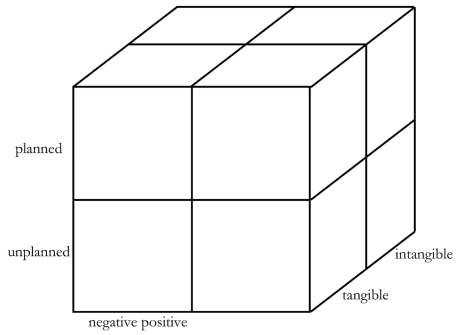

Figure 4 - Preuss' Legacy Cube

Of the three axes, planned/unplanned, tangible/intangible and negative/positive, it is the last that is most susceptible to being subject to the beholder's viewpoint. The same outcome delivered on that dimension might potentially be classed as positive or negative based on the stakeholder group addressed. For example, the clearing of the shanty towns (favelas) for the Brazil hosting of the FIFA World Cup in 2014 may have enabled the host nation's tourism and development organisations to present a more positive destination brand image but delivered a strongly negative outcome for the evicted residents. A duality reflected in the

media coverage that highlighted "two specific aspects of Brazil: natural splendours and social problems such as violence, inequality and social injustice" (Graeff et al., 2019, p. 11)

Hence the notion of an 'overall legacy' of each event is a series of interlocking 'cubes' representing the different stakeholder perspectives rather than a single overall outcome by which positive and negative outcomes can be classified. The differing value of legacy by stakeholder group is reflected in the critical assessment of Olympic legacy and its governance as a "tension between what is being done in the name of legacy, for whom, and at what cost and to what effect" (Girginov, 2012). Ideally, all the cubes could be weighted to represent an overall assessment of individual outcomes.

Each stakeholder group could review an outcome and make its own determination as to that outcome's value. So even within a shared single framework, outcomes would still be subject to stakeholder viewpoint, but at least they would have a common reference from which those determinations are likewise made explicit. Used proactively, the same framework can draw out differences in expectations of event outcomes and assist in forming consensus in investment, resourcing, and attention across stakeholders… and therefore a less contested outcome.

2.8.4 Stakeholder contributions and benefits

Building up from activities that impact on stakeholders to potential impacts and then potential legacies (impact being a time-bound and impermanent conceptual form of legacy), what measures could be used to report on the impact of any event to the key stakeholders? To avoid selective representation of an event, the model would need to have measures embedded in specific dimensions. The expectations of performance on measures should be seen in the context of the scale of the event, but this improves on the selective reporting to stakeholders (best bits only) or reliance on anecdotal outcomes. Neither provides a basis for a consistent rational assessment of events or the use of public funds.

Being able to establish the measures relevant to an event (or at least the weighting of importance to a specific event/host pair), then applying a scoring model of past events for which that data is available could be used to satisfy the identified stakeholders. The addition of a generic model that enables and weights input across different stakeholders could be used to screen the targeting of events, as many government agencies already undertake, but also provide connections to evidence from past editions to ground expectations of any one event's potential impact.

Event 'outcome scenarios' developed before an event would enable the creation of an 'impact trade-off' model. A well-defined and socialised set of scenarios established at or before event bidding would consider and account for differing stakeholder priorities within different circumstances and conditions. That process would empower event management (agents) to act upon in a way consistent with the resolved priorities and execute against a pre-planned scenario towards a revised range of outcomes. Then shifts in resources and priorities in light of a change of circumstances would be implemented in a structured manner as projected circumstances emerge, and hence avoid the 'uncomfortable knowledge' trap of inaction identified by Stewart and Rayner (2015).

Table 4 - Event stakeholder typology

ID	Stakeholder group	Attributes held	Attributes missing	Event stakeholders

1	Dormant	Power	Legitimacy + Urgency	Advisors, planners
2	Discretionary	Legitimacy	Power + Urgency	Residents, tax-payers, academics
3	Demanding	Urgency	Power + Legitimacy	Business, Suppliers
4	Dominant	Power + Legitimacy	Urgency	Event Owner, City as host
5	Dependent	Legitimacy + Urgency	Power	Athletes, competitors
6	Dangerous	Power + Urgency	Legitimacy	Politicians
7	Definitive	Power + Legitimacy + Urgency	-	Organising committee
8	Non-stakeholder	None	All	-

Scale dimensions of stakeholder engagement

The role of the resident and citizen is not fixed within the stakeholder mapping but is dynamic based on scale and perceived potential risks. Smaller-scale events where the government is only providing partial support or funding for events (and hence pose limited risk to the economy) are less likely to either seek or hear the resident's views before bidding for the event. Power is low even if legitimacy may be considered to be valid and hence classified as "discretionary stakeholders".

However, as the scale of the event increases, both the power of residents to influence decisions (for example votes against bidding for mega-event hosting) and their legitimacy as potential bearers of risk (for example the Montreal installation debt) also increases. Thus, moving residents from "discretionary" to "dominant" stakeholders in the bidding and hosting of events.

Temporal dimensions of stakeholder engagement

Stakeholder roles and influence can shift over time even within the lifetime of a single event hosting. The risks that organisers of events face in trying to elicit support for events is as noted by (Rocha, 2020) that "longitudinal findings show that, to gain support, organizers promise unattainable legacies, which then lead to dwindling support as they fail to deliver them." (p. 143). That failure to deliver the legacies of an event – especially if of significant scale – can impact support for future but unrelated smaller scale events by the same host. Hence some stakeholders who might have no or limited power or legitimacy before a failed event may achieve greater influence in future decisions by the poor delivery of prior events.

That can also be true where failure to realise legacies from previous events by other hosts empowers residents to prevent their cities from bidding for those events, based on the perceived risks compared to the benefits they believe will, or can be, achieved.

Stakeholders and event assessment

As noted above, the role and perspective of any stakeholder group in relation to an event cannot be assumed to be fixed, but can shift with the scale of the event being hosted and over the pre, hosting and post event timeframe. Hence stakeholder representation in legacy studies needs to be carefully defined to avoid simplistic treatment of stakeholder views across events, even by the same host. This is a caution that Koenigstorfer et al. (2019) also noted, concluding that specific and explicitly defined stakeholders are needed to "understand how event-attributed changes in structures and consequences are perceived and evaluated by different stakeholder groups."(p. 740). Given the dynamic roles stakeholders occupy in events, in assessing event outcomes, a clear and explicit identification of which stakeholder groups are represented in the assessment of outcomes is vital.

2.9 Model for all stakeholders - Zero-base legacy

Challenging the dominance of major events and their influence on the dimensions and perspective of current legacy frameworks by including recurrent events requires two fundamental shifts in outcome assessment.

The first is moving away from a binary classification of the 'legacy' or 'no-legacy' perspective that inherently embeds the presumption of legacy and the attendant dimensions that terminology carries. Instead, a more graduated viewpoint on the degree to which events of differing scales might contribute to outcomes and have an impact needs to start from the primary basis of a presumption of no legacy as its starting point, rather than the assumption that a legacy must exist which results in the creation of increasingly convoluted frameworks to justify an untenable position.

The second shift would be to the timeframe over which the outcomes can be assessed. The very extended timeframes afforded to legacies within existing frameworks make no account for recurrent events. That is, applying a multi-year time frame for the assessment of legacy for annually recurrent events would create overlapping periods. The assessment of the impact of any single event would therefore be confounded with prior and later years.

Adopting a ground-up or 'zero-based' (Chan, 2008) approach resets the understanding of legacy to start with the assumption of 'zero legacy' as the base case. Under this approach, the legacy of an event could be assessed as being accumulated in graduated increments in line with the scale and nature of the event's impacts on its host. The 'successful outcome' would therefore be relative to the expectations set in planning for and bidding for the event, rather than a subjective assessment or revisionist version of 'mission accomplished' claimed post hoc.

Fundamental to this 'zero-based' approach is an acknowledgement that theoretically an event can deliver 'zero legacy'. That acknowledgment runs counter to the rationale proffered on the role of legacy in offsetting costs against the shortfall in economic benefits, for why would a government ever invest public funds in hosting an event without at least the prospect of a legacy?

The answer to that cuts to the core of the narrative that has been built around the presumption that both a legacy is required and that shorter-term outcomes associated with

terminology such as 'impact' (and so dismissively framed as a negative outcome by the IOC) are therefore of little value.

The prior examples of the individual edition of a national/global sporting code, or a smaller scale event such as the aforementioned 'Pier to Pub' swim, serve as exemplars in considering a 'zero legacy' event and its contribution. In both cases, it could be argued that the only directly attributable impact is a temporary increase in the economic/business activity of local businesses and the event organisers through event attendance and ancillary activities by media around the event. This viewpoint accords more closely with the originating 'visitor economics' impact studies before their extension to include benefits that increase the value and justify the ever-increasing cost of major events (Davies et al., 2013). The impact of these 'zero legacy' events would not be assessed as some incremental contribution to the sporting infrastructure ascribed to the edition on a depreciation schedule. Instead, their contribution comes through direct activity such as event ticket purchases and incremental gains in visitors from outside the city to add new income into the local economy. Hence whilst these events deliver an identifiable impact, they fall outside the 'legacy' frameworks as described by much of the current literature, given their naturally short-term and recurrent nature.

Critical to the 'zero-based legacy or 'event impact' approach is a conceptual framework of event outcomes that would apply across events and that use the same measures across each event. While previously the pejorative use of the term 'impact' limited the expectations of what events can achieve, applying it to reduce the timeframe over which outcomes might be considered valuable allows for more events to be assessed. It thus avoids a bias towards large scale events where legacy expectations are both raised and required as a result of both the principal and agent incentives.

An alternate framework that adopts a zero-based approach would need to address the measurement of event activity outcomes in incremental levels across multiple categories or dimensions. The critical difference between this approach and the approach taken in the assessment of individual events is that the value ascribed to those activities would not need to be prescriptive. Critically each event host or owner could ascribe an extended or multiplier value to each quantum of a single activity without disrupting the consistency of the measurement underpinning the event assessment framework.

This approach is analogous to the current operation of the stock market with all investors having the same information but prepared to ascribe, at times, to significantly different values to earnings of companies in the same sector and category based on their future expectations of different companies. This is reflected in competing companies showing significant differences in their price/earnings (P/E) ratio despite operating in the same domestic market and under the same regulations (Shen, 2000).

Separating value from activity would allow a consistent zero-based legacy assessment framework to exist alongside tailored legacy assessment reports and economic impact analyses, and whilst the application of the measures remains open to misuse for political ends as noted by Crompton (2006), it would at least establish a clear line-of-sight to the underlying numbers across events rather than overreliance on the 'applied' results.

2.10 Classification of existing legacy approaches

To understand the context in which the literature is grounded, adopting a higher-level view of the literature reveals a different type of commonality in the approaches to event outcomes and assessment. Dinnie (2004) describes the value of this contextual overview as "a critical appraisal of the existing research on a topic. By classifying and evaluating the extant

knowledge base of a particular area [it] not only delineates the major themes and issues in the field but also identifies and develops avenues for future research"(p. 166).

Hence rather than classifying work by event assessed, research area or specific approaches to event evaluation looking at the context in which the studies are applied reveals four main classes based on their focus.

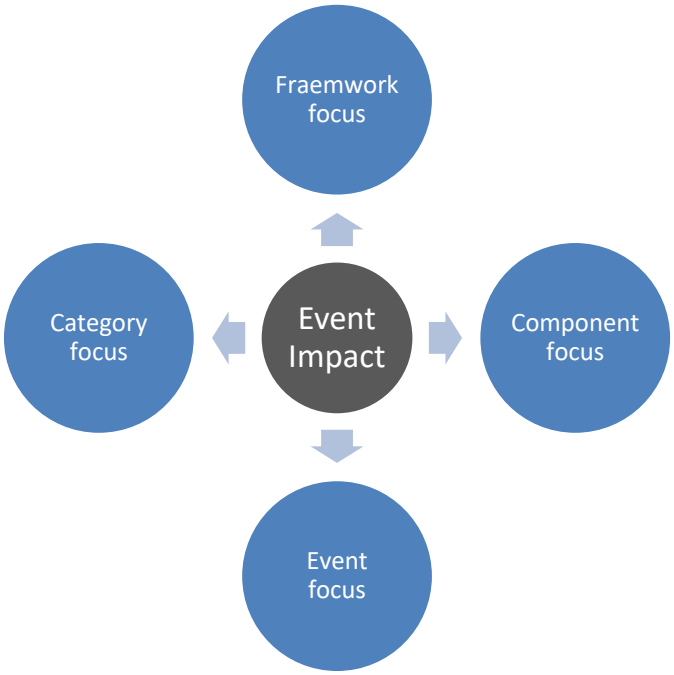

Figure 5 - Classification of existing legacy approaches

As noted in Figure 5 above, many of the studies reviewed in the literature appear to be able to be classified within one of four approaches to the consideration of impact of events. Namely:

- *Framework focus: Broad legacy frameworks that concentrate on trying to establish a comprehensive assessment of event legacy, proposing very extended timeframes of focus, fluid measures and primarily concentrate on mega-events, with limited application to smaller events*
- *Category focus: Frameworks that seek to include multiple areas of impact, but do not define measures needed to assess impact nor weighting between areas*
- *Component focus: Frameworks of event impact within a single area that may propose some measures but do not ascribe weight of importance to that area*
- *Event focus: Frameworks that provide extensive measures for assessing event impact but are locally tailored to a specific event and therefore lack comparability across events*

A brief review establishes a clearer understanding of the differences between each approach and their limitations in establishing a comparative model of event impact.

Legacy focus

The terminology of recent years for the associated and myriad of impacts the hosting of events can have has morphed to one of 'legacy' but without any agreed definition of either the term itself or its constituent components (Chappelet, 2012; Gratton & Preuss, 2008).

This lack of definition and consistency highlights the need for a new approach to this field of research.

Contributing to the confusion is that use of the term 'legacy' itself to describe the attendant benefits accrued to event hosts is both poorly defined and linguistically ambiguous, with inconsistent equivalents of the term in non-English languages (Thomson et al., 2013). In defining legacy as "any development that was created as part of the preparations for staging the [event], even if there is evidence that the development may have emerged in the fullness of time irrespective of the event" Essex and Chalkley (2003a, p. 95) acknowledge the tenuous nature of some claims in the legacy associations for events. Gratton and Preuss (2008) contend that legacy comprises "planned and unplanned, positive and negative, intangible and tangible structures created through a sport event that remain after the event"(p. 1924), which acknowledges the potential for positive and negative outcomes rather than a single 'positive-only' presumption. Dwyer et al. (2000) also noted the need for positive and negative outcomes from events.

However, the inclusion of 'unplanned' dimensions of legacy being able to be claimed post-hoc but unable to be anticipated when bidding for and committing to the cost of hosting the events undermines the confidence to consider a consistent framework for assessing and measuring event outcomes. Indeed, without a clear definition of legacy and the time over which attendant benefits of the event's hosting can be ascribed, the opportunity for confusion in determining a measure of an event within a host's event portfolio (Preuss, 2007) increases also. Attempts to define the dimensions of legacy have been proposed, for example, the five legacies proposed by Chappelet (2012): sporting, urban, infrastructural, economic and social. These contrast with the review of legacies models and the development of the authors own prior work as Preuss (2015) seeks to establish an alternate framework that presents the critical components of legacy to be infrastructure, knowledge, policy, networks, and emotions.

Amongst this discourse, even the most established and researched component of legacy frameworks (economic value of events), the assessment of economic impact from direct visitor spending shows concerns over methods of assessment and as noted "the methods used to research the economic impact of major events remain contested". (Davies et al., 2013, p. 31). This lack of agreement on the most substantively researched area of legacy exemplifies the similar lack of consensus over the measurement and contribution of the other impacts or legacies of events observed by Dickson et al. (2011) and that "not as much rigorous research has been done as might be expected given the claims made" (Horne & Manzenreiter, 2006, p. 11). That the current project determined to use the word impact in its title and avoid legacy is therefore not a semantic, but a deliberate choice of terminology. Impact connotes a time and event-bound constraint under which event benefits might be assessed and avoids resorting to claims of extended or 'uncoupled' benefits. In addition, it removes the permanence and lasting requirement from the dominant definitions of legacy that smaller events cannot expect to achieve.

By limiting the factors in the proposed framework to those directly attributable to a single event, and measurable as such, the approach avoids two key limitations discussed above from many of the prior legacy-focused models:

- *Over-reaching and conflating of event legacies with other potential effects in the intervening period in which an effect is being claimed.*
- *Smaller events that by nature of their scale do not require new infrastructure build can be considered against larger-scale events on a comparable and consistent basis.*

By the use of the term impact instead of legacy, smaller events avoid the pejorative labelling that legacy and its unconscious requirement to deliver a "permanent alteration to the urban environment" (Hiller, 2000, p. 440) on which they cannot compete. The continued dominance of legacy as the default viewpoint – whatever its final definition – may unconsciously be excluding smaller events unable to provide evidence of the required permanent improvements to find their place in the legacy narrative. However ill-defined, its pre-eminence in the literature may unreasonably be devaluing smaller events instead of providing them with a structured and comparative context in which they can rightly be assessed for their merit and benefit on a scale of similar-sized events, costs, and benefits.

Category focus

Despite progress in the specific dimensions of legacy – especially in the economic assessment of the impact of events (Preuss, 2005), there remains little consensus as to the relative importance of the different areas of event assessment in the wider context.

Reviews and discussions of different conceptual models (Chappelet, 2012; Cornelissen et al., 2011; Hiller, 2000; Preuss, 2015; Taylor & Edmondson, 2007) for legacy have critiqued and evaluated the merit and limitation of alternate components of legacy such as economic, social or community, environmental, political/policy, infrastructure, knowledge, health, and media.

Not readily evident from the reviews and alternate definitions of legacy, however, is the relative scaling of the benefits accrued to the event host across the individual components of legacy. Hence embedded within most proposed models of legacy there remains either an implicit assumption of equivalence or an undefined balance from where benefits might be considered to have been derived.

In working to provide a perspective on the importance of differing legacy elements (Mair & Whitford, 2013a) established that not all areas were indeed considered to be of equivalent importance by those researching in this field. Whilst providing some tangible evidence of differing scales of importance, the research does not address whether areas achieve ratings of higher perceived importance as a function of the area being previously under-researched or because of its perceived importance in benefit accruing to the host.

Hence this research also seeks to challenge the legacy development approach taken to date, whereby alternate models of legacy are prosecuted without explicit reference to the relative strength or importance of their constituent elements.

Single component focus

Some studies have focused on a particular type of event impact, which, as noted earlier, most frequently has been the investigation of economic impact. Unsurprisingly, this component of event impact has achieved so much attention, as costs in hosting events have risen. Hence hosts seek to establish the economic investment has been justified based on calculated economic return.

However, given the exploration of the lack of consensus on overall legacy in assessing event impact, scholars have progressed specific fields of research within the wider event studies.

Media holds a critical role in the promotion and as a source of revenue for events, hence an evaluation index for media performance is valuable (Zhao, 2012). It is also important to decode the content of that media in understanding if the promotional objectives of the host

are being achieved (Oldenboom, 2008), and if it leads to shifts in perception about the host (Knott et al., 2013) as an event legacy.

The impact of events on the local population has also drawn significant attention including investigations into the event legacy on sports participation (Reis et al., 2013), resident's 'quality of life' (Kaplanidou et al., 2013), development of community cohesion and social capital (Gibson et al., 2014), and connection between locals and the quality of the environment after the event (Collins et al., 2007). The contribution of each of these studies is a building of understanding of some of the ways each of these types of impact might be measured. However, they do not address how to combine the various types of event impact into a comprehensive assessment model or how to weigh the various components of event impact.

Specific event focus

Of the mega-events that dominate the sporting event landscape, the Olympic Games has been singularly successful in extending that dominance to outside the stadium and into research thinking and frameworks. An achievement aptly captured by Girginov and Hills (2008) in their expression "The Olympic Legacy Enterprise" (p. 2092)

During the development of the original Olympic Games Global Impact (OGGI) framework, academics created a preliminary list of 600+ measures before reducing it to 150 measures of impact (IOC, 2006). The limitation of this original model was in allowing the host to choose the measures that they feel best suit their event at that time. Whilst this offers a great deal of flexibility, it is at odds with the aim of this project to define a model that is consistent and repeatable and entirely comparable across different events held at any point in time.

The revised framework of the Olympic Games Impact (OGI) study now uses a "prescribed set of indicators to measure impacts across three topic areas or spheres of sustainability – socio-cultural, economic, and environmental" (UBC, 2013, p. 5); hence increasing the opportunity for comparability of editions of the same events. However, the additional requirement to measure these over 12 years in four reports has little relevance for any event and host outside a mega-event. That timeframe for assessment accords with the notion of permanence that is evident in the 'legacy' rhetoric. Indeed despite the word 'impact' in the title of the recent study, the International Olympic Committee (IOC) extends a bias to 'legacy' rather than 'impact' within its research by claiming that 'impact' has an inherently negative bias "more often regarded as implying an adverse effect or a damaging or destructive result" (IOC, 2015, p. 4) and embedding in the Olympic Charter "to promote a positive legacy from the Olympic Games to the host cities and host countries"(IOC, 2020, p. 17)

Finally, unstated but embedded within the frameworks that promote a single event focus are the benefits from a successful event that can also be accrued by the event owner. Tightly defining the terms of reference for assessing the event's impact may benefit the host but also accrues incremental value to the owner (Barney, 2003) across successful edition to successful edition. Differences in value to the event owner compared with the host will be considered as part of the analysis phase of this research project.

2.11 Framework structure for event assessment

The review of the literature demonstrates that the 'catch-all' definition of legacy has extended boundaries of what may or may not be included within the classification of event outcomes. Whether driven to justify increasing costs or expanded expectations of what

events should achieve for their investment, the definition of legacy has not been resolved and indeed remains 'contested' as a concept. While the degree to which this is due to a focus of legacy research on mega-events is not explicit, attempts to reframe outcomes to more direct and shorter-term impact have been resisted. Yet it is in reducing the 'legacy' burden on events that may enable it to achieve a degree 'less contestedness' and expand to include a wider range of events.

Indeed, emerging from the review are the criteria under which a framework needs to work to be more generally applicable. The four conditions identified in the review to address the missing elements of the legacy approaches to date are:

- *Applicable to events of different scales and types*
- *Address key areas of event impact within a single framework*
- *Define relative weightings for differing areas of impact*
- *Define measures that are relevant, important, and reliable.*

While the above conditions provide a solid basis for the development of a comparative framework under a structured approach to legacy, it is important to note that this approach may also exclude more subjective legacy elements from being included in the legacy dimensions being researched.

Having identified the constraints grounded in literature, the development of a framework structure that provides both consistency and flexibility is considered in the next chapter.

3 Development of the Conceptual Framework

While building on the domain knowledge already established, the development of a conceptual framework for assessing event impact needs to address the structural underpinning identified in the literature review before moving into substance and content with which it will be populated. To be enduring and widely applicable, it must also seek to capture the domain knowledge already established and remain grounded in it, but also allow for new research and tools to emerge in the discipline of event assessment.

This chapter considers the structural elements of the framework and the extent to which it may address issues identified in the previous chapter before moving into the content of the framework in the next chapter.

3.1 Development of the Event Impact Framework (EIF)

The articulation of the ambition to reduce 'contestedness' and to develop a more widely applicable and consistently applied framework requires four elements, hereinafter referred to as the Event Impact Framework (EIF) or framework. The four elements are as follows:

- *The first is that the framework needs to reflect the existing research and its representation of event impact integrates the measures of impact identified previously. That is, it remains grounded in the existing knowledge but seeks to build on it.*
- *The second is that in extending the framework to address a wider range of events than the dominant mega-events, it captures areas that may be underrepresented or even missing in the existing literature.*
- *Thirdly, the framework needs to be able to capture sufficient detail to be actionable and relevant; but not be burdened to the point of being so unwieldy as to not be able to be applied to a wider range of events (especially smaller events).*
- *Fourth is that the framework must be able to be consistently applied over time; but flexible in design to allow for the emergence of new dimensions and/or the ability to measure existing but underrepresented dimensions without disrupting the core model.*

3.2 Conceptual framework dimensions

Within a comparative framework of event impact, several dimensions are core to defining the scope and structural choices/constraints of the framework design. The key dimensions that the framework needs to address to allow for the comparison of outcomes across events are outlined in Table 5.

Table 5 - Conceptual framework dimensions

No.	Dimension	Role in Event Impact Framework
1	Scale of events	Ability to define events by a comparative scale and therefore the types of sporting events the framework is meant to address
2	Assessment dimensions	That the design allows for both depth (specificity) and breadth (universality) of the areas included within the framework.

3	Specification of measures	Determination of how measures are considered for inclusion in the framework and likewise those that might be excluded on an objective and consistent basis.
4	Weighting of dimensions	The relative weighting of impact elements has been unaddressed or underdeveloped in prior models but these must be explicit in comparing events of differing scales.
5	Timeframe of impact	The timeframe over which the impact assessment would be considered, which is relevant across events of differing scale, duration, and annual vs. one-off.

3.2.1 Defining events by comparative scale

As noted above, the Event Impact Framework (EIF) seeks to remain grounded and build upon existing domain knowledge and practice. A typology for the classification of a scale of events was developed by researchers at Sheffield Hallam University (Gratton et al., 2000), initially classifying events into a four-level scale (A-D). That model was later extended to add a Type E to recognise the impact of smaller local events (Wilson, 2006) and used as a five-level typology (Ramchandani, 2014) known as the SIRC event typology. This typology is presented in Table 6 below.

Table 6 - Sheffield Hallam SIRC Event Typology

MEGA	SUB MEGA		NON MEGA	
	National / International		Domestic	
Type A	Type B	Type C	Type D	Type E
Irregular, one-off international spectator events generating significant economic activity and media interest (e.g. Olympics, Football World Cup)	Major spectator events, generating significant economic activity, media interest and part of an annual domestic cycle (e.g. Open Golf, Wimbledon)	Irregular one-off major spectator/ competitor events generating limited economic activity (e.g. Grand Prix Athletics, World Badminton Championships)	Major competitive events generating little economic activity and part of an annual cycle of events (e.g. National Championships in most sports)	Minor competitor/spectator events, generating very little economic activity, no media interest and part of an annual domestic cycle of sport events. (e.g. local and regional events)

Using the SIRC established typology within the Event Impact Framework avoids the development of a new classification of events and grounds it within the existing literature and current researchers in the field. It is noteworthy that this approach to classifying the scale of events by alpha characters A - E is not unique to academic research only, but has been adopted in practice, albeit using a variation of the SIRC model gradings. For example, the City of Toronto in its "City of Toronto Standard Definitions for Special Events" (Deneau, 2016) initially used the A-E classification for its event classification, and in 2015,

Toronto's Mayor's Advisory Panel on International Hosting Opportunities added another grading to recognise the complexity and scale of "Mega-events" at a level above A in their typology (p. 14). In terms of scope for the framework the EIF should be sufficiently adaptive to apply to events across SIRC Types A – D at a minimum, and ideally, to be extended to Type E events in time.

3.2.2 Assessment dimensions

Core areas of event impact assessment have been built over decades, emerging from the early work of Ritchie (1984) on the assessment of 'hallmark events' across "economic, tourism/commercial, physical, socio-cultural, psychological, and political dimensions" (p. 4) dimensions. Building on that work, Burns et al. (1986) also included tourism and the impact on economic factors, breaking them into separate dimensions of 'transport, accommodation, restaurants, business' and also adding specific elements such as 'road accidents'.

The intervening years since those foundation studies have seen core areas developed and a myriad of sub-elements now contributing to the relevant research and knowledge with increasingly specific dimensions. Whilst by no means exhaustive, subsequent researchers have tended to address similar themes when considering event legacy, namely economic, sport, social, media, environmental, political, and reputational dimensions.

Hence given the breadth of subjects and impacts researched about events and their legacy, the framework elements for assessment of impact require that the dimensions:

- *Need to be a reasonable representation of knowledge established in the field*
- *Need to have sufficient breadth to address the core dimensions commonly assessed and used in studies of events of differing scale*
- *Represents the breadth of dimensions within a framework, which requires a means of grouping in a hierarchical structure for ease of comprehension and scope*

In seeking a balance between the depth (specificity) and breadth (universality) of the areas included within the framework, the framework cannot operate within a single dimension of event impact but needs to address multiple dimensions. A hierarchical approach that allows for flexibility across and within the framework's key dimensions should allow for differing events and their relative strengths to be included.

In a hierarchy, the top level of the framework would represent the foundational dimensions (pillars) of event impact. A more detailed and direct sub-area (drivers) should allow for both breadth of dimensions of impact to be considered and for depth into the specific elements that need to and can be measured (indicators), resulting in a consistent assessment of impact.

3.2.3 Specification of measures

A critical aspect of a measurement framework is defining which measures might be included and the principles by which a longer list of impact assessment metrics might be refined on an objective and consistent basis.

Measuring both Importance and Reliability

The assumption that a measure is important to event impact assessment is not immediately equitable or reliable. Availability, convention, and self-interest can skew the process of

measure selection towards existing measures without challenging their reliability as measures of event outcomes.

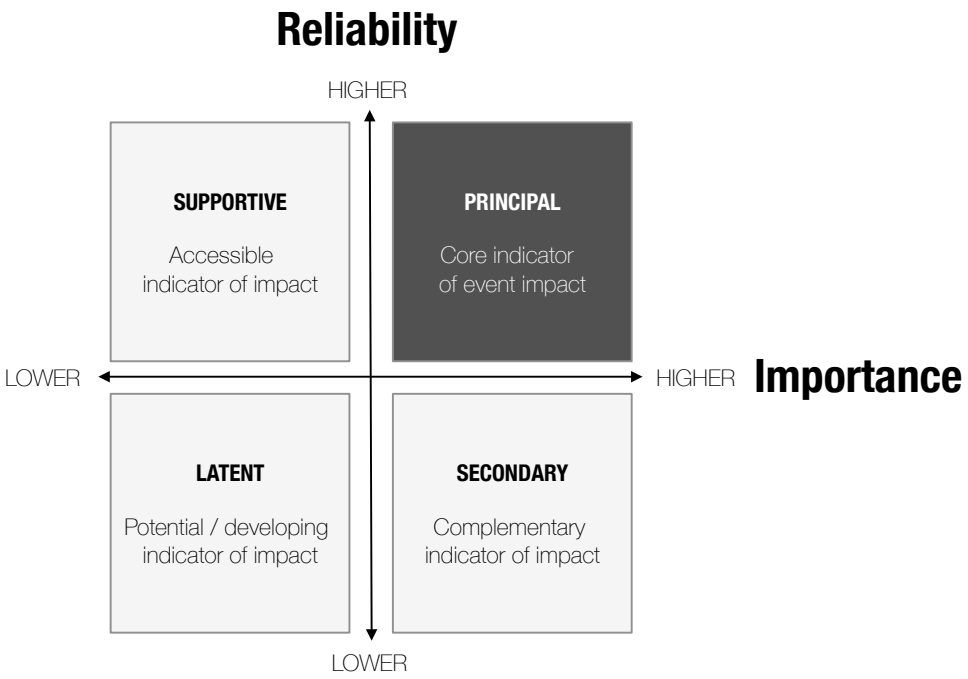

Figure 6 – Framing measure dimensions

As shown in Figure 6, assessing measures on dimensions of importance and reliability highlights the differences in the quality of the measure underpinning an event assessment. While the terminology and the descriptions of the quadrants were not included in the research and not provided to respondents, it proved useful in seeking to ensure the measures considered under each of the key dimensions were not biased by perceptions of convenience and availability.

Measure inclusion criteria

The following criteria could be applied consistently across all current and future indicators. The basis on which the final list of indicators is presented to respondents would be included in the survey to address concerns and ensure the indicators are rated on a common basis of understanding. This will acknowledge where the indicators were identified in the literature review but excluded from the framework list based on filtering criteria, for example, 'direct measures vs. percentage'.

Whilst many of the metrics identified in papers were expressed in percentages, these are also often duplicated with direct measures as counts for the same metric. For example, the breakdown of visitors by origin (domestic, national & international) is measured in direct counts and then converted to percentages. As visitor numbers are the primary metric of interest and the percentages are derived from direct counts, they are in essence duplicate measures. It is recommended to only include the direct measures of these metrics in the Indicator list.

An exception to this approach would be where a value is by convention provided as a percentage rather than a direct measure. Such as GDP (growth), unemployment (rate) etc.

Measure exclusion criteria

Measures of outcomes that rely on multi-year/decade induced effect were excluded from the proposed EIF. Their exclusion was determined based on three dimensions. The first is that these effects are predominantly only relevant to a very small subset of mega-events, while the EIF sought to address the ability to include a wider range of events. The second is to be able to assess the impact of any recurrent event without conflating its impact with either prior or subsequent editions. Finally, avoidance of conflating the event effect with other initiatives the host may undertake in the years post-event.

The weighting of legacy dimensions

The weighting of impact dimensions has been underdeveloped within existing models. As researchers have struggled to define 'legacy' and achieve agreement on its component elements, the issue of the relative weighting or importance of each of the elements has been unaddressed or underdeveloped.

The opportunity for a more prescriptive approach to event assessment allows for the establishment of consensus ascription of weights to different dimensions. A few of these exist in the literature but often only within a single jurisdiction or specific to a single event agency.

Availability bias in measurement

An implicit (and unconscious) assumption could be that the volume of attention on a given field of event research is reflective of its importance in event impact. This would be a rational but erroneous judgement on two bases.

Firstly, as noted previously, for the principal/agent relationships between event owner and host, and host and stakeholders, there are strong self-interest incentives to fund and guide research priorities to some dimensions over others.

Secondly, academia is built on the foundations of previous work and therefore is more likely to be representative of past attention rather than reflective of emerging research on the dynamic value of event impact within differing host environments, scales, and types.

Therefore, whilst it is critical to work from the literature to understand how the areas are currently located and how to define the components of each, it is vital to acknowledge that some areas may be either underrepresented or under researched in the literature as compared to the current and future state of event impact.

Timeframe of impact

The assessment of event impact on a decade or even multi-decade timeframes to determine legacy impedes the development of a comparative framework. A shorter timeframe enables the comparison of more frequent events to those on a quadrennial cycle.

As noted in the criteria for the selection and inclusion of assessment measures, each measure must be identifiably related to the event itself. Ideally, all events would be constrained to the year in which they are run, as this would place all events within a common time-bound level irrespective of the scale of the event within the SIRC typology.

It is noted however that the large scale multi-sport events require significant infrastructure and development over multiple years and has contributed to the extended timeframe within existing legacy models. An approach to recognising this extended development phase and yet maintaining a shorter impact time horizon on assessing event outcomes might be

resolved by accumulating the lead-up event impacts to be recognised in the year in which the event is hosted.

Likewise, proposals to include gains in GDP over extended periods (7-15 years) following the event pose both concerns for direct attribution to the event and the inability to conclude the event's outcomes within multiple quadrennial cycles. Hence, heavily extended timeframes that relate primarily to 'Type A' events found within the literature will not be included in the Event Impact Framework.

3.3 The emergence of key dimensions of a conceptual framework

The literature is replete with investigations of event impacts and includes numerous frameworks for capturing and assessing event legacy.

Table 7 - Comparison of research focus and legacy dimensions

	Legacy dimensions				
Research frameworks	Event size addressed	Impact dimensions	Measures specified	Weighting of dimensions	Timeframe of impact
1. Legacy framework	Mega-events	Multiple	-	-	Decades
2. Category framework	Multiple	Multiple	-	-	-
3. Single component framework	Multiple	Single	✓	-	-
4. Event framework	Single	Multiple	✓	-	Decades
Event Impact Framework	Multiple	Multiple	✓	✓	1 year

A summary of how event assessments have been presented in the literature in comparison to the EIF is presented in Table 7.

The contribution that this research seeks to make is to establish the potential for a framework that builds on those existing perspectives, by combining the dimensions found individually within the existing frameworks, but not all in one framework. That is, it considers events of different size, includes multiple dimensions of legacy and specifies measures by which outcomes might be assessed. In an extension to existing frameworks, the EIF also seeks to have weightings of perceived importance ascribed to the outcomes – something not evident in other frameworks. In seeking to develop a more inclusive framework across events, the timeframe over which impact might be considered is constrained to a single year, rather than the extended timeframe applicable mostly to mega-events.

3.3.1 Event Impact Framework structure

Having addressed the first two framework criteria identified in 3.1, the choice of structure for a comparative assessment framework is critical to achieving the two final criteria, namely:

- *Thirdly the framework needs to be able to capture sufficient detail but not be burdened to the point of being so unwieldy as to not be able to be applied to a wider range of events (especially smaller events).*
- *Fourth is that framework is flexible enough to be consistent over time allowing for the emergence of new dimensions or the ability to measure existing but underrepresented dimensions without disrupting the core model.*

Three broad approaches for the structuring of the assessment dimensions and hence the measurement of impact were considered. The focus was on meeting the need to provide clarity about the top-level areas being addressed and the ability to apply measures underpinning those areas.

		Measurement of impact	
		Single measure per area	*Multiple measures per area*
Impact Areas	*Single level*	**Single**	**Nested (Measures)**
	Multiple levels	**Nested (Levels)**	**Tiered**

Figure 7 – Dimensions in structural approaches

Three types of structures reviewed were single, nested, and tiered as depicted in Figure 7. Reviewing examples of each of these structures identified their advantages and limitations in the application to event measurement.

Single level structure

The simplest model for an event framework would be to use a single 'level' where each assessment area is given equal prominence and has only a single measure that underpins it.

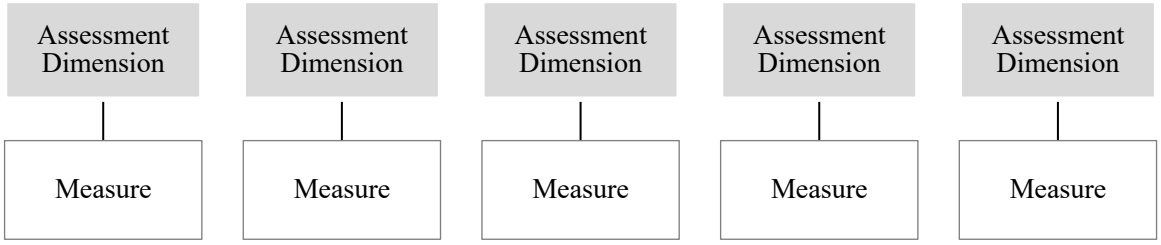

Figure 8 – Framework structure = Single Level

As shown in Figure 8, this is a simple but also simplistic viewpoint. The only method by which more measures can be added is through expansion of the dimensions at the top level.

Nested (Measures) structure

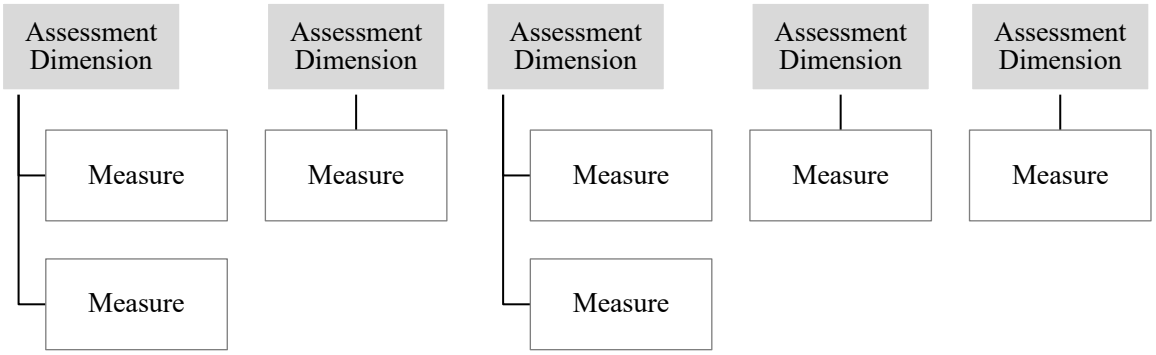

Figure 9 - Framework structure = Nested (Measures)

A nested structure (shown in Figure 9) advances on the single level structure by allowing for multiple measures within any single assessment area. While both are more comprehensive in the range of measures they can incorporate, the assessment areas remain on a single level and hence new dimensions of impact can only be included through expanding the top-level dimensions.

Nested (Levels) structure

The alternate version of a 'nested' structure groups the assessment dimensions into related areas rather than grouping the measures.

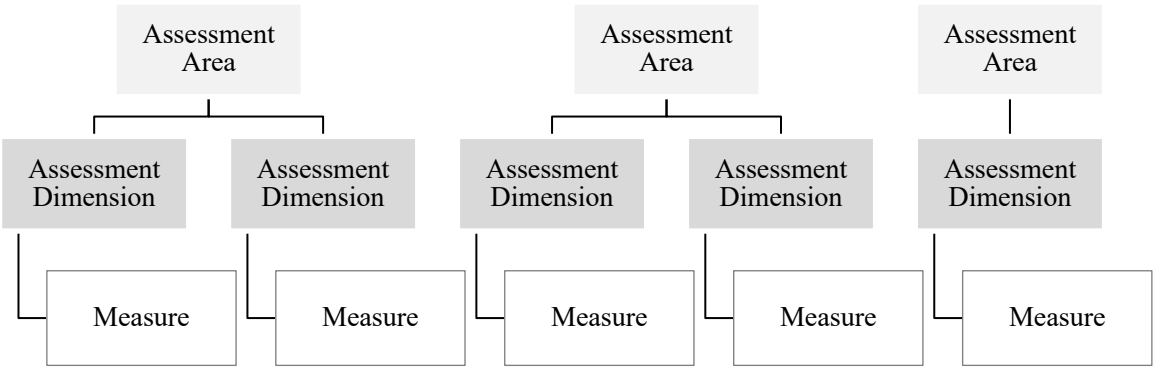

Figure 10 - Framework structure = Nested (Levels)

This 'dimension nested' version (shown in Figure 10) adds an additional layer of 'assessment area', which allows greater flexibility for grouping or expanding the assessment dimensions without requiring an expansion of the top-level structure. However, it does only allow for a single measure for each assessment criteria.

Tiered structure

The 'tiered' approach combined both versions of the nested structures into a single structure

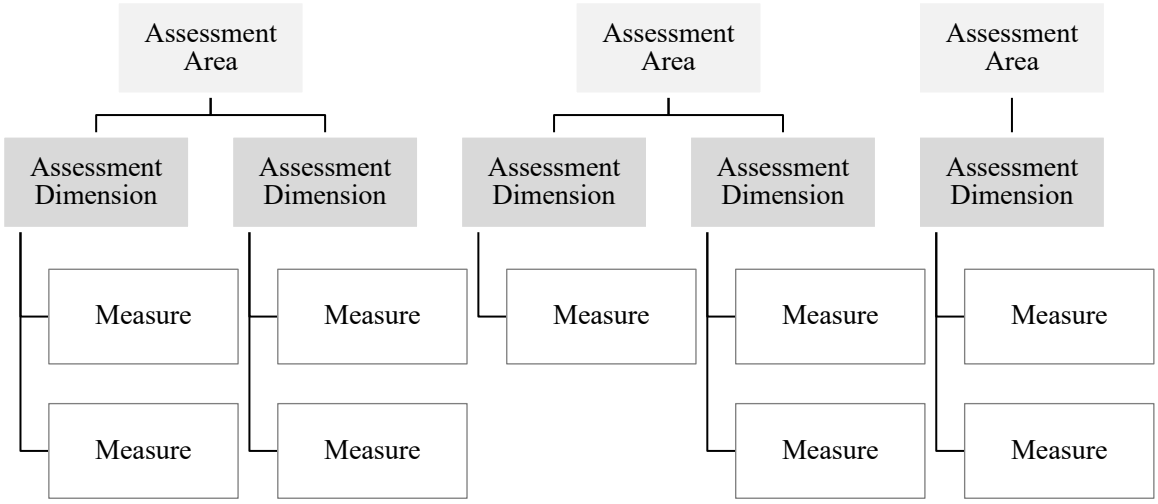

Figure 11 - Framework structure = Tiered

The 'tiered' structure as shown in Figure 11 achieves a balance in both the specificity of detail through the inclusion of multiple measures and the retention of the ability for the framework to be adapted, expanded, or refined over time within the key assessment areas.

This 'tiered' structure is particularly well suited to allow for the potential of underlying measures to be changed as data methods or model developments change, but without compromising the overall architecture of the model. That is, as better or improved measures of impact are developed, the framework is not disrupted but can be recalibrated and/or expanded.

		Measurement of impact	
		Single measure per area	*Multiple measures per area*
Impact Areas	*Single level*	**Single**	**Nested (Measures)**
	Multiple levels	**Nested (Levels)**	**Tiered**

Figure 12 - Target EIF structural choice

Hence, as highlighted in Figure 12, the target design for the Event Impact Framework was determined to be best fulfilled by the selection of the tiered structure approach. While the tiered terminology was not used within the research, simplified terminology was adopted to ensure communication was clear in delineating the levels of the framework.

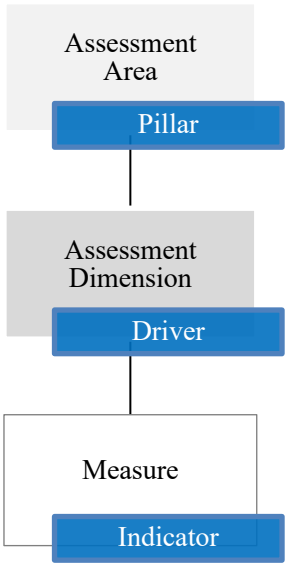

Figure 13 – EIF Tiered model terminology

The three terms used in the Event Impact Framework (as shown in Figure 13) to describe the three levels of the tiered model were as follows:

- *Assessment Areas = Pillars.*
  - *Major assessment area or theme that comprises multiple related dimensions*
- *Assessment Dimensions = Drivers*
  - *A dimension that focuses on a specific area of impact and contributes to the pillar*
- *Measure = Indicator*
  - *To enable respondents to consider traditional metrics and other areas such as social and intangible outcomes on an equivalent basis, the more general term of an indicator was used. There may be one or more within an assessment dimension.*

3.4 Legacy limitations

Subjective dimensions in legacy

A perceived challenge to the development of more consistent and objective measurements of event outcomes is the existing literature and models of legacy that include intangible and/or potentially subjective dimensions. An approach to resolve this would be to limit the breadth of dimensions to be addressed within the framework. However, that would not allow for comparison of events of differing scales, whereas as noted in the SIRC Event Typology, the outcomes shift in availability and importance when moving from global to domestic events. The unintended consequence of seeking to make the framework 'objective' would be to limit its applicability to only those events where the selected tangible dimensions are dominant and important enough to be measured. Alternatively, recognising the perceived limitations of social sciences to achieve reliable outcomes has been tackled already, and it does not require a limiting in the scope of the framework. Indeed Helmer and Rescher (1959) identified that "what matters is not whether or to what extent inexactitudes in procedures and predictive capability can eventually be removed; rather it is objectivity, i.e., the intersubjectivity of findings independent of any one person's intuitive judgment, which distinguishes science from intuitive guesswork however

brilliant"(p. 27). Helmer and Rescher's perspective is invaluable in that it highlights two critical aspects of the development of the Event Impact Framework.

The first is that the framework seeks to provide an objective method for the comparison of events – i.e. bringing more of 'science' from social science and less reliance on individual cases. The second is the avoidance of one person's intuitive judgment by ensuring the fieldwork brings together diverse viewpoints from as large a base as is practical. Building on those foundations, the accumulation of consistent datasets across events and time may allow for more predictive models to be built, not based on subjective or selective interpretations of legacy, but on a consistent and grounded set of measures already extant in the literature and applied in industry.

Assessment of negative outcomes

The development of a conceptual framework has been challenged by a question emerging from the literature on the nature of event legacy. As noted by Sant and Mason (2015) the term legacy is often associated with prospectively positive hosting outcomes "whereas negative legacies, such as overcrowding and environmental damage, are ignored by bid and event proponents."(p. 43). If an assessment framework is to capture the impact of events and is to be consistent and comparable, how should it handle the scoring of positive and negative impacts?

That an impact can have a positive and negative form was noted by Ritchie (1984), but in recent years more attention within the legacy discourse has been directed at an acknowledgement that event legacies can be negative as well as positive (Agha et al., 2012; Cashman, 2003; Thomson et al., 2013). If as described, a negative legacy is not merely the absence of achievement of planned positive outcomes, but rather a detrimental effect on the host (or sub-group within the host's stakeholders) then that is an outcome that can also be true of any long-term city development. That is, if event legacy is an acceleration of existing long-term plans as characterised by Agha et al. (2012), then the same negative outcome would also be derived from the same development projects undertaken with or without the event (Preuss, 2007). Hence the negative outcome cannot be considered to be special to the event, but rather linked to the development project itself and inherent in choices cities make in resources and development.

For instance, the development of a new dedicated train line to an airport from a city centre may already exist within a city's planning agenda, but its actual building is brought forward under the auspices of infrastructure needed to successfully host a major event. The underlying rationale for building the train line would be to make airline travellers' journeys more efficient and reduce congestion at peak times. The more successful the project is in attracting usage, the more it substitutes for services already provided in the market by reducing demand for existing transport options (for example taxis, buses, ride-shares) and their ability to generate an income. Hence the perspective on whether the train line development is a 'positive' or 'negative' legacy is not clearly delineated but rather dependent on the stakeholders impacted.

Critically, it is hard to identify any development that is without some negative impact on some stakeholders – even supposedly 'only positive' developments. If, for example, a government with finite resources chooses to invest in one area for critical improvement (building a new hospital), it then as a consequence must limit resources available for other positive or worthwhile initiatives (for example new schools).

This indirect negative impact has already been noted by (Siegfried & Zimbalist, 2006). The authors caution that where an event places a "public obligation for debt service, infrastructure maintenance, environmental remediation, incremental sanitation and security expense, possible cost overruns [would] generate a substantial budgetary problem [and as such] budgetary gaps must be filled either by cutting other government services or increasing taxes" (p. 423).

Given the binary outcome of positive or negative classification of a legacy proves insufficient, perhaps a more nuanced view of legacy would recognise the 19th Century writings of Jeremy Bentham and John Stuart Mill in the Utilitarianism logic of seeking the greatest happiness for the greatest number of people (Mill, 1859), accepting and acknowledging that universal positive good is rarely achievable.

Thus, the classification or conceptualisation of a legacy as being positive or negative without a clear and explicit statement of the viewpoint the stakeholders' are representing is at best overly-simplistic or arbitrary, and at worst deceptive by ignoring the differential nature of impacts across stakeholders. Hence a single outcome could be viewed as both positive and negative when moving from one stakeholder group to another.

Therefore, the determination to characterise legacies as being positive or negative cannot be made without acknowledging the stakeholder perspective implied in the classification.

It would therefore be better to measure the activity generated by the event and assess that against expectations for that event rather than simplistically seeking to categorise outcomes into positive and negative legacies, which is inherently subject to the perspective of those stakeholders viewing the outcome.

3.4.1 Legacy typology limitations

While drawing on the literature guided the structure of the EIF in resolving the general limitations noted, the Event Impact Framework would be most effective if it was also able to contribute to resolving specific limitations within existing approaches to legacy. Hence the table below considers how the proposed EIF as structured might apply against the legacy typology and their identified limitations.

Table 8 - Resolving legacy typology limitations

Typology Focus	Typology limitation	EIF approach to address limitation
1. Legacy framework	Frameworks that concentrate on trying to establish a comprehensive assessment of event legacy, proposing very extended timeframes of focus, fluid measures and primarily concentrate on mega-events, with limited or no application to smaller events	Reducing the scope of effect and timeframe over which outcomes are assessed to allow for assessment of events leveraging existing infrastructure, and/or annually recurrent events. Consistent dimensions for comparability of outcomes across editions and hosts.
2. Category framework	Frameworks that seek to include multiple areas of impact, but do not define measures needed to assess impact nor weighting between areas	Measures to be identified in the literature related to each area of impact and be tested for importance and reliability.
3. Single component framework	Frameworks of event impact within a single area that propose some	Explicit weightings of perceived importance in event assessment to be

	measures but do not ascribe weight of importance to that area	applied to each of the areas to establish relativity of different areas.
4. Event framework	Frameworks that provide extensive measures for assessing event impact but are locally tailored to a specific event and therefore lack comparability across events	Consistent measures and areas to be applied across events. Tailoring is in the performance on measures of each event, not by their exclusion from the framework.

The summary of key limitations for each of the four legacy typology areas is shown above in Table 8. As noted the consistency in the structure, the application of explicit weights and a reduced timeframe all contribute to allowing the EIF to reduce some of the identified concerns and expand the range of events to which it might be applied.

Applying content to the structure

It is foundational that the structure for the Event Impact Framework should provide a consistent hierarchy of elements, but that it also allows for future refinement and advancement. The tiered approach selected meets that criterion. In addition, the determination to shorten the time horizon for assessing outcomes is a deliberate choice for the framework to apply to a wider range of events.

As noted in the review of stakeholder incentives, the shortened timeframe does not imply effects cannot be longer-term, but rather that those should be considered the exception rather than the baseline. Likewise, the purpose of the EIF is to provide a consistent and comparative set of measures in assessing event outcomes. Any monetary value placed on those outcomes is for the stakeholders to determine in terms of their own needs, development goals, and funding models.

Finally, while is important that the framework structure seeks to address identified limitations within existing legacy approaches, the core value of the EIF is in determining the content that the framework should include at each level of that structure – in this case the Pillars, Drivers and Indicators. The following chapter seeks to identify those key dimensions, establish evidence of its application in the current literature around event outcomes and assessment, and compile potential measures for consideration in the Delphi study phase of the research.

4 Foundations of Framework Content

Building on the structure established in the Development of the Conceptual Framework, this chapter outlines the process for the initial testing of the major impact areas (Pillars) to be considered as the top-level anchors of the framework, then reviewing the sub-areas (Drivers) that comprise them. The review of Drivers is not intended as a comprehensive analysis of each of the 30 Drivers covered but instead highlights different aspects and viewpoints concerning event outcomes. Finally arising from the review of Pillars and Drivers are the measures (Indicators) that were found related to the Pillars from which the final selection for testing was made.

4.1 Development of core content

The process of determining the key areas for assessing events was undertaken as a multi-stage process… building from the initial literature review and engaging expert panels to a review of the outcomes on an iterative basis.

While the framework intended to provide a comparative basis for a wider range of events, the initial work was to determine the boundary scope of events to be included in the research.

Excluding non-sporting events

The addition of non-sporting events (such as festivals and cultural events), which are often included in the same research domain, would implicitly move the Event Impact Framework from a 'sports event centric' view to a more generic 'host centric' view of events.

Whilst considered, it was decided to limit the research scope of the model to include only sporting events, as the definition of legacy and its component elements needed to be constrained to maintain a common core within the already contested space. This approach would also eliminate the burden of finding and recruiting non-sport experts to the panel to ensure their views were fairly represented.

It was acknowledged that potentially expanding the Event Impact Framework footprint to include other types of events would be possible and even desirable in the future, as it would potentially increase its relevance to hosts and enable them to compare a wider range of events within their event portfolio.

4.1.1 Initial screening of impact areas

The development of the content for review in the framing stages of the Event Impact Framework started with mapping the Pillar framework into terminology for searching with the literature and establishing relevant papers for extraction. The identification and definition of key search terms for academic papers refined the testing of search terms and the citation and incident frequencies.

The development of collation protocols and recording of papers to ensure traceability of the initial measures found to be linked to key authors and papers. The process of manual review of all extracted papers to identify and map indicators into a preliminary dataset for consideration by the working group.

Testing structure and content within workshops

The process of refining and confirming the primary model dimensions included the participation and contributions from experts including academics, event practitioners and representatives from international sports federations.

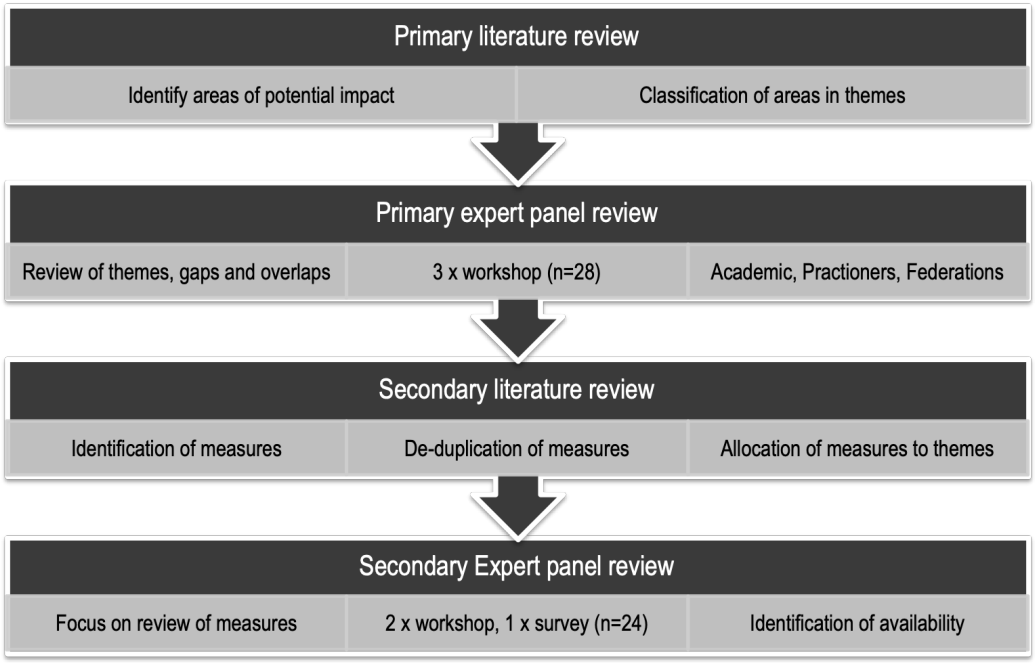

Figure 14 - Process of primary model determination

In the development phase of identifying the core dimensions of the framework as outlined in Figure 14, the term 'event outcomes' was used throughout the workshops and pre-testing survey. It was selected as a common term that was also agnostic as to whether the outcome should be classed as a legacy or an impact of an event.

4.2 Proposed framework model

4.2.1 Core emergent themes

From the workshops conducted, the emerging consistency of views was on seven broad areas of event impact – although noting that these also required clear definitions of their constituent elements to be accepted. There was also some diversity of views as to whether 'Tourism' was a pillar in its own right or a sub-element (Driver) of Economic impact.

Table 9 - Emerging themes of impact areas

No.	Theme	Description	Sub-elements
1.	Economic	The economic impact of an event to its host through direct economic contributions	Tourism, Infrastructure, Direct event, Security, Business activity, Sponsorship
2.	Media	The media impact of an event defined by the scale of media output, social media activity and media presence	Media output, Audience size, Sponsors output, Media presence, Media accessibility
3.	Sport	The sporting impact of an event through it its contribution to the	Sporting infrastructure, Event scale, Event prestige,

		development of the sport and future events	Reputation of the sport, Sport expansion,
4.	Social	The social impact of an event in terms of advancement of social development and well-being	Community cohesion, Volunteering, Goodwill/Pride, Social engagement, Health and Active living
5.	Brand	The branding impact of an event in terms of development of host destination image and event brand image	Host attention, Image building – Host & Event, Destination attractiveness, Host capability reputation
6.	Environment	Environmental impact of an event in terms of its environmental footprint	Energy efficiency, Consumption footprint, Resource utilization, Building design leadership
7.	Political	The political impact of an event in contribution to soft-power relations and host profile	-

Of the emerging themes identified, it was recognised that the most difficult to consistently or accurately measure is political. There exists no agreed definition or scale of measurement for political impacts and it tends only to relate to Type A or Type B events in the SIRC typology. Some commentary is provided at the end of the chapter but this dimension was excluded from the final framework due to the lack of both objectivity and visibility inherent in this dimension.

4.2.2 Interaction between areas of impact

Whilst the framework necessitates the determination of discrete areas of impact, the workshop discussions also noted the interrelationships between areas of impact as represented in Figure 15 below.

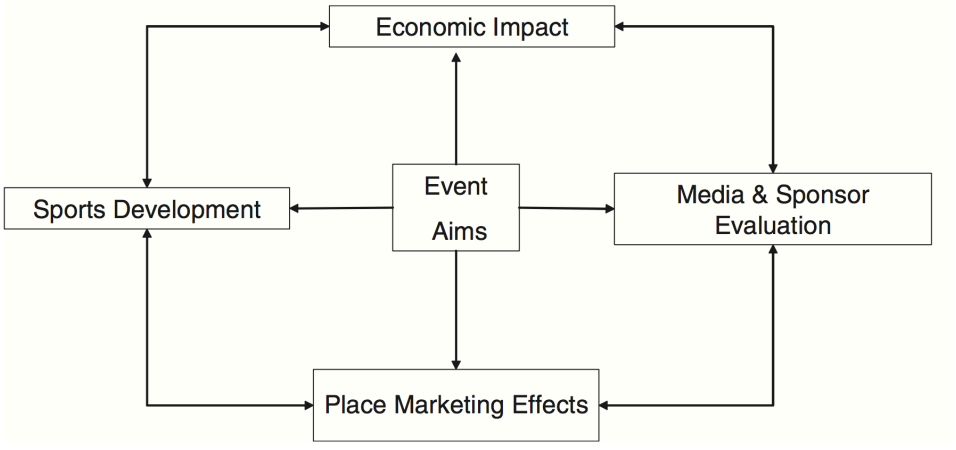

Figure 15 - Balanced Scorecard in evaluating events

That acknowledgement of inter-relationships reflects earlier work in reviewing reports of economic impact where Gratton et al. (2006b) highlighted the need for a 'balanced scorecard' view (Figure 15) of evaluating event impact. Noting that the scorecard shares four of the six final pillars that emerged from the workshops (missing Social and

Environment), it does affirm that maximum impact from events relies on achieving strongly in multiple areas… not a single dimension alone.

This reinforces that the framework should address multiple dimensions of event outcomes and the fieldwork panel needs to bring together a wide range of expertise to address the breadth of dimensions being covered.

4.3 Review of Pillar and Driver foundations as related to event impact

The following sections 4.4 to 4.10 undertakes a review of each major impact area (Pillars) and its sub-elements (Drivers) and the evidence for their role in assessing event impact.

Each section is structured as follows:

- The background about the Pillar and a listing of the Drivers within that Pillar. Where relevant, additional notes from the review process are also included where they provide guidance and/or clarification of the scope of the pillar.
- A short review of each of the drivers and research on how they relate to event impact. A comprehensive literature review and how they interrelate to relate to each other is not attempted but where pertinent, relationships between drivers is noted.
- A listing of the potential measures found within the literature and practitioners' models that were considered for inclusion in the Delphi study for rating by respondents. The final list of measures rated is found in 6.7 "Indicator Ratings".

The review sought to identify evidence for pillars and drivers within the current literature and is not presumptive in assigning any weighting of effect in their identification.

4.4 Pillar - Economic impact

Economic impact was defined for the purposes of this study as the 'economic scale of an event through direct economic contributions'. The importance of the economic impact in the hosting of events is noted by Horne (2007) in highlighting that "a review of the enormous amount of literature on the socio-economic, socio-cultural, physical and political impacts of Olympic Games concludes that 'economic benefits are the prime motive' for interests involved in hosting them (Malfas et al., 2004: p. 218)".

Providing a deliberately constrained definition of economic outcomes to those being directly identifiable contributions is emblematic of the determination to consider how smaller events can and are measured and include them within the framework dimensions. This 'direct contributions' view avoids the extended economic impact favoured by larger-scale events in 'economic impact reports' – in which the use of multipliers and induced impacts are common – especially where investment in infrastructure is requisite for the event hosting.

Additional constraints and clarifications on the economic impact from the literature review were likewise provided to ensure a common understanding in the assessment of the conceptual framework.

- *Employment is a flow-on outcome of the scale of the event and operational spend; it is not a driver in and of itself*
- *Additional attracted FDI (Foreign Direct Investment) is seen to be, in most cases, a flow-on effect rather than a driver*
- *Sponsorship rights value is accrued as a source of revenue to the rights holder and is reflected in their spending on the event*

- *Spending by partners on activating their sponsorship is incremental to direct event spending by the organizing committee*

4.4.1 Economic Drivers

Examples of economic sub-elements ('drivers') include event attendance and event-led tourism, direct event spending (operational, capital, security), infrastructure and asset development, increased economic activity for existing businesses, spending by sponsors on activation programmes (not sponsorships rights value).

These economic drivers are summarised below in Table 10 and literature insights on each are discussed in the following sections.

Table 10 - Economic Drivers and descriptions

Pillar	Driver	Description
Economic	Tourism	The economic activity generated by visitors including during and around the event
Economic	Infrastructure	New or upgraded capacity and development of long-term assets: airport, housing, road, transport, stadia etc.
Economic	Direct event	Operational spending by lead organising body on the planning, development, promotion and hosting of the event.
Economic	Security	Spending on venue security, public spaces and visiting athletes/players, teams, officials
Economic	Business activity	Increase in economic activity for existing businesses as a direct result of the event.
Economic	Sponsorship	Spending by sponsors and organisers on supporting activities

Economic impact overview

Despite the significant attention the assessment and research on the economic impact are given, the majority of research in the literature relies on the ex-post analysis of events. Researchers have found that the ex-ante predictions of great economic benefits (largely focused on tourism and infrastructure) are generally grossly overstated (Crompton, 2006; Davies et al., 2013) and that effects on employment and wages are usually small, if sometimes positive. The researchers reported that this is due to several factors, primarily the far too large multipliers estimated, the crowding-out effects of tourism and spending not taken into account, and the increases in wages, supply, and demand.

There are attempts at more precision in the preparation of formal impact reports when as Jolly (2013) notes in "assessing the economic impact of the 2006 Games, it was necessary to determine whether expenditure on infrastructure was the direct result of hosting the Games or whether it would have been incurred anyway." (p. 40).

The intent to separate the outcome attributable to the event from other factors was sound, however, in the final report the treatment of capital works "assumes that all of the capital expenditure has been undertaken due to the 2006 Games on the basis:

- *there is no substantive evidence suggesting that the capital works expenditure would have been undertaken anyway and*
- *due to this uncertainty, there would be a considerable degree of arbitrariness implicit in any assumptions relating to brought forward capital expenditure." (KPMG, 2006, p. 142).*

The basis for the inclusion of capital works in the report was because there was no evidence it would have been undertaken anyway. This is at odds with Essex and Chalkley (2003) when proposing that outcomes may be claimed for an event "even if there is evidence that the development may have emerged in the fullness of time irrespective of the event" (p. 5). Both approaches attribute the infrastructure to the event but from contrary positions.

The framework does not seek to address these contradictions directly nor look to value the outcomes. Instead by focussing on more dimensions and measures that are more widely applicable to a range of events, the results will align more directly with operational outcomes and the uncertainty of induced effects and the treatment of capital works will be limited to the events of sufficient scale and investment to warrant the specialist analysis.

4.4.2 Tourism

Tourism, as the largest and primary driver of the economic impact of events, underpins the rationale of established models and event assessment cases. But an assumption that the 'larger the scale of the event, the larger the tourism effect' is challenged by the research that finds "It is not necessarily the more expensive events that yield the most benefits: the type and, importantly, timing (seasonality) of the mega-event, and the countries participating in the event all impact on the 'success' of these events, measured in terms of tourist arrivals." (Fourie & Santana-Gallego, 2011, p. 1369).

Whether those arriving contribute economically to the long-term (multiple years post-event) or just the short-term (during the event) tourism, Solberg and Preuss (2007) find that the long-term effects of tourism are greater in larger cities where economies of scale also mean investments in infrastructure are more likely to profitable. That is, larger cities can recoup the value of the investment in the infrastructure, hence linking scale, infrastructure and visitor attraction to the event tourism effect.

By contrast, (Allmers & Maennig, 2009) argue that the positive impacts of tourism during World Cup events are largely negligible for cities that already have large tourist bases (France, Germany). However, they also predict that the benefits might be far greater in countries/cities that do not normally experience large amounts of tourism.

Seemingly at odds initially, both views can be reconciled through the lens of the measures that underpin the assessment of the impact (mostly economic) of tourism. Summarising their findings on the economic contribution of tourism through events Gratton et al. (2006b) succinctly identify the key drivers as: the ability to attract people to the host area, the total number of spectators; and the total number of days of competition and the ability for visitors to extend their stay.

The difference between the two perspectives may therefore be that the opportunity to add incremental tourism is greater (or more readily observed) in host locations with lower historic levels; and/or hosting events in those areas more distant encourages tourists to extend their stay to take advantage of the time and costs already invested in attending. Hence, measuring the underlying tourism numbers remains a reliable baseline of impact and is independent of the ascribed effect which appears dependent on the host size and location.

In addition to the positive contributions to the economic impact of tourism, Gratton et al. (2006b) provide a caution concerning an event's prestige; in that "the economic impact is not necessarily a function of the status of an event in world sporting terms" (p. 57). Whilst the status of an event might be assumed to be a strong predictor of tourism arrivals, further research by Fourie and Santana-Gallego (2011) found that while "on average, mega-sporting events increase predicted tourism by roughly 8% in the same year… there is, however, large disparities between the types of event; the Summer Olympics, FIFA World Cup and, to a lesser extent the Cricket World Cup and Lions Tour, all seem to have a significant positive impact on tourism, while the Winter Olympics and the Rugby World Cup do not" (p. 1369).

Whether that status does have a strong predictor effect on the sponsorship value to the owner, accrual of media rights or ability to attract the sport's elite athletes are not captured within the narrower view of direct economic impact but could be identified within a framework that includes those dimensions.

The boundaries on what should be included within the tourism driver are highlighted by Crompton (2006) who identified that "Economic impact attributable to a tourism attraction relates only to new money injected into an economy […]. Only those visitors who reside outside the jurisdiction and whose primary motivation for visiting is to attend a tourism attraction or who stay longer and spend more time there because of it should be included in an economic impact study"(p. 70).

At times, the factors that impact adding 'new money' into the economy may not be within the control of the event organisers. (Li et al., 2011) found that for the Beijing 2008 Games "international tourism was predicted to be positive in the *ex ante* estimation, this impact was analysed to be negative in the *ex post* estimation" (p. 292). It was noted that the decrease in international tourism arrivals and expenditure "was different from previous Olympics, which brought a tourism boom to the host country in the Olympic year [with] new rules on tightening visas was one of the most crucial factors" (p. 292).

4.4.3 Infrastructure

The building of infrastructure that lasts long after the event is often the most tangible of event legacies. However if not productively used those assets also are subject to claims of being 'white elephants' in the post-event assessments (Alm et al., 2016; Death, 2011; Majumdar, 2012; Westerbeek, 2010). More specifically, the issue of infrastructure and events encompasses two key dimensions of event legacy.

The first is that infrastructure and events often lack clarity in the assignment of costs related to the event itself. Costs may be assigned to the event in ways that conflate operational costs and revenues with the capital infrastructure investments made by host cities on venues, transport and other developments. To avoid seeing the 'cost' element subverted, Preuss (2004) emphasises that "it is impossible and even wrong to state the overall effect of different Olympics with a single surplus or deficit. The true outcome is measured in the infrastructural, social, political, ecological and sporting impacts a city and country receive from the Games"(p. 26). That said, the most damning evidence for cost over-runs in meeting the infrastructure requirements of mega-events comes from Flyvbjerg et al. (2016), who determined that "All Games, without exception, have cost overrun. For no other type of megaproject is this the case" (p. 15). Not sounding a sufficiently salutatory warning against hubris in bidding for these events in that "to decide to take on one of the most costly and financially most risky type of megaproject that exists, something that many cities and nations have learned to their peril"(p. 2).

Reflecting the separation of operational hosting and capital asset expenditure, the IOC now requires prospective bidding cities "to differentiate the Urban Development Budget from the Games Organisation Budget"(IOC, 2021a, p. 27) where "upgrades of existing sport facilities or renovation of public infrastructure… should not be considered as Games-specific spending" (p. 27). Notwithstanding the IOC's direction to include event facility upgrades within the city development plans, the second aspect of infrastructure legacy is the practical utilisation of the assets following the event's conclusion.

Both types infrastructure assets that are specific to the event (for example stadia) and assets built with the event as the core impetus but adding to the host's built environment (for example transport, airport etc.) can suffer from underutilisation following the event's conclusion. To avoid leaving 'white elephants' (costly and poorly used assets), evidence from Allmers and Maennig (2009) shows that the building of new facilities for events has the best return when spread out to capture more population per venue – particularly when smartly placed to integrate with local needs. That, however, is not a panacea in itself, as Coates and Humphreys (1999) further detail that historically the building of sporting arenas and placement of local sporting teams in US cities with lower average wealth has no impact on average wage growth, in those cities.

Investigating whether the lack of benefit is generalisable beyond specific stadia within the US market to large scale, one-off events, Jasmand and Maenning (2008) researched the effects of the 1972 Munich Olympics, where 73% of the budget was spent on construction. The hypothesis was that this investment would translate into long-term higher employment and wages in the regions that hosted the games, as compared to similar regions that did not. The results showed that while higher employment did not materialise, higher wages did emerge over the next 16 years to the end of the study period in 1988. By contrast, this was found to not be the case in other Olympics, where studies on Los Angeles 1984 and Atlanta 1996 found no significant increase in either employment or wages.

In contrast to the work of Jasmand and Maenning (2008), Kasimati and Dawson (2009) found a more optimistic view that the 2004 Athens Games had long-term beneficial impacts on the economy, accruing approximately 1.5% GDP growth 1997-2005 in the pre- and immediate post Games periods, with an extended view falling to 0.5% over 2006-2012, as well as decreasing unemployment (p. 405).

E. Kasimati, P. Dawson / Economic Modelling 26 (2009) 139–146

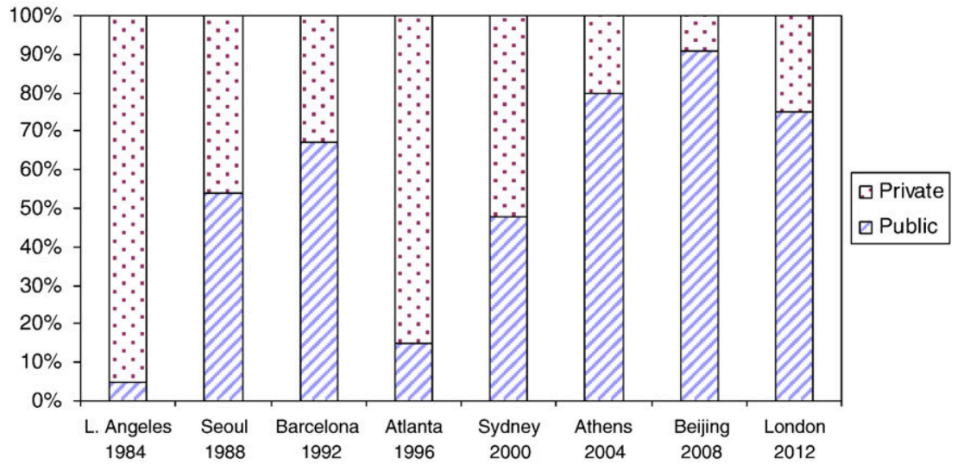

Figure 16 - Proportion of Private vs. Public funding - Olympic Games 1984-2012

The differences in findings between studies on different editions of the same event could arise from differences in the balance of Public vs. Private funding as seen in Figure 16 - Proportion of Private vs. Public funding - Olympic Games 1984-2012. Alternatively, the differences in findings could be due to the techniques used to analyse the economic impact of 'special events', specifically input-output models, which Dwyer et al. (2005) argue are limited and tend to overestimate the benefit and underestimate the cost of these events.

4.4.4 Direct Event Spending

Event organising committees are responsible for significant direct spending in the planning and delivery of an event. As noted by Chappelet (2018), the "operating budget includes all costs inherent to the operation of the Games (accommodation, travel, accreditation, security, decoration, salaries, mandates, etc.). It also has to include the cost of any temporary sports facilities that might be necessary" (p. 4).

In looking at the calculation of economic impact from events, Mules and Dwyer (2005) draw particular attention to the role of direct event spending in noting "it needs to be emphasized that organizers and sponsor expenditure must be considered alongside visitor expenditure" (p. 339). This is a finding that many studies overlook, with the result that "neglect of this category of event-related expenditure can result in a large underestimation of economic impacts." (p. 339).

Looking beyond tourist spending for economic impact, Davies et al. (2013) studied the direct expenditure by event organisers with local suppliers and contractors. The result showed "across 22 major sports events on average, organisational spending accounts for 17 per cent of the total direct expenditure impact attributable to an event".

Further research testing the reliability of economic impact forecasts Ramchandani and Coleman (2012) found that the contribution of organiser spend compared to tourism expenditure varies from as low as 9% to 93% of all event expenditure.

At an absolute level, the impact of organisers' direct event spend "depends on revenues that they receive, the main sources being participant registration fees, public and private sponsorship, and financial support in kind."(Dwyer et al., 2000, p. 180).

Smaller and regional events do not necessarily create the cost burden of major events, but their spending impact is also limited to their funding sources. In the example of the Comrades Marathon Saayman and Saayman (2012) note "Entry fees are used by the organisers in hosting the event [and if the] prize money is excluded, total spending by organisers amounts to almost ZAR7 million" (p. 227). This value equates to approximately 10% of the total direct spending, which is in line with the lower end of values observed by Ramchandani and Coleman (2012).

Looking downstream at the impact of organiser spending, Hotchkiss et al. (2003) found evidence that organisers' direct spending on hosting Atlanta '96 did provide long-term increased employment and wage growth compared to similar non-venue or near-venue counties. The three areas identified as leading to this growth were:

- *Short term employment benefits from Direct Event Spending on goods and services*
- *Job training provided by the government as a result of hosting the Olympics, creating long-term additional worker skills*
- *Investment in facilities and infrastructure*

This created, all told, 17% more jobs in VNV (venue or near the venue) counties post-Olympics. The authors also note the outcomes showed significantly less effect on driving wage growth than the effect seen in raising longer-term employment in the VNV counties.

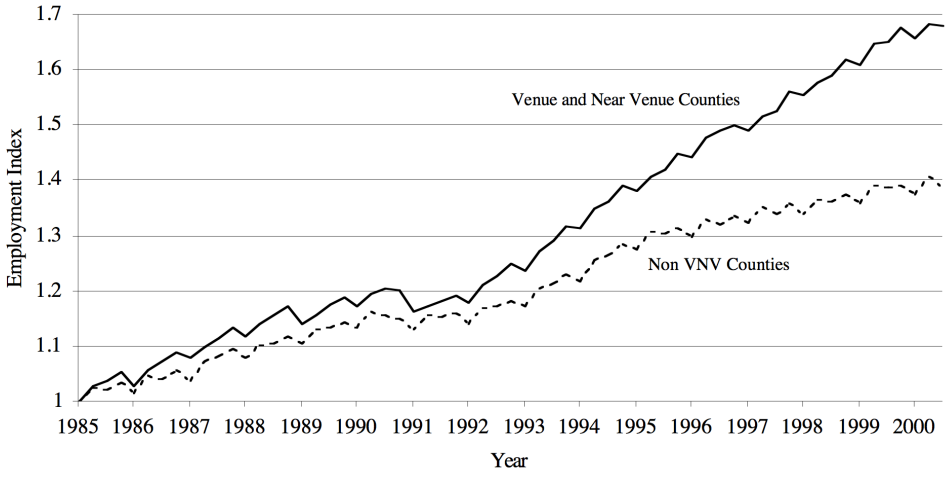

Figure 17 - VNV Employment Growth Effect

For events to deliver a positive outcome, reliance on visitor spending in developing economies is vital. However, despite the positive effect seen from Atlanta, the benefits of hosting a major sporting event do not always overcome the event investments made by the hosts. Humphreys and Prokopowicz (2007) conclude that in hosting the UEFA European Cup, despite the high expected impact of visitor spending, the costs imposed on Poland and Ukraine for Euro 2012 creates a net-negative in economic impact.

4.4.5 Security

Thinking of mega-events more dispassionately as projects, or megaprojects as Flyvbjerg et al. (2016) term them, then it is unsurprising that "the budget estimation is a challenge in the planning of mega-events." (Girgin & Tasci, 2019, p. 259). What is less widely recognised is the rising complexity of estimations, "especially due to increasing security cost [which leads to] underestimating the cost of a megaevent and overestimating the funding"pp259.

So while major events are promoted as a global platform for the host, it is that same capacity for generating the "enormous international media attention that many mega-events receive, [that makes them] increasingly become targets for the activities of protest and terrorist groups" (Jago et al., 2010, p. 230). Hence despite the already increased security and surveillance in place at Sydney 2000, the terrorist attack on 11 September 2001 "has resulted in the cost of security provided for mega-events increasing enormously and if such costs continue to escalate, they risk becoming a disincentive for destinations even bidding to host mega-events." (p. 231).

Some organisers have managed to shift the cost from the event's books, with Matheson (2006) noting that the Salt Lake Winter Games budget failed to include the "millions of dollars of additional security provided [at] no cost to the local organizing committee"(p. 3) However that cost pales given for "the 2004 Summer Games, the government in Athens spent $1.5 billion on security alone."pp3

The headline numbers on security spending for mega-events distracts from a more practical perspective. That is, within events of differing scales, security costs are a significant portion of the event budget. Even in historically low-risk environments like Australia, the

budget projection for the Gold Coast 2018 Commonwealth Games showed 'Security (Event & Public)' was costed at GBP£69.3m, or 16% of the total event budget. That represented not an insignificant proportion of the budget given the combined costs for GOLDOC Operating costs and Games Administration totalled 21% (CGF, 2011, p. 36).

The cost increase in security is driven by the increased presence and use of technology. Since the Sydney 2000 Games, the continued presence of security hardware and enduring security legislation and the effect is seen in "state-based security policies… to increasingly security conscious event precinct design, to progressively intrusive crowd management practices" (Toohey & Taylor, 2012, p. 12). Such surveillance has "entrenched a greater acceptance of enhanced sport security practices"(p. 12).

That embedding of security into event practices means elevated security spending has become an enduring aspect of event hosting. Thinking of security only in terms of direct costs may, however, mask unintended effects on other dimensions of event impact. As the cultural practices of security "subordinates all other activity within its domain" Bajc (2007, p. 1569) obverses "the general public was discouraged from participating in various ways and only ticket holders were able to attend"(p. 1569). This exclusion of the general public is at odds with efforts to foster a greater social engagement of the wider community around events and restrictions may also place limits on the touted economic benefits of spending around event venues.

4.4.6 Business Activity

The effect of events is not limited solely to the initial injection of spending by the event, sponsors, and tourists, but the multiplier effect that stimulates the host location economy. The "expenditure of visitors and organizers/sponsors stimulates economic activity and creates additional business turnover, employment, household income and government revenue in the host community" (Mules & Dwyer, 2005, p. 342). The 'ripple effect' in an economy from the stimulation is an indirect impact, with the expenditure re-circulating through the economy as "the affected firms purchase inputs from other business operators. These other businesses, in turn, purchase inputs from other firms and so on" (p. 343).

That focus on external income stimulating local businesses is consistent with Peeters et al. (2014), who in studying lower-income countries hosting major events with the hope of increasing tourism flows noted "a key ingredient to turn the organisation of a major international sporting event into a successful economic strategy for developing countries is the attraction of large numbers of rich, foreign tourists" (p. 290). It is in servicing the high-value tourist demand that business activity is driven. Activity that would not be created within the local economy without the event.

To stimulate that 'ripple effect' in business activity from an event, 'leakages' need to be minimised. Rather than simply seeking a greater number of visitors as a measure of success, Geland (2003) advises "the most important indicator is the proportion of local products in the total value of goods and services purchased by tourists. If the impact of tourist spending is to be maximized, this proportion must be as high as possible." (p. 423)

For a single event to create a significant lift in business activity depends very much on the scale of the event host's economy. There is a perverse effect that hosts with highly developed and strong economies are best positioned to afford a mega-event. They are also likely to see that the impacts of these events are statistically insignificant at the GDP level. Nevertheless, positive outcomes flow from the event that would not have occurred otherwise (Kavetsos & Szymanksi, 2009).

In trying to stimulate the indirect impact of the 2002 Manchester Commonwealth Games on the regional economy, a specific programme was developed to enable local businesses, known as the 'Prosperity Project'. Smith and Fox (2007) note the "business club created as part of the project was particularly successful. Business club members were able to get involved in the supply chain activity [and] tender for, and win, around £45 million of sub-contracting work" (p. 1136).

That 'business engagement' strategy has seen "Sheffield and Manchester 's municipal government [attract] major sports events [which] have played a significant part in the economic and social revitalization of decaying industrial urban areas." (Zagnoli & Radicchi, 2009, p. 57).

4.4.7 Sponsorship Activation

Initially, sponsorship can appear as an asymmetric marketing investment "with fixed outlays up front, and unclear benefits to be realized in the future." (Nickell & Johnston, 2019, p. 61). Despite the perceived uncertainty for sponsors, attracting sponsorship for events is a critical part of event funding, as hosts seek to offset their costs through corporate partner investments. For their part, companies seek to increase the effect and the Return on Investment (ROI) of their sponsorship 'sunk cost' through not only assessment of the quality and reach of the event, but by additional investment before and during an event. This is referred to as 'sponsorship activation', meaning "to interact with consumers in ways that improve their experience with the brand" (Chavanat et al., 2009, p. 668). The authors note "to optimize the sponsorship investment it is essential to activate the sponsor's brand."

So rather than falling out of favour, International Event Group found that "sponsorship spending is projected to grow 4.5 percent in North America this year - to $24.2 billion - and 4.9 percent globally, to $65.8 billion" (IEG, 2018).

Without considering the additional investment of 'sponsorship activation', the rights fees paid to event organisers understates sponsorship spending on sports and events. Indeed rights fees are at best only half the value sponsoring organisations need to invest, with marketers expecting to "spend an average of $2.20 on activating sponsorships for every $1 spent on rights fees" (IEG, 2018).

When considering even a relatively modest 2:1 ratio on activation to rights fees, the value of sponsorship spending around an event is given perspective when considering 'The Olympic Partners Programme' revenues from 1985-2016. Below Figure 18 shows that despite only a modest increase in the number of global partner brands from 9-12 over that period, the value of the rights has increased from $96m to over USD $1bn (IOC, 2019), implying a 'sponsorship activation' around IOC events by partners in the order of $2bn each quadrennial cycle.

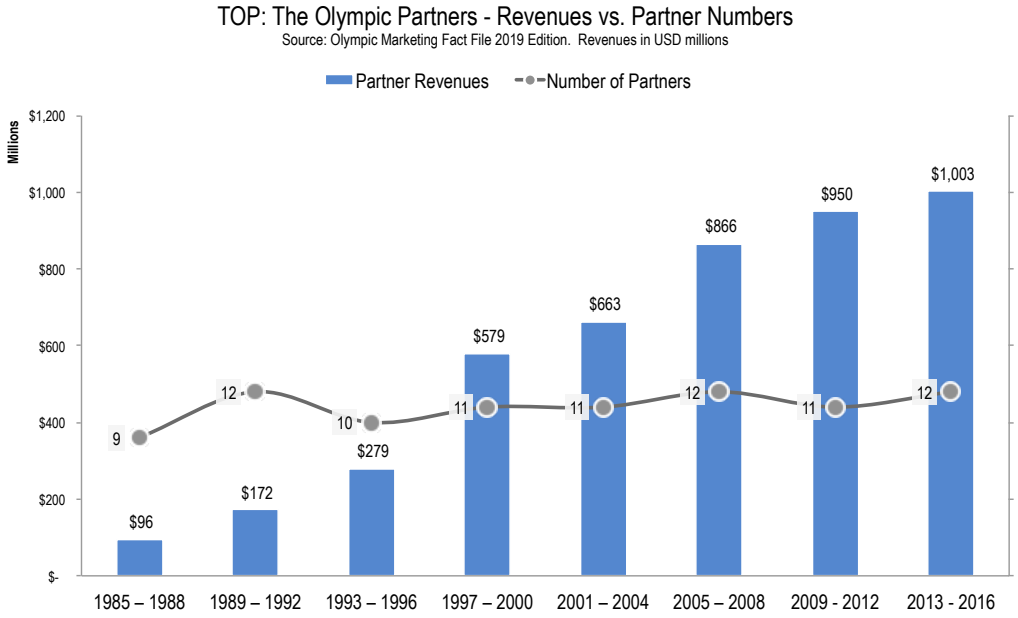

Figure 18 - TOP: The Olympic Partners - Revenues vs. Partner Numbers

Creating 'marketing spaces' (both physical and virtual) where sponsors can engage and enhance event attendees' experience is foundational to achieving the sponsors required ROI. (Donlan & Crowther, 2014) suggest that the sponsor's role is central enough that the "event strategy should thus be infused by the strategic sponsorship objectives and carefully integrated with wider marketing activities and customer experience management"(pp293) of the event.

That idea of creating literal 'spaces' for sponsors is not without critics. Allowing major Olympic sponsors to move their brands into public spaces at the exclusion of any existing advertising is seen as a permitted 'takeover of the urban realm' under the Host City Contract (Eick, 2010).

Despite concerns, sponsor-based activities in public spaces at events show an increasing trend towards Live Sites, Fan Parks or Fan Zones (McGillivray & Frew, 2014). The use of Live Sites are promoted as places of active engagement for fans to gather and celebrate but are closer to being constructed spaces that flow from the corporate structuring and securitisation of these events (Cornelissen, 2011).

Whether such practices are indeed the "commercialization and commodification of (public) space" (Eick, 2010, p. 279), for sponsors they are seen as crucial in making sure their brands are visible, engaging, and strongly bound with the event (Mazodier & Quester, 2014), and by doing so avoid misascription of event sponsorship so they can realise the value of the event association. (Chavanat et al., 2009).

But the reach of physical public spaces that sponsors can occupy during events is not limited even to the host city or country. During the 2010 South Africa World Cup, there were six designated Live Sites across the globe known as the International FIFA Fan Fest™. These were hosted by the event official global sponsors including Coca-Cola, Emirates, Hyundai and Sony (Rowe & Baker, 2012). Such sites extend both the size of the potential audiences the brands can directly engage with through the events and the impact of the 'activation investment' being made to multiple sites across markets.

4.4.8 Displacement and crowding-out

In addressing the economic dimensions of event impact, it would be neglectful not to acknowledge factors limiting economic gains from event hosting, especially in reference to the case of major and/or mega-events.

While identifying the economic impact of tourism in the host location, Baade and Matheson (2004) note that drawing on the same measures in adjacent destinations can reveal a wider view of economic effect. With reference to Sydney 2000, they note Sydney's hotel occupancy around the time of the 2000 Olympic Games was close to 100%, up from around 49% the same time in 1999. Critically however, hotel occupancy in Melbourne and other Australian cities was significantly down 17% in the same period. Hence, the hosting of a mega-event in a city may actually act as a redistribution of wealth amongst local cities in addition to the operational and infrastructure funding in the host location.

As noted the effect of bringing a large scale event to a host destination may appeal to some residents, while others seek to avoid it. The effect of 'crowding out' can be seen in locals changing travel and spending habits, or travellers avoiding the host location while the event is on or residents purposefully leaving the area to avoid the event. (Agha & Rascher, 2016). If the net effect is for economic activity to fall below its regular level, then gains from an event have been negated.

Ideally, the displacement of residents who seek to leave their city during the period of the event to avoid the expected crowding from an influx of visitors should be offset by incremental gains in business activity ascribed to an event. But those gains do not always materialise.

At Gold Coast 2018, while sports venue attendance was strong, "local businesses have been stunned by the lack of foot traffic and commerce." (Ransom, 2018). Some of that lack of expected activity was 'self-inflicted' as "hotels that raised their rates in expectation of a windfall have hastily cut them and many have less than half their usual occupancy". Others could have been avoided, including the "Gold Coast organizing committee (GOLDOC) which urged local residents to avoid driving in a sprawling city [that has] turned away locals and tourists alike".

In relation to the 2000 Sydney Olympic Games, reflecting the same issues found with crowding out of visitor spending, "some non-residents, who might have visited the country, decide not to do so because of congestion and high prices during the event's period"(Baade & Matheson, 2004, p. 346). Likewise, regarding the NCAA Basketball Final Four in the US – the hosting of 'March Madness' was worth approximately $500 million a year in TV rights in 2002 and attracted 50,000 visitors but Matheson and Baade (2004) found that although gross spending around the event is high, it is always compensated by lowering spending elsewhere. The net result is very close to zero over a long-term study.

Even if successful in attracting the level of activity projected, a central tenet of the benefits accrued from spending by visitors' arrivals in the event location and enticing them to extend their stay is the increased activity for businesses providing services needed. Cautioning on this approach, Geland (2003) argues that volume is not enough in that "neither the number of spectators nor the size of spending necessarily provides a reliable measure of economic impacts" (p. 423). At a more granular level, Daniels et al. (2004) identified that the income benefits from increased tourism surrounding events mostly go to those in service industries (largely hotel and lodging, food and drink, and recreation) and usually benefit those on lower incomes. But rather than hire more staff, companies have typically paid overtime to

their existing staff or hired temporary event-only staff, creating a windfall for those employees but not a net creation of new employment roles.

4.4.9 Economic measures

As a result of the review of the research literature and practitioners' contributions, 63 potential economic measures were initially identified (Table 11).

The measures were reviewed for inclusion in the research model according to the criteria outlined in 3.2.3 'Specification of measures'. The final measures chosen for rating by respondents are listed in Table 78 - Measure ratings: Economic.

Table 11 - List of potential economic impact measures

Construction spending ($)	Construction: Improved/new infrastructure (count)	Construction: Transportation infrastructure ($)
GDP growth: current year ($)	GDP growth: current year (%)	Income growth: City before/after event ($)
Income growth: Individual ($)	Income growth: National ($)	Income growth: Household ($)
Income growth: Venue region vs. National ($)	Job creation: Event hosting direct (Count)	Job creation: Pre/Post from event (Count)
Visitor Length of stay: Average number of days (n)	Media spend: Advertising dollars ($)	Organizational host expenditure ($)
Resident expenditure $	Retail sales: Total spent over total event ($)	Retail sales: Post-event novelty effect ($)
Retail sales: Activity (Count)	Visitor: Room nights (Count)	Visitor: Room nights total ($)
Visitor: Room nights average ($)	Spectator admissions (Count)	Ticket sales: Post-event sporting events ($)
Ticket sales: Total sales ($)	Ticket sales: Total sales (Count)	Visitor distance: average (km)
Visitor expenditure: Total spent over total stay ($)	Visitor expenditure: Total spent per day, per visitor ($)	Visitor expenditure: Total all visitors ($)
Visitor numbers: Total (count)	Visitor numbers: Local (count)	Visitor numbers: Regional (count)
Visitor numbers: National (count)	Visitor numbers: International (count)	Visitor numbers: Per day (count)
Economic impact: Nation/Region/City	Income flows: Nation/Region/City	Jobs created: Nation/Region/City
Wages: Nation/Region/City	Taxes: Nation/Region/City	Staff: Organising committee
Attendance: Total	Attendance: Overseas	Attendance: Domestic (nation)
Attendance: Domestic (region)	Attendance: Domestic (city)	Ticket Sales by % of capacity
Unique attendance: Total	Unique attendance: Overseas	Unique attendance: Domestic (nation)
Unique attendance: Domestic (region)	Unique attendance: Domestic (city)	Costs: Full breakdown
Revenues: Full breakdown	Funding: Full breakdown	Budget: Full breakdown

4.5 Pillar - Sport

Sport impact from an event was defined in terms of the size and duration of the event, its prestige or importance to competitor rankings, and its public reputation. It included the importance of the contribution to the development of the sport both in terms of infrastructure and reinvestment in supporting the expansion and future growth of the sport.

4.5.1 Sport Drivers

Examples of sport sub-elements ('drivers') include numbers and breadth of participants, quality of performances during the event, support and enforcement of 'clean sport', positive sport infrastructure, talent pathway legacy and development of the sport, and interest in future editions of the event.

These sports drivers are summarised in Table 12 and literature insights on each are briefly discussed in the following sections.

Table 12 - Sport Drivers and descriptions

Pillar	Driver	Description
Sport	Event scale	Participant numbers and breadth - gender, nationalities; regions in which it is the dominant sport.
Sport	Event prestige	Quality of performances, quality and standard of the competition/players, records, prize-money and ranking points
Sport	Reputation of the sport	Testing of competitors, support for 'clean sport', promotion of 'Fair Play', and intolerance for poor behaviour by competitors
Sport	Sporting infrastructure	National/international standard infrastructure, development of officials, talent pathway, and identification programs as legacy
Sport	Sport expansion	Development of innovative formats, reinvestment in sport, and growth of interest in future editions

In defining the scope of sport impact, a rich array of potential dimensions were noted in the literature review. Within the scale of the event, the breadth of participants may include designations of gender, age groups, or the country of origin of participants etc.

The classification of quality of performances referred to medals won, goals scored in competition as well as ranking points accumulated, recognition of records set and whether competitors had reached various 'finals' stages. As noted earlier in the interrelation of framework dimensions, the size of sponsorship and rights deals are primarily driven by the popularity (attendance and media scale) of the event. Hence while the popularity of the event is related to the quality of the 'on-field' performance, it is not included in this pillar. Likewise, the significant value of the sponsorship rights is often used in offsetting the costs of hosting the event and hence are included in the Economic pillar.

Recognising that ranking points, finals achieved, and medals won are used by national sporting bodies under 'own the podium' models for sport funding and development. However, linking changes in levels of national government investment in elite sport

development may occur from an event but can be difficult to consistently isolate that value to specific events.

4.5.2 Event scale

The scale of an event is one of the key criteria by which types of events are classified. This is often about the audience size it attracts (see 'Audience size' in the Media pillar) and the number and diversity of the participants.

There is a direct link between event scale and positive economic outcomes for tourism and business activity. This is reflected in the observation "that tourism from participating countries increases more than tourists from countries not participating in the mega-event." (Fourie & Santana-Gallego, 2011, p. 1364). To that end, broad representation within the event participants, and the opportunity to expand that breadth, can have a demonstrably discriminating effect as a tourism attractor.

While large scale events attract media attention and sponsorship attention, those events can be limited in their impact by the application of restrictive host agreements. The ambition to host global events may be an attractor but "highly formalised host agreements negatively impacted the host destinations' respective abilities to leverage the event." (Kelly, 2019, p. ii). The result of global event owners using their preferred global partners meant "reduced opportunities for local suppliers (both to benefit financially and also to gain event experience from hosting large-scale events); limited display of local culture; and economic leakage from the local economy" (Kelly, 2019, p. 137).

While mega and global events dominate references to event scale, an overlooked perspective is the event scale relative to the host location. Seeking to attract smaller-scale events (in absolute number of participants) to mid-size communities brings significant economic and skills benefits and a reason "that brings people to a city that they may not otherwise visit"(Veltri et al., 2009, p. 126).

Even within the scale of an event, the mix of participants and competitive balance of participants impact its long-term interest and viability. As Zheng et al. (2019) has noted "outcome uncertainty underpins competitive balance and competitive balance is important in attracting the engagement of fans, supporters, consumers, broadcasters, merchandisers, sponsors and the long-term morale of the majority of participants" (p. 54). The authors note that some sports have adapted rules and structures to improve the competitive balance but broadly the endemic lack of "competitive balance for female competitions does not augur well for the promotion of female events in many sports [resulting in increasing] market share of gold medals and medals for traditional 'powerhouse nations' " (Zheng et al., 2019, p. 54).

4.5.3 Event prestige

Not all sports competitions have a specific performance benchmark such as a world record time or distance. Most are described as 'rank-order tournaments', where participants are rewarded for their performance relative to other competitors. In such competitions, there is no absolute benchmark for performance, but "athletes consider the rules of competition, costs, and incentives and choose actions to optimise against the efforts of their competitors" (Westmattelmann et al., 2021, p. 689). The quality of the tournament and performances is a key factor in an event's ability to attract the best players to compete against each other and/or to offer the best rewards.

Hence event status, prestige, and prize money are all factors on which athletes choose to compete. Prior research has noted such a relationship in road running where Maloney and McCormick (2000) identified "two separate responses of runners to wages. First, higher prizes are associated with faster times for individuals already in the race. Second, higher prizes attract a faster field." (p. 118). That ability to pay higher prizes was also linked to prestige, in noting "prize money and prestige are linearly correlated" (p. 118).

A detailed 'labour supply' analysis of US golf tournaments by Hood (2016), posed three questions at the centre of every host's planning when seeking to develop their event:

1. *"Does the prize money and status of one tournament have an effect on the participation rates of the tournaments surrounding it in the schedule"*
2. *"Do increases in the purse increase the participation rate?"*
3. *"Is the previous participation and performance of a golfer a good indicator of his future participation?"*

Since the tournament organisers are prohibited from paying 'appearance fees', this type of tournament provides a relatively clean 'labour supply' market to test the influence of an event's status on participation choices. And as Hood (2016) notes "the answer to all three questions is yes." (p. 289).

The prestige of an event influences fan behaviours as well, creating a virtuous circle of attention and attraction. Chiu et al. (2019) reflect this view when referring to the FIFA World Cup as "individuals prefer to attend this prestigious mega-event not only because it is an unforgettable experience, but also because it can increase their social identity and group status" (p. 1408). Event quality of performances and status is therefore reflected in elevating personal status of those attending, increasing demand and reinforcing the event's prestige. That 'halo' effect of world-class events and performances creates a driver not only for athletes and sponsors but for hosts as "elite sport performance and the staging of sports mega-events are seen as two potential sources of international prestige." (Grix & Houlihan, 2014, p. 576). That motivation to associate with the highest echelon of sport is a strong driver and it was "in large part the desire for international prestige that led Brazil to successfully bid to stage the world's two largest mega sporting events in close succession" (Grix & Houlihan, 2014, p. 576).

The event's status also impacts the involvement of other stakeholders, where mega-events attract higher involvement from not just spectators but also volunteers in part because people feel more pride in being associated with events of global status than might be derived from lesser-known events (S. Kim et al., 2015). Hence, the impact and status of an event are not just in its attractiveness, but also in ensuring the event garners the support it needs for successful delivery. As noted by Lechner and Solberg (2021) "cancellation of the championship would have caused a severe loss of prestige for many stakeholders, not only those that were directly involved in the preparations" (p. 22).

4.5.4 The reputation of the sport

Rules in sport are meant to provide a fair and safe environment in which athletes can compete. This extends from on-field competition to performance enhancements, financial structures, and the behaviour of players and fans. In identifying the core concepts of 'fair play', Serrano-Durá et al. (2021) found two emergent themes. "The first is related with rule breaking and financial inequality of sports clubs and, the second is linked to inequality in

classification systems (for questions of gender, diversity and the planning of competition schedules)" Serrano-Durá et al. (2021, p. 648).

A different approach to disaggregating the 'fair play' concept into specific tiers of principles by Motoarca (2015) determined three levels as shown in Table 13

Table 13 - Three Types of Fair Play

Fair Play (FP)	Description
FP-1	Principles without which the game of soccer as it is intended to be played could not take place.
FP-2	Principles that are not directly connected to the structure of the game, but whose purpose is to promote an atmosphere of cordiality, respect, and equal opportunity on the pitch (and sometimes outside of it).
FP-3	Principles are not constitutive of the game, and neither are they essential for the acceptability of soccer as a worthy social endeavour. Instead, they are rules of etiquette.

"Three Types of Fair Play" from (Motoarca, 2015, p. 125)

Establishing an ideal form of the purpose and ethics of sport, the IOC states that "the goal of Olympism is to place sport at the service of the harmonious development of man, with a view to promoting a peaceful society concerned with the preservation of human dignity"(IOC, 2020, p. 11). But how closely does that goal match the reality of modern sport?

"The reality of sport is quite different from its ideal. More often than not gambling, cheating, match fixing, diving, doping, and so on make the front pages rather than the back" (Murray, 2012, p. 587). The front pages are filled not just with illicit competition actions, but also with off-field behaviours, because of the public status afforded to them by their sport. Concerns about the negative reputational impact - disrepute - extends to behaviours outside of the sport when as noted by Turner and McCrory (2003) "sports personalities were role models and that the use of marijuana brought sport into disrepute" (p. 378). In mitigating potential reputational damage, club and national bodies can inadvertently compound the issues. In response to public scandals at the Cronulla Sharks Rugby League team in 2009 "the perceived protection provided by the club [led] the public to believe that there was a cultural problem at the club" (George, 2009, p. 38). That perception had a direct commercial impact with sponsors withdrawing as they "did not want their brand damaged by their continued association with a team and sporting body whose reputations had already been damaged" (p. 38).

Sports' aspirational transformative power is undermined by doubts about the authenticity of competition and performances and the complicity of authorities in protecting athletes' misbehaviours. This damage is not just in a direct effect on the sport. but broadly erodes sport and events' potency as tools of political 'soft-power', as that "gap between sporting reality and sporting idealism is perhaps one of the reasons why sports-diplomacy has been sporadic in the past"(Murray, 2012, p. 588).

4.5.5 Sporting infrastructure

The discourse on sporting infrastructures from events often attracts the 'white elephant' criticism of stadia built and then under-utilised following the hosting and leaving a burden rather than a positive legacy.

That thinking of sports infrastructure only in terms of tangible assets is described by Kavetsos and Szymanksi (2009) in "the economic regeneration benefits from the construction of Olympic facilities in East London and the additional sporting provision there" (p. 192). This formed some of the key outcomes underpinning intentions to bid for the London 2012 Olympics.

The benefits derived from the development or redevelopment of fixed sporting infrastructure for an event differs across potential hosts because "an event can have positive impacts on the cities which have an undeveloped sport infrastructure and which have high unemployment, and less so in cities with a developed infrastructure and low unemployment" (Malfas et al., 2004, p. 214).

Researchers have noted that it is not only hard asset legacies that exist but "sport legacy consists of image, infrastructure, knowledge about coaching, and the idea that sport is itself a cultural good" (Preuss, 2015, p. 3). This inclusion of 'soft infrastructure' outcomes is indicative of a wider perspective that sport infrastructure, sport knowledge, and sport networks can all develop as sport legacies from events (Sallent et al., 2011).

Even within the traditional definition of infrastructure assets the increasing adoption of temporary infrastructure has been a significant development in venue management practices around the planning and staging of sport events (Taylor & Edmondson, 2007). This means even if the traditional hard asset used for an event is temporary, the 'soft' assets and practices can be a direct outcome of the event hosting.

Sallent et al. (2011) reinforce the existence of and distinction between hard and soft infrastructure noting "we refer to physical capital when an event generates new sports equipment and infrastructure, or we speak of human capital referring to the expertise, abilities and competence of the individuals involved in tourism or sports practices" (p. 401). A cautionary note follows that "however, as relationships are a much less tangible concept it has not attracted the same amount of academic attention" (p. 401).

4.5.6 Sport expansion

A simple narrative on events' outcomes is the opportunity they afford to develop and expand a sport's footprint. That may take the form of deepening the participation/players based within existing markets, the opportunity to extend the sport's audience into new regions, or providing the opportunity for developing areas to engage in sport as a community.

It is acknowledged that the "capacity to draw large audiences in order to continue to profile and grow the sport is still a major consideration for any sporting organization" (Turner, 2007, p. 355). At the global level, the driver to grow the sport may be the protection of the market position against emerging competitors. However, the same drive to grow is seen at the local community level, even if not motivated by commercial imperatives. "We're really passionate about ocean swimming. That's why [the event] started, to get more people ocean swimming. To grow the sport" (Kelly, 2019, p. 88).

'Increased sport participation' has been touted as a legacy of major events. The 'demonstration effect' posits that by seeing elite sport, others will be inspired to become more active themselves (implicitly, in the sport they are watching). However, the effect of elite sport is found to be limited in that it "may have the potential to increase participation frequency in sport, and perhaps to re-engage lapsed participants [but bringing] new participants into sport is not likely to be successful"(Preuss, 2018, p. 105).

More direct community-based programmes such as the Football Foundation of South Africa (FFSA), established from the hosting of the 2010 FIFA World Cup, shows "that community-based programmes catalysed by mega-events can make important developmental contributions [as] local stakeholders have successfully connected with international corporations and have drawn down upon the latter's financial and human resources"(Bek et al., 2018, p. 451). Acknowledging that "aspects of the project's success can be argued to be spatially and temporally specific there are core elements of good practice" (p. 451) that can be applied to better ground the impact of global events in the local communities.

While historically Olympic "exhibition sports were included on a trial basis, a number of which became medal-sports, such as badminton, handball, baseball and taekwondo" (Stefani, 2020, p. 85), that pathway is no longer open. After seeking to fix the number of Sports Federations at 28, the Olympic Agenda 2020 allows that "an Olympic Organizing Committee (OOC) could request adding new sports on a one-off basis, as long as those new sports were organized by a recognized SF and as long as an athlete capping procedure was adhered to, including new and continuing sports" (Stefani, 2020, p. 87) In effect, more sports come at the depth and breadth of athletes competing. Future growth and expansion of individual sport federations rely not on mega-events but their own championships and events where participant numbers and entries are controlled by the sport's own governing body. For example, as a sport steeped in (even hampered by) tradition, the introduction of T20 (short format) cricket was a disruption more than an innovation, but one that sought to reverse the sport's decline and attract new and weakly held fans to the sport (Fujak et al., 2020). Its success has led other sports to follow it with their own short format events and competitions including Fast5 netball, Rugby 7's, GolfSixes, AFLX, and Fast4 tennis, as they seek to expand their audiences and/or reengage existing and younger fans.

Even having a mega-event as the pinnacle of the sport is not enough to buffer a sport from increasing global competition for audiences and participants. In endorsing a new event format for athletics, Sebastian Coe as President of the IAAF highlighted the vital role new events and formats have in a sport remaining relevant. Noting "we need innovation and more opportunities for our athletes to interact with fans and show their personalities – and Nitro Athletics is a great example of what can be done and what needs to be done to revolutionise how we present our sport and how our fans connect with the sport and the athletes". (World Athletics, 2017)

4.5.7 Sport measures

As a result of the review of the research literature and practitioners' contributions, 53 potential sport measures were initially identified (Table 14).

The measures were reviewed for inclusion in the research model according to the criteria outlined in 3.2.3 'Specification of measures'. The final measures chosen for rating by respondents are listed in Table 82 - Measure ratings: Sport

Table 14 - List of potential sport impact measures

Number of association member countries (count)	Infrastructure: Development of elite sport infrastructure	Infrastructure: Development of sport infrastructure
Infrastructure: Increased sport space	Infrastructure: Negative effect on mega-event development	Infrastructure: Urban regeneration

Sport development Increased sport funding ($)	Sport Development: Elite level	Sport products: Increased retail sales ($)
Athletes (count)	Officials (count)	Competing nations (count)
Sports (count)	Disciplines (count)	Medal events (count)
Prize money total ($)	Number of venues	Event prestige / ranking
Sporting infrastructure	Athlete Achievement: Development pathway	Athlete Achievement: Records set
Number of Disciplines (Count)	Number of Nations	Number of Sports (count)
Number of Venues (total)	Number of Venues (Existing)	Number of Venues (New)
Number of Venues (Renovated)	Number of Venues (Temporary)	Number of Officials
Total event duration (days)	Total number of competition days	Total Number of Competitors
Total Number of Competitors - Elite	Total Number of Competitors - Non-Elite	Total Number of Continents
Total Number of Female Competitors	Total Number of Male Competitors	Total Number of Sessions
Total Venue Capacity	Number of competitors excluded (drug testing)	Number of competitors excluded (game infringements)
Official charity partner presence	Commitment to Fair Play	Event ranking points classification
Increase in domestic association memberships (count)		

4.6 Pillar - Media

Media impact was described as the media scale of an event, defined by media presence and output.

In defining the scope of media impact from the literature review, it is noted:

- *The number of media organisations and staff is an indicator of the ability to generate original content and is likely to be proportional, but not directly linked, to the volume of overall coverage – and to some extent, social media volume.*
- *Prominence and accessibility of event content are influenced by differences in 'Mainstream vs. niche interest', 'free to air vs. pay vs. online', and 'live vs. delayed & highlights'*
- *The size of sponsorship and rights deals are driven by the popularity (attendance and media scale) of the event, but the majority of those rights are spent on the hosting of the event and are therefore in the economic pillar.*

4.6.1 Media Drivers

Examples of media sub-elements ('drivers') include the volume of coverage via press, broadcast, online and social media; the number of media organisations; production and technical staff; media prominence, access ('free to air vs. pay) and timing (live vs. delayed &/or highlights); and media engagement, reach, and consumption across media channels.

These media drivers are summarised in Table 15 and literature insights on each are briefly discussed in the following sections.

Table 15 - Media Drivers and descriptions

Pillar	Driver	Description
Media	Media output	Content volume (broadcast hours, column inches, posts/tweets/views); the number of broadcast rights holders with access to the content
Media	Audience size	Audience (event attendance, broadcast audience, news and online)
Media	Sponsors' output	Sponsor content generation; the number of sponsors, suppliers, and partners; longevity of partnerships
Media	Media presence	On-site accredited media personnel (journalists, TV media, technical staff, and PR)
Media	Media accessibility	Accessibility to media coverage (e.g. 'free to air vs. pay vs. online'; 'live vs. highlights')

4.6.2 Media output

The importance of the volume of broadcast content is highlighted by the IOC in their selection of metrics in the 'IOC Marketing Fact File' with significant growth in content noted even in the last six years. Olympic Broadcasting Services (OBS) provides Rights-Holding Broadcasters (RHB's) with venue content. The scaling of that content from the original hours across the RHB's global footprint is reflected in PyeongChang 2018 which saw 5,600 generated by OBS multiplied as "the IOC's global broadcast partners made more

coverage available from PyeongChang than any previous Olympic Winter Games, offering a combined 157,812 hours – an increase of 38 per cent from Sochi 2014" (IOC, 2019, p. 30).

The importance placed on the ability to generate that broadcast content is reflected in its inclusion in guarantees by the host. Drummond and Cronje (2018) seemingly places it alongside such foundational requirements as personnel safety noting "the government had to provide a series of undertakings and guarantees [including] the safety of FIFA delegates and guaranteeing telecommunications infrastructure, specifically for the international broadcast centre" Drummond and Cronje (2018, p. 61).

Within the media content generated across digital and traditional mediums, TV exposure remains a key channel to publicity of an event, with event owners recognising the relative value of different channels. "A news feature on a huge TV channel is worth infinitely much more than one thousand tweets" (Jutbring, 2014, p. 33). Since then, the growth of interactive and 'second-screen' participation increases rather than diminishes the value of the broadcast audience for event owners. "Our television viewers, they are the key stakeholders for us […] traditional media, TV, for instance, is still dominating." (Jutbring, 2014, p. 33).

It is not just the scale of media content that has shifted, but the non-competition behaviour of the players themselves under the influence of media demands. Reflecting on the changes in football goal celebrations over time Schirato (2013) notes that "the incorporation of football into the wider field of entertainment […] means that celebrations are now more or less required to be attention-grabbing, strongly affective and dramatic, because by and large the game sells itself (to fans, the media, business) as a passionate activity" (p. 13).

The power of broadcast media to influence perceptions is not limited to the sporting content alone but ancillary content presented during those broadcasts. Affirmed by legislative work to limit what can be included in broadcasts where regulations "have recognized the power of the broadcast media by enacting bans on all alcohol advertising in those media. In other jurisdictions, bans on the broadcast media pertain to specific beverages, print and outdoor media advertising rather than the full marketing mix" (Casswell & Maxwell, 2005, p. 346).

4.6.3 Audience size

The scale of the audience for an event is a key dimension in the classification of events. The first three levels of the Sheffield Hallam SIRC Event Typology classifications (Mega and Non-Mega) Types A, B and C are all described firstly as 'international or major spectator events' (Gratton et al., 2000; Gratton et al., 2006b) , denoting the importance of the audience in the event profile and scale.

For the global mega-events, the scale of that audience is seen in the IOC reported 3.6bn viewers for London 2012 and 2.1bn for Sochi 2014, with the claimed viewership down slightly for Rio (3.2bn) in 2016 and PyeongChang 2018 (1.92bn) (IOC, 2019, p. 25).

Increasing TV exposure has been considered the main driver of value for, and from, events. Seeing "television viewers, [as] the key stakeholders for us, the higher audience we have, the higher value the event will have and then we'll be able to negotiate a better contract with our sponsors and the television" (Jutbring, 2014, p. 33). That 'higher value' from the scale of the audience is critical to fulfilling existing commitments and for future development. The lag effect of success plays out in driving future rights revenues, with Rio 2016 broadcast rights at 2,868 (USD millions) being an increase in 300m over London 2012 (IOC, 2019), despite achieving lower actual viewership numbers.

While the attention is drawn to the mega-events for the scale of their global reach, the issue of audience size and reach are amongst key metrics for all events. For example, "the 2009 World Men's Handball Championship (WMHC09), hosted in Croatia featured over 400,000 spectators in 110 games, 1.58 billion in the cumulative TV audience, and 1,420 hours of media coverage on 90 channels" (Pranic et al., 2012, p. 237).

For smaller sports and small market share teams, the development of increased exposure is recognised as a vital pathway to building an attachment to their brands. This is noted by Doyle et al. (2013) where "sport marketers must place an increased effort on increasing their visibility and maximising their exposure. Small market share teams may wish to negotiate deals with popular venues and media outlets to ensure that team matches are broadcast to a larger number of people" (p. 294).

The use of social media to multiply that content and add incremental gains in the audience increases audience reach further still. Noting that even for mega-events with well-honed distribution infrastructure, opportunities still exist for leveraging emerging technologies. "The embedded nature of social media as a component of the London 2012 Olympics communication channels enabled the use of a wide variety of tools, ranging from mobile platforms to online and conventional broadcast media that reached an audience of 4.8 billion people around the world" (Halliwell & Freeman, 2014, p. 3).

While mega-events dominate media reach from their output, other sports that already have strong content volume are concerned with increasing their audience scale and reach. Rowe and Miller (2021) note that even historically successful US domestic sports "are seeking to cultivate new consumers, especially those in the more affluent parts of Europe. This strategy highlights the increasing inability of single nation-state markets to thrive in a globalizing MediaSport environment" (p. 255).

4.6.4 Sponsors' output

"Sponsorship involves an exchange between a sponsor and the event property; the property sells the right to associate with the event to the sponsor" (Mazodier & Quester, 2014, p. 17). However, the success of the sponsorship does not start only at the event but in building sustained associations over time. "A consumer's prior experience with a sponsor or event, or both, can trigger cognitive and affective responses that can impact the processing of sponsorship messages" (Cornwall, 2008, p. 32).

Sponsorships are a key source of event revenues, with The Olympic Partners Program (TOP) delivering 20% of IOC revenues at over USD $1bn for 2013-2016 (IOC, 2019). But in return and to ensure exclusivity for its partners, the IOC apply a vigorous defence of the rights to associate themselves with the global Olympic brand. The fees paid to the event rights holder to associate the brand with the event is not where the impact of sponsors output is generated from the event. What is actually included in those 'sponsorship rights' is not fixed and is often a complex mix of 'rights' that may or may not include any exposure assets for the brand. The exposure of the sponsor via logos and messages to audiences across channels is a key determinant of sponsorship effectiveness (Y. Kim et al., 2015) it as an advertising goal and hence sponsors must make additional investments to 'leverage' their rights for them to be effective. Even a sponsor that plans to spend the same amount on marketing their association as they paid for their rights. is behind the market, as competitors "spend an average of $2.20 on activating sponsorships for every $1 spent on rights fees" (IEG, 2016). Generating a return from that additional spend is "consistent with relationship marketing theory, the findings support the association between exposure and attitude toward the event sponsor and its products [and] suggests that repeat attendance at a sponsored event

is related to attendees' enhanced perceptions of the title sponsor" (Lacey et al., 2007, p. 251).

The fact that non-rights holding brands seek to insinuate themselves into an association with the event or property provides evidence about the effectiveness of the sponsorship relationship. This tactic, known as 'ambush marketing', "describe[s] the practice by non-sponsor firms, including direct competitors of sponsors, to mislead the audience about their identity and status as official sponsors. Ambush marketers act as if they were sponsors in order to derive the same benefits as official sponsors" (Mazodier et al., 2012, p. 194). The tactics of the ambush marketer such as "heavy media scheduling and outdoor saturation at the event location are common clues of ambush activities, [with] their common aim [being] the formation of the mistaken belief that the ambusher's brand is also associated with the event in some official capacity" (Mazodier et al., 2012, p. 194). While legal remedies through 'clean stadium' agreements are often seen as unwelcome demands by large event holders, without them the value of their sponsorships would be discounted if able to be ambushed. An approach to reduce the influence of ambush marketers has been the use of an 'official sponsor; designation, but that alone is not enough, with Brownlee et al. (2018) stating "an official sponsor designation, without additional activation, is limited in differentiating sponsors from non-sponsors" (p. 152).

So while sponsors spend and promotion of the brand-event association is essential, the event owner can also contribute value by showcasing its sponsors, because "knowing that "this type of event is sponsored by this firm" or that "this firm sponsors this type of event" are equally valuable to the marketing manager" (Cornwall, 2008, p. 31)pp31.

4.6.5 Media presence

There are two key perspectives on the presence of media at events. The first is the ability to generate first-hand media content for consumption and redistribution, as noted in the earlier driver of "Media output". The second is the additional contribution to the economic impact of the event on the local economy that can be overlooked in tourism models of event impact.

Certainly, at a mega-event, the scale of media presence is a substantial demand on accommodation and hosting infrastructure as with "London 2012 Olympic Games, there were forecast to be 13,000 broadcast journalists, 7000 print journalists, 12,000 non-accredited media" (McGillivray, 2013, p. 102). But for a framework to be applicable across events, the drivers need to be relevant to events of different scales. Despite not attracting the scale of media attention and presence of the global events, Barajas et al. (2016) identified that for even a regional 2-day event of Rally Ourense, "81 media organizations covered the event. This includes 20 newspapers, 9 of them specialized in sport, 16 television stations or TV producers, 7 radio stations, 10 photography agencies, 5 press agencies and 23 Internet press. In total 209 persons from these media were registered and 124 of them came from outside the city." (p. 128). This acknowledgment of the scale of media presence is not new it being included amongst the key metrics reported when describing the success of the 2009 World Men's Handball Championship in Split, Croatia, noting the event "featured 24 competing nations, 110 games, 10,000 local volunteers, 1,568 media personnel, and over 400,000 spectators, including 10,000 fans" (Pranic et al., 2012, p. 238).

Some have suggested that direct media presence is under pressure, with Bakardjieva et al. (2012) projecting the mass media and institutional gatekeepers are being circumvented by citizen reporters and commentators who provide first-hand, real-time coverage and non-hegemonic interpretations" (p. i). That disruption of 'citizen journalism' has not been borne

out, but rather "the combination of digital content and word-of-mouth from the social web … causes a steady amplification of the information being discussed" (Halliwell & Freeman, 2014, p. 4) and extends the content's reach. Hence, the effect of social media platforms is more in increasing the 'velocity of coverage' and as a multiplier of original content, as stringent media rights continue to restrict access to and/or could create legal protections to generations of original content around events to accredited personnel.

4.6.6 Media accessibility

For each event owner, the method and channels of access to event content – reflexively synonymous with broadcast coverage – often seeks to find a balance between 'free-to-air' access (which maximises the event's potential reach) and 'pay TV' rights (which provide significant revenues but to a restricted audience). Fragmentation of the media landscape and global audiences have increased the complexity of rights negotiation. In delivering the content into the channels/rights mix, some events have taken control of the content generation and packaging. Examples include the development of the Olympic Broadcasting Services (OBS) and by Red Bull in their end-to-end event packaging and management.

While the core focus is often on the broadcast content, the packaging of highlights offers the opportunity to raise the profile of the event and reach an expanded audience. Despite this, as Jutbring (2014) observes, "potential news coverage is not considered as important by the key stakeholders, including the DMO, as the TV broadcast, despite that the news media represent editorials spanning from international news agencies to sport blogs, via radio and newspapers" (p. 33). But access to media content is not purely one of economics. The shared experience of sporting moments is seen as integral to the cultural landscape, and as such is deserving of special protections, at times referred to as 'anti-siphoning'. Here the term 'siphoning' means taking a previously 'free-to-air' event and only making it available behind a paywall, where previously it had been available to the public at no direct cost. As Reddin (2017) notes, in Australia the Broadcasting Services Act 1992 "give free-to-air broadcasters a protected right over pay-TV broadcasters to first purchase the broadcasting rights of sports on the anti-siphoning list, which are chosen on the grounds of cultural significance" Reddin (2017, p. 55). Even before the advent of streaming services and internet-based distribution became challengers to the mainstream media, researchers noted that "the acceleration of mega sporting event documentation (or reporting) has gathered pace, with greater availability (and democratisation) of media-making technologies and a more sophisticated network of distribution" (McGillivray, 2013, p. 99).

This more sophisticated network of distribution has impacted legislators in their definitions of what qualifies a service as being 'accessible', when seeking to enact protections to ensure "certain events of national interest are available to view live and for free, by the widest possible audience"(Ofcom, 2019, p. 1). Indeed as the regulator for communications in the UK, Ofcom's (Office of Communications) 2019 report determined to update its definition of a 'qualifying channel' to include internet delivery for 'listed events'. Acknowledging that "individuals can receive television channels by means of broadcast TV as well as the internet on their main screen - the household's principal television screen, or, in the absence of that, a computer or handheld device" (Ofcom, 2019, p. 1) bringing recognition of the shift to mobile and handheld devices are a key part of how broadcast events are consumed in the modern era.

4.6.7 Media measures

As a result of the review of the research literature and practitioners' contributions, 46 potential measures were initially identified (Table 16).

The measures were reviewed for inclusion in the research model according to the criteria outlined in 3.2.3 'Specification of measures'. The final measures chosen for rating by respondents are listed in Table 80 - Measure ratings: Media.

Table 16 - List of potential media impact measures

Advertising rights ($)	Advertising total spend ($)	Audience size (individual viewers reached)
Audience size (total markets viewership potential)	Audience size (total viewership cumulative)	Broadcast rights revenues (% increase)
Broadcast rights revenues ($)	Media agencies present (count)	Media coverage (news items)
Media broadcast coverage (total hours)	Media impact (awareness of event)	Media impact (interest in visiting)
Media infrastructure (quality of facilities)	Media infrastructure (size of facilities)	Accredited media representatives (count)
Newspaper coverage (word count)	Newspaper coverage article count)	Social media - Facebook interactions
Social media - Google searches	Social media - Sentiment (positive vs negative) ratio	Social media - Twitter mentions
Social media (Facebook followers)	Sponsorship advertising ($)	Sponsorship advertising (hours)
Television ratings (% share)	Television ratings (count)	TV Hours produced (count)
TV hours watched: free-to-air	TV hours watched: pay-TV	TV Nations (count)
Broadcasters: total	Broadcasters: total live	Broadcasters: total delayed, etc
Media exposure value	Website: video streams (count)	Website: unique visitors (count)
Website: total visitors (count)	Website: page impressions (count)	Accredited media: domestic/international split?
Twitter followers (count)		

4.7 Pillar - Social

Social impact was described as the social scale of the event in terms of the advancement of social development.

In defining the scope of social impact from the literature review, it is noted:

- *Mass participation in grassroots sport or activity is considered more closely related to social impact than sport*
- *Promotion of an active lifestyle may include opportunities for community participation in lead-up events*

- *Outreach programs in human development embrace children's rights and development through sport, within that scope*

4.7.1 Social Drivers

Examples of social sub-elements ('drivers') vary widely in the literature but include perceptual elements of identity ('feel-good', public engagement and support) as well as more tangible outcomes, including educational or cultural outreach programmes, inclusiveness through community involvement in hosting (volunteering), and promotion of health and well-being through 'quality of life' legacies, for example, green space, health, safety and security.

These social drivers are summarised in Table 17 and literature insights on each are briefly discussed in the following sections.

Before exploring these details, it is important to mention that the earlier noted limitations in the binary representation (positive, negative) of individual impacts and legacies, is challenged in relation to social impacts from events. Arcodia and Whitford (2006) note that such a representation creates a contradiction when applied to the concept of social capital, observing that "while the literature recognizes that negative socio-cultural impacts may, and do, occur, the concept of negative social capital is an oxymoron because social capital does not develop in a community experiencing negative impacts" (p. 15).

Hence a framework like that proposed in the EIF that addresses impacts from the 'zero-budget' baseline perspective avoids this 'oxymoron'. The absence of the development of positive social capital (zero-based) is evidence of negative impacts when the baseline expectations (and presumed cases) is for an improved social capital as a result of the event.

Indeed, that all legacies in social capital are based on gains from the pre-event state is implicit in most models, but explicit in the EIF with the added caution of impacts and legacy outcomes needing to be thought of in relation to specific stakeholder groups before being determined as positive or not.

Table 17 - Social Drivers and descriptions

Pillar	Driver	Description
Social	Community cohesion and quality	Community shared and non-sporting infrastructure and better 'quality of life' (green spaces, health, safety, security, transport).
Social	Health and Active living	Health, well-being & active lifestyle, increased grassroots sport participation, community sporting spaces.
Social	Volunteering	Increased capacity including recruitment, management, training and skills development, visitor engagement.
Social	Goodwill/ Pride	Shared pride, sense of wellbeing and confidence amongst hosts.
Social	Social engagement	Community outreach programs e.g. educational, cultural, human development, children's rights and development through sport.

4.7.2 Community cohesion and quality

The development of academic methods for the testing of event impact on social structures signifies both the importance of this area and identifies key dimensions found to underpin the concepts of social impact.

The Festival Social Impact Attitude Scale (FSIAS) was developed to measure residents' perceptions of the social impacts from events (festivals in this case). During testing of the FSIAS by Delamere et al. (2001), the secondary factor analysis found two subfactors for each of the two core dimensions - social benefits and social costs. "In the instance of Factor 1- Social Benefits - the identified subfactors related to the general areas of "Community Benefits" and "Cultural/Educational Benefits". For Factor 2, the two-factor solution related to the areas of "Quality of Life Concerns" and "Community Resource Concerns" (p. 17)

As already noted in the Economics pillar, events and tourism are tightly connected. This draws further attention to the identification by Marcouiller (1997) that the impacts of tourism on discrete communities include "developing a local sense of place, community pride or image, and local quality of life" (p. 351). That notion of 'sense of place' links also to the Brand pillar, where the development of a distinct identity is a key outcome of event hosting.

Fredline et al. (2003) build on Marcoullier's observation of 'community pride' and establish it as a proxy for a measure of social capital; "sense of community pride, well-being, and stability is an indicator of the social capital of a community" (p. 25). This is reflected not just that social capital as a vital element in maintaining a healthy and vibrant civil society (Onyx & Leonard, 2000) but extends the role of social capital by stating that "social capital is a proxy for social cohesion" (Gibson et al., 2014, p. 113). There is consistency in those perspectives with that Fredline et al. (2003), when in considering how to assess the social impact of events, stated that the "emphasis in the 47-item scale was on the potential of a festival to develop social capital-type benefits and on costs associated with disruption and burden on the community" (p. 28).

The Social Impact Perception (SIP) scale (Small and Edwards, 2003) was developed as a tool to measure residents' perceptions of the social impacts of events. Building on that later, Small (2007) adds further clarity by drawing out that the social impacts should be divided into 'personal' and 'community' levels, noting "those that have a community-level impact include the three factors of "community cohesion and identity," "community growth and development," and "behavioral consequences." (p. 53). But the impact of sporting events on a local community may in fact be inverse to its scale, with Taks et al. (2015) observing small events have greater "potential for tighter social networks and connectedness of the local population with the event [which] makes non-mega events significantly different from mega events with regard to their effect on local communities."(p. 4)

In reviewing empirical work on Quality of Life (QOL) for residents and events, Kaplanidou et al. (2013) note that "while the topic of resident attitudes and perceptions toward expected event impacts has been adequately researched, […] there is a paucity of research with respect to the impact of events on QOL." (p. 633). Supporting research shows visitors rather than detracting from QOL can enhance it for residents (Chhabra & Gursoy, 2009), and that events should seek to create "satisfying social encounters between hosts and tourists to increase the impacts of tourism on residents' QOL"(Carneiro et al., 2018, p. 22).

4.7.3 Health and Active Living

Increasing sport participation as a legacy of major events is a benefit often promulgated by supporters in event bids (Frawley & Van den Hoven, 2014). It is attractive as a proposed legacy, given the attendant benefits of participation in sport or an increasingly active lifestyle has on the individual and the community, as the "value of participation in sports for individuals and society has long been established" (Girginov & Hills, 2008, p. 2097).

Indeed, the positive gains in physical and mental wellbeing from more active living at the personal level, coupled with the avoidance of rising health costs by public and private providers alike, makes a compelling case for a sports legacy. Unfortunately the results are not as conclusive as researchers have found the "legacy of the hosting the Olympics on increasing general sport participation is still open to debate" (Agha et al., 2012, p. 138). This has been in part due to a lack of clear and consistent measurement, as noted by Veal et al. (2012) in previous studies, even in nations with well-established sporting eco-systems like Australia. Against that background, it remains unsurprising that in planning forward for the London 2012 Games ambition to increase and sustain increased participation levels acknowledged "Olympic impact on sports participation, within the host city and more generally is reported to be positive only anecdotally. More detailed research has been largely inconclusive" (London Assembly, 2007, p. 47).

The LOCOG took a systemic approach to develop and invest in sports participation before the Games rather than relying on an ephemeral legacy of 'Inspiring a Generation'. Contending that in terms of sport, physical activity, or health legacy "no previous Olympic Games has raised participation levels in sport and physical, would not be entirely correct" (Homma & Masumoto, 2013, p. 1463) reflects on the quality of research hampering definitive answers being achieved. LOCOG's more proactive approach sought to resolve the outstanding elements noted by Veal et al. (2012) that "when reliable and consistent participation data are available, the question of causality in the context of the wider sport development and participation system remains to be addressed" (p. 155).

4.7.4 Volunteering

The role of volunteers in the legacy of events has two key perspectives: the impact of volunteers on the outcomes of the event, and the impact on the volunteers themselves as a result of their participation in the event.

Even without the presence of major events, governments recognise the value of volunteering for its potential contribution to society as both an input to community functioning and an output of community cohesion.

The notion of increased cohesion and connection is articulated as 'social capital', which Putnam (2000) defines as "social organization such as networks, norms, and social trust that can facilitate coordination and cooperation for mutual benefit" (p. 2). The specific role of volunteering within that development of social capital is noted by Welty Peachey et al. (2015) where "volunteering has the potential to produce a form of engagement and social participation, which creates social bonds and contributes to social capital" (p. 28).

As noted by Chen (2013), despite changes in government during the lead-up to the London 2012 Olympic Games, the value and role of volunteering in building deeper social connections was successively supported (Table 18) and its role in longer-term social contribution articulated; "for people who are not from London after sixteen days of the London 2012 Olympic Games [impact] seems to be restricted to intangible legacies, such as volunteerism"(p. 89).

Table 18 – Government commitment to volunteering leading up to London 2012

	Strategy	Goal	
Labour Government's volunteering priorities	Bringing the UK together through cultural and volunteering activities.	Using the London 2012 Games to inspire more people to engage in volunteering work and help other people.	DCMS 2008
Coalition Government's volunteering priorities	Encouraging and enabling people to play a more active part in society through volunteering	To get more people volunteering for their local communities.	DCMS 2010

Adapted from (Chen, 2013)

Importantly those intangible legacies are not created in a vacuum, but rather from deepening of existing social ties. The event is the platform, as Girgin and Tasci (2019) described, is the "reinforcement of the social fabric of the community through better quality of personal relationships derived from the extensive volunteer efforts frequently associated with the event" (p. 254).

4.7.5 Goodwill/Pride

Some of the foundational event legacy research on the intangible outcomes included dimensions of excitement and pride brought through being the focus of international attention (Burns & Mules, 1986). Despite that early positioning, later research by (Horne, 2007) determined the impact of hosting mega-events on the attitudes and beliefs of the residents of the host country to be one of the "known unknowns" (p. 88). Since then, work developed with more specific attention "devoted to the social utility of mega-events such as psychic income (i.e., the 'feel good factor'), which may translate into a sustainable legacy for a host society such as social capital (i.e., community connectedness)" (Gibson et al., 2014, p. 113)

While during an event feelings of "pride and self-actualisation associated with playing host to a major event" (Fredline, 2005), a challenge noted by Gibson et al. (2014) is that in measuring pride and goodwill, "temporal effects should be considered in future studies because attitudes before, during, and after the event could be quite different" (p. 114).

Perhaps it has been the inability to assign a monetary value to 'goodwill' that means it may not be universally recognised, despite its early recognition. However, as de Nooij and van den Berg (2018) observed "politicians with the ambition to host a mega sport event because it brings fun, pride, and happiness to the people could be making a sound decision, while economists are still insufficiently capable of capturing this effect in their cost–benefit calculations" (p. 69). But more than transient 'excitement' or 'fun', a successful event can deliver a legacy of self-belief. This was observed in relation to the 2010 FIFA World Cup in South Africa where Knott et al. (2015) found "The successful hosting of the event led to an increased national pride and self-confidence, with residents regarding themselves and their nation as more competent and capable of delivering large projects and overcoming infrastructural challenges" (p. 54).

4.7.6 Social engagement

Rather than follow the 'economics first' narrative around event legacy, Allen (2013) noted "that sport can be used as a tool to support and help realize social and economic

development." (p. 407) in reference to South Africa's hosting the 2010 FIFA World Cup. A perspective consistent with the London 2012 plan that also draw on the importance of human outcomes when seeking "to achieve a legacy in sport, health, education, tourism and business" (Chen, 2013, p. 12). This built and expanded on the earlier work on London 2012 legacy planning which had noted "that the delivery of any Olympic legacy will be contingent on enhancing health and well-being, housing, education, employment and arresting anti-social behaviour" (Girginov, 2012, p. 555).

In 'Sport for Development and Peace', Kidd (2008) dissects the social contribution of sports into its effect and the audiences in impacted, noting "the use of sport and physical activity to advance reconciliation and intercultural communication in regions of conflict ('sport for peace'); and the use of sport and physical activity to [focus] on strengthening basic education, public health, community safety and social cohesion and helping girls and women, youth-at-risk, persons with HIV/AIDS (PWA) and persons with disabilities (PWD) in LMICs" (p. 373). The latter is specifically aligned to the UN Millennium Development Goals. While much of the social legacy narrative focuses on mega-events, Kidd's focus brings sport back to a more specific and localised impact that meets a community's needs. The impact of smaller-scale events in more isolated areas is not less important. Indeed, as Bertella (2014) notes, small-scale events "[…] can fulfil a socio-cultural and psychological function, thereby reinforcing a community's sense of identity and enhancing local networks" Bertella (2014, p. 132).

When developed to work with, not be imposed on their host communities, sporting events can "enhance individual and collective capacities, improve efficacy, create social capital and, where poverty is implicated, promote social and economic justice and wellbeing" (Schulenkorf & Edwards, 2012, p. 380). In the development of that social capital, Gibson et al. (2014) credits (Putnam, 2000) for framing the role of active participation in community groups and its effect on social cohesion. Specifically concerning sport events, they note "his ideas are pertinent to nations seeking social legacies through sport events as he envisages collectivity and engagement being generated through participation in community groups and activities" (p. 116).

In contrast to event tourism strategies that focus on bringing external people to the destination, engagement strategies seek to take the event into communities to activate the 'collectivity' of communities. One such example is The Festival Melbourne, which was associated with the 2006 Commonwealth Games. Despite facing "a low budget and the constraints set by the CGF, the Festival exceeded organizer expectations and [by] incorporating open-air Games broadcasting, celebrations of Australia's Indigenous communities and other cultural exhibitions, the Festival utilized space within Melbourne city as well as in outlying and regional locations, building on the Games-inspired sense of domestic cohesion and pride across the host state of Victoria" (Byrne, 2014, p. 1210).

The perspective of residents can be remarkably resilient. Even if the promised economic legacies fail to materialise, the sense of cohesion from hosting an event can remain. In studying the attitudes of residents of Vancouver and Ottawa to different legacy categories at the 2010 Winter Games, Karadakis and Kaplanidou (2012) noted that in both cities, the evaluation of economic legacy decreased as time progressed as performance levels became evident during and after the event. From starting higher in the pre-event phase, participants reset their expectations of importance ranking and concluded that residents "felt that psychological legacies met expectations, while economic legacies failed to meet expectations over the three timepoints" (p. 260).

4.7.7 Social measures

As a result of the review of the research literature and practitioners' contributions, 50 potential social measures were initially identified (Table 19).

The measures were reviewed for inclusion in the research model according to the criteria outlined in 3.2.3 'Specification of measures'. The final measures chosen for rating by respondents are listed in Table 84 - Measure ratings: Social'.

Table 19 - List of potential social impact measures

Community benefit: Improved cultural and education legacy	Community benefit: Increased volunteer participation legacy	Community Benefit: Interest at grassroots level for sport
Community benefit: Investment into sport	Community Benefit: Social camaraderie	Community benefit: Social cohesion
Community benefit: Volunteering, Social responsibility	Create Jobs: Pre-Post increase in employment	Enthusiasm: Local identity and community spirit
Enthusiasm: Shared community spirit toward an event	Increased activity: Personal activity level	Increased Participation: General Population (%)
Increased Participation: General Population (Count)	Identity impact: influence of event on national identity	Participation rates: At the community level
Participation rates: Impact through event attendance	Participation rates: Via watching sport as a facilitator	Perceived benefit for children: Socialisation and cohesion
Perceived benefit for children: Youth activity participation	Perceived benefit for children: Youth sport development	Quality of Life: Pre/Post resident attitude
Social Capital: Amongst event host communities	Social Capital: Change in pre/post event	Social Capital: Derived from hosting event
Social Capital: Improved tourism networks host communities	Social Capital: Sense of Pride from hosting	Social Capital: Sense of Pride from infrastructure legacy
Social Capital: Supporting a winning team	Social Inclusion: Amongst different groups in community	Social Inclusion: Diversity inclusion programme
Socialisation: Experience into sport participation	Sport Development: Community level	Tourism: Increased 'legacy' of tourist destination pre-post
Tourism: Relationship of sport to tourism sector	Tourism: Visitor levels into host community (pre-post)	Well-Being: General population health
Volunteers: total	Volunteers: international	Volunteers: domestic
Public support: nation/region/city	Event awareness: nation/region/city	Infrastructure: Negative community benefit (cost/benefit)
Infrastructure: Negative use of community space	Capacity utilisation of event	

4.8 Pillar - Brand

Brand impact was described as the branding scale of the event in terms of the development of the host and event image.

In defining the scope of brand impact from the literature review, it is noted:

- *As few host places or events are completely unknown before an event, the brand impact is attributed to a change in brand as a result of hosting the event.*
- *A strong brand image is of economic value in enabling the host to compete for future growth in tourism, trade (business and FDI) and talent attraction; elements not captured in the economic pillar of direct spending.*
- *Assessment of the sentiment (positive vs. negative coverage) and message content of the media, not just its volume, is within the brand rather than the media pillar*
- *Development of the event's image is key to raising the impact of, and interest in, future editions of the event*

4.8.1 Brand Drivers

Examples of brand sub-elements ('drivers') include the building of profile and presence of the host and event in media; familiarity, associations and perceptions of host destination and the event; and the use of iconic elements, landmarks; and people for enduring image, expression of host's unique culture, or place.

These brand drivers are summarised in Table 20 and literature insights on each are briefly discussed in the following sections.

Table 20 - Brand Drivers and descriptions

Pillar	Driver	Description
Brand	Host attention	Recall of event location, landmarks, and the strength of association between event and host
Brand	Image building – Host	Positive associations of the host location (e.g. climate, safety, cleanliness, fun, friendly etc.)
Brand	Image building – Event	Positive associations with the event (e.g. popularity, excitement, prestige, atmosphere, quality, experience)
Brand	Host attractiveness	Interest in host location as a place to visit, to work, to invest
Brand	Host reputation	Reputation for the successful hosting of events and ability to attract future events

Branding is a key intangible asset of event hosting. The opportunity to present the host as a desirable destination for both tourism and longer-term investment are factors in placing events within development strategies. The more significant the event the more the effect is presumed to be on building the host's brand. "Under this approach, the integration of brands capitalizes on the heightened awareness generated by a major event and focuses on the development of positive experiences for the visitor that synergies between brands can generate." (Zhang et al., 2018, p. 112).

4.8.2 Host attention

The 'showcase effect' (Hiller, 1989) has made sport mega-events valuable international promotional opportunities, and hence there is a growing interest to host them (Horne & Manzenreiter, 2006). That perspective is echoed by Gratton et al. (2006b) in noting that "cities staging major sports events have a unique opportunity to market themselves to the world" (p. 44). But the authors also note the intensity of attention on the host is not driven by the city but by competition for content. "Increasing competition between broadcasters to secure broadcasting rights to major sports events has led to a massive escalation in fees for such rights which, in turn, means broadcasters give blanket coverage at peak times for such events, enhancing the marketing benefits to the cities that stage them" p. 44)

That competition for content has also meant the use of events as a means of promoting and positioning a host city or nation has become a highly congested and contested space. Hence the recall of event location, identifying or iconic landmarks, and the strength of association between event and host become critical components of the attention. (Knott et al., 2015)

It is harder to recall now, but the 1992 Games promise to 'put Barcelona on the map' (**Error! Hyperlink reference not valid.**) was not a given at the time. Most major event bids contain some statements about the profile of their city and that of the event. In their bid for the 2010 edition, New Delhi stated "hosting a sporting event at a scale such as the Commonwealth Games is a matter of international prestige for the country, and is bound to boost brand India" (CGF, 2010). Despite the prediction, whether the net brand effect was damaging, neutral or positive was left unanswered in the absence of an agreed method for assessing the brand outcome.

While powerful broadcast media is a key channel for conveying information about a destination, it is not the only pathway. In research on Macao as a Grand Prix event host, McCartney et al. (2008) found "social and personal were the largest sources of information on Macao (57%) […with] print and broadcast media (43%) (p. 60). It is noted that these channels "played a larger role in feeding information on Macao [than] print and broadcast media where tourism marketing and advertisement budgets are generally allocated" (p. 60).

While events and especially mega-events can drive elevated attention to the host destination, two cautionary notes are made about that attention. The first is that of course "hosts seek to maximise their media coverage and value, but the intense media scrutiny likewise exposes any problems to a global audience as the host destination can't control how their image will be portrayed" (Swart et al., 2017, p. 236). This is an important reminder to hosts that attention and positive image building are separate elements in assessing event outcomes.

The second caution for hosts is that while events are an effective vehicle for driving attention to the event location, Bob et al. (2019) observed that "most of the awareness was oriented towards the attributes of the event itself and did not seemingly transcend to other components of the brand image of [the host city]" (p. 8). This observation explicitly separates attention, host image and event image as discrete components to be managed in event hosting.

4.8.3 Image building – Host

The role of sporting events, especially large-scale events as a catalyst for "The image of an industrial city had to be renewed and replaced by another one. The celebration of the 1992 Olympic Games was the perfect opportunity and providing the ideal 'flagship' event to develop a new image for Barcelona" (**Error! Hyperlink reference not valid.**).

That sentiment of a sporting event as a catalyst for image renewal has continued to hold sway. As Li et al. (2011) found, successfully building destination image underpins successful outcomes on other legacy dimensions. They noted "if the Beijing Olympics went smoothly, it would be more likely to create a good image of China as a tourism destination. More international tourists would travel to China after the Olympics, which would enhance tourism legacies" (p. 292).

In assessing event outcomes, separating the level of attention an event generates from the image of the host is not just semantic. In an era of ubiquitous media and competition for a 'share of eyeballs', the volume of media content generation can be driven to focus on fulfilling broadcast commitments.

In highlighting the value of imagery embedded within that content, McCartney et al. (2008) position media content not simply as broadcasting a sport or event, but as a foundational element of communicating a place to a remote audience. They credit "Gunn's (1972) pioneering "dimorphic theory" of image [which] suggested that the image of a destination can be formed through either an actual visit (organic) or by externally received (induced) information from sources such as broadcast and print media and word of mouth" (p. 183).

While an image can be formed from 'induced' sources, "telecasting plays an active role in defining, shaping and changing national images around the world [but it does not] automatically translate into an image held by audience members [who] lack direct experience of, or information about, a particular nation" (Zeng et al., 2011, p. 321). But just as host destinations have become more cognisant of the power of event coverage to 'showcase' their place, so have the event owners and media partners. Indeed in seeking to extract the value from the media rights paid the "TV broadcast is considered valuable [with] little possibility for the [host] to influence the negotiated process of how the TV broadcast is planned" (Jutbring, 2014, p. 35). A host that recognises this should look for other pathways on which to build their image or consider how to embed the desired image elements into the event layout, structure, or delivery.

Positive elements of the host location (for example, climate, safety, cleanliness, fun, friendly etc.) are attributes that hosts, broadcasters and event rights holders would all seek to include in the content. Needing to combat perceptions of 'afro-pessimism', Allen et al. (2013) found "the 2010 FIFA World Cup presented South Africa …with a unique opportunity to showcase its development, competency, natural beauty, as well as its culture and diversity on one of the largest global platforms. (p. 2002)"

That power to reposition a host through an event is attractive and hence "in the field of international relations, there is growing evidence that states seek to host sports mega-events because, above all, they believe it will enhance their 'international prestige' " (Grix, 2012, p. 289)More directly, states seek "the connection between sports, soft power, and public diplomacy especially in the context of trying to improve a country image" (Dubinsky, 2019). The challenge for event assessment is the 'behind closed doors' nature of diplomacy and its timeline makes measurement of any direct effect very opaque.

A cautionary note on the effect a single event can have on a host is reflected in the observation from Swart et al. (2017) that while "a positive media impact [can] exist in the short-term, intended media legacies are impacted (and can be overwhelmed) by prevailing political, social and economic conditions in the host destinations" (p. 241). Acknowledging the capacity of external factors to disrupt a positive media effect is compounded when seeking to extend the timeframe over which the impact can be claimed "making it complex

and challenging to isolate and claim any specific legacy effects years after the event" (p. 241).

4.8.4 Image building – Event

In researching how the relationship between host and event worked in building brand image, Xing and Chalip (2006) identified that the degree of fit between a hosting destination and a sport event should influence the degree to which transference of brand image is effective, and reported three key findings. First, that the fit between destinations and events did not necessarily result in better evaluations of event and destination brand images. Second, sporting events had some impact on destination brand image, but the reverse was not true. Third, destination brand image did affect intentions to visit the host region.

'Fit' has become central to the viability of events and has come to challenge the researchers' second finding. The host image does matter to the event, especially in light of human rights concerns over numerous Olympic and FIFA editions between 2010 – 2022. The importance of how the event is hosted and by whom was summed up in a warning to event owners and sports associations by David Grevemberg, CEO of the Commonwealth Games Federation. He declared "the legitimacy of sport has spiralled negatively downward…due to corruption, deceit, broken promises, and the adverse impact of major events on host communities and their citizens. We have reached a tipping point …where we can either rebuild our relevance and resonance or swiftly become irrelevant and obsolete" (Waldron, 2016).

The absence of recurrent events from many legacy frameworks overlooks the time and investment needed to build an event brand. Identifying the distinct advantage of recurrent editions, Parent et al. (2012) highlight that "in contrast to the brand creation process for one-off sports events, it appears that recurring event organizers can allow themselves time to refine their brand after event execution and before their next event" (p. 156).

4.8.5 Host attractiveness

The promise of events on the positioning of the host for future gain is not new. As highlighted by Gratton et al. (2000) from a 1995 National Heritage Committee investment case report, "In addition the favourable publicity which can follow from a successful event may increase the attractiveness of a city, raise its profile overseas, and enable it to attract an increasing number of tourists" (p. 20). That perspective is shared even amongst strong tourism and investment destinations, which plan "leveraging strategies [to] deliver the three key legacy outcomes: tourism, trade and investment, and multicultural engagement" (Fairley et al., 2016, p. 467).

More foundational than simply drawing attention to a host, "increasingly, sports events are part of a broader strategy aimed at raising the profile of a city [and] in the case of many cities, is invariably linked to strategies of urban regeneration and tourism development" (Gratton et al., 2006b, p. 44). Anticipation of the success of events in building the host's attractiveness is why prospective hosts "aggressively compete for the opportunity to host [major events] in an effort to attract local and international investment, enhance the city's image, and generate revenue" (Sant & Mason, 2015, p. 42).

Not all cities do or should seek out major events. This might be because it is either not economically viable, or they wish to avoid the 'aggressive competition' and potentially overpaying for an event. By focussing on long-term talent attraction and avoiding major events and assumed tourism legacies, potential hosts can be successful in making their city attractive through a portfolio of 'right-sized' events.

Writing on what makes a city attractive to the 'creative class' of human capital talent, Florida (2003) found "people are not moving to these places for [the] sports stadiums, freeways, urban malls, and tourism-and-entertainment districts [they look for] high-quality experiences" (p. 9). It is a portfolio of events that together deliver a calendar of 'high-quality experiences' that makes a destination attractive to talent, rather than relying on a single event. While a sound strategy, it is not a binary 'either/or' for tourism and talent; hosts don't have to trade-off between the two. Westerbeek and Linley (2012) found that "when cities have a portfolio of events, the city is more likely to be seen as attractive both as a place to visit and as a place to live and work" (p. 202). Attracting talent underpins long-term growth because places grow "by their ability to attract people from the outside." (Florida, 2003, p. 10). However if hosts can leverage the right mix of event types and scale they can hope to maximise the city's long-term attractiveness.

One benefit of committing to larger-scale events is in positioning the host as an attractive location and capable host for future events. Thus ensuring "the ability to attract more national and international events to the host city is crucial in order to make efficient use of the resources generated" (Chen, 2013, p. 128). This means that building the broader hosting reputation during an event is vital, and directly connects 'attractiveness' to an economic imperative to maximise existing investments.

4.8.6 Host reputation

Following various criticisms of Athens 2004, Beijing 2008 and Rio 2016 editions of the Olympic Games, the IOC has awarded three consecutive editions to previous host nations Tokyo 2020, Paris 2024 and Los Angeles 2028 – four editions, if counting 2032 Brisbane as within Australia's broader hosting history. The pattern shows evidence of owners seeking a 'safe haven' strategy, where the reputation and capacity to host the Games is certain and the hosts less controversial. As noted by Heslop et al. (2013), "Mega-event organizers understand the importance of protecting and enhancing the mega-event corporate reputation and brand image in host site decisions" (p. 7). The reputation effect of a 'safe' host is a sentiment echoed in capital market responses to adverse events where "studies have found that high reputation is a benefit because of the stock of social capital and goodwill it generates" (Zavyalova et al., 2016, p. 253).

"However while sport mega-events put nations and their capability to host them in the global spotlight that same focused attention by the media can prove to be a double-edged sword for [sport mega-event] host destinations" (Swart et al., 2017, p. 235)pp235. Event rights holders may seek to mitigate the risk of poor execution by imposing formalised regulations "to ensure events are implemented in accordance with strict parameters"(Kelly et al., 2019, p. 123). Inversely, those "formalised regulations can restrict the ability of host cities to leverage benefits from an event" (Kelly, 2019, p. 49). Hence, cities seeking to bid for events need to balance both the opportunity to showcase their capability to host an event of that scale and work within the restrictions imposed that can limit the ability to leverage it for local outcomes.

A successful hosting history can be used to offset some more restrictive elements and open leverage opportunities. The reputational benefit from demonstrated capability in event hosting can give confidence to rights holders when "successful in their first attempt at hosting the event, they have some leeway […] should the event organizers have the opportunity to host the event again" (Parent et al., 2012, p. 155). In particular a strong hosting reputation can become a leverageable asset for a host given that "positive experiences of one host cannot easily be transferred to another, because the reputation

effects of an event are highly location specific" (de Nooij & van den Berg, 2018, p. 76). The implied uncertainty for rights holders may dissuade them from considering other hosts and still expecting the same or better outcomes.

4.8.7 Brand measures

As a result of the review of the research literature and practitioners' contributions, 57 potential brand measures were initially identified (Table 21).

The measures were reviewed for inclusion in the research model according to the criteria outlined in 3.2.3 'Specification of measures'. The final measures chosen for rating by respondents are listed in Table 86 - Measure ratings: Brand.

Table 21 - List of potential brand impact measures

Associations: Change in Foreigner non-visitor views of the host nation	Associations: Change in Foreigner visitor views of the host nation	Associations: Event brand attributes (%)
Associations: Host attribute attributes (%)	Associations: Pre/Post host image attributes (%)	Associations: Pre/Post host image attributes (rating)
Attitude: Impact of event involvement on event associations	Attitude: Impact of event involvement on host associations	Awareness: Event recall (%)
Awareness: Host recall (%)	Brand boycotting: Loss of sales from product boycott	Image: Host city attributes
Image: Transfer of image attributes from event to host	Change in host brand equity	Change in host city brand image
Change in intention to visit	Change in nation brand image	Sponsor brands recall
Athlete brand financial valuation	Event brand financial valuation	Host brand financial valuation
Brand-event fit: product/event brand congruence	Host-event fit: host/event brand congruence	Business attractiveness of host (change %)
Positive media coverage (% growth)	Volume change in media coverage (%)	The volume of media coverage (2years+)
National Pride: Host nation self-regard	Legacy perceptions of the host place	Sponsorship: Event influence on intention to purchase
Sponsorship: Value of brand/athlete exposure ($)	Sponsorship: Post-event sponsor product purchases ($)	Visitor experience: Intention to recommend (NPS)
Visitor experience: Intention to revisit	Associations: Event brand attribute ratings (%)	Associations: Host Brand attributes ratings (%)
Brand image: Host city attributes (% growth)	Change in host brand equity (% growth)	Change in intention to visit (% growth)
Increase in foreign brand spending within the host city ($)	The volume of media coverage	Number of Attendees at sponsor events (in-country)
Number of Attendees at sponsor events (in host city)	Number of Attendees at sponsor events (international)	Number of Sponsor Events (in-country)
Number of Sponsor Events (international)	Number of Sponsor Events (in host city)	Reputation: Change in Foreigner non-visitor views of the host

Reputation: Change in Foreigner visitor views of the host	Sponsorship: Increase in sponsor product purchases ($)	Visitor experience: Intention to revisit (%)

4.9 Pillar - Environment

Environment impact was described as the environmental scale of an event in terms of its environmental footprint.

In defining the scope of environmental impact from the literature review, it is noted:

- *Legacy of higher environmental building performance for event infrastructure acts as a model for driving higher standards in future public, commercial and residential buildings.*
- *Mitigation of additional waste is treated as a product of increased consumption from an event rather than a separate item*
- *Land and resource utilisation would include the redevelopment of brownfield sites; protection and development of green spaces and waterways/wetlands clean-ups etc.*
- *Transport strategies applied during an event including public versus private; energy/emission efficient modes etc. mitigate against raised environmental footprint.*

4.9.1 Environment Drivers

Examples of environmental sub-elements ('drivers') include efficiency in energy, water, CO^2, land and resource utilisation during the event, mitigation of the impact of increased consumption and travel during the event, positive legacies of best practice and improved technologies.

These environment drivers are summarised in Table 22 and literature insights on each are briefly discussed in the following sections.

Table 22 - Environment Drivers and descriptions

Pillar	Driver	Description
Environment	Energy efficiency and quality	Use of renewable sources of energy, recyclable and repurposed materials. Improved air quality during and after the event.
Environment	Consumption footprint	Reduced consumption and carbon footprint - public transport, electric vehicles, the proximity of the venue to accommodation, food miles and avoidance of landfill waste
Environment	Resource utilisation	Reuse, redevelopment or regeneration of wasteland and/or waterways, improved natural habitat & biodiversity, avoidance of habitat and wetland loss
Environment	Building design leadership	Energy & environmentally advanced event buildings, design leadership, raised building practices and labour market skills

The evolving field of environmental impacts has moved from an afterthought where "impact assessments…are almost always economic, with any environmental plan then relegated to minimizing the impact of an event" (Jones, 2008, p. 347), to a prime position in the 'greening of events' (Mallen et al., 2010).

However, the inclusion of environmental impact within the event impact dimensions also attracts accusations of 'greenwashing'(aimed at many CSR initiatives by large corporations), rather than being seen as serious environmental strategies and mitigations.

Nevertheless, progress has been made. Rather than simply 'greenwashing', the term 'event greening' better describes the twin disciplines of "mitigation of the direct environmental impact or 'footprint' of the event (including the carbon dioxide emissions, as well as waste created, water and energy used, biodiversity threatened, etc.), and the potential of the event to catalyse a broader societal and political shift towards more sustainable pathways" (Death, 2011, p. 101).

In a prescient assessment of the developing field of the environmental impact of events, Jones (2008) concluded "assessing the environmental impact of major sporting events is important and will become increasingly so as SD [Sustainable Development] commitments become enshrined in the constitutions of sports organizations and in the statutes of the public sector bodies often asked to support such events" (p. 355). The release of ISO 20121:2012 "Sustainable Events" (ISO, 2012), and its reconfirmation in 2017, is evidence of progress towards a structured approach to events and their environmental impact.

The two largest global events – FIFA World Cup and Olympic Games – now both require environmental plans in bid proposals. However, as Preuss (2013) notes, they differ in focus even there, as "FIFA only focus is on reducing the footprint of the event. In contrast, the IOC tries to reduce the long-term environmental harm by requesting ―a concept of sustainable development and ―assessment of the environmental impacts and legacies" (p. 3586). But as noted above, their pronouncements are not without their critics, with Boykoff and Mascarenhas (2016) denouncing such bid requirements as hollow as "IOC luminaries have veered historically toward "greenwashing" – the duplicitous practice of voicing concern for the environment and claiming credit for providing solutions while doing the bare minimum, if anything at all" (p. 2).

4.9.2 Energy efficiency and quality

The first stage of environmental measurement by events have been the local operational impacts. As noted by Collins and Cooper (2016), "environmental strategies developed by festivals have to date predominately focused on addressing the local environmental impacts of their activities and operations, such as increased use of renewable energy on site, waste minimisation and recycling, water conservation, and reducing noise and light pollution" (p. 3). Tackling the operational aspects is not new, and neither is it limited to infrequent events and festivals. Indeed, perhaps it is even more critical for more frequent events such as highly attended domestic competitions. Since 2005, Manchester City Football Club have been actively working on the improving waste management, sourcing more food locally and working on making it easier for fans to use public transport to get to games even installing "an 85 m wind turbine [] to power the stadium [with any excess energy] made available to local residents" (Collins & Flynn, 2008, p. 753). However, being a globally followed club should not be a requirement to undertake improvements. Scale should not be a barrier as proved by smaller clubs like Gloucestershire's Forest Green Rovers, who have gained recognition by FIFA and the United Nations as 'the world's greenest football club' (UN, 2018) for its initiatives related to renewable energy, locally sourced foods and the development of a new 'Eco Park' stadium built of sustainable wood. (FGR, 2021)

4.9.3 Consumption footprint

When thinking beyond the measurement of footprint to reducing the environmental impact of events, "practical and concrete physical flows of materials and energy are important, for example, fuel, energy and resource inputs and wastes and emission outputs" (Korhonen et al., 2018, p. 547) This is a challenge for events. Events require significant energy and physical movements of materials and people. The economics of event tourism is attractive but "unless the idea of bringing a lot of people together in one place is abandoned, it is hard to see how this environmental impact of travelling can be avoided, although it may be possible to mitigate or offset it" – which makes it "necessary to be able to measure the carbon footprint" (Chappelet, 2018, p. 25).

The adoption of progressively 'greener' events trying to look beyond 'offsetting', is reflected in the Sydney 2000 bid documents that sought work on "energy conservation and use of renewable energy sources; water conservation; waste avoidance and minimization; protecting human health with appropriate standards of air, water and soil quality; [and] protecting significant natural and cultural environments" (Gold & Gold, 2013, p. 3530). The recently awarded Brisbane 2032 has gone further again and promised a 'climate positive' Games (Climate Council, 2021) addressing energy, infrastructure, travel and transport.

4.9.4 Resource utilisation

The wider consideration beyond just a 'carbon footprint' considers that an extended view of impacts should be the goal. Moving from "carbon footprint, which measures the total amount of greenhouse gas (GHG) emissions [whereas] the ecological footprint is able to provide a more comprehensive assessment of a festival's environmental impacts as it includes an analysis of GHG emissions and also different land use pressures."(Collins & Cooper, 2016, p. 3).

Renewing and remediating land areas as part of events, goes from 'minimising operational footprint" to improving on the event environment as a legacy. "Selecting the Stratford site for the London 2012 Olympic and Paralympic Games provided an opportunity to regenerate a rundown area of historical industrial development and dereliction, as well as remediate the significant levels of contamination that had accumulated over a century and a half." (Hellings et al., 2011, p. 5). Recognition land usage and adapting events to avoid habitat loss and biodiversity impacts are not new. Three decades ago at the 1992 Winter Olympics, "the spectator areas were sited to take account of the most sensitive sites... special care was taken to avoid damage to the peat landscape, with pistes narrowed to 4 m rather than the usual 7 m, and tracks were re-aligned to avoid humid zones… felled tree trunks were removed by helicopter to be used throughout the site and the low-impact track vehicles were used in order to reduce damage to the wetland flora" (May, 1995, p. 271).

Contributing to social regeneration through the redevelopment of unused or underused land seeks to leverage investment in an event for long-term benefit. There is however a trade-of between disruption and sustainable benefits. To avoid displacing existing communities, Rogerson (2016) notes "increasingly cities bidding for major events have sought to find spaces where the impact on existing communities and functions is limited – usually vacant spaces within or at the edge of cities – but in turn this makes the ability to adapt and integrate such facilities for community use thereafter more challenging" (p. 507).

4.9.5 Building design leadership

Whilst events have added sustainability to the operational execution, some sporting facilities are being designed with an external endorsement of their environmental credentials as a target outcome. " 'Leadership in Energy and Environmental Design' (LEED) is a green building certification program that awards ratings based on points accumulated across areas including location and transportation, water efficiency, energy and atmosphere, materials and resources, and indoor environmental quality" (LEED, 2021). The Nationals Stadium in Washington, DC was the first LEED-certified 'new build', with Chicago's Solider Field awarded the same certification following its refit. (Ammon et al., 2016).

A barrier to including sustainability practices within event infrastructure is that some criteria do not apply easily to sports precincts and their intensive but short bursts of use, compared to other city infrastructure. However, Nationals Stadium (Washington DC) was challenged to achieve green transport requirements "to provide for 5 per cent of peak demand, so they created the popular "bike valet" service [which is] even being expanded this year, due to popular demand." (USGBC, 2016).

Design leadership on large scale stadia is welcome but such stadia still have relatively low utilisation compared to community-based sports facilities. Rather than simply measure 'operational footprint', the new La Trobe University Sports Park Stadium achieved "a 6 Star Green Star rating from the Australian Green Building Council [for] building material performance, natural ventilation, sustainable energy generation, water storage and reuse, and building management and operational strategies" (Campbell, 2021, p. 52).

This shift towards 'regenerative design' and greater consideration for reflecting the usage pattern for sports infrastructure may be required to encourage others to seek formal recognition and certification for design and infrastructure building, even if already embedded in existing practices.

4.9.6 Environmental measures

As a result of the review of the research literature and practitioners' contributions, 52 potential environmental measures were initially identified (Table 23).

The measures were reviewed for inclusion in the research model according to the criteria outlined in 3.2.3 'Specification of measures'. The final measures chosen for rating by respondents are listed in Table 88 - Measure ratings: Environment.

Table 23 - List of potential environment impact measures

Air Pollution (Air Quality - % gases)	Air Pollution (Pollutants %)	Air Pollution (Greenhouse gases (t))
Air pollution Total Co2 emissions (t)	Air pollution Reduction - energy efficient systems)	Air Pollution (Co2 emission - Visitor's travel)
Air Pollution (% reduction from better traffic flow)	Animals (Disruption of migration routes)	Biodiversity protection (% lost)
Construction impact (% environmentally friendly materials)	Ecological construction (Remake of old industrial sites)	Habitat loss (Forest, Wetlands cleared for construction)
Infrastructure (New structure, city beautification, higher standards)	Infrastructure (Upgrade to ecological materials)	Knowledge (Implementation of environmentally friendly practices)

Public transportation - passenger miles (total)	Public transportation - passenger journeys (count)	Recycling (New recycling infrastructure)
Recycling (Environmental awareness)	Renewable Energy used (%)	Urban renewal (Event areas %)
Visitor travel (Mode + distance at event)	Visitor travel (Mode + distance to event)	Visitor travel (congestion impact)
Waste - Renewable Energy systems	Waste - Total collected (t)	Waste - Recycled (%)
Waste (Waste-processing and compost-production plant)	Waste - Visitor garbage (t)	Water - Recycled/Recovered %
Water Quality (pollutants total)	Water environment (Public Perception)	Water environment (Reduction of habitats)
Air pollution Reduction - via energy efficient systems	Knowledge (Implementation of environmental practices)	Public transportation - visitor passenger miles (total)
Public transportation - visitor passenger journeys (count)	LEED Certification Status	LEED score - Water efficiency
LEED score - Energy and atmosphere	LEED score - Green infrastructure & buildings	LEED score - Indoor environmental quality
LEED score - Location and transportation	LEED score - Materials and Resources	LEED score - Sustainable sites

4.10 Political

Of the emerging themes identified through the literature review and discussion of the framework dimensions in 4.1, the dimension most challenging to capture objectively is Political.

The immediate concern was the absence of an agreed definition or scale of measurement although that might be able to be overcome in the longer term by proposing and getting agreement on elements for both. More fundamental is relevance and timeframe for political impact is at odds with trying to establish a framework that includes more events rather than excludes. As Cornelissen (2008) noted it is not all events that matter but "major sport events have reached particular significance in the contemporary global political economy: they have become important means whereby states seek to profile themselves for foreign investors… and channels through which states may seek to communicate messages to counterparts" (p. 481). The sporting events that register that strongly in a global context are major events which are Type A or Type B events in the Types A-E of the SIRC typology. Hence this dimension of political impact would only be relevant to a limited number of events and would actually reduce the ability to make comparisons across events.

Of further concern for the inclusion of political as a pillar is a recognition that "these events are seen as a means by which governments can achieve various objectives [including] external (for example, reimaging, re-branding, or international status) and internal (for example, nation-building or societal transformation)" (Gibson et al., 2014, p. 114). To include the political outcomes sought as measures of the political dimension will likely be confounded with dimensions already captured in other pillars of the framework, for example, Brand ('Host reputation') and Social ('Social cohesion').

By contrast, the level of Foreign Direct Investment (FDI) a nation attracts is a strongly quantifiable dimension that is not confounded within other pillars. Despite that strong objective dimension, there remains a challenge in crediting individual events with specific FDI decisions, and even then it would still only be relevant to a limited number of events of sufficient global standing and scale.

'Soft Power' and Diplomacy

One area not as liable to confusion with other pillars is the role in politics and sports of their use as opportunities for nations to develop or use 'soft power'. The term 'soft power' coined by Joseph Nye describes "the ability to affect others by attraction and persuasion rather than just coercion and payment" (Nye, 2017, p. 2). As noted above for politics more broadly there is an absence of objective and available measures for 'soft power' that can be linked to specific events consistently. This is further complicated by nuances in framing the impact of sport and diplomacy as noted by Jackson (2013) with the need to "consider diplomacy *within* sport, diplomacy *for* sport, and diplomacy *through* sport"(p. 176).

While there is recognition of the sport and politics are inextricably intertwined "sport continues to occupy a rather ambiguous position within the context of politics, foreign policy and diplomatic relations" (Jackson, 2013, p. 274). The lack of clear and consistent measures that are attributable to individual sports events means that politics is not included as a pillar in the development of the Event Impact Framework.

4.11 Assessing content importance

The identification and application of content and substance to the structure grounded the conceptual framework within the literature and research on event impacts across six pillars and the 30 drivers that underpin those areas of event impact. In addition, the measures referenced in relation to those dimensions were also captured. This anchors the framework from a 'bottom up' perspective of granular measurement.

Despite the work in developing the framework to this stage, there is no implicit or explicit importance or weighting placed on any of the pillars, drivers, or indicators in their contribution to event assessment. There were some indications from the formulation phase that different areas were perceived to have differing weights of importance within the working group, but there was neither an attempt to formalise that nor was the group of a sufficient scale for it to be reliable.

Hence the next stage of the development of the framework was to establish relative weightings for each of the Pillars and their sub-areas of drivers in contribution to assessing event outcomes. Being consistent with the outlined structure, the indicators were also planned to be assessed for their importance and reliability as the basis for measuring event outcomes.

Given the contestedness of the field and the diversity of views, the development of weightings was planned to engage a global expert panel to work through a consensus forming 'Delphi' technique as outlined in the 'Research Methods' chapter that follows.

5 Research methodology

As noted in the literature review, the terminology for the associated and myriad of impacts the hosting of events can have on the organizer has morphed to one of 'legacy' but without any agreed definition of either the term itself or its constituent components (Gratton & Preuss, 2008). The domain of event legacy was shown to meet the criteria established by Gallie (1956) for an Essentially Contested Concept.

That the concept of legacy is well established but remains contested presents significant challenges when seeking to research and establish a framework for comparison of event impact. A robust research design and methodology would need to resolve three key challenges:

- *Firstly, the research needs to be able to establish a 'consensus' viewpoint in a domain that is in a contested space and without an agreed definition.*
- *Secondly, the research method need to maintain sufficient breadth across event outcome areas without losing specific detail on the items requiring responses.*
- *Thirdly, it needs to distinguish between responses from different respondent perspectives without compromising anonymity or biasing responses.*

To achieve this, the design combined the well-established Delphi method as the overarching structure for consensus forming, and the use of 'direct' method questions which are noted as being well-suited to 'anchor free' ratings. These conditions were met throughout the study.

5.1 Research project structure

5.1.1 Project partnership

The Global Sport Impact (GSI) project was a partnership established in 2015 between Victoria University and Sportcal UK that sought to explore how the impact of sporting events of different scales and types might be assessed. The collaboration was important to Sportcal as they were keen to understand the current state of research in the field of event impact to inform future developments of their Global Sports Index tool.

Sportcal was the lead agency on the survey and communications between the expert panel and the research company fielding the study. Participation in the research was voluntary and completed at each respondent's discretion. From the project, Victoria University received anonymised third-party data output from each round of the study, the use of which was approved by the Victoria University Human Research Ethics Committee (VUHREC) under project ID: HRE21-126.

The project collaboration was concluded in Q4 2018. There is no current agreement in place for a follow-on project.

5.1.2 Overview of participating institutions and participants

Sportcal Global Communications Ltd. (London, UK)

Founded in 1991 to provide a calendar of sporting events around the globe, Sportcal has become a leading provider of sports market intelligence. The company delivers sports analysis and insight to its base of global clients, with sports market intelligence spanning 190 nations on 27,500 annual events each year in over 185 sports.

The participants in the research phase were employees of Sportcal from management, research, and client roles, as well as a specialist external adviser.

Victoria University (Melbourne, Australia)

Victoria University is a large higher education and research institution based in Melbourne, Australia. It conducts undergraduate and postgraduate teaching and research and is ranked 9th globally for Sport Science by the Academic Rankings of World Universities (ARWU).

The University has built strong industry partnerships with national sporting organisations including the Australian Sports Commission, Sport Australia, Tennis Australia, Swimming Australia, Sport Integrity Australia, and with other sport universities such as Beijing Sport University (China), German Sport University (Germany), Real Madrid Graduate School (Spain), Auckland University of Technology (NZ), and Loughborough University (UK).

Lightspeed GMI (London, UK)

Lightspeed GMI (a division of Kantar Group) is a globally integrated research organization specializing in online platforms to provide research and insights. In addition to its global audience panels, Lightspeed GMI also provides custom research to address specific client needs. In April 2019, Lightspeed and all other group subsidiaries were unified under the Kantar brand name.

5.1.3 Research Ethics

The structure of the research project and the receipt of third-party data from an independent external provider limited the potential for ethical exposure in the investigation of the research area. The primary focus was to ensure the research fieldwork maintained academic standards for 'informed consent' in communication with the expert panel, and that the fieldwork data was to be fully anonymised before delivery from the research agency.

5.1.4 Research model selection

Despite the contested nature of the field, the goal was to resolve the expert panel's views on how impact might be measured and critically weighted. A 'Delphi study' approach was identified and selected as a recognised "group facilitation technique…designed to transform opinion into group consensus" (Hasson et al., 2000, p. 1008)

Critically it is not just the transformation of opinion, but acceptance of the consensus that is powerful. Particularly in seeking to reduce the 'contested' status of the field, the application of the Delphi approach has shown "the method produces useful results which are accepted and supported by the majority of the expert community" (Czinkota & Ronkainen, 2005, p. 122).

A further element in the design considerations is that researchers have recognised classes of Delphi studies including Rauch (1979), "who worked out three ideal types of the Delphi method - classical, policy, and decision Delphi" (Florian, 2000); and as an applied approach of idea-generation or as a judgment function (Häder & Häder, 2000). This research is focussed on the latter, or classical type of Delphi, as it seeks to elicit the consensus judgement of the expert panel, which as is noted requires "greater responsibility because it must select and formulate items that will be evaluated by the experts" (Nowack et al., 2011, p. 1607).

That responsibility echoes the admonition of Zyphur and Pierides (2019) for quantitative researchers to consider "the purpose and consequences of their actions. This includes how

variables are defined, decisions about which analytic strategy to use, what counts as observation or measurement"(p. 50). Hence the selection of a quantitative Delphi study in the form of a 'judgement function' may exclude some subjective variables or elements that are not compliant with that model. Any exclusion does not connote any lesser value as an event outcome, but rather reflects the constraints within the model selected for this research.

The Delphi method is a widely used and cited technique and has underlying principles but is not pre-designed in a structure that relies on the researcher to consider the method required to align with the panel's engagement and achieve robust and reliable outcomes.

For this research, the design required consideration of dimensions such as:

- *Number of rounds required*
- *Content of each round*
- *Maintaining content relevance*
- *Definition of consensus*
- *Minimizing respondent fatigue*
- *Specific question types to be used*

These will be reviewed in the outline of the methodology applied.

5.2 Research method

The Delphi method established at the Rand Corporation in the 1960s is a "technique to apply expert input in a systematic manner using a series of questionnaires with controlled opinion feedback. key features are a preservation of anonymity in the expert panel's responses and iteration of the questionnaires" (Linstone & Turoff, 2011, p. 1712)

Those principles of a systematic approach require an "iterative multistage process, designed to transform opinion into group consensus" (Hasson et al., 2000) to be developed, and is especially effective where the "problem does not lend itself to precise analytical techniques but can benefit from subjective judgments on a collective basis" (Linstone & Turoff, 2002). A description that accurately captures the underlying challenge of decoding a 'contested concept'.

That process of collating subjective judgments by "expert input in a systematic manner using a series of questionnaires with controlled opinion feedback. Key features were preservation of anonymity in the expert panel's responses and iteration of the questionnaires." (Linstone & Turoff, 2011).

Even though 'iteration' is a vital component of the Delphi process, there is no set definition of how many rounds a study must contain. The minimum (initial and a single iteration) is not unusual, with Nowack et al. (2011) noting "the majority of the identified studies conducted two Delphi rounds. Five studies conducted a third round and only one conducted a fourth Delphi round." (p. 1611)

The iterative process, whilst central to the approach, is alone not a sufficient condition for a study to qualify as a 'Delphi study'. Rowe and Wright (1999) determined that for a qualitative study to be called 'Delphi', it must possess four core elements: "anonymity, iteration, controlled feedback, and statistical aggregation of group response" (p. 354).

5.2.1 Anonymity and informed consent

Several potential issues stem from variations in the understanding and degree of anonymity being applied when conducting a Delphi study. The concept of 'informed consent' is fundamental to research ethics (Israel & Hay, 2008) and requires that "research participants need to understand that they are authorizing someone else to involve them in research and what they are authorizing" (Faden et al., 1986, p. 605).

Conforming with Delphi protocols requires that anonymity "ensures that experts can express their opinion without perceived social pressure of other survey participants." (Strauss & Zeigler, 1975, p. 254). However, the consent given to participate in a Delphi study may hold differing interpretations of 'anonymity' for participants. Anonymous may be as being synonymous with unidentified, nameless, or unknown, and may or may not align with the researcher's expectation or use of anonymity in the study.

Participants are usually sought as content experts and when giving their consent may interpret a promise of anonymity as the literal understanding of that term – that is full anonymity (Ogden, 2008) where the researcher can identify neither the participant nor their responses. That is, a fully blind study. In this case, the consent accords closely with the common understanding of the term.

It is not uncommon for studies to be conducted under a state Ogden (2008) refers to as 'partial anonymity'; whereby both the participants and their responses are known to the researcher but not identifiable in the findings through the aggregation of results. For potential participants, 'partial anonymity' is closer to confidentiality. That is, the "information shared with researchers will not be disclosed in a way that can publicly identify a participant or source" (Kushner, 2005, p. 75) – rather than being fully unnamed or unidentified per se.

The challenge is that 'partial anonymity' is often in common usage within Delphi studies – creating the potential for misalignment in expectations between the participants and the researcher who assumes expectations are shared. The assumption erodes the power of gaining informed consent. To meet both conditions of informed consent the participant must be both voluntary and to "know what they are authorizing" (Faden et al., 1986, p. 605).

Table 24 – Levels of anonymity as applied to Delphi studies

	Full Anonymity	Partial Anonymity	Staged Anonymity
Description	Participant responses are not individually identifiable by the researcher or other experts.	The researcher can identify responses from each participant. Experts are unknown to each other.	Participants start as fully anonymous but move later to open discussions.
Use	Useful with a small group of experts where the risk of being able to identify individual responses and/or other participants is high.	Larger panel size and diverse participation base and where the researcher is outside and independent of the expert panel.	Scenarios work that moves from individual responses for concept development to open debate for reaching consensus.
Research benefit	Participants are free to make and change	Can identify the source of outliers and	Open and robust discussion, shortened

	responses over iterations without repercussion. The researcher can be from within the expert network.	direct follow-up iterations on expediting consensus. Reduces iterations and respondent fatigue.	timeframes, fewer iterations. New concepts [creative synthesis] are added during discussions.
Challenges	Minimised exposure to issues of data protection. Extended iterations could impact the validity of responses from the panel.	Ensuring respondents do not feel targeted or pressured to conformity through focussed iterations but genuine panel consensus.	Participants' prior responses are exposed to critique by others. Power structures may shift research dynamics and integrity.

This study used Full Anonymity as identified in Table 24 above and was compliant with the conditions Rowe and Wright (1999) identified for a Delphi study, namely anonymity, iteration, controlled feedback, and statistical aggregation of group response.

Research informed consent

As noted above, the explicit and shared understanding of the conditions of anonymity and consent need to be defined for each study. That explicit description for the basis of participation strengthens the degree to which 'informed consent' could be considered as given - no matter which form of the Delphi technique is used.

On each communication sent to the panel members, the following statement of consent was included: "Your participation in this study is completely voluntary. Survey responses will be strictly confidential and data from this research will be reported only in the aggregate. Your information will be coded and will remain confidential."

All communications – both initially emailed survey invitations and each reminder invitation - confirmed the voluntary basis for their participation including the receipt of communications from the project and the option to be removed from any future communications.

Hence the following protocols were actioned.

- *All respondents choosing to complete the survey were shown the statement of voluntary participation before being required to click a confirmation link to access the survey.*
- *Even once started, participants were advised that they could choose to stop and withdraw at any time without completing the survey.*
- *Participants who had started but not completed the survey could also return at a later time until the round was closed, removing the pressure to complete in a single pass.*
- *Any questions or concerns from participants were collected and actioned through the email address gsi.experts@sportcal.com, which was provided on the invitation screen.*
- *Any requests received to be excluded from current or future survey contact were redirected to the research company for removal from the mailing database.*

5.2.2 Definition of consensus

A defining quality of the iterative Delphi process is to seek a narrowing in the responses towards a shared viewpoint. A clear understanding of what constitutes a consensus might

be expected to be defined before starting the study. Unfortunately, a "failure to offer an interpretation of the meaning of consensus is an important omission in many examples of Delphi studies." (Powell, 2003, p. 379).

A lack of anchoring expectations for participants raises concerns of increased respondent fatigue, as they are unclear on when consensus is nearing and question iterations will cease. However, Delphi iterations do not have to continue until an arbitrary consensus is reached. Instead, the vital element is that the "process is usually reiterated until stability in the responses is attained, but not necessarily when consensus is achieved" (von der Gracht, 2012, p. 1527). From an ethical and research quality viewpoint, participants need to be made aware that while an agreed consensus is the targeted outcome, their participation is not without practical boundaries in that process. Hence to provide certainty to participants, this research defined a fixed number of rounds required of the expert panel and aligns with von der Gracht's advice of looking for 'stability in the responses; rather than 'when consensus is achieved'.

By defining both a 'consensus range' and a fixed number of iterations, clear boundaries were provided for their granting of 'informed consent'. The focus of the study was not on continuous iterations until exhaustion of variations, but rather testing the degree to which consensus could be achieved within a constrained but compliant Delphi model.

5.3 Research design

As the research was a 'judgement function' application of the Delphi approach, the work in determining the primary dimensions of the model was based on a multi-stage collection and refinement process.

5.3.1 Research objectives

The core objectives of the Delphi study were threefold:

- *To establish a hierarchy of importance and through the weighting of the Pillars;*
- *To establish the most important Drivers of impact; a hierarchy of importance through the weighting of the Drivers within each Pillar;*
- *To establish which are the most important and reliable Indicators for assessing event impact*

5.3.2 Scoping workshops

As outlined in the previous chapter, the development of the core research content was grounded in the primary source literature review of 200+ papers on differing dimensions of event impact and meta-analyses of the literature (Mair & Whitford, 2013a; Thomson et al., 2019), which established a list of 300+ potential measures.

Representatives from academic institutions, government, event practitioners, and sports associations met in two workshops conducted by Sportcal in London, UK and Lausanne, Switzerland to discuss the literature identified dimensions of event impact.

The outcome of the workshops was an agreement that the major elements of event impact could be represented overall by six core areas (Pillars): Economic, Social, Sport, Media, Brand, and Environment, noting that Tourism and Infrastructure were also considered as potential Pillars.

5.3.3 Traditional vs. 'Real-time' Delphi

Over the 40 years since its development, tools have evolved from paper-based to online moderated discussions, and recently to "round-less" or 'real-time Delphi' (Gordon & Pease, 2006). A real-time study offers the advantage of a significantly reduced timeframe and lower dropout rates from participants (Aengenheyster et al., 2017) without lowering the quality of the outcomes.

Consideration was given to this approach, but engaging a global panel across time zones without disadvantaging any group made a coordinated real-time program difficult.

The selected method for Delphi delivery was:

- *An online study consisting of three rounds.*
- *Launched on pre-determined intervals*
- *Smart routing of questions based on previous answers to reduce fatigue and maintain quality of responses*

5.3.4 Delphi study scope

Constraining items for rating

The background work in establishing current impact measures produced resulted in an unworkable number of potential Indicators to be rated in the Delphi process.

The combined effort of the literature review and the input from the working group identified an initial list of 321 potential variables distributed across the six Pillars, of which 188 were classified to be direct versus 133 indirect measures of impact.

Table 25 - Number of potential Indictors per Pillar

Pillars	Potential Indicators identified	Indicators included in final research
Economic	63	40
Media	46	39
Sport	53	37
Brand	57	35
Social	50	39
Environment	52	33
Total	321	223

As noted in 3.2.3 "Specification of measures", the focus on the attributable impact from a single event meant the exclusion of multi-year/decade induced measures of legacy presented in some conflicting models in the literature.

Despite reducing the Indicators consistently across all pillars, at that scale the study still represented a significant burden for respondents and potentially would have negatively affected the quality of responses through respondent fatigue.

Managing respondent engagement

For respondents to accurately weight all Pillars, Drivers and in particular, all 200+ Indicators were determined to be unworkable. If pursued, it was likely to compromise the quality of the research if implemented in a blunt approach.

The immediate concern in seeking to present all dimensions and measures for all respondents to rate was the potential for high respondent fatigue and hence poorer quality of responses. Such a demand would also lead to lower overall completion rates and a smaller usable sample base when partial completes were excluded. Likewise, the ability to reengage respondents to participate in follow-on rounds within the iterative Delphi approach would be severely compromised, as would the opportunity to engage the panel in any future studies on the same area.

Finally, while all respondents had expertise in some aspects sport and sport events, their most relevant expertise would not be factored into responses if asked to rate all items. Under a randomised model, the topics most aligned to their expertise might only appear after having been fatigued by other items in which they held limited expertise or experience

Reducing respondent fatigue

Several alternative methods were considered for avoiding respondent fatigue without compromising the study quality or purpose. All respondents might still see and rate the Pillars at the top level, hence reducing the work for respondents was in either limiting how many Pillars they would rate in detail, or limiting the detail they would see under each Pillar.

By reducing the number of Pillars shown to individual respondents on which they rate the detailed variables would significantly reduce the work required. Methods for the allocation of a subset of Pillars considered included a simple randomised allocation and allotment based on self-nominated areas of expertise. That is, for respondents to work on areas where they have felt they have the greatest experience and expertise.

The alternate pathway was to show all six Pillars to respondents but provide them with a subset of variables to rate within each Pillar. That subset could be a randomised subset of 15-20 indicators from the Pillar total, or a pre-processing by a select panel to choose 15 Indicators within each Pillar as being important to include in the model, with the respondent's rate for importance being the highest frequency Indicators selected.

The determination was made that to maximise engagement without reducing the scope of variables was to allocate respondents to Pillars based on self-nominated expertise. The treatment of each Pillar as a workstream (communities of expertise) for Round 2 and Round 3 and maintaining all measures was seen as the most likely successful strategy for retaining the study's full academic rigour within a practicable execution.

5.3.5 Workstreams within expert panel

The development of workstreams within the expert panel was established through the participants own nominations. The limiting of respondents to their field of expertise is not a new approach. Indeed Grupp et al. (2000) noted that "the group of respondents is knowledgeable in the entire field, but one cannot assume that their expertise covers all sub-areas and individual topics…it is, therefore, advisable to determine the expertise for each individual assessment task separately" (p. 58).

Hence during Round 1, all respondents were asked to nominate which two areas (Pillars) they would like to work on in participation in future rounds. Based on their self-nominated areas of expertise, each respondent received invitations to Rounds 2 and 3 in which they were shown their two nominated Pillar areas and their underlying Drivers and Indicators.

Workstream allocation

An alternative to the simple nomination of two areas was for respondents to nominate up to three (3) areas of expertise and then assign each respondent to two (2) of their nominated areas. This approach might allow the research to balance the number of participants in each workstream and help to maintain sample sizes in each workstream for Round 2 and Round 3 of the study.

Whilst sound in principle, the considerations of efficiency of delivery, independence of the process, and having respondents missing being able to respond to one of their core knowledge areas, resulted in the simpler approach being favoured. Addressing sample depth could be achieved in future studies by expanding or targeting greater breadth in the range of respondents invited.

5.3.6 Choice of rating methods

At the core of the development of a consensus was the rating of the event impact elements Pillars and Drivers and assessing Indicators across the workstreams. Several common question styles were used in the survey design including 5-point Likert scale and ranking choices, but the choice of the rating methods within those question styles was important to the quality of outcomes.

Within the rating and ranking questions, two techniques were identified as being particularly appropriate for the research: (1) Constant Sum and (2) Best/Worst Scaling (BWS). Both are classed as direct methods of relative rating or 'ratio scaling' techniques and were applied to tasks requiring the relative weighting of areas concerning impact. That these methods are both classed as 'direct' methods were important for consistency considering the methods applied within the Delphi framework.

In testing the reliability of Constant Sum and Best/Worst Scaling methods against other ratio scaling techniques, Louviere and Islam (2008) found "high agreement within direct or indirect methods, but less agreement between direct and indirect methods". In particular, the authors found "inferences derived from indirect measures appear to be susceptible to context effects"(p. 910). Given the already contested nature of the topic of legacy, the use of two direct methods was selected for (1) internal consistency between approaches, and (2) to avoid any susceptibility to "context effect".

Constant Sum Method

The Constant Sum Method was used for developing the relative weightings for both the overall Pillars and the relative weightings of Drivers under each of the Pillars.

The constant sum or 'fractionation' method was first proposed in the 1940s by Metfessel (1947) as a method for measuring comparative judgments. This was further developed and validated during the 1950s by Comrey (1950); Dudek and Baker (1956).

In the study, respondents were provided with an allocation of 100 points in which they could apply in any proportion to represent their view of the importance of each dimension in assessing event impact. While any fractional weighting could be applied across the dimensions, the total sum of the weightings would always equal 100 points – a constant

sum. To avoid zero values being recorded, a minimum value of one (1) point the lowest allowable value to be allocated to any Pillar or Driver. Hence a maximum of 95 points was possible for any Pillar and a maximum value of 95-97 points for any Driver. Scores for the relative weightings were presented back to respondents in the subsequent round, with the constant sum method reapplied to retest consensus scores.

Best/Worst Scaling (BWS) Method

Best/Worst Scaling (BWS), first proposed by Finn and Louviere (1992), was the other direct 'anchor-free' method used in the research study. As a discrete choice experiment (DCE) method, BWS does not use category-based scales. This is particularly valuable if respondents are required to make choices between items where no reliable rating scales are available.

In addition, BWS has been noted to overcome rating scale limitations in that "rating scales do not force respondents to discriminate between items, allowing them to state that multiple items are of similarly high importance" (Flynn & Marley, 2014, p. 3). By contrast "BWS addresses these issues by valuing items within a random utility framework" (p. 3).

The use of BWS in the Delphi study was in comparing the relative importance of specific Indictors against each other – a task requiring discrimination between items without an anchoring scale or guide.

5.4 Survey instrument

The core elements of the study being researched (and hence the data that was anticipated to be made available) established over the rounds of the study were as follows:

- *Assessment of Pillar importance and consensus weightings*
- *Assessment of Driver importance and consensus weightings*
- *Testing of Indicator ratings of importance and reliability*
- *Panel profile and demographics*

5.4.1 Delphi survey structure

The approach in developing the weighting content for the Event Impact Framework was to conduct the research in three rounds. Each round would have a specific focus and seek to develop a consensus viewpoint and/or Delphi stages and content.

Table 26 - Dimensions of research by Pillar

	Pillars	Drivers	Indicators
Economic	1	6	40
Media	1	5	39
Sport	1	5	37
Brand	1	5	35
Social	1	5	39
Environment	1	4	33
Total	6	30	223

Definitions for research

To ensure consistency and agreed on understanding (not necessarily the same as agreement) as to how named items were being defined in this context, all Pillar and Drivers were defined and shown in context for the respondent whenever they were required to apply a weighting to their relative merit. (Appendix 0 - Appendix 1 - Pillar definitions provided to respondents).

While practically useful in not assuming a shared interpretation of any named item, the clarity in definition also contributed to Informed Consent in that respondents were fully informed as to the nature of items they were required to rate.

5.5 Delphi content by Rounds

5.5.1 Overview of survey content by round

In overview the three stages of the Delphi study were:

Table 27 - Overview of Round 1-3 content

Round	Round 1	Round 2	Round 3
Respondent panel	All participants	By workstream from Round 1	By workstream from Round 2
Respondent profiling	Region location Role Areas of Expertise	-	-
Pillars	Allocation of Pillar weightings	Consensus on Pillar weightings	
Drivers	-	Allocation on Driver weightings	Consensus on Driver weightings
Indicators	-	Rating of Importance of Indicators	Rating of Reliability of Indicators
Workstream weightings	-	-	Cross-workstream rating of indicator importance

As outlined in Table 27 each round sought to determine a key aspect on the weighting or rating of event outcomes. In Round 1, the focus was on establishing the initial Pillar weightings and having respondents identify their areas of expertise for allocation to workstreams in future rounds. In Round 2, respondents reviewed and confirmed Pillar weightings and gave their input on the Driver weightings and Indicator importance ratings. In the final round (Round 3), respondents reviewed and confirmed Driver weightings for their workstreams and rated Indicators for their reliability.

5.5.2 Delphi Survey - Round 1

The first round sought to establish:

- *Scope of Event Impact Framework*
- *Weighting on Impact Pillars*
- *Identification of expertise*

The core contents of Round 1 were as follows:

Stakeholder Priorities

That both Event Hosts and Event Owners were identified as dominant stakeholders in the classification of stakeholders, respondents were asked to reflect on the priorities each might place on event outcomes. Respondents were presented with the six pillars identified and asked to rank them in descending order of importance Event Hosts and for Event Owners.

Pillar Weighting

In Round 1 the Pillars were presented for allocation of weightings of importance to event impact across the six identified Pillars. Respondents were asked to indicate how important each area is to actual event impact by allocating weightings to each Pillar. To assist with consideration of the weightings, a descriptor of each Pillar was provided to ensure clarity of meaning and scope.

As noted earlier regarding the Constant Sum question design, respondents were provided with an allocation of 100 points that could be applied in any proportion to each Pillar to represent their view of its importance in assessing event impact. No Pillar could be given a weighting higher than 95 points or lower than one point.

As the top-level elements and key to anchoring understanding of the perspectives on sport event impact, all respondents were allowed to provide weightings on the Pillars.

Event scope

Having noted the dominance of mega-events within the research domain and the relative paucity of equivalent research on recurrent events, respondents were allowed to identify which future events they would want to be prioritised in the development of an Event Impact Framework.

Respondents were able to identify up to 6-10 events from a list of nine (9) event categories. This task was not core to the Delphi study but was included to create variety in the tasks required of respondents and to retain their engagement.

Expert workstreams

Respondents were asked to self-nominate their two strongest areas of expertise from the list of six Pillars that formed the basis for workstream participation in Rounds 2 and 3 should they continue with the research.

5.5.3 Delphi Survey - Round 2

As noted above respondents receive invitations to Round 2 of the study based on their expertise rating from Round 1.

Hence all references below are based on those selections and describe the elements each respondent will address in Round 2 albeit the specific content will vary respondent by respondent.

The second round seeks to establish:

- *Consensus on Pillar weightings*
- *Weightings on workstream Drivers*
- *Rating of Indicators*

The core contents of Round 2 were as follows:

Pillar review of weightings

In line with the iterative process of the Delphi method, respondents were presented with the results of the Pillar weightings established in Round 1. Those scores were the "Group Consensus Range", which identified the 60% of respondents scores within the 20th-80th percentile range and the median score for each Pillar. Alongside those results, the individual's weightings from Round 1 were provided for review within the context of the group results. Respondents were then offered the opportunity to consider their initial weightings and enter updated weightings for Round 2.

The presentation of the weightings from Round 1 showed where a respondent's weights vary from the median, providing additional perspective for the respondent in considering whether to adjust their previously ascribed weights.

There was no requirement for a respondent to change their weightings in Round 2. Regardless of whether their weightings remain unchanged or were altered for Round 2, their rationale for the weightings provided was captured.

Weighting of Drivers

As noted in the initial description of the model, Drivers form the second tier of the weightings in the model. In Round 2, for each of the workstreams selected, the Drivers were presented for allocation of weightings of importance to event impact across the Drivers within that workstream. Respondents were asked to indicate how important each area is to event impact by allocating weightings to each Driver. Again, a descriptor of each Driver was provided to ensure clarity of meaning and scope.

As noted earlier in 5.4.7, for the Constant Sum question design, respondents were provided with an allocation of 100 points that could be applied in any proportion to each Driver to represent their view of its importance in assessing impact within that Pillar. No Driver could be given a weighting higher than 95 points or lower than one point.

Rating of Indicators - Importance

As noted earlier, the extensive list of potential Indicators created a challenge in the rating of all indicators by all respondents. The considered design and allocation of respondents to specific workstreams reduced that burden to a manageable level within each workstream. The list of Indicators was further reduced by filtering based on relevant and direct measures of impact as outlined in 3.2.3 Specification of measures.

Each respondent was presented with a randomised list of Indicators to rate the importance of each measure in assessing the potential [workstream] impact of an event. Ratings of Importance were on a five (5) point Likert scale where 1 = "Not important" and 5 = "Extremely important"

Unlike the Pillars and Drivers, there was not a consensus sought for the Indicators. The rating of the Indicators for importance was only presented to respondents in Round 2. The

responses from Round 2 were to identify the 24 highest rated Indicators (based on mean score) and to be presented to respondents in Round 3. The second rating in Round 3 was to assess the Indicator's reliability as a measure rather than to develop a consensus on its importance.

5.5.4 Delphi Survey - Round 3

As noted above, respondents receive invitations to Round 3 of the study as a continuation of their workstream allocation issued for Round 2.

Hence all references below are based on the relevant workstream and describe the elements each respondent will address in Round 3, albeit the specific content will vary respondent by respondent.

The third round seeks to establish:

- *Consensus on workstream Driver weightings*
- *Rating of Indicators for reliability as measures*
- *Cross-workstream rating of Indicators*

Driver review of weightings

In line with the iterative process of the Delphi method, respondents were presented with the results of the Driver weightings established in Round 2. Those scores were the "Group Consensus Range", which identified the 60% of respondents scores within the 20^{th}-80^{th} percentile range and the median score for each Driver. Alongside those results, the individual's weightings from Round 2 were provided for review within the context of the workstream results. Respondents were then afforded the opportunity to consider their previous weightings and enter adjusted weighting values for Round 3.

The presentation of the weightings from Round 2 showed where a respondent's weights vary from the median, providing additional perspective for the respondent in considering whether to adjust their previously ascribed weights, but no requirement was made for a respondent to change their weightings in Round 3.

Unlike with the Pillar weightings, respondents only completed the weightings of the Drivers within their workstream and were presented with two workstreams each for which weightings were required.

Rating of Indicators - Reliability

Addressing a dimension of event impact that received some negative or cautious attention is the reliability of the numbers claimed as measures of event scale or impact. Ideally, the Indicators within the Event Impact Framework need to be both Important (as established in Round 2) and Reliable to be of the most value as measures of, or proxies of, the Drivers and hence the Pillars. Under this study, the term Reliability was the rating of the accuracy and consistently available for assessing the [Pillar] impact.

As noted above, the Round 3 variables for Reliability rating were comprised of the 24 highest importance rated Indicators (importance mean score rating from Round 2). Each Indicator presented was rated for whether it is consistently measured and available for assessing the [workstream] impact on a five (5) point scale. Ratings of Reliability were on a five (5) point Likert scale where 1 = "Not reliable" and 5 = "Extremely reliable"

Again, unlike the Pillars and Drivers, there was no consensus sought for the Indicators. The rating of the indicators for Importance (Round 2) and Reliability (Round 3) were not repeated.

Cross-workstream indicator rating

Each group of measures were rated within the relevant workstream. As respondents only participated in two workstreams, there was no opportunity for respondents to consider the importance of highly rated measures from across all six workstreams. Being able to test the inter-rater rankings when not all respondents have seen all the items required a controlled design for the items to be rated within the BWS (noted in 5.3.6) method.

By creating a subset of items drawn from across each of the Pillars provided an opportunity for respondents to consider the items they had rated in the context of items from other workstreams. However, for the ratings within the BWS method to be representative, each element needed to be shown against all other elements and all shown to a respondent an equal number of times. The Balanced Incomplete Block Design (BIBD) is an experiential design method to ensure that the items being presented for rating within the BWS meets this criterion (Marden, 2019). The terminology of BIBD refers to a design in which the total number of elements to be rated is broken into a set of 'blocks'. While each block is 'incomplete' (in terms of only including a subset of the elements), across all the blocks the appearance of each element is 'balanced'. That is, each item appears as a choice against all other items and all items appear an equal number of times.

Initially, the intention was to select the two highest-rated ($\bar{x}$ Importance) Indicators from each of the six pillars for a total of 12 elements for ranking in the BWS method. However, there are specific combinations in which the number of elements (referred to as 'treatments' = v) can form a balanced design in incomplete blocks. It is not possible to form a BIBD using 12 elements. Drawing on Fisher and Yates (1963) statistical tables, the closest compliant form was found to be that of 13 elements. Hence, including three Indicators from the Sport pillar and two Indicators from each of the remaining five pillars allowed a BIBD that could be resolved.

The testing of the compliance of that form of BIBD using the R software 'crossdes' programming package (Sailer, 2015) confirmed that the elements were considered 'balanced' in that "each treatment appears equally often in the design [and] the number of concurrences of treatments i and j is the same for all pairs" (pp13).

The final design was comprised of 13 Indicators drawn from across the six Pillars (v treatments). Each measure is repeated (r) = 4 times, within 13 blocks (b) with each block showing 4 options (k observations) from which respondents could choose.

Table 28 – Best/Worst Scaling (BWS) element combinations

	[Option 1]	[Option 2]	[Option 3]	[Option 4]
Block 1	1	3	4	12
Block 2	5	6	9	12
Block 3	7	8	11	12
Block 4	1	2	5	7
Block 5	3	6	7	10
Block 6	4	5	8	10

Block 7	2	3	8	9
Block 8	3	5	11	13
Block 9	2	4	6	11
Block 10	1	9	10	11
Block 11	4	7	9	13
Block 12	1	6	8	13
Block 13	2	10	12	13

As noted above, the compliant combinations in the 13 Indicator form of BIBD were developed in programming language R 'crossdes' package from which the Table 28 – Best/Worst Scaling (BWS) element combinations were generated. The result from running the 'crossdes' program confirmed that "the design is a balanced incomplete block design w.r.t. rows."

Based on the balanced design achieved the measures were presented in 13 blocks of four items. For each block respondents selected the measure they considered as the most important and the measure they considered least important measure of the four presented. Across the combinations of measures shown in the 13 blocks, each measure was shown to a respondent an equal number of times and all measures were shown in comparison to each other measure.

Pillar weightings by event typology

At the conclusion of Round 3, an optional extension question was made available. The question allowed respondents to provide their perspective on how the weightings of importance for the six Pillars might vary across events of different scale. To ensure a common understanding of event scale, the five Type A – E classifications of events in the SIRC event typology and their descriptions were presented.

Respondents allocated Pillar weightings using the same 100 point constant sum method used for the Pillars in Round 1 and Round 2 such that each event type totalled 100 points. The question presented all five event types on a single screen and was limited to a single round of weightings as it was optional and core to Delphi method consensus testing.

5.5.5 Calculations used in the analysis

Measure Impact Score

A combined construct of Impact Score was given to each measure in the workstream to provide a discriminating overall score for rated measures.

Impact Score [$\overline{x}$ I * $\overline{x}$ R] = $\overline{x}$ Importance * $\overline{x}$ Reliability.

For the top 24 measures for $\overline{x}$ Importance from Round 2 and were represented for Reliability ratings in Round 3, the Impact Score is a direct product of the two mean scores. Measures outside the top 24 $\overline{x}$ Importance measures were not made available for rating for Reliability in Round 3, hence a proxy Reliability measure was developed.

The proxy measures used the minimum $\overline{x}$ Reliability awarded from within the workstream and are applied to 'unrated' Reliability values. The use of the 'minimum rating' rather than

'average rating' for Reliability was selected to minimise the chance of overstating the Impact Score within the workstream context.

Framework internal reliability

To understand to what degree the allocation of weights ascribed to the Pillars and Drivers (top-down) by respondents might also reflect the importance and reliability ascribed to measures of impact (bottom-up) additional calculations were developed.

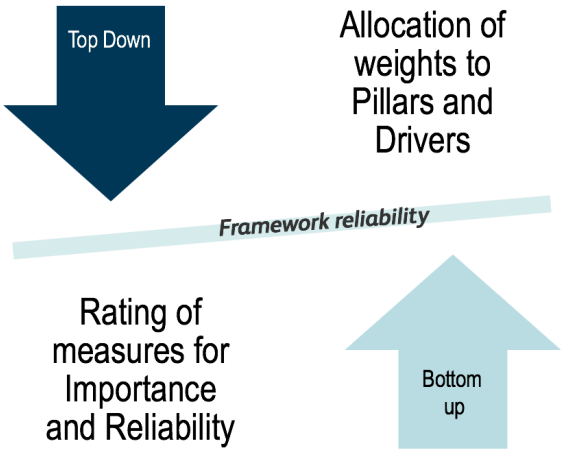

Figure 19 - EIF Framework reliability model

As described in Figure 19 the testing sought to identify the agreement between methods of assigning importance within the framework. The direct allocation of weights (Driver $\bar{x}$) was held as the independent variable, against which are range of Indirect weights established by the Importance and Reliability ratings of the measures might be tested.

Table 29 - Internal reliability tests

Test	Typology	Indirect Construct	Correlation (r)
Test #1	I	Importance ($\bar{x}$) only	Mean scores of Importance ratings
Test #2	IR^{24}	Importance ($\bar{x}$) * Reliability ($\bar{x}$) (top 24)	Importance * Reliability means scores for the 24 measures rated for reliability
Test #3	IR^{All}	Importance ($\bar{x}$) * Reliability ($\bar{x}$) (All)	Importance * Reliability means scores for All measures through use of proxy scores

The tests as shown in Table 29 were intended to explore the ability to generate similar weights for Drivers based on the direct weightings of Drivers against an imputed score for weightings of Drivers derived from the ratings of measures for importance and reliability. Noting that as Indicators were not specified as explicitly aligning to individual Drivers, judgement was used to guide how measures were assigned and to recognise that different allocations could alter the resultant scores.

5.6 Research fieldwork

The fieldwork of surveying the expert panel was conducted over eight (8) weeks.

At the start of each round of the study, an initial invitation was emailed to all active panel participants (i.e. those not requesting removal and had qualified for that round), and reminder invitations were also sent to advise of closing dates for survey participation.

Each participant completing the survey received both a confirmation screen at the end of the survey and a follow-up thank you email confirming their results would be included in the aggregated analysis of findings.

Expert panel

Under the project, Sportcal made available its subscriber mailing database to the independent research provider. Their subscribers are global industry and government sports associations, international federations, and practitioners.

When combined with publicly available lists of academics publishing and presenting on research in sport management, the breadth and depth of expert panel established would not have been achievable without an industry partner's cooperation.

Survey research management

Within in each round the members of the panel received an initial invitation, reminders were sent during the round to those who had not yet completed the survey, and a thank you email was sent to each respondent completing the round confirming their input.

Table 30 - Respondent communications by round

Round	Round 1	Round 2	Round 3	Round 4
Invitations	2	1	1	1
Reminders	3	3	3	2
Thank you	1	1	1	1

Respondent panel size

In a review of panel sizes used in 32 prior Delphi studies, Giannarou and Zervas (2014) found "18 studies of used the opinion of less than 20 experts […] with the number ranging between 30 and 50 participants in 5 studies and between 50 and 100 in 4 more. Also, there are 5 studies which used an even greater number of participants, i.e. >100." (p. 67)

In that context, the panel size of n=182 from Round 1 of the study established it at the higher end of Delphi studies in scale.

Survey response numbers by round

Table 31 - Survey response rates by round

Round	Round 1	Round 2	Round 3

Invited	840	182	96
Partial complete	112	n/r	n/r
Did not complete	546	86	49
Completed	**182**	**96**	**47**
Completion rate	21.7%	52.8%	49.6%

n/r = Not recorded

As shown in Table 31, the initial completion rate of >20% for Round 1 provided a strong base group of n=182 for the initial weightings of the Pillars. In subsequent rounds, each saw approximately half the panel continue with the result of n=47 completing Round 3.

Limitations in fieldwork data

To maintain full anonymity for respondents, no identifying codes that might be used to link respondent results between Rounds was provided in the anonymised data by the fieldwork company. The tracing of individual respondent's ratings across rounds prevented testing consensus changes between rounds for both Pillars and Drivers. Likewise, it prevented the profiling of respondents by role, age, tenure, etc., in Rounds 2 and Round 3 to be utilised as variables in their responses to Driver or Indicator ratings.

6 Results

6.1 Approach to data analysis

The online survey instruments for the three Rounds of the Delphi study were developed as a linked series but fielded individually. The overall approach taken in analysing the anonymised data sets generated from the survey instruments was consistent across Rounds.

Analysis of quantitative survey data

The output files from each round of the Delphi study were delivered from the research provider as a separate Excel spreadsheet file with one row per respondent and a column for the response to each question.

To compare responses by different groups (for example region, role or area of expertise), the data was imported into SPSS statistical analysis software.

Survey data generated in this study was either nominal/categorical (for example respondents chose an area of expertise, role, tenure) or numeric values (for example Pillar weightings, rating of measure importance). The main types of analysis were comparison of mean Pillar weightings by respondent demographics, and identifying similarities and differences between Rounds for the weightings allocated to Pillars and Drivers.

Standard descriptive statistics calculated for Pillar and Driver weightings include Median, Minimum, Maximum, Interquartile Range (IQR), Mean, Standard Deviation, and the top and bottom quintile results.

As Delphi is a method for moving towards a consensus, the statistics of mean scores and standard deviations were used as applied by Holey et al. (2007) for "comparing movement between Delphi rounds as a measure of both stability and convergence"(p. 53). Recognising those measures represent the expert panel views with "the mean, as a measure of central tendency, represents the group opinion of the panel. The standard deviation, as a measure of spread, represents the amount of disagreement within the panel" (Greatorex & Dexter, 2000, p. 1018). The demonstration of stability in panel responses would be represented by little to no variation in the mean score between Delphi rounds. The demonstration of convergence towards a consensus viewpoint would be expected to be shown in a narrowing of the standard deviation in response values between rounds.

When comparing mean Pillar weightings by respondent demographics, a t-test was used. This compares the mean scores of two groups of respondents when respondent choices are based on a numerical rather than a categorical scale (Pallant, 2020). When comparing the weightings between Delphi rounds, three statistical tests were applied: a t-test, the Mann-Whitney U test, and Levene's test.

t-tests used to compare the mean scores of two different groups of respondents when respondent choices are based on a numerical rather than a categorical scale. When two groups are the same participants responding on two different occasions a paired sample t-test is used to evaluate whether there is a significant difference in the mean values of the responses (Pallant, 2020).

Mann-Whitney U test is the non-parametric equivalent to the t-test. It converts responses to ranks and then compares the medians of the two groups. This test, unlike the t-test, does not assume the normality of distribution in the data. The null hypothesis is that the two groups have the same median values and the alternative hypothesis is that the two groups have different median values (Sheskin, 2003).

Levene's Test for Equality of Variances was used to test for differences in the standard deviation between Delphi rounds. Levene's test is a test for homogeneity of variance that is less sensitive than alternative tests (such as the Bartlett test) to departures from normality (Gastwirth et al., 2009).

6.2 Delphi Panel Respondent Profile

To preserve the anonymity of respondents, a limited number of profiling questions were included in Round 1 of the Delphi study. Each was designed to provide some key dimensions on the profile of the overall panel, but not so detailed that an individual could be identified from their responses.

6.2.1 Panel profile by Age

Table 32 - Expert panel profile by Age

Age (grouped)	Respondents (n)
18-24, 25-34	29
35-44	51
45-54	63
55-64, 65+	38
Prefer not to answer	1
Total	182

Table 32 shows the profile of the respondent panel by Age group. Without controlling for Age, it is noteworthy that there is a solid representation of ages across the panel having grouped those under Age 35 in a single cohort. That the ages continue through to Age 65 reflects the time required to develop expertise and that the study might benefit from the panel's depth of experience.

6.2.2 Work role

Which best describes your area of work?

Table 33 - Role description

Work role (grouped)	Respondents (n)
Academic	65
Government (city) Government (local) Government (national)	22
International Sports Federation	31
Finance / Banking / Professional Services / Advisory	18
National Sports Association Sports development	13
Event management	16

Event owner Other	
Media broadcaster Media buying News & Journalism Sponsorship / Sponsor management	17
Total	182

Having provided 20+ potential roles for respondents, Table 33 identifies the work role most strongly represented within the panel as coming from Academia, then representatives from International Sports Federations and Government roles. The overall breadth of roles brings differing perspectives as sought in a Delphi study.

6.2.3 Tenure

How long have you been working in the role/field of expertise you currently hold?

Table 34 - Tenure in field of expertise

Tenure (grouped)	Respondents (n)
Less than 1 year, 1-2 years, 3-5 years	50
6-10 years, 11-15 years	64
15 years+	68
Total	182

Table 34 shows the time in which the respondents had been in their role. This was sought to understand their depth of experience beyond what could be implied from their ages. The results showed the largest group to contribute was not those new to events and sport, but those with 15+ years of experience. That said, the differences across the three groups were well balanced between recent entries into the sector and more experienced individuals.

Table 35 - Tenure by Age Group

	Age group			
Tenure	18-34	35-44	44-54	>55
5 years or less	47%	29%	18%	6%
6-15 years	9%	48%	33%	9%
More than 15 years	0%	9%	49%	43%
Total	16%	28%	35%	21%

Surmising from the Age profile of the panel shown in Table 35 that the Delphi study might benefit from the depth of expertise, Table 35 makes explicit the intersection of Age and Tenure in their field of expertise. Understandably, the youngest age group 18-34 dominate (47%) those with a tenure of 5 years or less, but there are older participants also relatively new to the field and hence contribute new perspectives. Those with a tenure of 6-15 years

centre on the two mid age groups of 35-44 (48%) and 44-54 (33%) respectively. Those with more than 15 years tenure is grounded in the two oldest age groups 44-54 (49%) and >55 at 43%.

6.2.4 Panel profile – Role location

Where are you based for your role?

Table 36 - Location base for role

Global region	Respondents (n)
Africa	3
Asia	14
Europe	112
Middle East	7
North America	22
Oceania	20
South America	4
Prefer not to answer	0
Total	182

With senior roles in sport administration and events operating across markets and even globally, Table 36 shows where respondents are based for their roles rather than the operating scope of that role. The previously representation of Academics and International Federations contributes to the strong European presence in terms of location but does not necessarily reflect a bias in the breadth of research or responsibilities towards that market. Next are North America, Oceania and Asia at similar levels, and that respondents are from each region attests to the global profile of the panel.

6.2.5 Panel profile – Role responsibility

What is the breadth of your role's responsibility?

Table 37 - Breadth of responsibility in role

Responsibility breadth	Respondents (n)
Local	12
Regional	15
National	43
International	102
Prefer not to answer	10
Total	182

Having separated the notion of where a role is located as seen in Table 36 from the breadth of its responsibilities, Table 37 identifies the scope of those roles.

The results show the majority of respondents hold international responsibilities (56%), with national and regional the next most common. The benefit of this respondent panel profile is that it balances the need for a wider global perspective with a more localised and targeted national and even local perspective.

6.2.6 Expertise selections

The research design considered the nomination of respondents' areas of expertise and allocated them to specific workstreams as outlined in 5.3.5 "Workstreams within expert panel". As noted, each respondent could nominate two areas (of the six pillars) to which they felt most qualified to contribute to in further rounds.

Table 38 - Nominated areas of expertise

Area of Expertise	Selections (n)	% of respondents
Economic	81	22%
Sport	108	30%
Media	44	12%
Social	64	18%
Brand	60	16%
Environment	7	2%
Total	364	100%

Table 38 - Nominated areas of expertise shows the responses of the 182 respondents. Noting that as respondents could each nominate two areas of expertise, the total number of selections is 2 x 182 = 364.

In responding to the self-nominated areas of expertise, Sport (30%) and Economic (22%) were the two most commonly selected within the panel.

Noteworthy is the lack of selection of Environment (2%), which within popular narrative garners significant attention but is either an area that respondents felt was not within their field of expertise… or while they may have been familiar with the development, it was not strong enough to be selected in their 'top two' areas.

Social (18%), Brand (16%) and Media (12%) formed the mid-range of the selection choices.

6.2.7 Cross-incidence of expertise selections

The breakdown of the percentage of respondents within each of the expert areas nominated in the column title.

Table 39 - Cross-incidence of expertise selections

	Economic	**Sport**	**Social**	**Media**	**Brand**	**Environment**
Economic	-	10%	2%	4%	6%	0%
Sport	10%	-	12%	4%	4%	0%
Media	4%	4%	1%	-	3%	0%
Social	2%	12%	-	1%	3%	1%

Brand	6%	4%	3%	3%	-	1%
Environment	0%	0%	1%	0%	1%	-
	22%	30%	18%	12%	16%	2%

Table 39 identifies the workstream combinations selected by respondents from their two choices as a percentage of respondents. Working across the expertise areas shows some disciplines are more common than others.

The two most frequent combinations were Sport + Social (12%) and Economic + Sport (10%). The remaining combinations were all less than 10%, showing breadth in the combinations selected. Importantly, while the intersection of combinations may have achieved <5%, the sample base used for the analysis is within each area of expertise, i.e. the response numbers in Table 38 - Nominated areas of expertise.

6.3 The Weighting of Pillars – Round 1

A central element of the Delphi Study was to develop an explicit weighting of the impact of the pillar dimensions on sports events.

6.3.1 Overall weightings – Round 1

Table 40 - Initial weightings of Pillars from Round 1

Pillar	Median Pillar Weighting	Mean Pillar Weighting	St. Dev. Pillar Weighting	Min. Pillar Weighting	Max. Pillar Weighting
Economic	21	23.6	10.12	5	60
Sport	20	23.0	11.45	2	70
Media	15	15.1	5.80	3	35
Social	15	14.2	6.75	2	35
Brand	15	14.4	6.90	2	50
Environment	10	9.7	5.03	1	25

Table 40 identifies the initial importance weightings given by the panel to the six Pillars of event impact. In this initial round of the study, all respondents provided weightings across each of the pillars irrespective of their nominated areas of expertise.

A minimum value of '1' and a maximum value of '95' was available for each Pillar, with a total required to equal '100' across the six pillar weightings.

Within the mean scores for the Pillars, there are three groupings of the Pillars. Economic ($\bar{x}$ 23.6) and Sport ($\bar{x}$ 23.0) are given the highest weights for importance. The next tier includes Media ($\bar{x}$ 15.1), Social ($\bar{x}$ 14.2), and Brand ($\bar{x}$ 14.4) which are given similar ratings. Environment ($\bar{x}$ 9.7) was given the lowest rating for importance in assessing event impact.

The range between the minimum and maximum weightings showed a large disparity in views… particularly notable for Economic (5-60) and Sport (2-70), with Brand (2-50) also showed a wider range than Social and Media, which shared the same tier for mean scores.

The maximum scores for Sport and Economic are both more than three Standard Deviations away from the mean score for those pillars, indicating a non-normal distribution. Some respondents provided outlier values at the top end of the range compared to the rest of the panel. This 'high-end outliers' distribution was also the case with other pillars but not to the same degree.

6.3.2 Differences in weightings of pillars by expert workstream – Round 1

As noted in the methodology "5.5.1 Overview of survey content by round", the research design required respondents to give their responses to the Pillar weightings before presenting them with an opportunity to nominate their areas of expertise. This was done to minimise the potential for priming their responses to the initial Pillar weightings.

Table 41 - Differences in mean pillar weightings by workstream

	Mean Pillar Weight		Respondent count		Expertise bias	
Pillar	Workstream not selected	Workstream selected	Workstream not selected	Workstream selected	Workstream selection differences	*p*Value
Economic	21.5	26.2	101	81	+4.7	0.002*
Sport	20.2	24.9	74	108	+4.7	0.006*
Media	14.1	18.0	138	44	+3.9	0.001*
Social	12.4	17.5	122	60	+5.1	0.001*
Brand	13.5	16.3	118	64	+2.8	0.009*
Environment	9.7	9.9	175	7	+0.2	0.919
Total			728	364		

* Indicates *p*Value statistically significant at 95% confidence level.

The results in Table 41 shows the differences in mean weighting scores for each of the Pillars between those that did not nominate that Pillar as one of their two areas of expertise (Workstream not selected) against those that had nominated that Pillar (Workstream selected).

The difference in mean scores between the two groups shows that respondents consistently ascribed higher weightings to Pillars that they considered is within their area of expertise as compared to those that did not. That bias towards their own Pillars was found to be statistically significant for five of the six Pillars. Economic, Sport, Media, Social, and Brand all showed bias in perceptions of importance by those who later nominated them as areas in which they had experience and/or expertise. The final Pillar of the six, Environment, showed a similarly high rating towards those who nominated it for their workstream, but the difference and sample base was too small for the result to be statistically significant.

Despite the statistical differences between the groups on five of the six pillars, the three tiers of weightings noted from Table 40 is consistent within each group. That is, Tier 1 (Economic, Sport), then Tier 2 (Media, Social, Brand), and then Tier 3 (Environment).

6.3.3 Differences in weightings of pillars by role – Round 1

Within the range of respondent roles, the largest single cohort shown in 'Table 33 - Role description' was those from academia. This single group was large enough to test whether there are differences in the weighting of pillars by Academics vs. [Not] Academics.

Table 42 – Differences in mean pillar weightings by workstream, Role = Academic

	Mean Pillar Weighting		Respondent count		Mean Pillar Weighting
Workstream	Workstream not selected	Workstream selected	Workstream not selected	Workstream selected	Workstream selection differences
Economic	22.71	25.63	41	24	+2.92
Sport	20.29	22.56	24	41	+2.27
Media	13.44	17.50	57	8	+4.06
Social	13.54	17.81	28	37	+4.27
Brand	13.21	13.67	47	18	+0.46
Environment	11.27	10.50s	63	2	-0.77
Total			260	130	

The results in Table 42 show the differences in how respondents who nominated 'Academic' for their role rated the Pillars for importance, split by their areas of expertise. The similar 'tiers' of Pillar weightings were evident, but Brand and Environment were closer in weightings as compared to the overall panel.

While again the differences between the areas of 'Workstream selected' and 'Workstream not selected' showed differences in mean scores, the largest differences were in the Tier 2 pillars of Social ($\overline{x}$ +4.27) and Media ($\overline{x}$ +4.06), followed by Tier 1 pillars of Economic ($\overline{x}$ +2.92) and Sport ($\overline{x}$ +2.27).

6.3.4 Differences in mean pillar weightings by role [Not] Academic.

Respondents from all other roles were put into a single cohort of 'Not Academic' to test the workstream differences in Pillar weightings.

Table 43 - Differences in mean pillar weightings by workstream, Role = [Not] Academic

	Mean Pillar Weighting		Respondent count		Mean Pillar Weighting
Workstream	Workstream not selected	Workstream selected	Workstream not selected	Workstream selected	Workstream selection differences
Economic	20.70	26.50	60	57	+5.80
Sport	20.20	26.40	50	67	+6.20
Media	14.60	18.20	81	36	+3.60
Social	12.00	17.10	90	27	+5.10
Brand	13.60	17.40	75	42	+3.80

Environment	8.90	9.60	112	5	+0.70
Total			468	234	

Table 43 shows a similar pattern to the overall panel in the grouping and ordering of the three tiers of Pillar weightings. Again, differences were evident between those who had selected the workstream compared to those who had not, but the pillars on which the differences were largest were not those seen in the Academic weightings Table 43.

Within the non-Academic cohort, the largest differences were for Sport ($\overline{x}$ +6.20), Economic ($\overline{x}$ +5.80) and Social ($\overline{x}$ +5.10).

6.3.5 Significance differences between Groups in Round 1

In Table 41 statistically significant differences were found between respondents working in academic roles and non-academic roles.

Testing significance in differences between Pillar weightings by Academic vs. [Not] Academic

Table 44 – Differences in Academic vs. [Not] Academic Mean Pillar Weightings

Pillar	Academic mean score	[Not] Academic mean score	Differences mean score	*p*Value
Economic	23.78	23.50	+0.28	0.859
Sport	21.72	23.76	-2.04	0.255
Media	13.94	15.70	-1.76	0.049*
Social	15.97	13.16	+2.81	0.007*
Brand	13.34	14.98	-1.64	0.125
Environment	11.25	8.89	+2.36	0.002*

* Indicates *p*Value statistically significant at 95% confidence level.

Table 44 shows the differences between Academic and [Not] Academic cohort weightings in Pillar importance for the assessment of event impacts. Noting that overall the same 'tiers' of the pillar weightings observed was evident for both groups even within those statistically significant different weightings.

Concerning the weighting for Social ($\overline{x}$ +2.81) and Environment ($\overline{x}$ +2.36), the Academic cohort placed more weight on their importance compared to non-academics. Conversely, the [Not] Academic cohort placed a higher weight on the importance of Media ($\overline{x}$ +1.76) as compared to Academics.

Other differences were noted in Sport and Brand but were found not to be significant. Despite disparity on other pillars, it is noteworthy that both groups were very close on their weighting applied to the Economic pillar.

Testing significance in differences between Pillar weightings by Tenure

Testing whether how long respondents had spent in the sector might influence their perceptions of the importance of pillars, the tenure classifications noted in Table 34 were grouped into three cohorts of Group 1 (<5 years), Group 2 (5-10 years) ,and Group 3(> 10 years).

Table 45 - Differences in Tenure (grouped) Mean Pillar Weightings

	Group 1 mean score	Group 2 mean score	Group 3 mean score	Differences		*p*Value	
Tenure Group	<5 years	5-10 years	>10 years	Grp 1 - Grp 2	Grp 1 - Grp 3	Grp 1 - Grp 2	Grp 1 - Grp 3
Economic	23.12	22.45	25.04	+0.67	-1.92	0.699	0.332
Sport	19.74	23.75	24.78	-4.01	-5.04	0.043	0.013*
Media	16.34	15.14	14.07	+1.20	+2.27	0.289	0.046*
Social	16.10	14.42	12.50	+1.68	+3.60	0.209	0.006*
Brand	14.18	14.63	14.34	-0.45	-0.16	0.698	0.905
Environment	10.52	9.61	9.26	+0.91	+1.26	0.326	0.175

* Indicates *p*Value statistically significant at 95% confidence level.

Table 45 shows the differences in Pillar weightings across the three Tenure cohorts, using Group 1 (< 5 years) as the reference baseline and identifying the differences of Groups 2 and 3 in comparison.

There was one significant difference in that Group 1's weighting applied to Sport ($\overline{x}$ -4.01) was significantly lower than that of Group 2. Those differences are larger when considering the Group 3 (>10 years) ratings of pillars against Group 1 (< 5 years). While Group 1 placed less weight on Sport ($\overline{x}$ -5.04) than Group 3, they rated Social ($\overline{x}$ +3.60) and Media ($\overline{x}$ +2.27) as significantly higher than Group 3.

Overall, the three groups were most consistent in their weighting of Brand, with differences of only ($\overline{x}$ -0.45) and ($\overline{x}$ -0.16) for groups 2 and 3 respectively.

6.4 Pillar weightings – Round 1 and Round 2

Through the iterative design of the Delphi study, the six Pillars were presented to all respondents for weighting twice.

In Round 1, all respondents provided their initial weightings on the relative importance of each of the six Pillars. In Round 2, their initial weightings scores along with the results from Round 1 were provided for respondents to review and to guide the adjustment or confirmation of their Round 1 weightings. i.e. Round 2 aimed to test the degree of consensus on the Pillars.

6.4.1 Pillar weightings by round

Table 46 – Pillars– Round 1, Initial weightings

Round 1	Median	STD DEV	Max	Min	Mean	IQR	Bottom quintile	Top quintile
Economic	21	10.12	60	5	23.6	14	15	30
Sport	20	11.45	70	2	23.0	11	15	30
Media	15	5.80	35	3	15.1	10	10	20
Social	15	6.75	35	2	14.2	10	10	20
Brand	15	6.90	50	2	14.4	10	10	20
Environment	10	5.03	25	1	9.7	6	5	15

In addition to the initial Pillar weighting results shown in Table 40, the results in Table 46 expand on those to show the Interquartile Range (IQR), the Bottom quintile (bottom 20% of scores), and the Top quintile (top 20% of scores). As noted previously, comparing the maximum and minimum scores given for each Pillar against the standard deviations from the mean indicates a small number of outlier results for most pillars, but especially for Economic, Sport, and Brand.

The cut-off scores for the top and bottom quintile scores highlights the 'three tiers' in the weightings noted above. Without the outlier results, Economic and Sport were brought into a smaller range but still receive higher weightings than Media, Social, and Brand, which are brought onto the same level with Environment remaining lower than the other Pillars.

Table 47 - Round 1 Pillar Consensus Range presented in Round 2

Round 1	Median	Consensus Range
Economic	21	15-30
Sport	20	15-30
Media	15	10-20
Social	15	10-20
Brand	15	10-20
Environment	10	5-15

Table 47 shows the Median and the Consensus Range (the 60% of responses between the Bottom and Top Quintiles) that were presented to respondents in Round 2.

Table 48 – Pillars – Round 2, Revised weightings

Round 2	Median	STD DEV	Max	Min	Mean	IQR	Bottom quintile	Top quintile
Economic	25	9.02	60	1	24.0	5	20	30

Sport	20	11.14	70	1	21.5	10	15	25
Media	15	4.75	25	1	14.6	10	10	20
Social	15	6.19	35	2	16.0	10	10	20
Brand	11	6.33	40	2	13.0	5	10	17
Environment	10	6.16	50	1	10.9	10	5	15

Table 48 shows the summary results of weightings for each of the six Pillars following the second round of weightings by the panel.

In Round 2, respondents were presented with the data in Table 47, as well as their initial weightings from Round 1. In that context, respondents could consider their initial weightings for each Pillar against the Consensus Range and determine if they would want to reaffirm their initial weightings or revise them.

The results Round 2 show the highest weightings ascribed to Economic ($\bar{x}$ 24.0) and Sport ($\bar{x}$ 21.5) with the second tier of Pillars was confirmed, as was Social ($\bar{x}$ 16.0), Media ($\bar{x}$ 14.6), and Brand ($\bar{x}$ 13.0). The lowest weighting was again allocated to Environment ($\bar{x}$ 10.9).

The maximum and minimum scores reveal that some respondents chose to reaffirm ratings from Round 1, giving weightings at the higher and lower end of the ranges and outside the Consensus Range presented, in particular Sport (1-70) and Economic (1-60).

Consensus on Pillars

The results for both scoring rounds were tested to identify if there had been any change in the overall Pillar weightings from Round 1 to Round 2.

Table 49 - Pillars consensus significance (All respondents)

	Round	N	Mean	t-test significance	Mann-Whitney significance#	Std. Deviation	Levene's test significance^
Economic	1	182	23.60			10.12	
	2	96	23.97	0.767	0.533	9.02	0.110
Sport	1	182	23.03			11.45	
	2	96	21.52	0.292	0.145	11.14	0.314
Media	1	182	15.07			5.78	
	2	96	14.60	0.472	0.694	4.75	0.048*
Social	1	182	14.16			6.76	
	2	96	16.01	0.027*	0.007*	6.19	0.227
Brand	1	182	14.40			6.90	
	2	96	12.97	0.093	0.078	6.34	0.516
Environment	1	182	9.73			5.03	

2	96	10.93	0.083	0.073	6.16	0.386

* Indicates statistically significant at 95% confidence level.
^Levene's test for equality of variances. # Mann-Whitney U test for equality of distribution

The results in Table 49 shows the differences in mean weightings applied in Round 1 and Round 2 to each of the pillars.

Tests of significance were twofold. The first was testing for changes in the mean scores given to each of the pillars, and the second was testing for changes in the variations of the weightings ascribed to each pillar.

The results show the Social pillar was found to be the only pillar to have had a significant change in the mean score from Round 1 ($\bar{x}$ 14.16) to Round 2 ($\bar{x}$ 16.01). In the Media pillar, there was a tightening of the consensus with significant narrowing of the Standard Deviation around its mean weighting. No other pillars showed statistically significant changes in the weightings or variances around the mean from Round 1 to Round 2

6.4.2 Filtering for outliers

It was noted that the range of values ascribed in Round 1 (Table 46) and Round 2 (Table 48) showed a small proportion of values that were more than three Standard Deviations from the mean scores for Pillars. Classifying those values as outliers reflects the understanding of an outlier as a data point or value in a data set that is seen to be inconsistent with the rest of the observations. "In most instances inconsistency is reflected in the magnitude of an observation (i.e. it is either much higher or much lower than any other observation"(Sheskin, 2003, p. 399). Such values were identified as outliers within the core panel responses and were removed for retesting of the panel responses and calculation of mean scores and standard deviations. The filtering of outliers resulted in the exclusion of 4-15 responses from individual Round 1 pillars and exclusion of 4-7 responses from individual Round 2 pillars.

Reducing the outliers from the Pillar weighting allocation sets.

Table 50 - Pillars consensus range (excluding outliers)

	Round	N	Maximum value	Minimum value	Filter upper limit	Filter lower limit	Filtered N
Economic	1	182	60	5	45	5	170
	2	96	60	1			90
Sport	1	182	70	2	45	5	169
	2	96	70	1			89
Media	1	182	35	3	30	5	167
	2	96	25	1			91
Social	1	182	35	2	<=30	>=5	178
	2	96	35	2			92
Brand	1	182	50	2	30	5	170
	2	96	40	2			90

Environment	1	182	25	1	25	>=5	170
	2	96	50	1			90

Table 50 shows the resulting sample base for each of the Pillars from excluding outliers at the upper and lower end of the range of weightings assigned in Round 1 and Round 2.

Consensus on Pillars – Retesting after removal of outliers

Table 51 - Pillars consensus significance (excluding outliers)

	Round	Filtered N	Mean	t-test significance	Mann-Whitney significance#	Std. Deviation	Levene's test significance^
Economic	1	170	22.21			7.605	
	2	90	23.22	0.261	0.298	6.449	0.035*
Sport	1	169	21.63			7.994	
	2	89	19.92	0.067	0.091	6.556	0.008*
Media	1	167	15.81			4.990	
	2	91	15.19	0.286	0.466	4.136	0.016*
Social	1	178	14.24			6.484	
	2	92	16.23	0.013*	0.000*	5.511	0.078
Brand	1	170	13.91			5.364	
	2	90	12.86	0.125	0.138	4.989	0.400
Environment	1	170	10.02			4.510	
	2	90	11.01	0.089	0.050*	4.291	0.556

* Indicates statistically significant at 95% confidence level.
^Levene's test for equality of variances. # Mann-Whitney U test for equality of distribution

In a refinement to the tests of differences in pillar weightings between Round 1 and Round 2 results as seen above in Table 41, the results in Table 51 shows the difference in Pillar weightings from Round 1 to Round 2 when outliers are excluded from the data sets.

The result is that in addition to the tightening of the weighting in Media observed earlier, this refinement showed the Economic and Sport pillars also significantly narrowed in the range of views expressed around the mean scores. There was not, however, any significant change in the mean scores between the rounds for those same pillars.

Consistent with the earlier findings, the Social pillar again showed a significant shift in the mean score between rounds, increasing in weighting in Round 2, but not a narrowing of the views accompanying the change. The two remaining pillars (Brand and Environment) did not show a significant change in either mean or standard deviation between rounds despite the exclusion of a similar number of outliers as for the other pillars.

6.4.3 Final weightings adjusted for workstream bias

The emergence of the significant expertise bias influenced the final pillar mean weightings as the proportion of experts in each workstream was not consistent (see Table 38). The following table shows the final Pillar weights having been adjusted for the proportions of expert/non-expert ratings in each workstream.

Table 52 – Balanced Overall Pillar Weights

Pillar	Overall Mean Score	Adjusted mean score	Mean score impact
Economic	23.97	23.59	-0.38
Sport	21.52	22.99	+1.47
Media	14.60	15.04	+0.44
Social	16.01	14.08	-1.93
Brand	12.97	14.48	+1.52
Environment	10.93	9.71	-1.22

Table 52 shows the Pillar weights having been adjusted for the bias demonstrated within each workstream. The effect of balancing for bias has the smallest result on the Economic ($\bar{x}$ -0.38) and Media ($\bar{x}$ +0.44) pillars, with larger differences in the remaining pillars.

Both the unadjusted overall mean scores and the adjusted mean scores show the pillars cluster into three groups or 'tiers' of weightings, but those tiers are clearest in the balanced results. Tier 1 includes Economic ($\bar{x}$ 23.59) and Sport ($\bar{x}$ 22.99), Tier 2 as the mid-tier level includes Media ($\bar{x}$ 15.04), Social ($\bar{x}$ 14.08) and Brand ($\bar{x}$ 14.48), with Environment forming Tier 3 at ($\bar{x}$ 9.71).

6.5 Driver Weightings – Round 2 and Round 3

As noted in the research method, Round 2 was the first opportunity for respondents to provide an assessment of the weighting of the Pillar sub-elements, the Drivers.

These were presented to respondents in their workstreams in Round 2 once having completed the consensus ratings of the Pillars. In Round 3, the Drivers were provided to respondents for their Driver consensus weightings in the same way they had completed for the Pillars in Round 2.

6.5.1 Driver weightings – Economic

Respondents who nominated "Economic" as an area of expertise in Round 1 were directed into the "Economic workstream". The weighting of importance required a minimum value of '1' and a maximum value of '95' with a total of '100' points required across the drivers presented.

Table 53 - Economic Drivers – Round 2, Initial weightings

Workstream: **Economic**	N	Median	STD DEV	Max	Min	Mean	Bottom quintile	Top quintile
Tourism	47	25	10.80	65	10	23.9	15	30

Infrastructure	47	15	7.12	30	5	16.6	10	20
Direct event spending	47	20	8.04	50	10	21.1	15	25
Security	47	10	4.85	20	1	9.00	5	10
Business activity	47	15	5.41	25	5	15.1	10	20
Sponsorship activation	47	10	7.31	40	1	14.3	10	20

Table 53 identifies the initial importance weightings given by the members of the Economic workstream to the six Drivers within the Economic Pillar.

Within the mean scores for the Drivers, there are three levels of assigned weightings. Tourism ($\bar{x}$ 23.9) and Direct Event Spending ($\bar{x}$ 21.1) were given the highest weights for importance. The next tier is Infrastructure ($\bar{x}$ 16.6), Business Activity ($\bar{x}$ 15.1), and Sponsorship Activation ($\bar{x}$ 14.3), which were given similar ratings. Security ($\bar{x}$ 9.0) was given the lowest rating for importance in assessing event economic impact.

Table 54 - Round 2 Economic Drivers, Consensus Range presented in Round 3

Workstream: **Economic**	Median	Group Consensus Range
Tourism	25	15-30
Infrastructure	15	10-20
Direct event spending	20	15-25
Security	10	5-10
Business activity	15	10-20
Sponsorship activation	10	10-20

Table 54 shows the Median and the Consensus Range (the 60% of responses between the Bottom and Top Quintiles) that were presented to Economic workstream respondents in Round 3.

Table 55 – Economic Drivers – Round 3, Revised weightings

Workstream: **Economic**	N	Median	STD DEV	Max	Min	Mean	Bottom quintile	Top quintile
Tourism	19	25	7.72	50	15	24.7	20	27
Infrastructure	19	15	5.65	30	5	17.6	15	20
Direct event spending	19	20	4.49	25	10	20.8	18	25
Security	19	10	4.27	20	1	9.40	5	10
Business activity	19	15	4.23	20	9	14.4	10	20
Sponsorship activation	19	10	5.86	25	1	13.1	10	19

Table 55 shows the summary results of weightings for each of the six Drivers following the second round of weightings by the Economic workstream in Round 3.

In Round 3, respondents were presented with the data in Table 54 as well as their previous weightings from Round 2. In that context, respondents could consider their initial weightings for each Driver against the Consensus Range and determine if they would want to reaffirm their initial weightings or revise them.

The results Round 3 show the highest weightings ascribed to Tourism ($\bar{x}$ 24.7) and Direct Event Spending ($\bar{x}$ 20.8) remain the top two Drivers in the workstream, with Infrastructure ($\bar{x}$ 17.6), Business Activity ($\bar{x}$ 14.4), and Sponsorship Activation ($\bar{x}$ 13.1) the next three in respective order. The lowest weighting was again allocated to Security ($\bar{x}$ 9.4).

The maximum and minimum scores reveal that some respondents chose to reaffirm their initial ratings from Round 2, giving weightings at the higher and lower end of the ranges and outside the Consensus Range presented. At the upper range was Tourism (50), with Infrastructure (5), Security (1) and Sponsorship Activation (1) at the lower end of the weightings.

Drivers Consensus – Economic

The results for both scoring rounds were tested to identify if there had been any statistical change in the overall Driver weightings from Round 2 to Round 3.

Table 56 - Driver Consensus testing - Economic

Workstream: **Economic**	Round	N	Mean	Differences in Mean	Std. Deviation	Differences in Std. Deviation
Tourism	2	47	23.89		10.801	
	3	19	24.74	+0.85	7.723	-3.08
Infrastructure	2	47	16.64		7.115	
	3	19	17.58	+0.94	5.650	-1.47
Direct event spending	2	47	21.11		8.044	
	3	19	20.79	+0.32	4.492	-3.55
Security	2	47	9.02		4.848	
	3	19	9.42	+0.40	4.273	-0.58
Business activity	2	47	15.06		5.407	
	3	19	14.42	-0.64	4.234	-1.17
Sponsorship activation	2	47	14.28		7.312	
	3	19	13.05	-1.23	5.864	-1.44

* Indicates statistically significant at 95% confidence level.
The absence of an * indicates no statistically significant differences in either mean or standard deviation.

The results in Table 56 shows the differences in mean weightings applied in Round 2 and Round 3 to each of the Drivers.

Tests of significance were twofold. The first was testing for changes in the mean scores given to each of the pillars, and the second was testing for changes in the variations of the weightings ascribed to each pillar.

While the results show that for all six Drivers the change in standard deviation was negative i.e. were nominally a narrower range of results in Round 3. Despite the consistency in direction of change, testing showed none were statistically significant differences in means or standard deviations between Round 2 and Round 3.

6.5.2 Driver weightings – Sport

Respondents who nominated "Sport" as an area of expertise in Round 1 were directed into the "Sport workstream". The weighting of importance required a minimum value of '1' and maximum value of '95', with a total of '100' points required across the drivers presented.

Table 57 - Sport Drivers – Round 2, Initial weightings

Workstream: **Sport**	N	Median	STD DEV	Max	Min	Mean	Bottom quintile	Top quintile
Event scale	54	16	8.26	50	1	18.0	10	22
Event prestige	54	20	10.24	60	3	22.4	15	30
Reputation of the sport	54	20	7.15	40	5	21.3	15	25
Sporting infrastructure	54	20	7.84	40	9	20.6	15	30
Sport expansion	54	15	8.28	45	1	17.8	10	25

'Sport' workstream weightings shown in Table 57 are more even in the distribution of mean scores across the drivers than those in the Economic workstream seen above in Table 55.

The sheer size of the event was not seen as the most important driver of event impact, with Event Scale ($\bar{x}$ 18.0) only marginally ahead of the opportunity to develop the sport as noted in Sport Expansion ($\bar{x}$ 17.8).

Respondents placed more weight on other aspects of events with Event Prestige ($\bar{x}$ 22.4) marginally ahead of Reputation ($\bar{x}$ 21.3) and Sporting Infrastructure ($\bar{x}$ 20.6). It is also worth noting that Event Prestige attracted the most diverse results (STDDEV 10.24) of the five Drivers tested.

Table 58 - Round 2 Sport Drivers, Consensus Range presented in Round 3

Workstream: **Sport**	Median	Group Consensus Range
Event scale	16	10-22
Event prestige	20	15-30
Reputation of the sport	20	15-25
Sporting infrastructure	20	15-30
Sport expansion	15	10-25

Table 58 shows the Median and the Group Consensus Range (the 60% of responses between the Bottom and Top Quintiles) that were presented to Sport workstream respondents in Round 3.

Table 59 – Sport Drivers – Round 3, Revised weightings

Workstream: **Sport**	N	Median	STD DEV	Max	Min	Mean	Bottom quintile	Top quintile
Event scale	28	18.5	7.16	40	5	17.5	10	20
Event prestige	28	20	6.84	40	12	23.3	20	30
Reputation of the sport	28	20	4.95	30	10	20.8	19	25
Sporting infrastructure	28	20	6.52	40	10	20.4	15	25
Sport expansion	28	17.5	5.71	32	10	18.0	15	23

Table 59 shows the results of weightings from Round 3 after reviewing their initial weightings for each Driver against the Group Consensus Range and determining if they would want to reaffirm their initial weightings or revise them.

The value of a research design that protects anonymity is evident in that the maximum and minimum values allocated to each Driver were at or outside the upper and lower values of the Group Consensus range. That is, respondents felt they were able to contribute their views without being compelled to conform to the group.

Consistent with the Round 2 results, Event Prestige, Sport Reputation, and Sport Infrastructure have ascribed the highest mean scores, with Event Scale and Sport Expansion slightly below those other Drivers.

Drivers Consensus – Sport

The results for both scoring rounds were tested to identify if there had been any statistical change in the overall Driver weightings from Round 2 to Round 3.

Table 60 - Driver Consensus testing - Sport

Workstream: **Sport**	Round	N	Mean	Differences in Mean	Std. Deviation	Differences in Std. Deviation
Event scale	2	54	17.96		8.264	
	3	28	17.50	-0.46	7.157	-1.11
Event prestige	2	54	22.37		10.243	
	3	28	23.29	+0.92	6.836	-3.41
Reputation of the sport	2	54	21.30		7.150	
	3	28	20.82	-0.48	4.945	-2.21
Sporting infrastructure	2	54	20.56		7.840	
	3	28	20.43	-0.13	6.523	-1.32
Sport expansion	2	54	17.83		8.278	
	3	28	17.96	+0.13	5.706	-2.57

* Indicates statistically significant at 95% confidence level.
The absence of an * indicates no statistically significant differences in either mean or standard deviation.

The results in Table 60 show the differences in mean weightings applied in Round 2 and Round 3 to each of the Drivers.

Two elements are noteworthy within the results. The differences in mean scores between Round 2 and Round 3 revealed relatively small shifts indicating no large-scale changes to weightings in Round 3. Additionally, no statistically significant differences in mean scores were found, indicating the Round 3 results were a refinement of, rather than a revision of, the initial ratings given.

As seen in the Economic workstream, all Sport Drivers demonstrated a consistent directional change of a reduced Standard Deviation. That is, the trend was toward a narrowing in the range of views towards a consensus. This meant that despite the limitation of sample size, none were found to be statistically significant.

6.5.3 Driver weightings – Media

Respondents who nominated "Media" as an area of expertise in Round 1 were directed into the "Media workstream". The weighting of importance required a minimum value of '1' and a maximum value of '95' ,with a total of '100' points required across the drivers presented.

Table 61 - Media Drivers – Round 2, Initial weightings

Workstream: **Media**	N	Median	STD DEV	Max	Min	Mean	Bottom quintile	Top quintile
Media output	23	20	8.16	45	10	23.7	20	30
Sponsors output	23	20	5.56	30	10	18.3	12	20
Media presence	23	20	5.60	30	10	19.1	15	22
Media accessibility	23	16	8.52	40	1	18.6	15	24
Audience size	23	20	8.92	40	5	20.4	10	25

Table 61 identifies the initial importance weightings given by the members of the Media workstream to the five Drivers within the Media Pillar.

Within the mean scores for the Drivers, Media Output is given the highest weighting ($\bar{x}$ 23.7) followed by Audience Size ($\bar{x}$ 20.4). The mean scores of the remaining three Drivers were narrowly ranged including Media Presence ($\bar{x}$ 19.1), Media Accessibility ($\bar{x}$ 18.6), and Sponsors Output ($\bar{x}$ 18.3). Media Presence and Sponsors Output also show the smallest values in Standard Deviation (σ 5.60 and σ 5.56 respectively), which indicates stronger alignment within the workstream on their weightings as compared to the other Drivers.

While the overall upper and lower quintile range values are closer as compared to Economic and Sport, the minimum scores for Media Accessibility (1) and Audience Size (5) indicate some strong individual disparity within the workstream. Scores at that level indicated little or no weight being placed on those elements, which for Audience Size is unusual given the importance placed on it by the rest of the workstream and its role as a driver of media value.

Table 62 - Round 2 Media Drivers, Consensus Range presented in Round 3

Workstream: **Media**	Median	Group Consensus Range

Media output	20	20-30
Sponsors output	20	12-20
Media presence	20	15-22
Media accessibility	16	15-24
Audience size	20	10-25

Table 62 shows the Median and the Group Consensus Range (the 60% of responses between the Bottom and Top Quintiles) that were presented to Media workstream respondents in Round 3.

Table 63 – Media Drivers – Round 3, Revised weightings

Workstream: **Media**	N	Median	STD DEV	Max	Min	Mean	Bottom quintile	Top quintile
Media output	9	20	6.69	30	10	21.8	18	28
Sponsors output	9	20	4.57	22	10	16.9	13	20
Media presence	9	20	3.91	30	15	20.6	20	20
Media accessibility	9	20	4.66	30	15	19.2	15	20
Audience size	9	20	8.02	35	10	21.6	15	27

Reviewing the results in Table 63 shows that the results from Round 3 are consistent with the initial weightings in Round 2. Again, Media Output ($\bar{x}$ 21.8) and Audience Size ($\bar{x}$ 21.6) featured as the highest scoring Drivers but they remain only marginally ahead of Media Presence ($\bar{x}$ 20.6) and Media Accessibility ($\bar{x}$ 19.2). As the weighting process is a 'constant sum' approach, any improvement in mean scores comes from down weighting other drivers. In this case, Sponsors output ($\bar{x}$ 16.9) attracted a lower mean score than it did in Round 2.

Drivers Consensus – Media

The results for both scoring rounds were tested to identify if there had been any statistical change in the overall Driver weightings from Round 2 to Round 3.

Table 64 - Driver Consensus testing - Media

Workstream: **Media**	Round	N	Mean	Differences in Mean	Std. Deviation	Differences in Std. Deviation
Media output	2	23	23.65		8.155	
	3	9	21.78	-1.87	6.685	-1.47
Sponsors output	2	23	18.26		5.561	
	3	9	16.89	-1.37	4.567	-0.99
Media presence	2	23	19.13		5.603	
	3	9	20.56	+1.43	3.909	-1.69
	2	23	18.61		8.516	

Media accessibility	3	9	19.22	+0.61	4.658	-3.89
Audience size	2	23	20.35		8.922	
	3	9	21.56	+1.21	8.017	-0.91

* Indicates statistically significant at 95% confidence level.
Absence of an * indicates no statistically significant differences in either mean or standard deviation.

The results in Table 64 shows the differences in mean weightings and standard deviation of those weightings applied to each of the drivers in Round 2 and Round 3 by the Media workstream participants.

While the changes in mean scores between rounds are larger than for the Sport workstream, they remain within two points of the original weightings. Testing the mean scores found no statistically significant differences between the weightings allocated in the two rounds.

One consistent theme observed across the workstreams was that while the mean weightings remain consistent from round to round, there is a trend towards a reduction in the Standard Deviation for the weightings. While no changes were statistically significant, the consistency in direction for all drivers suggests that with a larger sample base a significant movement towards consensus might be observed and claimed.

6.5.4 Driver weightings – Social

Respondents who nominated "Social" as an area of expertise in Round 1 were directed into the "Social workstream". The weighting of importance required a minimum value of '1' and maximum value of '95', with a total of '100' points required across the drivers presented.

Table 65 - Social Drivers – Round 2, Initial weightings

Workstream: **Social**	N	Median	STD DEV	Max	Min	Mean	Bottom quintile	Top quintile
Community cohesion and quality	31	25	6.34	40	10	24.2	20	30
Volunteering	31	15	5.88	30	5	16.3	10	20
Goodwill / Pride	31	20	8.59	50	5	17.5	10	20
Social engagement	31	22	6.64	35	5	23.1	20	30
Health and Active living	31	20	8.00	35	5	19.0	15	25

Table 65 identifies the initial importance weightings given by the members of the Social workstream to the five Drivers within the Social Pillar.

Like the weightings observed in the Economic workstream, the mean scores show two drivers immediately attracting higher ratings than the remaining drivers. Community Cohesion ($\bar{x}$ 24.2) and Social Engagement ($\bar{x}$ 23.1) are given the highest weightings and attract relatively low Standard Deviation scores of (σ 6.34) and (σ 6.64) respectively. The only driver that showed even stronger initial agreement in the weighting was Volunteering (σ 5.88), but it was also given the lowest weighting by the workstream ($\bar{x}$ 16.3).

The upper and lower quintile range values are closer than for the Economic and Sport drivers, with all except Community Cohesion still recorded at least one value of five (5), and in the case of Goodwill/Pride, also attracting the highest value of 50 points.

Table 66 - Round 2 Social Drivers, Consensus Range presented in Round 3

Workstream: **Social**	Median	Group Consensus Range
Community cohesion and quality	25	20-30
Volunteering	15	10-20
Goodwill / Pride	20	10-20
Social engagement	22	20-30
Health and Active living	20	15-25

The Median and the Group Consensus Range results from Round 2 (as shown in Table 66) were presented to Social workstream respondents in Round 3.

Table 67 – Social Drivers – Round 3, Revised weightings

Workstream: **Social**	N	Median	STD DEV	Max	Min	Mean	Bottom quintile	Top quintile
Community cohesion and quality	21	25	3.83	30	20	23.8	20	25
Volunteering	21	15	5.09	25	5	16.0	10	20
Goodwill / Pride	21	20	5.24	25	5	17.8	10	20
Social engagement	21	20	4.58	30	15	22.6	20	25
Health and Active living	21	20	6.72	35	5	19.9	15	25

Table 67 shows the results following the second round of weightings by the Social workstream. These reflect the results of reviewing their initial weightings for each Driver against the Group Consensus Range and determining if they would want to reaffirm their initial weightings or revise them.

Consistent with prior workstream results. the value of anonymity is evident in that some of the weightings in Round 3 included values outside the upper and lower values of the Group Consensus range from Round 2. Equally, it is vital to recognise the expression of diverse viewpoints on individual drivers did not disrupt the value of trying to seek an overall consensus. This is reflected by Community Cohesion ($\bar{x}$ 23.8) and Social Engagement ($\bar{x}$ 22.6) again received the highest overall ratings and Volunteering ($\bar{x}$16.0) anchoring the lower end of the weightings for the workstream.

Drivers Consensus – Social

The results for both scoring rounds were tested to identify if there had been any statistical change in the overall Driver weightings from Round 2 to Round 3.

Table 68 - Driver Consensus testing - Social

Workstream: **Social**	Round	N	Mean	Differences in Mean	Std. Deviation	Differences in Std. Deviation
Community cohesion and quality	2	31	24.16		6.336	
	3	21	23.76	-0.40	3.833	-2.50*
Volunteering	2	31	16.29		5.878	
	3	21	15.95	-0.34	5.094	-0.78
Goodwill / Pride	2	31	17.52		8.590	
	3	21	17.76	0.24	5.243	-3.34
Social engagement	2	31	23.06		6.638	
	3	21	22.62	-0.44	4.577	-2.06
Health and Active living	2	31	18.97		8.002	
	3	21	19.90	+0.93	6.722	-1.28

* Indicates statistically significant at 95% confidence level.
The absence of an * indicates no statistically significant differences in either mean or standard deviation.

The results in Table 68 above show the differences in both mean weightings and changes in Standard Deviation applied to the drivers by the Social workstream participants.

The changes in mean scores between rounds are small compared to other workstreams, with the largest difference of Health and Active Living ($\overline{x}$ +0.93) still within a single point of the original weightings. Not surprisingly given their round-to-round stability, testing the mean scores found no statistically significant differences between the weightings allocated in the two rounds.

The directional trend to narrowing of Standard Deviations seen in other workstreams is reflected again in the Social workstream except for 'Community cohesion and quality'. While there was little change from the original mean score weighting ($\overline{x}$ -0.40), the driver had a strong narrowing of views around that score (σ -2.50), where the Levene's test for equality of variances showed a significance of 0.042. This provided a significant example of consensus formation that is only otherwise seen as a strong consistent trend in the other drivers.

6.5.5 Driver weightings – Brand

Respondents who nominated "Brand" as an area of expertise in Round 1 were directed into the "Brand workstream". The weighting of importance required a minimum value of '1' and a maximum value of '95', with a total of '100' points required across the drivers presented.

Table 69 - Brand Drivers – Round 2, Initial weightings

Workstream: **Brand**	N	Median	STD DEV	Max	Min	Mean	Bottom quintile	Top quintile
Host attention	32	15	6.33	30	2	15.1	10	20
Host attractiveness	32	20	6.14	30	5	19.5	15	25

Image building – Host	32	20	6.19	40	10	21.1	16	25
Image building – Event	32	25	9.08	50	10	24.2	16	30
Host reputation	32	20	8.87	50	3	20.1	15	29

The initial importance weightings provided by the members of the Brand workstream to the five drivers that underpin the Brand Pillar are shown in Table 69 above.

Unlike Economic and Sport workstreams where two drivers attracted higher ratings than the other drivers, Brand is closer to the Media workstream in that both attracted the strongest weighting to a single Driver. In the case of the Brand workstream, that was 'Image building – Event' ($\bar{x}$ 24.2) and for Media it was 'Media Output' at ($\bar{x}$ 23.7). However, that highest weighting also attracted the widest range of views (σ 9.08) from within the workstream.

Other Drivers were quite closely balanced with 'Image building – Host' ($\bar{x}$ 21.1) followed by Host reputation ($\bar{x}$ 20.1) and Host attractiveness ($\bar{x}$ 19.5).

Reflecting on the allocation of scores, the lowest weighting given within the brand workstream was given to 'Host attention' ($\bar{x}$ 15.1), which is often cited as a key rationale given for event hosting. Conversely, high weighting on Event Image building reveals a more event owner focus within the workstream participants.

Table 70 - Round 2 Brand Drivers, Consensus Range presented in Round 3

Workstream: **Brand**	Median	Group Consensus Range
Host attention	15	10-20
Host attractiveness	20	15-25
Image building – Host	20	16-25
Image building – Event	25	16-30
Host reputation	20	15-29

The Median and the Group Consensus Range results from Round 2 (as shown in Table 70) were presented to Brand workstream respondents in Round 3.

Table 71 – Brand Drivers – Round 3, Revised weightings

Workstream: **Brand**	N	Median	STD DEV	Max	Min	Mean	Bottom quintile	Top quintile
Host attention	15	15	4.71	25	10	16.0	10	20
Host attractiveness	15	20	3.78	30	15	20.0	19	20
Image building - Host	15	20	4.23	25	10	20.0	19	25
Image building - Event	15	25	5.92	30	10	23.0	20	30
Host reputation	15	20	6.60	30	10	21.0	15	30

Table 71 shows the results following the second round of weightings by the Brand workstream after determining if they would want to reaffirm their initial weightings or revise them having considered the Group Consensus Range.

The same pattern is evident with 'Image building – Event' ($\bar{x}$ 23.0) given the highest weighting and 'Host attention' ($\bar{x}$ 16.0) remaining the lowest rated, with the remaining Drivers are closely grouped ($\bar{x}$ 20.0 – 21.0).

Drivers Consensus – Brand

The results for both scoring rounds were tested to identify if there had been any statistical change in the overall Driver weightings from Round 2 to Round 3.

Table 72 - Driver Consensus testing - Brand

Workstream: **Brand**	Round	N	Mean	Differences in Mean	Std. Deviation	Differences in Std. Deviation
Host attention	2	32	15.06		6.334	
	3	15	16.00	+0.94	4.706	-1.63
Host attractiveness	2	32	19.53		6.138	
	3	15	20.00	+0.47	3.780	-2.36
Image building – Host	2	32	21.09		6.187	
	3	15	20.00	-1.09	4.226	-1.96
Image building – Event	2	32	24.22		9.079	
	3	15	23.00	-1.22	5.916	-3.16
Host reputation	2	32	20.09		8.870	
	3	15	21.00	+0.91	6.601	-2.27

* Indicates statistically significant at 95% confidence level.
The absence of an * indicates no statistically significant differences in either mean or standard deviation.

The results in Table 72 shows the differences in mean weightings and standard deviation of those weightings applied to each of the drivers in Round 2 and Round 3 by the Brand workstream participants.

While the trend towards a reduction in the Standard Deviation for the weightings is consistent with all other workstreams, again the limitation in sample base precludes a definitive move toward the consensus being claimed.

While the same limitation applies to changes in the mean scores for each of the individual Drivers, it also reflects a consistency with the first weightings provided. Looking more widely at the changes in mean scores across the Brand workstream, the Drivers are moving into an overall narrower range. The between rounds change in the highest-rated Driver 'Image building – Event' ($\bar{x}$ -1.22) and the lowest Driver 'Host attention' ($\bar{x}$ +0.94) narrows the range from 9.16 to 7.00, with the other three Drivers still bounded within the reduced range.

6.5.6 Driver weightings – Environment

Respondents who nominated "Environment" as an area of expertise in Round 1 were directed into the "Environment workstream". The weighting of importance required a minimum value of '1' and a maximum value of '95', with a total of '100' points required across the drivers presented.

Table 73 - Environment Drivers – Round 2, Initial weightings

Workstream: **Environment**	N	Median	STD DEV	Max	Min	Mean	Bottom quintile	Top quintile
Energy efficiency and quality	5	25	4.18	30	20	24.0	20	26
Consumption footprint	5	30	11.40	50	20	31.0	24	34
Building design leadership	5	20	7.42	30	10	21.0	18	26
Resource utilisation	5	25	4.18	30	20	24.0	20	26

With only four Drivers within the Environment Pillar, the values of the means scores are higher than those for the other workstreams where weights were allocated across five Drivers (six in Economic). The weightings given by the members of the Environment workstream to the four drivers are shown in Table 73 above.

Despite the reduced number of Drivers and higher scores, Environment is similar to the Brand and Media workstream with the strongest weighting directed to a single Driver. In the case of the Environment workstream, that was 'Consumption Footprint' ($\bar{x}$ 31.0). This Driver also attracted the widest range of views (σ 11.40) from within the workstream.

Other Drivers were more closely balanced with 'Energy efficiency' and 'Resource utilisation' at ($\bar{x}$ 24.0), followed by 'Building design' ($\bar{x}$ 21.0).

Table 74 - Round 2 Environment Drivers, Consensus Range presented in Round 3

Workstream: **Environment**	Median	Group Consensus Range
Energy efficiency and quality	25	20-26
Consumption footprint	30	24-34
Building design leadership	20	18-26
Resource utilisation	25	20-26

The Median and the Group Consensus Range results from Round 2 (as shown in Table 74) were presented to Environment workstream respondents in Round 3.

Table 75 – Environment Drivers – Round 3, Revised weightings

Workstream: **Environment**	N	Median	STD DEV	Max	Min	Mean	Bottom quintile	Top quintile
Energy efficiency and quality	2	27.5	3.54	30	25	27.5	26	29

Consumption footprint	2	35	7.07	40	30	35.0	32	38
Building design leadership	2	17.5	3.54	20	15	17.5	16	19
Resource utilisation	2	20	7.07	25	15	20.0	17	23

The results of the second round of weightings by the Environment workstream are shown in Table 75. Similar results to Round 2 are evident. 'Consumption Footprint' ($\bar{x}$ 35.0) again attracted the highest weighting, followed by 'Energy efficiency' ($\bar{x}$ 27.5) and 'Resource utilisation' ($\bar{x}$ 20.0), with 'Building design' ($\bar{x}$ 17.5) again receiving the lowest weighting.

Drivers Consensus – Environment

The results for both scoring rounds were tested to identify if there had been any statistical change in the overall Driver weightings from Round 2 to Round 3.

Table 76 - Driver Consensus testing - Environment

Workstream: **Environment**	Round	N	Mean	Differences in Mean	Std. Deviation	Differences in Std. Deviation
Energy efficiency and quality	2	5	24.00		4.183	
	3	2	27.50	+3.50	3.536	-0.65
Consumption footprint	2	5	31.00		11.402	
	3	2	35.00	+4.00	7.071	-4.33
Building design leadership	2	5	21.00		7.416	
	3	2	17.50	-3.50	3.536	-3.88
Resource utilisation	2	5	24.00		4.183	
	3	2	20.00	-4.00	7.071	+2.89

* Indicates statistically significant at 95% confidence level.
The absence of an * indicates no statistically significant differences in either mean or standard deviation.

A key factor in reviewing the results differences between Rounds 2 and 3 in Table 76 is to note that fewer respondents nominated Environment as an area of expertise in Round 1 than for the other Pillars. This curtails the inferences that might be drawn despite the larger changes in both mean scores and standard deviation as compared to other workstreams. That is confirmed in that despite the larger values, there is no statistical difference in means or standard deviations between Round 2 and Round 3 for each of the above Drivers.

Despite only two respondents completing the final consensus for Round 3, the trend to a narrowing of standard deviations with three of the four drivers showed the same direction change as the 26 drivers in the prior workstream.

Hence, it is only under very the constrained circumstances of the Environment workstream that the only instance of a widening of standard deviation from Round 2 to Round 3 across all 30 drivers tested is observed with 'Resource utilisation' (σ +2.89).

6.6 Summary table of scores for all Pillars and Drivers

Given the final results of the Delphi rounds for both pillars and drivers were analysed within each workstream a table summarising all results under each pillar is presented below.

Table 77 - Summary table of scores for all Pillars and Drivers

	Consensus Mean Score^		Consensus Mean Score^
Pillar - Economic	**23.59**	**Pillar - Sport**	**22.99**
Tourism	23.89	Event prestige	22.37
Direct event spending	21.11	Reputation of the sport	21.30
Infrastructure	16.64	Sporting infrastructure	20.56
Business activity	15.06	Event scale	17.94
Sponsorship activation	14.28	Sport expansion	17.83
Security	9.02		
Pillar - Media	**15.04**	**Pillar - Social**	**14.08**
Media output	23.65	Community cohesion & quality	24.16
Audience size	20.35	Social engagement	23.06
Media presence	19.13	Health and Active living	18.97
Media accessibility	18.61	Goodwill / Pride	17.52
Sponsors output	18.26	Volunteering	16.29
Pillar - Brand	**14.48**	**Pillar - Environment**	**9.71**
Image building – Event	24.22	Consumption footprint	31.00
Image building – Host	21.09	Energy efficiency and quality	24.00
Host reputation	20.09	Resource utilisation	24.00
Host attractiveness	19.53	Building design leadership	21.00
Host attention	15.06		

Table 77 above collates the results from the final consensus scores for the six pillars from Table 52 and the mean scores from each of the workstreams for each of the 30 drivers. The drivers are presented within each of the relevant pillars in order of descending mean score.

6.7 Indicator ratings

As noted in the methodology, the measures of event impact were rated on two dimensions in consecutive rounds – Importance and Reliability.

In the context of this study, the definitions provided to respondents against which to make the assessment were as follows:

- *Importance – Rating the importance of each measure in assessing the potential [Pillar] impact.*
- *Reliability – Rating the reliability with which each measure is consistently measured and available for assessing the [Pillar] impact.*

The first dimension 'Importance' was rated during Round 2 following the allocation of weightings to Drivers. The subsequent ratings of Reliability were completed in Round 3. Each set of measures was rated within its relevant workstream.

In the tables below, the results of each measure on both Mean and Standard Deviation are reported. Consistent with 5.5.3 "Rating of Indicators" in the research method, all presented measures were rated for Importance, with the highest importance rated measures then represented for a rating of Reliability in Round 3. As a combined construct of the two dimensions tested, the Impact Score [$\bar{x}$ I * $\bar{x}$ R] = $\bar{x}$ Importance * $\bar{x}$ Reliability is also included in each table.

As noted for the top 24 measures for $\bar{x}$ Importance, the Impact Score is a product of the two rated mean scores. For measures outside the top 24 $\bar{x}$ Importance were not presented for reliability rating. Hence instead the minimum $\bar{x}$ Reliability achieved within the workstream is awarded to each of the measures without a Reliability value, and used to establish a proxy Impact Score.

6.7.1 Indicator ratings: Economic

In Round 2, a total of 40 measures were presented to Economic Workstream respondents for rating their importance in assessing event impact. In Round 3, a subset of 24 measures was represented for Reliability in assessing event impact.

Table 78 - Measure ratings: Economic

Workstream: Economic	Importance (I) n=47		Reliability (R) n=19		Impact Score
	Mean	Std. Dev.	Mean	Std. Dev.	[$\bar{x}$ I* $\bar{x}$ R]
Total Number of tickets sold	3.98	0.921	4.58	0.507	**18.2**
Visitor numbers: Total (count)	4.34	0.668	3.95	0.705	**17.1**
Average stay ticketed spectators (days)	4.06	0.895	4.16	0.501	**16.9**
Ticket sales: Total value of sales ($)	3.81	0.876	4.42	0.507	**16.8**
Visitor: Room nights (Count)	4.28	0.615	3.84	0.602	**16.4**
Total Operational Spend ($)	3.98	0.872	4.00	0.667	**15.9**
Visitor expenditure: Total all visitors ($)	4.38	0.768	3.63	0.895	**15.9**
Visitor expenditure: Average spend per day, per visitor ($)	4.36	0.605	3.53	0.905	**15.4**
Visitor numbers: International (count)	4.40	0.771	3.47	0.905	**15.3**
Visitor Length of stay: Average number of days (n)	4.19	0.770	3.63	0.684	**15.2**
Total Number of tickets available	3.57	0.903	4.26	0.872	**15.2**
Total unique number of ticketholders	3.79	0.907	3.95	0.705	**15.0**
Event funding inflows (from outside host) ($)	4.00	0.834	3.74	0.872	**15.0**
Visitor expenditure: Total spent over total stay ($)	4.30	0.778	3.42	0.902	**14.7**
Visitor: Room nights cost average ($)	3.74	0.846	3.89	0.658	**14.5**
Visitor numbers: Per day (count)	3.81	1.014	3.79	0.787	**14.4**
Total Capital Spend (incl. infrastructure) ($)	3.77	0.960	3.79	0.976	**14.3**
Visitor numbers: National (count)	3.83	0.868	3.63	0.895	**13.9**

Workstream: Economic	Importance (I) n=47		Reliability (R) n=19		Impact Score
	Mean	Std. Dev.	Mean	Std. Dev.	[$\bar{x}$ I* $\bar{x}$ R]
Visitor numbers: Regional (count)	3.72	0.926	3.58	0.838	**13.3**
Total number of attendees fanzone (international)	3.64	0.987	3.47	1.073	**12.6**
Retail sales: Total increased spend during event ($)	3.77	0.937	3.32	0.885	**12.5**
Jobs created: Event host direct (n)	3.79	0.999	3.26	1.098	**12.4**
Income growth: Venue region ($)	3.60	0.993	2.79	1.032	**10.0**
Jobs created: Nation/Region/City (n)	3.62	1.074	2.63	1.065	**9.5**
Visitor numbers: Local (count)	3.53	1.158	-	-	**9.3**
Hosting media spend: Total advertising dollars ($)	3.51	0.856	-	-	**9.2**
Total number of attendees fanzone (in country)	3.45	0.775	-	-	**9.1**
Average stay accredited personnel (days)	3.40	0.948	-	-	**8.9**
Total number of attendees fanzone (in host city)	3.32	0.980	-	-	**8.7**
Spectators: Number of Nations	3.30	0.998	-	-	**8.7**
GDP growth: current year (%)	3.28	1.192	-	-	**8.6**
Total of Accredited personnel (competitors, officials, media)	3.19	0.924	-	-	**8.4**
Total Number of unticketed spectators	3.09	1.195	-	-	**8.1**
Income growth: per Household ($)	3.04	1.197	-	-	**8.0**
Resident expenditure at event ($)	3.00	1.142	-	-	**7.9**
Accreditations: Number of Nations	2.96	0.955	-	-	**7.8**
Security costs: Total ($)	2.83	1.110	-	-	**7.4**
Security expenditure: Host government ($)	2.77	1.146	-	-	**7.3**
Staff size: Organising committee (FTE)	2.77	1.088	-	-	**7.3**
Security expenditure: Organising Committee ($)	2.74	1.170	-	-	**7.2**

The results in Table 78 highlight the importance of capturing not only the importance of a prospective measure but also the confidence with which that measure can be relied upon.

Ordering the potential measures provided by descending value of their Impact Score ($\bar{x}$ Importance * $\bar{x}$ Reliability) shows that the measure with the highest mean score for Importance in Round 2 'Visitor numbers: International' ($\bar{x}$ 4.40) only rates as the ninth strongest once the reliability of that measure is factored in.

The overall highest Impact Score of 18.2 was achieved by the 'Total Number of tickets sold', despite only rating as the tenth highest-rated for Importance ($\bar{x}$ 3.98). However, it did receive the highest Reliability rating ($\bar{x}$ 4.58).

Table 79 – Rating disparities: Economic

Rated Measures - Economic	Importance ($\bar{x}$)	Reliability ($\bar{x}$)	R $\bar{x}$ - I $\bar{x}$
Total Number of tickets available	3.57	4.26	**+0.69**
Ticket sales: Total value of sales ($)	3.81	4.42	**+0.61**

Total Number of tickets sold	3.98	4.58	**+0.60**
Visitor expenditure: Total spent over total stay ($)	4.30	3.42	**-0.88**
Visitor numbers: International (count)	4.40	3.47	**-0.93**
Jobs created: Nation/Region/City (n)	3.62	2.63	**-0.99**

A summary of the three largest differences (positive and negative) between the Reliability rating and Importance is shown in Table 79. The results show measures that are common to almost all events and involve a basic automated count are amongst those more strongly reliable than a simple rating of importance would imply. Each of the three largest positive [R $\bar{x}$ > I $\bar{x}$] differences are related to ticketing and while not weak on importance, it is [+0.60 to +0.69] even stronger on ratings of Reliability.

Conversely, the measures that showed the largest negative were visitor spending levels [-0.88], the number of visitors arriving for the event [-0.93], and realising new job creation [-0.99], and are rated much higher on importance than can be relied upon in assessments. Given how strongly the rationale for the 'visitor economy' effect is used in the assessment of events, the results underline the value in understanding both the importance and reliability when considering assessment dimensions.

6.7.2 Indicator ratings: Media

In Round 2, a total of 39 measures were presented to Media Workstream respondents for rating their importance in assessing event impact. In Round , a subset of 24 measures was represented for Reliability in assessing event impact.

Table 80 - Measure ratings: Media

Workstream: Media	Importance (I) n=23		Reliability (R) n=9		Impact Score
	Mean	Std. Dev.	Mean	Std. Dev.	[$\bar{x}$ I* $\bar{x}$ R]
Audience size (individual viewers reached)	4.48	0.593	3.67	1.000	**16.4**
Broadcast rights revenues ($)	3.91	1.041	4.11	0.928	**16.1**
Total Number of Broadcast Nations	4.22	1.043	3.78	1.093	**16.0**
Audience size (total viewership cumulative)	4.39	0.722	3.44	1.014	**15.1**
Media impact (awareness of event)	4.13	0.626	3.56	0.726	**14.7**
Total Number of Hours Broadcast	4.00	0.905	3.67	1.118	**14.7**
Social media - Facebook interactions	3.74	0.752	3.89	0.601	**14.5**
Total Reach	4.30	0.703	3.33	0.866	**14.3**
Broadcasters: total	3.74	1.010	3.78	0.972	**14.1**
TV Hours produced	3.65	0.885	3.78	1.394	**13.8**
Total Cumulative Audience	4.26	0.752	3.22	1.202	**13.7**
Media coverage (news items) (hours)	4.09	0.949	3.33	0.866	**13.6**
TV hours watched: free-to-air	4.04	0.878	3.33	0.866	**13.5**
Peak Viewing Audience (count)	4.17	0.717	3.22	0.833	**13.4**
Total Number of Unique Visitors	3.87	0.869	3.44	0.726	**13.3**

Workstream: Media	Importance (I) n=23		Reliability (R) n=9		Impact Score
	Mean	Std. Dev.	Mean	Std. Dev.	[$\bar{x}$ I* $\bar{x}$ R]
Total Number of Broadcasters	3.65	0.885	3.56	1.236	**13.0**
Peak Market Share (%)	3.96	0.706	3.22	0.833	**12.8**
Advertising total spend ($)	3.78	0.850	3.33	1.118	**12.6**
Social media - Sentiment ratio	3.74	1.096	3.33	1.000	**12.5**
TV hours watched: pay-TV	3.74	0.864	3.33	1.000	**12.5**
Media exposure value ($)	3.83	0.834	3.22	0.833	**12.3**
Global Rank (Peak)	3.78	0.736	3.22	0.833	**12.2**
Audience size (total viewership potential)	4.04	1.107	3.00	1.000	**12.1**
Average Audience Rating (TARP)	3.91	0.793	3.33	0.707	**11.7**
Sponsorship advertising spend ($)	3.57	0.896	-	-	**10.7**
Website: total visitors	3.52	0.898	-	-	**10.6**
Total Number of Page Impressions	3.52	0.846	-	-	**10.6**
Website: video streams (hours)	3.52	1.039	-	-	**10.6**
Media infrastructure (quality of facilities)	3.48	0.898	-	-	**10.4**
Newspaper, news media articles (count)	3.48	0.846	-	-	**10.4**
Social media - Google searches	3.48	0.846	-	-	**10.4**
Social media - Twitter mentions	3.48	0.898	-	-	**10.4**
Media agencies present (count)	3.43	0.992	-	-	**10.3**
YouTube Subscribers	3.39	0.783	-	-	**10.2**
Twitter Followers	3.35	0.885	-	-	**10.1**
Accredited media representatives (count)	3.26	0.915	-	-	**9.8**
Total Number of Programs	3.26	0.864	-	-	**9.8**
Facebook Likes	3.17	0.937	-	-	**9.5**
Instagram Followers	3.00	0.905	-	-	**9.0**

The results in Table 80 order the potential Media measures by their Impact Score [$\bar{x}$ I * $\bar{x}$ R]

The measure with the highest mean score for Importance in Round 2 ‘Audience size - individual viewers reached’ ($\bar{x}$ 4.48) also achieved the highest overall Impact Score [16.4]. By comparison, the highest Impact Score achieved within the Economic Measures was [18.2] for ‘Total Number of tickets sold’, with ‘Visitor: Room nights (Count)’ achieving an Impact Score of [16.4] as the fifth highest placed measure in that workstream.

The higher overall scores have a strong ‘traditional media’ focus around audience size, broadcast content, and global reach, with only one item relating to social media included in the top ten of Impact Scores. ‘Social media – Facebook interactions’, which achieves an Impact Score of 14.5 through a balance on both rating dimensions. Despite this, other social media related measures were not rated as important enough to make the top 24 cut-off for

the secondary rating on reliability. Indeed, six of the lowest ten importance ratings were given to social media related measures.

Table 81 - Rating disparities: Media

Rated Measures - Media	Importance ($\bar{x}$)	Reliability ($\bar{x}$)	R $\bar{x}$ - I $\bar{x}$
Broadcast rights revenues ($)	3.91	4.11	**+0.20**
Audience size (total viewership cumulative)	4.39	3.44	**-0.95**
Peak Viewing Audience (count)	4.17	3.22	**-0.95**
Total Reach	4.30	3.33	**-0.97**
Total Cumulative Audience	4.26	3.22	**-1.04**
Audience size (total viewership potential)	4.04	3.00	**-1.04**

For the Media measures, the disparity in ratings of Importance and Reliability are not evenly balanced. As shown in Table 81, the highest positive result [R $\bar{x}$ > I $\bar{x}$] is for 'Broadcast rights revenues' with a slim [+0.20], but there are more negative [R $\bar{x}$ < I $\bar{x}$] results with strong scores for Importance diluted by lower Reliability scores.

Despite the strong presence of Audience measures in the overall Impact Scores, they also achieve the highest negative results [-0.95 to -1.04] as the veracity of audience measures is challenged. The results reflect commentary on inflated or unverifiable audience numbers (which are important for determining media value) claimed by event owners as the workstream assessed their reliability. These differences consistently focus on Audience size despite the ratings of Importance and Reliability being conducted over separate rounds and the order or measurement items randomised.

6.7.3 Indicator ratings: Sport

In Round 2, a total of 37 measures were presented to Sport Workstream respondents for rating their importance in assessing event impact. In Round 3, a subset of 24 measures was represented for Reliability in assessing event impact.

Table 82 - Measure ratings: Sport

Workstream: Sport	Importance (I) n=54		Reliability (R) n=28		Impact Score
	Mean	Std. Dev.	Mean	Std. Dev.	[$\bar{x}$ I* $\bar{x}$ R]
Number of Nations	3.85	0.899	3.96	0.999	**15.2**
Total Number of Competitors - Elite	3.81	0.892	3.96	0.881	**15.1**
Event ranking points classification	3.85	0.810	3.89	0.832	**15.0**
Number of association member countries (count)	3.72	0.899	4.00	0.816	**14.9**
Infrastructure: Development of sport infrastructure	3.91	0.708	3.75	0.887	**14.7**
Sport development: Increased sport funding ($)	3.78	0.984	3.86	0.848	**14.6**
Total Number of Competitors	3.72	0.834	3.89	0.875	**14.5**
Infrastructure: Development of elite sport infrastructure	3.74	0.828	3.71	0.713	**13.9**
Athlete Achievement: Records set	3.57	0.716	3.86	1.177	**13.8**

Workstream: Sport	Importance (I) n=54		Reliability (R) n=28		Impact Score
	Mean	Std. Dev.	Mean	Std. Dev.	[$\bar{x}$ I* $\bar{x}$ R]
Infrastructure: Increased sport space	3.61	0.834	3.75	0.887	**13.5**
Total event duration (days)	3.52	1.077	3.61	1.031	**12.7**
Total Venue Capacity	3.41	0.714	3.68	1.056	**12.5**
Medal events (count)	3.26	0.873	3.79	1.197	**12.4**
Number of Venues (Renovated)	3.33	0.890	3.71	0.854	**12.4**
Increase in domestic association memberships (count)	3.52	1.059	3.50	1.072	**12.3**
Number of Venues (Existing)	3.37	1.051	3.64	0.826	**12.3**
Total Number of Continents	3.33	1.116	3.68	1.020	**12.3**
Total number competition days	3.43	0.983	3.57	1.069	**12.2**
Athlete Achievement: Development pathway	3.65	0.850	3.29	0.854	**12.0**
Number of Venues (total)	3.20	0.939	3.57	0.879	**11.4**
Commitment to Fair Play	3.65	0.994	3.11	1.066	**11.4**
Number of competitors excluded (drug testing)	3.30	1.127	3.39	1.227	**11.2**
Infrastructure: Urban regeneration	3.33	1.064	3.29	0.897	**11.0**
Infrastructure: Negative effect mega event development	3.31	1.061	2.68	0.945	**8.9**
Number of Venues (New)	3.20	0.998	-	-	**8.6**
Number of Sports (count)	3.17	0.966	-	-	**8.5**
Prize money total ($)	3.17	1.060	-	-	**8.5**
Total Number of Female Competitors	3.17	0.927	-	-	**8.5**
Total Number of Competitors - Non-Elite	3.15	0.998	-	-	**8.4**
Number of Disciplines (Count)	3.11	0.904	-	-	**8.3**
Sport products: Increased retail sales ($)	3.00	0.911	-	-	**8.0**
Total Number of Male Competitors	2.89	0.965	-	-	**7.7**
Total Number of Sessions	2.85	0.833	-	-	**7.6**
Number of Venues (Temporary)	2.81	0.913	-	-	**7.5**
Number of Officials	2.70	0.816	-	-	**7.2**
Number of competitors excluded (game infringements)	2.69	0.987	-	-	**7.2**
Official charity partner presence	2.61	1.054	-	-	**7.0**

The results shown in Table 82 are the rating of potential Sport measures by their Impact Score [$\bar{x}$ I * $\bar{x}$ R].

Unlike the Economic workstream, none of the Sport measures achieved a mean score for Importance above $\bar{x}$ 4.0 in Round 2. However, as both workstreams share a similar low end mean score, the Sport scores are as a result more compressed in the range of Importance ratings given.

The highest-rated measure is 'Infrastructure: Development of sport infrastructure' ($\bar{x}$ 3.91) noting the sport-specific focus rather than broader infrastructure seen in the Economic

Pillar. This measure is only fifth in terms of Impact Score (14.7) for Sport with a slightly lower Reliability ($\bar{x}$ 3.75). The highest overall Impact Score - 'Number of Nations' (15.2) - benefits from its simplicity with Reliability ($\bar{x}$ 3.96) improving its Importance ($\bar{x}$ 3.86). Overall, the impact scores for Sport reflect a stronger focus on elite competition (number of elite competitors and event ranking) and formal association with the sport ahead of the number of event participants or duration of the event itself.

The impact of the event on future sport development demonstrates a commitment to Fair Play as additional sport spaces and talent pathways and were all rated in the top 24 items for Importance, with a simpler measure of 'Increased sport funding ($)' achieving a higher overall Impact Score (14.6) than the more detailed or specific initiatives.

Table 83 - Rating disparities: Sport

Rated Measures - Sport	Importance ($\bar{x}$)	Reliability ($\bar{x}$)	R $\bar{x}$ - I $\bar{x}$
Medal events (count)	3.26	3.79	**+0.53**
Number of Venues (Renovated)	3.33	3.71	**+0.38**
Number of Venues (total)	3.20	3.57	**+0.37**
Athlete Achievement: Development pathway	3.65	3.29	**-0.36**
Commitment to Fair Play	3.65	3.11	**-0.54**
Infrastructure: Negative effect mega-event development	3.31	2.68	**-0.63**

As seen in the Economics workstream, the measures for their Reliability are rated higher than their Importance, and the inverse is also true. Within the Sport measures, the disparity in ratings of Importance and Reliability are fairly balanced. As shown in Table 83, the differences (positive and negative) are of a similar scale.

Simple 'count' measures are seen to be more Reliable but of lesser importance in terms of event impact as compared to the longer-term development of sport. As noted above, the 'Development Pathway' and 'Fair Play' are seen to be important but are weaker on the reliability assessment. The measure with the largest negative disparity is the avoidance of 'white elephant' infrastructure [-0.63], which is very close in importance to 'Number of venues (Renovated)' but as a post-hoc outcome is harder to capture until after the effect has been observed.

6.7.4 Indicator ratings: Social

In Round 2, a total of 38 measures were presented to Social Workstream respondents for rating their importance in assessing event impact. In Round 3, a subset of 24 measures was represented for Reliability in assessing event impact.

Table 84 - Measure ratings: Social

Workstream: **Social**	Importance n=31		Reliability n=21		Impact Score
	Mean	Std. Dev.	Mean	Std. Dev.	[$\bar{x}$ I* $\bar{x}$ R]
Perceived benefit for children: Youth activity participation	4.23	0.884	3.62	1.071	**15.3**
Event awareness: Nation/Region/City	3.97	0.836	3.81	0.602	**15.1**

Workstream: **Social**	Importance n=31		Reliability n=21		Impact Score
	Mean	Std. Dev.	Mean	Std. Dev.	[$\bar{x}$ I* $\bar{x}$ R]
Community benefit: Investment into grassroots sport ($)	4.16	0.860	3.62	1.024	**15.1**
Public support for event: Nation/Region/City	4.16	0.898	3.52	0.814	**14.6**
Community benefit: Interest at grassroots level for sport	4.03	0.836	3.52	0.928	**14.2**
Volunteers: Total involved	3.77	0.845	3.76	0.995	**14.2**
Perceived benefit for children: Youth sport development	4.10	0.944	3.38	1.071	**13.9**
Participation rates: Impact through event attendance	3.81	1.078	3.57	0.746	**13.6**
Social Capital: Sense of Pride from hosting	3.90	0.790	3.48	1.078	**13.6**
Enthusiasm: Shared community spirit toward event	4.10	0.651	3.29	0.956	**13.5**
Total Attendance - Education/Cultural/Festivals (in host city)	3.74	0.930	3.52	0.928	**13.2**
Community benefit: Improved social cohesion	4.23	0.717	3.10	0.944	**13.1**
Activity participation rates: At community level	3.84	1.036	3.33	0.796	**12.8**
Increased activity participation: General Population (%)	3.84	0.860	3.29	0.902	**12.6**
Social Capital: Sense of Pride from infrastructure legacy	3.94	0.854	3.19	1.030	**12.6**
Community benefit: Improved cultural and education legacy	3.81	0.910	3.30	0.954	**12.6**
Sport Development: Community level facilities	3.84	0.688	3.24	0.995	**12.4**
Social Capital: Supporting a winning team (pride)	3.61	0.989	3.43	0.870	**12.4**
Perceived benefit for children: Socialisation and cohesion	3.71	0.902	3.29	1.189	**12.2**
Social Capital: Change in pre/post event	3.94	0.727	3.00	1.140	**11.8**
Quality of Life: Pre/Post resident attitude	3.87	0.763	3.00	1.049	**11.6**
Increased activity: Personal activity level	3.68	1.013	3.00	0.894	**11.0**
Social Inclusion: Diversity inclusion programme	3.77	0.884	2.90	0.889	**10.9**
Well-Being: General population health	3.68	0.979	2.43	1.028	**8.9**
Total Attendance - Education/Cultural/Festivals (in country)	3.61	1.022	-	-	**8.8**
Community benefit: Increased volunteer participation rate	3.52	0.769	-	-	**8.6**
Social Capital: Improved tourism networks in host communities	3.52	0.851	-	-	**8.6**
Total Number of Education/Cultural/Festivals (in host city)	3.52	1.061	-	-	**8.6**
Awareness of event - unique number of website visitors	3.48	0.851	-	-	**8.5**
Total Attendance - Education/Cultural/Festivals (international)	3.42	1.089	-	-	**8.3**
Job creation: Pre-Post increase in employment	3.29	1.216	-	-	**8.0**
Participation rates: via watching sport	3.26	0.999	-	-	**7.9**
Awareness of event - number of page impressions	3.26	0.815	-	-	**7.9**
Total Number of Education/Cultural/Festivals (in country)	3.19	1.078	-	-	**7.8**
Total Number of Education/Cultural/Festivals (international)	3.06	0.892	-	-	**7.4**
Number of Nations	2.97	1.048	-	-	**7.2**
Official charity partner activities (count)	2.81	0.833	-	-	**6.8**

Workstream: **Social**	Importance n=31		Reliability n=21		Impact Score
	Mean	Std. Dev.	Mean	Std. Dev.	[$\bar{x}$ I* $\bar{x}$ R]
Number of Female Volunteers	2.68	1.013	-	-	**6.5**

The results shown in Table 84 are the ratings of potential Social measures ordered by their Impact Score [$\bar{x}$ I * $\bar{x}$ R].

Similar to the Sport workstream, the Social workstream saw the most highly rated measure only achieve an overall Impact Score of less than 16.0, with 'Youth activity participation' achieving a combined score of [15.3].

However, where no measure achieved an Importance rating above 4.0 for the Sport workstream, both 'Youth activity participation' and 'Improved Social cohesion) were given Importance ratings of ($\bar{x}$ 4.23) under the Social workstream. The overall Impact Scores for Social are therefore more limited by their Reliability ratings than the Importance ratings. That is evident when assessing the differences between the two rating scales, shown in the table below.

Table 85 - Rating disparities: Social

Rated Measures - Social	Importance ($\bar{x}$)	Reliability ($\bar{x}$)	R $\bar{x}$ - I $\bar{x}$
Volunteers: Total involved	3.77	3.76	**-0.01**
Quality of Life: Pre/Post resident attitude	3.87	3.00	**-0.87**
Social Inclusion: Diversity inclusion programme	3.77	2.90	**-0.87**
Social Capital: Change in pre/post event	3.94	3.00	**-0.94**
Community benefit: Improved social cohesion	4.23	3.10	**-1.13**
Well-Being: General population health	3.68	2.43	**-1.25**

For the Social measures, the disparity in ratings of Importance and Reliability are not balanced across those that are positive and negative. As shown in Table 85,Table 83 there are no measures for which Reliability is greater than Importance (R $\bar{x}$ ≻ I $\bar{x}$). The smallest difference within the workstream's measures is 'Volunteers: Total involved' at [-0.01], which whilst negligible in itself, is noteworthy in that all other measures show an increasing disparity between the ratings.

Those measures showing the largest disparity include measures that were strongly rated for Importance but for which research methods and or consistency of measurement hampered the reliable collection of data. Those included 'Improved social cohesion' and measures of 'social capital' and 'social inclusion'. While 'general population health' is well tracked in many developed economies, it may be the explicit linking of those measures to a specific event that sees it achieve the widest difference in rating scores [-1.25].

In context, the maximum difference between Reliability and Importance ratings of [-1.25] as seen in the Social measures is wider than any in prior workstream results, where the largest disparities were Economic [-0.99], Sport [-0.63], and Media [-1.04] respectively.

6.7.5 Indicator ratings: Brand

In Round 2, a total of 35 measures were presented to Brand Workstream respondents for rating their importance in assessing event impact. In Round 3, a subset of 24 measures was represented for Reliability in assessing event impact.

Table 86 - Measure ratings: Brand

Workstream: **Brand**	Importance n=32		Reliability n=15		Impact Score
	Mean	Std. Dev.	Mean	Std. Dev.	[$\bar{x}$ I* $\bar{x}$ R]
Media impact: Volume of media coverage	4.28	0.888	3.60	0.828	**15.4**
Brand Awareness: Event recall (%)	4.09	0.689	3.73	0.799	**15.3**
Brand Awareness: Host recall (%)	3.97	0.740	3.80	0.775	**15.1**
National Pride: Host nation self-regard	3.81	0.896	3.67	0.816	**14.0**
Media content: Volume change in media coverage (%)	4.00	0.718	3.47	0.915	**13.9**
Brand Associations: Pre/Post host image attributes (%)	3.84	0.920	3.53	0.834	**13.6**
Brand Impact: Change in nation brand image	4.06	0.716	3.33	0.976	**13.5**
Place branding: Legacy perceptions of host place	4.03	0.782	3.27	0.873	**13.2**
Media content: Positive media coverage (% growth)	4.22	0.659	3.07	0.961	**13.0**
Brand image: Host city attributes (% growth)	3.97	0.822	3.20	1.082	**12.7**
Brand Impact: Change in host brand equity (% growth)	3.81	0.859	3.33	0.816	**12.7**
Brand Associations: Event brand attribute ratings (%)	3.66	0.653	3.40	0.910	**12.4**
Host-event fit: host/event brand congruence	3.88	0.793	3.13	0.834	**12.1**
Visitor experience: Intention to revisit (%)	3.78	1.008	3.20	0.862	**12.1**
Reputation: Change in Foreigner visitor views of host	4.03	0.782	3.00	0.845	**12.1**
Visitor experience: Intention to recommend (NPS)	3.94	0.948	3.00	0.756	**11.8**
Brand Associations: Host Brand attributes ratings (%)	3.66	0.701	3.20	0.941	**11.7**
Brand Impact: Change in intention to visit (% growth)	3.69	1.061	3.13	0.915	**11.6**
Brand value: Event brand financial valuation	3.84	0.847	3.00	1.000	**11.5**
Reputation: Change in Foreigner non-visitor views of host	3.88	0.793	2.93	0.704	**11.4**
Brand value: Host brand financial valuation	3.63	0.833	2.93	0.884	**10.6**
Investor brand: Business attractiveness of host (change %)	3.56	0.801	-	-	**10.4**
Number of Attendees at sponsor events (international)	3.53	0.879	-	-	**10.3**
Sponsorship: Value of brand/athlete exposure ($)	3.47	0.671	-	-	**10.2**
Number of Attendees at sponsor events (in host city)	3.34	0.937	-	-	**9.8**
Increase in foreign brand spending within host ($)	3.31	0.965	-	-	**9.7**
Sponsorship: Increase in sponsor product purchases ($)	3.28	0.772	-	-	**9.6**
Brand recognition: Sponsor brands recall	3.25	0.842	-	-	**9.5**
Number of Sponsor Events (international)	3.22	0.832	-	-	**9.4**
Brand boycotting: Loss of sales from product boycott	3.19	0.931	-	-	**9.4**

Workstream: **Brand**	Importance n=32		Reliability n=15		Impact Score
	Mean	Std. Dev.	Mean	Std. Dev.	[$\bar{x}$ I* $\bar{x}$ R]
Brand value: Athlete brand financial valuation	3.19	0.931	-	-	**9.4**
Sponsorship: Event influence on intention to purchase	3.19	0.896	-	-	**9.4**
Number of Attendees at sponsor events (in country)	3.09	0.995	-	-	**9.0**
Number of Sponsor Events (in country)	3.09	0.963	-	-	**9.1**
Number of Sponsor Events (in host city)	3.09	0.893	-	-	**9.1**

The results shown in Table 86 are the rating of Brand measures ordered by their Impact Score [$\bar{x}$ I * $\bar{x}$ R].

As in the Sport and Social workstreams, no Brand measure achieves an overall Impact Score of more than 16.0, with 'Media impact: Volume of media coverage' achieving the highest score of [15.4]. Simple recall metrics drawing on 'brand awareness' with the ability to recall the 'Event' [15.3] and recall the 'Host' [15.1] respectively were the next highest Impact Scores.

While the volume of media coverage is vital to events achieving their desired brand impact, there is a noteworthy difference in the 'Change in media' measures. The 'Volume change in media coverage (%)' is rated ($\bar{x}$ 4.00) for Importance achieved an overall Impact Score of [13.9]. By contrast, the more specific element within that volume of 'Positive media coverage (% growth)' was given the second highest rating for Importance ($\bar{x}$ 4.22), but it was also perceived to be less reliable as a measure and hence only achieves an Impact Score of [13.0].

In addition to the media and recall measures, the development of host self-image and changes in brand associations with the host were amongst the other highest rated measures. The Impact Scores around the experience of visitors and their likelihood to revisit or recommend at [12.1] and [11.9] placed them outside of the top 15 rated measures, despite the rhetoric on 'tourism legacies' being claimed amongst lasting effects.

Table 87 - Rating disparities: Brand

Rated Measures - Brand	Importance ($\bar{x}$)	Reliability ($\bar{x}$)	R $\bar{x}$ - I $\bar{x}$
Brand Awareness: Host recall (%)	3.97	3.80	**-0.17**
Brand Awareness: Event recall (%)	4.09	3.73	**-0.36**
Media impact: Volume of media coverage	4.28	3.60	**-0.68**
Visitor experience: Intention to recommend (NPS)	3.94	3.00	**-0.94**
Reputation: Change in Foreigner visitor views of host	4.03	3.00	**-1.03**
Media content: Positive media coverage (% growth)	4.22	3.07	**-1.15**

As seen in the Social workstream, the differences in ratings of Importance and Reliability within the Brand workstream are likewise not evenly balanced between positive [Reliability > Importance] and negative [Reliability < Importance].

As shown in Table 87, there are no Brand measures for which Reliability is greater than Importance (R $\bar{x} \not> I \bar{x}$). Where within the Social workstream the smallest difference was for 'Volunteers: Total involved' at [-0.01], for the Brand workstream the lowest difference was [-0.17] for 'Brand Awareness: Host recall (%)'. The measure previously identified for its strong Importance rating 'Positive media coverage (% growth)' ($\bar{x}$ 4.22) is notable also for the largest difference in the measures at [-1.15].

Also noteworthy is that 'Media impact: Volume of media coverage', which despite registering a Reliability rating [-0.68], was lower than its Importance rating but still achieved the highest Impact Score overall. Likewise, the potential of events to impact visitor willingness to recommend or change their views on the host is confirmed in their Importance ratings but was challenged by the lower perceived Reliability of the same measures, showing [-0.94] and [-1.03] respectively.

6.7.6 Indicator ratings: Environment

In Round 2, a total of 33 measures were presented to Environment Workstream respondents for rating their importance in assessing event impact. In Round 3, a subset of 24 measures was represented for Reliability in assessing event impact.

The assessment of the results of the Environment workstream and its comparison with other workstream results remains limited by the low number of respondents identifying as having expertise in the workstream area.

Table 88 - Measure ratings: Environment

Workstream: **Environment**	Importance n=5		Reliability n=2		Impact Score
	Mean	Std. Dev.	Mean	Std. Dev.	[$\bar{x}$ I* $\bar{x}$ R]
Habitat loss (Forest, Wetlands cleared for construction)	4.80	0.447	4.50	0.707	**21.6**
Waste - Recycled (%)	5.00	0.000	4.00	0.000	**20.0**
Renewable Energy used (%)	4.60	0.548	4.00	0.000	**18.4**
Ecological construction (Remake of old industrial sites)	4.40	0.548	4.00	1.414	**17.6**
Air pollution Total CO^2 emissions (t)	4.80	0.447	3.50	2.121	**16.8**
Waste - Visitor garbage (t)	4.80	0.447	3.50	2.121	**16.8**
LEED score - Energy and atmosphere	4.20	0.837	4.00	1.414	**16.8**
Visitor travel (Mode + distance to event)	4.60	0.548	3.50	0.707	**16.1**
Water environment (Reduction of habitats)	4.60	0.894	3.50	0.707	**16.1**
Construction impact (% environmentally friendly materials)	4.40	0.894	3.50	0.707	**15.4**
Public transportation - visitor passenger journeys (count)	4.40	0.894	3.50	0.707	**15.4**
Water - Recycled/Recovered %	4.40	0.548	3.50	0.707	**15.4**
Visitor travel (Mode + distance at event)	4.80	0.447	3.00	0.000	**14.4**
Waste - Renewable Energy systems	4.80	0.447	3.00	0.000	**14.4**
Public transportation - visitor passenger miles (total)	4.00	1.000	3.50	0.707	**14.0**
Biodiversity protection (% lost)	4.40	0.894	3.00	0.000	**13.2**
Infrastructure (New structure, city beautification, higher standards)	4.40	0.894	3.00	1.414	**13.2**

Workstream: **Environment**	Importance n=5		Reliability n=2		Impact Score
	Mean	Std. Dev.	Mean	Std. Dev.	[$\bar{x}$ I* $\bar{x}$ R]
Urban renewal (Event areas %)	4.00	1.000	3.00	0.000	**12.0**
Animals (Disruption of migration routes)	4.20	0.837	2.50	0.707	**10.5**
Recycling (Environmental awareness)	4.20	1.304	2.50	0.707	**10.5**
Recycling (New recycling infrastructure)	4.20	0.837	2.50	0.707	**10.5**
Air Pollution (% reduction from better traffic flow)	4.80	0.447	2.00	0.000	**9.6**
Air pollution Reduction - via energy efficient systems	4.80	0.447	2.00	0.000	**9.6**
Visitor travel (congestion impact)	4.40	0.548	2.00	0.000	**8.8**
LEED score - Water efficiency	4.00	0.707	-	-	**8.0**
LEED score - Green infrastructure & buildings	3.80	0.447	-	-	**7.6**
LEED score - Location and transportation	3.80	0.447	-	-	**7.6**
LEED score - Sustainable sites	3.80	0.837	-	-	**7.6**
Infrastructure (Upgrade to ecological materials)	3.60	0.894	-	-	**7.2**
Knowledge (Implementation of environmental practices)	3.60	1.140	-	-	**7.2**
LEED Certification Status	3.60	0.548	-	-	**7.2**
LEED score - Materials and Resources	3.60	0.548	-	-	**7.2**
LEED score - Indoor environmental quality	3.40	0.548	-	-	**6.8**

The results shown in Table 88 are the rating of Environment measures ordered by their Impact Score [$\bar{x}$ I * $\bar{x}$ R].

As noted above, there are significant limitations to drawing conclusions based on the Impact Scores, especially the very high values at the top end of the range. The Environment workstream results show nine measures with an Impact Scores greater than 16.0 when multiple workstreams did not have any that high is a function of the low number of respondents.

Whilst remaining cognisant of that limitation, the highest impact scores in the Environment workstream included both active measures for positive impact and measures of avoidance of negative effects from events. Like other workstreams, the measures tend to focus on those that are simple and direct measures especially around the operating footprint of the event such as waste produced and % recycled, % of renewal energy used, or CO^2 production. Also noteworthy are the measures that seek to capture avoidance of negative impacts including natural habitat loss, and more positively, the reuse of existing land to prevent new losses.

Despite events and sport infrastructure being seen as an opportunity to apply and promote sustainable building practices many of the formalised standards were not seen to be rated highly for importance. The exception was for 'LEED score - Energy and atmosphere' which aims for buildings with reduced energy demands through conservation and efficiency gains – which are more again directly quantifiable measures.

Table 89 - Rating disparities: Environment

Rated Measures - Environment	Importance ($\bar{x}$)	Reliability ($\bar{x}$)	R $\bar{x}$ - I $\bar{x}$
Habitat loss (Forest, Wetlands cleared for construction)	4.80	4.50	**-0.30**
Ecological construction (Remake of old industrial sites)	4.40	4.00	**-0.40**
Public transportation - visitor passenger miles (total)	4.00	3.50	**-0.50**
Public transportation - visitor passenger journeys (count)	4.40	3.50	**-0.90**
Air pollution Reduction - via energy efficient systems	4.80	2.00	**-2.80**

As with the Social and Brand workstreams, even within the limitations acknowledged, none of the Environment measures that are shown in Table 88 has ratings of Reliability greater than their importance. For completeness with other workstreams, a summary of some of those differences are included and shown in Table 89.

Of interest is how a measure like 'habitat loss' might be rated as more Reliable than the number of journeys taken by visitors on public transport. Perhaps it is more reflective of the scale of work required to implement visitor usage collection via a 'visitor pass' rather than the reliability of the analysis itself.

The large disparity between Importance and Reliability for 'Air Pollution Reduction' is evidence of both recognition of the vital need to ensure clean air and the distributed nature of the problem across sites and systems, something LEED and other building standards are seeking to attract more attention to.

6.7.7 Cross-workstream measure ratings

As noted in the Research Method, each group of measures were rated within individual workstreams. An opportunity for respondents to consider the importance of highly rated measures from across all six workstreams was provided at the end of Round 3.

In keeping with the BWS method design, the measures were presented in blocks of four. Each measure was shown to a respondent an equal number of times and in comparison to all the other measures when rating the most important and least important within each block presented.

Table 90 - Cross workstream BWS measure ratings

Pillar	Measure	Most Important (count)	Least Important (count)	**BWS Net Score**
Economic	Visitor expenditure: Total all visitors ($)	78	-24	**54**
Media	Audience size (total viewership cumulative)	74	-23	**51**
Social	Community benefit: Investment into grassroots sport ($)	68	-21	**47**
Economic	Visitor numbers: International (count)	68	-24	**44**
Sport	Development of sport infrastructure	66	-25	**41**
Media	Audience size (individual viewers reached)	60	-20	**40**
Brand	Media content: Positive media coverage (% growth)	50	-27	**23**
Social	Community benefit: Improved social cohesion	52	-41	**11**

Brand	Brand Awareness: Event recall (%)	39	-46	**-7**
Sport	Total Number of Competitors - Elite	29	-52	**-23**
Sport	Event ranking points /classification	13	-76	**-63**
Environment	Waste - Recycled (%)	9	-115	**-106**
Environment	Air pollution Total CO^2 emissions (t)	5	-117	**-112**

In coding the results of the BWS test, each count of a measure being rated as 'most important' was given a value of +1.0. Conversely, when rated as 'least important' rating given a value of -1.0. The count results for each measure are shown in Table 90 and are ordered in descending 'BWS Net Score'.

The results show several different aspects from the cross-checking of measures underpinning the Pillars. The diversity of the respondents' views was evidenced in that all measures received ratings of both 'most important' and 'least important', with the difference in Net Scores generated in the frequency at which they were ascribed to those ratings. Also noteworthy was that the measures from three different pillars achieved the top three BWS Net Score ratings. Indeed, four different pillars were represented in the top five Net Scores.

There is a consistent theme between the measures rated in BWS and the pillar weightings allocated in Round 1 and Round 2. In both cases, Economic measures are given higher ratings/weights and Environment is rated at the bottom of the values in both approaches.

6.8 Framework internal reliability

As within each workstream, weights were assigned to Drivers in a Direct method. Measures were also rated to provide an Indirect method for understanding the importance of different areas of that pillar on assessing event impact.

Tests of Framework Reliability - Direct vs. Indirect measures

Table 91 - Tests of Framework Reliability

			Driver Weightings vs. Measure ratings	
Test	**Typology**	**Indirect Construct**	**Correlation (r)**	**Correlation (r^2)**
Test #1	I	Importance ($\bar{x}$) only	0.4088	0.1671
Test #2	IR^{24}	Importance ($\bar{x}$) * Reliability ($\bar{x}$) (top 24)	0.5396	0.2912
Test #3	IR^{All}	Importance ($\bar{x}$) * Reliability ($\bar{x}$) (All)	0.7044	0.4962

Tests for the degree to which similar weights for the importance of Drivers can be achieved via direct allocation of weightings on Drivers against imputed scores derived from the ratings of measures for importance and reliability were developed. The results in Table 91 show the correlation between the direct allocation of weights for the Drivers (mean weights) against different constructs of Measures ratings.

The results show that correlation scores vary over a wide range from a low in Test #1 (I) of $r^2 = 0.1671$ to a strong correlation of $r^2 = 0.7764$ from Test#4 (pIR^{All}).

As the 'top down' Driver weightings are constant for tests #1 - #3, the differences in correlation outcomes is level of detail on the 'bottom up' measures. As noted Test #1 (I) used only the Importance ($\bar{x}$) for the measures and showed the lowest correlation with the 'top down' weightings. Test #2 (IR^{24}) used both Importance ($\bar{x}$) and Reliability ($\bar{x}$) for the top 24 items in each pillar and improved $r^2 = 0.2912$. Using the same dimensions as Test#2 but extending to include all measures in the workstream via the proxy scores for Reliability for unrated items Test #3 (IR^{All}) achieved an $r^2 = 0.4962$.

6.9 Stakeholder priority outcomes

To understand the degree to which the outcomes of events might reflect shared goals between the dominant stakeholders the perceived priorities of Event Hosts and Event Owners was tested within the expert panel.

Table 92 - Event Host vs Event Owner priorities

	Event Host Priorities	Event Owner Priorities	Host-Owner		
N=182	Ranking Mean score^	Ranking Mean Score^	Difference of Means	t value	Significance
Economic	2.09	2.46	-0.368	-2.864	.005*
Sport	3.04	2.44	+0.604	4.331	.000*
Brand	3.16	2.74	+0.418	3.249	.001*
Media	3.59	2.87	+0.720	7.016	.000*
Social	3.79	4.85	-1.055	-9.394	.000*
Environment	5.32	5.64	-0.319	-3.922	.000*

^Mean score for ranked choices. Maximum (highest priority) score = 1, Minimum score (lowest priority) = 6

The results of the ranked priorities are shown in Table 92 noting that a lower score indicates a higher priority rank was given. The findings reveal significant differences in the mean scores of all pillars, indicating the expert panel perceive a strong misalignment in priorities between the two primary stakeholders in the event.

The priorities for the Event Owner are closely clustered across Sport (2.44) and Economic (2.46) to Brand (2.74) and Media (2.87). By contrast the priority scores given to Social (4.85) and Environment (5.64) are a step change different. For Event Hosts the priority is given to Economic (2.09) but then Sport, Brand, Media and Social are more evenly ranked. Environment is the lowest priority for both groups.

6.10 Pillars weights by Event Typology

As noted an optional extension question was available at the end of Round 3 which allowed respondents to provide their view on how the Pillars might be weighted differently across events of different scale. The SIRC event typology was provided to categorise the different scale of events. A total of 16 respondents undertook the optional question.

Table 93 - Mean Pillar weights by event scale

N=16	Economic	Sport	Media	Social	Brand	Environment
Type A	24.37	19.50	15.81	16.69	12.56	11.06
Type B	23.63	19.00	16.56	15.00	14.69	11.13
Type C	19.88	23.25	16.25	15.31	13.75	11.56
Type D	13.69	26.88	14.69	17.81	13.25	13.69
Type E	10.63	31.25	9.69	23.75	11.19	13.50

The results in Table 93 show that differing weights being applied by the respondents. In line with the typology the weighting of Economic impact declines from Type A (24.37) to Type E (10.63). The table uses Type A events are the baseline score and identifies differences between the mean scores given for each Pillar under Types B – E.

Table 94 - Differences in Pillar mean weights from Type A events

N=16	Economic		Sport		Media		Social		Brand		Environment	
Type A $\bar{x}$	24.37		19.50		15.81		16.69		12.56		11.06	
	$\bar{x}$ Diff.	Sig	$\bar{x}$ Diff.	Sig	$\bar{x}$ Diff.	Sig	$\bar{x}$ Diff.	Sig	$\bar{x}$ Diff.	Sig	$\bar{x}$ Diff.	Sig
Type B	-0.75	0.418	-0.50	0.451	+0.75	0.418	-1.69	0.202	+2.13	0.024*	+0.06	0.898
Type C	-4.50	0.004*	+3.75	0.018*	+0.44	0.752	-1.38	0.343	+1.19	0.222	+0.50	0.590
Type D	-10.69	0.000*	+7.38	0.002*	-1.13	0.589	+1.13	0.499	+0.69	0.509	+2.63	0.028*
Type E	-13.75	0.000*	+11.75	0.006*	-6.13	0.003*	+7.06	0.016*	-1.38	0.214	+2.44	0.097

Despite the small respondent base, there are several significant differences in the pillar weights ascribed revealed under the paired t-test shown in Table 94.

Using the SIRC Type A event as the baseline, there is no difference in the weightings of the Economic pillar between that and the Type B event, but by contrast all event Types C- E were progressively and significantly lower in their weightings. In the Sport sillar the same pattern is evident with Types C-E all statistically different from the Type A event but in this pillar, they attract a progressively higher weighting.

Additional differences for the Type E as the smallest scale events are that it is lower on Media weight but given higher weights on Social pillar compared to Type A events, while Types B-D were not found to show significant differences on the same pillars.

7 Discussion of findings

In this chapter, the quality of the research base is briefly reviewed and then the findings from the Delphi study are drawn together and discussed in the context of the key research questions. The key characteristics of the framework and its applications are reviewed, and limitations, potential advances and future developments arising from this research are also discussed.

7.1 Study base and panel profile

Research dataset

In seeking to use the data set resulting from the Delphi study, it was important to understand its sufficiency and the integrity on which to base the research analysis. In a review of panel sizes used in 32 prior Delphi studies, Giannarou and Zervas (2014) found "18 studies used the opinion of less than 20 experts […] with the number ranging between 30 and 50 participants in 5 studies and between 50 and 100 in 4 more. Also, there are 5 studies which used an even greater number of participants, i.e. >100." (p. 67). In that context, the respondent panel size of n=182 from Round 1 of the study established it at the higher end of Delphi studies in scale and a robust sample base. The completion response rate (the number of surveys completes divided by the total number of invitees) of 20% is consistent with expectations for an online survey emailed to an identified panel of participants (Sammut et al., 2021). As respondents who started but only partially completed the survey were excluded by the research fieldwork supplier, only data from fully completed responses from each round were included in the final files provided. This ensured there was no missing data within the files and all responses were able to be used in the analysis phase.

Composition of the expert panel

As noted in the Research Methods, the anonymising of data files provided prevented the linking of responses from round to round. Hence, it was not possible to profile respondents by role, age, tenure, etc. in Rounds 2 and Round 3 or evaluate the impact of these demographics on the responses to Driver and Indicator ratings. Whilst this is a limitation of the data set, considering the smaller sample sizes of Round 2 and Round 3, there was less likelihood of statistically robust insights that might have been able to be drawn from such analysis, even had the data been available.

Within the Round 1 data set, there were more respondents from Europe than other regions, but given that is where the majority of international sporting associations are based, that result was not unexpected. A moderating factor on the balance of regional representation was that all international regions were represented to some extent, and that over 50% of respondents held roles with international responsibilities, lessening the importance of the location of the role as a factor.

Three areas that were shown in the results to elicit significant differences in perspectives on event outcomes were tenure, role, and expertise. Concerning tenure, the depth of experience respondents brought was solid, with a third of respondents nominating a tenure in the industry of 15 years or longer. In terms of industry roles, academia comprised around one-third of the respondent base, with representatives from International Sports Federations and Government roles the next largest cohorts. The final area of expertise provided solid representation for five of the six Pillars, with lower representation under Environment and Sport and Economic the areas of expertise most often nominated. The effects of these

dimensions on the weightings applied to the Pillars are considered and potential implications are discussed further below.

In summary, given the respondent panel was not controlled for composition, a solid basis for a Delphi study with good representation across role, region, responsibility, tenure, and age of respondents was achieved. This provided a cross-section of the target audience and a broad representation of perspectives across the global sport sector from which the insights have been derived.

Classification of Delphi outcomes

The underlying premise of a Delphi technique is to elicit a consensus viewpoint from a group of opinions. In terms of classifying the outcome of this study in terms of Delphi typology, the typology framework of Greatorex and Dexter (2000) has been applied.

Table 95 – Exemplar Delphi study typology

Type	Description
A	Stable opinion and stable high disagreement
B	Stable opinion and stable moderate agreement
C	**Stable opinion with increasing to strong agreement**
D	**Stable opinion with increasing to moderate agreement**
E	Stable opinion with `high moderate high' agreement
F	Stable opinion with `moderate high moderate' agreement
G	Opinion change between 1st and 2nd appearance followed by stability
H	**Increasing (or decreasing) opinion change**
I	Opinion change between 1st and 2nd appearance and between the 1st and 2nd appearance for the 3rd appearance

From (Greatorex & Dexter, 2000, p. 1019).

For consistency and clarity, the typology for the description of Delphi outcomes shown in Table 95 is used as a reference in this discussion. In their analysis, Greatorex and Dexter (2000) identify Type D as the most common outcome, then Type C and Type H as the next most frequent, but at nearly one third the rate of Type D.

Assigning a type of consensus to the areas analysed in the research should not be interpreted as an outcome being a ‘right answer’, or ‘right ‘type’ of consensus. The principle applicable on testing of all dimensions is that where the degree of existing consensus is examined, it is to determine the degree of agreement/disagreement on that dimension, not to force conformity to a particular viewpoint.

7.2 Weightings of event impact

Event impact weightings

The results of this study into whether an expert panel can agree on the relative importance of the components of event impact were strongly supportive that this can be achieved.

The results showed that of the six core pillars that comprise the top level of the Event Impact Framework (EIF), the weightings ascribed by panel respondents fall into ‘three tiers’. The pillars that were ascribed the highest mean score weighting were Economic and

Sport, ahead of the next tier of three pillars that comprised Media, Social and Brand, with the final pillar Environment attracting the lowest Pillar weighting. This tiered ordering effect was reconfirmed in the second round, in which the underlying mean scores for five of the six pillars did not vary significantly. The exception to that was the Social pillar, which saw an increase in its mean score in the second round, but not enough to challenge the overall pillar tiers.

The use of mean scores as the method of determining the consistency of opinion through Delphi group judgements draws on the work of Huber and Delbecq (1972). Their work compared aggregation rules and found that the use of arithmetic means of group members' continuous scale ratings increased accuracy and reduced error over other methods (p. 173). If the absence of significant changes in mean scores between rounds is an indicator of the consistency in group opinion (Greatorex & Dexter, 2000), then the findings from this study challenge the notion that contestedness around the concept of legacy must remain a persistent condition, despite the lack of progress towards an agreed definition to date. When presented with clear components of shorter-term event impact (rather than extended and 'catch-all' generalisations), a stable overall opinion on the relative importance of each of the event impact components was able to be formed in a single round, which was confirmed by retesting in a second round. Hence this research makes a solid contribution to the field in showing that there exists a stable consensus agreement on the relative value of the components of legacy, despite a lack of agreement on its overall definition.

The second important test within the results was to understand the extent of agreement or disagreement within the panel around the mean scores. Referencing the standard deviation as the indicator of the extent of agreement around those weightings shows a small narrowing of agreement in the weightings selected between rounds for most pillars, but not enough to be significant. The exception to that was the Media pillar, which showed a strong narrowing of consensus around its mean weighting between the first and second rounds. Taken together, the consistency in the level of agreement and the stable opinions expressed suggests that an underlying foundation of consensus operates but in the past this has not been measured using a consistent approach across disciplines.

Importantly research has sought to ensure all dimensions identified in the literature and included in the model are represented in the analysis and discussion of results. The goal is not to apply a 'reductionist' approach by removing 'low weighted' dimensions from the final framework but rather explore the breadth of views on all dimensions explicitly. Indeed by ensuring the inclusion all dimensions with explicit weightings in the final framework allows any prospective host or stakeholder group to review and ascribe different weighting or valuation on the outcomes of their own events.

In terms of classifying the Pillar weighting results across a two-round consensus process, the elements displayed Type B features, that is, "stable opinion with stable moderate agreement" under the Greatorex and Dexter (2000, p. 1019) typology.

Effect of outliers on consensus agreement

As noted, the Pillar weightings applied showed a stable opinion and a tendency towards a narrowing of consensus between Pillar rounds. However, within those results, a few respondents provided extreme values for the weights of individual Pillars at both the high and low ends of the scale. Values more than three standard deviations from the mean were classified as outliers and excluded, then the rest of the panel responses were retested. Without those extreme values, it was observed that there was a statistically significant narrowing of the standard deviations for three pillars between rounds. The relevant pillars

exhibiting this trend were those given the highest weightings, namely Economic, Sport, and Media, with Social already exhibiting stronger agreement between rounds without outlier exclusion.

As there was no significant change in the mean scores of the relevant three pillars, the result of removing outliers was not to shift the consensus weights applied to the Pillars but rather to reveal a strengthening agreement on the weights ascribed. Considering this improvement in the degree of agreement after removal of outliers, the overall Delphi outcome might be classified as a Type C study, which reflects a "stable opinion with increasing to strong agreement" (Greatorex & Dexter, 2000, p. 1019).

In finding improved agreement, it is also important to recognise that the exclusion of outliers must per se reduce the standard deviation of any round. However, it does not automatically result in any significant impact on the degree of agreement shown between two independent rounds that both have outliers excluded. In addition, the filtering of outliers was shown not to change the consensus opinions but only the level of agreement around them. The design of the Delphi study allowed for the outliers to be gathered both through anonymity and not forcing conformity in weightings. Likewise, the strong sample size allowed for outliers to be excluded without compromising the depth of the remaining sample base, enabling the retesting of both opinion and agreement in each round.

The existence of the outliers is not unwelcome. Their identification and the revealing of their impact in diluting the level of agreement inherent in the Pillar weightings can be seen as encouraging. The development of the EIF and its underlying tests shows that perhaps the perceived level of constestedness on the concept of legacy is in part exacerbated by outliers. That is, their effect was to dilute levels of agreement and mask an underlying shared core of agreement on the relative value and contribution of areas of event impact.

EIF brings explicit weightings to the Triple-bottom line of event assessment

The development of the EIF did not seek to create new dimensions on which to assess events, but rather to establish a framework that made explicit the weightings of the importance of those dimensions. Not having started with a preconceived notion as to the framework design or component dimensions, reflecting on the degree to which the resultant EIF aligns with existing event assessment approaches makes explicit both its contribution to the field and its support in deepening current practice.

The EIF aims to inform hosts as to how to assess and compare potential and current events across multiple dimensions consistently. The judgement of the value placed on those outcomes can be ascribed by each host dependent on their capacities and development strategies. Hence, rather than contrast the EIF with theoretical frameworks of event legacy, the reference point should be using existing event impact studies being conducted by industry practitioners and issued to host cities. The Triple Bottom Line (TBL), namely Economic, Social, and Environmental outcomes, is a common classification used in impact and bid planning including reports prepared for the most recent Olympic Games city Brisbane, who sought "to determine the potential economic, social and environmental benefits of hosting a 2032 Games in Queensland." (KPMG, 2021, p. iv). The same approach has been used when assessing ex-post event impact expressed in monetary value (Grant Thornton, 2010; KPMG, 2006, 2021; PWC, 2005).

Rather than proposing to replace this established framework, the EIF aligns with that existing structure but seeks to be more specific as to the components of each of the 'triple bottom line' areas and how each contributes to event outcomes.

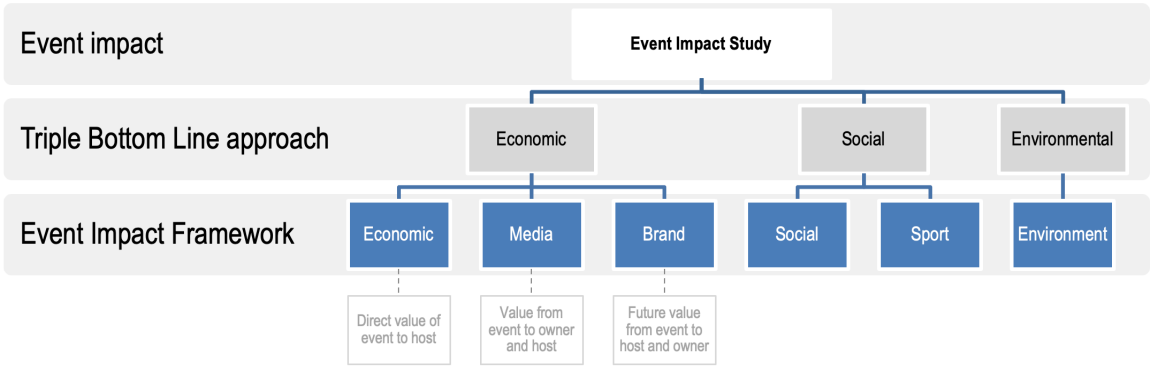

Figure 20 - EIF and Triple Bottom Line

The representation of the alignment between the EIF and TBL as outlined in Figure 20 shows the EIF Pillars to be sub-elements of three core areas of the TBL's Economic, Social, and Environmental dimensions. This approach of aligning existing event outcomes to Triple Bottom Line dimensions was applied by Bob and Swart (2010) in surfacing the social legacies of major events in the context of the dominant economic narrative. In noting that the achievement of event outcomes "depend on whether planning to leverage benefits occur, the allocation of resources to generate benefits and political will"(p. 74), they pre-empt the issues of planning, alignment and leverage discussed further below.

The clearest alignment between the approaches was on the Environment dimension, where there was a 1:1 fit between the two models. The EIF adds two important dimensions to the representation of outcomes under the TBL. The first was that the results of the Delphi study delivered specific weightings to each of the six Pillars. For example, while Social was shown as equivalent to Economic and Environmental outcomes in representation in the TBL, the EIF does not assume prominence is synonymous with importance. The second was adding the more detailed view under the TBL areas that can be grouped in summary form in the familiar TBL representation. The grouping of Sport and Social Pillars into the TBL's Social factor is not a contrivance but rather shares a perspective with the Norwegian Sports Policy in recognising that "Sport and physical activity have positive effects on both individuals and society in general. There is no clear distinction between the intrinsic value of physical activity and the social significance of sport" (Norway Government, 2019). The higher weighted Sport Pillar combined with the Social Pillar brings the overall weighting of the TBL's Social dimension closer into line with the TBL's Economic dimension as compared to it being categorised solely as Social. Likewise, the clustering of the EIF's Media and Brand along with Economic highlights the different sources of economic value from events. If the balance in event value moves more towards intangible assets, the detail offered by the EIF to represent that change would be hidden at the aggregated TBL Economic level alone. By explicitly uncovering the weights being applied to different areas of event outcomes, a host can also make a better assessment of its fit with their development strategy. This concept of recognising differential outcomes from events reflects earlier work in developing event profiles to "aid in the comparison between diverse legacies that may occur in different fields, such as sport, infrastructure". (Dickson et al., 2011, p. 285).

As well as the observed alignment between the EIF and the TBL approach and the additional information provided by the EIF model, the results of the EIF Delphi study

provided further important contributions to the notion of the TBL assessment of event outcomes. Firstly, the two highest weighted pillars of Economic and Sport from the EIF research align to different elements of the TBL dimensions Economic and Social respectively. This supports the TBL approach as an overarching model, rather than reinforcing an overreliance on only one of the 'bottom lines'. The second insight is the paucity of relative weight ascribed to the third element of the 'triple', that is Environment. If the stable results are an indicator of group consensus, the current view of event outcomes is closer to a 'double bottom-line' than a triple. Environmental impacts have not yet established equivalent weight in event assessment but are still hampered by "environmental plans being relegated to minimizing the impact of an event" (Jones, 2008, p. 347).

It is acknowledged that in the last two years the weighting of the third Environment dimension may have changed, as awareness and measurement of its components improve. This evidences another advantage of the EIF for testing impact; being able to re-run the study periodically to confirm any measurable changes in sentiment and attention in event outcomes over time.

7.3 Sources of bias within pillar weightings

Given the difficulty of developing an agreed definition on legacy as discussed earlier, the ability to quantify the diluting effect of outliers was best seen as a start towards reducing contestedness of the concept. However, it is also important to identify and consider what other factors influence persistent differences in views. The seeking of sources of bias within the average viewpoint echoes the issue posed by Murphy et al. (1998) in asking "to what extent is the average a good representation of the group judgement when identifiable subgroups within the group differ in their judgements?" (p. 37). Hence the key profiling demographics were analysed to see if they represented different perspectives and to determine if there was evidence of significant bias within the respondents' weightings.

Expertise bias

It is recognised that those with expertise may have an informational advantage in particular fields in assessing potential outcomes, but they may also display unconscious personal preferences that impact their objectivity, leading to a position where they are "both better informed and more biased" (Li, 2017, p. 60). Given the respondents were asked to self-identify their areas of expertise in relation to sport event outcomes, the research provided an opportunity to understand the trade-off between being informed and the potential for bias in ascribing areas of importance to assessing event outcomes.

The results showed that those who identified themselves with some expertise in an area consistently ascribed higher weightings to Pillars than those that did not identify as having similar expertise. This bias was found to be statistically significant for five of the six Pillars (Economic, Sport, Media, Social, and Brand). The Environment pillar did not have a sufficient expertise base for the result to be statistically significant. The consistency of the bias observed limited the likelihood that the pattern is a random artefact, but rather that respondents were unaware of how those outside their expert domain perceived the same variable. That lack of external calibration of their field reflects the observation that "even the most knowledgeable and perceptive expert may nonetheless show the patterns of miscalibration" (Koehler et al., 2002, p. 7).

The implications for hosts seeking to assess the potential impact of target events for their city is that they need to remain cognisant of the expertise composition of those making the assessment. Hence ensuring expertise diversity in assessing event impact is a key aspect of

avoiding a bias dominating deliberation on event assessment or bid determinations. That diversity is a vital consideration for two key aspects of event assessment. The first is to ensure that value is neither overstated in some areas and undervalued or overlooked in others. The second is to ensure consistency between event assessments. Without such consistency, different expertise profiles may introduce variations that are not due to differences in the events under study, but rather due to differences in the expert groups undertaking the studies.

Hence while "no expert, regardless of their domain experience, is immune to bias"(Wilson et al., 2020, p. 1217), this research sought to contribute to reducing contestedness through the development of explicit measurement and exposure of biases within the expert groups' views rather than trying to "limit bias by seeking impartial evaluators [which] may reduce the quality" of assessments (Li, 2017, p. 60).

Despite the bias detected in the respondent base, it is noted that the net effect of combining experts and non-experts is that the ratings of Tier 1 (Economic, Sport), then Tier 2 (Media, Social, Brand), and then Tier 3 (Environment) remained stable, illustrating the stability of a combined viewpoint.

Academic vs. Industry view

As the largest single cohort and a group that provides a critical external perspective on the field, comparing the responses of Academics against the [Not] Academic group showed both share a common ordering effect by the same 'tiers' of pillar weightings. That is, both agreed that the Pillars should be weighted in the order Tier 1 (Economic and Sport), Tier 2 (Media, Social and Brand), and Environmental as a third tier. There were, however, statistically significant differences evident within the numeric values of the weightings nominated, with the Academic cohort placing significantly higher weightings on the Social and Environment dimensions. The high weighting placed on those specific areas is consistent with Mair and Whitford (2013a) who, when testing the ranking of future areas of event research, found academics placed Socio-cultural and community impacts and Environmental impacts as two of their top three priorities. By contrast, the [Not] Academic saw Media as being most important. While receiving some Academic attention already, addressing the value placed on Media by the industry as an area of future research would make academic research more aligned and relevant to the sector. Likewise, if Academics have a case as to why Social and Environmental aspects need to be given greater attention in event outcomes by the industry, then more work remains to convince and embed this in the industry's priorities.

Hence despite a broad agreement on the overall weights, revealing the bias within the cohorts manifests Gallie's view that a concept can both have a shared usage but different views on how that concept is valued and used (Gallie, 1956). Reducing contestedness over the concept of event outcomes means achieving less disparity in the weightings of Pillar components. This has been achieved at least within the Economic pillar, where the two cohorts ascribed almost identical weights to that Pillar. Interestingly it was also Economic Impacts that Mair and Whitford (2013a) identified as the other item that Academics ranked as an area of future research. So at least on this element of event outcomes, industry and academic perception of attributed value and future interest are well aligned, but work remains for the other Pillars.

Industry tenure effect

In a globally competitive market for sport events, the ability to adapt and shift with changing demands and drivers of consumption is vital. The time and experience needed to become a leader may be seen as an advantage in navigating a changing marketplace. In less globalised times, the effect of extended tenure on senior decision making was found to run counter to the agility needed to compete today as "long-tenured executives will tend to have persistent, unchanging strategies and strategies that conform closely to industry averages" (Finkelstein & Hambrick, 1990, p. 486). But even within more agile and innovative cultures, the effect of extended tenure on CEOs is to shift towards an "increased [focus on] innovation from exploiting existing technologies, rather than exploring emergent technologies" (Li & Yang, 2019, p. 546).

In the context of the EIF study, investigating the impact of tenure showed that those with the deepest industry experience placed much greater weight on the role of Sport in event impact than those newer to the sector. Instead, the newer group ascribed greater weight to the areas of Media and Social impact than their longer-tenured counterparts. While it appears axiomatic that 'sport' remains core to sport events, the new entrants to the sector may be less wedded to the sport aspect and instead view it as content for media where already "sport is unusually dependent on commercial media-financed, impossible-to-repeat live events"(Rowe, 2020, p. 704) and social connection. The younger tenure bias appears closer to the Marshall McLuhan dictum of 'the medium is the message' (McLuhan, 1964), with events tilting more towards being a 'medium' for other messages including social campaigns on representation, diversity, and equality.

Balancing the heritage of sport with the current global and social demands requires developing new strategies to navigate and innovate the event space to remain relevant in the global marketplace. The market may not reward exploitative innovation in the long-term by those looking simply to mine the heritage positions of a sport. Instead, as with other sectors, it may reward exploratory innovations that move with and ahead of emerging issues and drivers. From this research, the identification of the potential for bias at each end of the industry tenure continuum is noted. Understanding the differing perspectives around valuing event outcomes might provide a less contested way forward to maximise event impact across shared dimensions.

Reducing bias in event assessment and bids

In detecting sources of bias within the overall panel opinion echoes Murphy et al. (1998) in asking "To what extent is the average a good representation of the group judgement, when identifiable subgroups within the group differ in their judgements?" (p. 37). Considering the work on the development of the EIF as a process as a judgment function (Häder & Häder, 2000), the work on improving human judgement by Nobel laureate Daniel Kahneman is also highly relevant. The differences in judgements occur not only in bias but rather the two errors in judgements can be defined as "bias is the average error in judgments [and] by contrast, noise is the variability of error" (Kahneman & Sibony, 2021, p. 2). Both of those factors – bias and noise - were captured in the Delphi study. Bias was uncovered within specific groups in their higher mean scores attributed to their areas of expertise, as well as noise in the standard deviations around those mean scores.

Although bias was found, this does not prevent the average weights of the EIF to be a good representation of the group as a whole. Indeed, to find bias within expert opinions is not unexpected, as Li (2017) notes "evaluators with expertise in a particular field may have an informational advantage [but] they may also have personal preferences that impact their

objectivity" (p. 60). In testing to what extent the bias of individual groups may have disrupted the overall results, the analysis showed that it did not prevent a consistent weighting of mean scores from being achieved across the two rounds and diverse experts. Indeed the process of improving judgements is not to exclude the expertise but to ensure sufficient diversity in the panel as to take advantage of the 'wisdom of crowds', whereby the group's judgement outperforms its members (Surowiecki, 2005). The Delphi process sought to ensure this outcome using its strong sample base.

The second dimension of differences in judgements (noise) was the variability in views. "Members of a specialty are more likely to advocate techniques that involve their specialty." (Murphy et al., 1998, p. 38) and hence be prone to 'over-weighting' the importance of their own area in comparison to the areas outside their core field. By explicitly identifying 'over-weightings' as outliers to the group's judgements, the process of capturing and accounting for them renders the group views as more accurately representative. In reducing future noise, the development and sharing of the EIF results provide an 'anchoring effect' that may lead to more representative views by lowering the tendency to wide variations within future judgement groups (Tversky & Kahneman, 1974).

7.4 Drivers of event impact

A key strength in the design of the EIF was to establish a nested structure to allow for sub-elements, or Drivers, of each Pillar to be assigned weightings for their relative importance. Unlike the top-level Pillars, the Drivers were rated within each Pillar's workstream. As members of each workstream had nominated themselves as having experience or expertise in the area, the weightings for Drivers did not add further to the expertise bias already identified.

Establishing the degree to which consensus occurred on ratings of Driver importance within each workstream, the same tests of significance on mean scores and standard deviations as applied to Pillar weightings were used. This approach again tested the degree of stable opinion and any movement toward or away from a consensus view on the importance of each Driver.

Economic Drivers

Of all the Drivers underpinning the Economic pillar, it was not surprising that Tourism was accorded the highest weighting, since the success of events is often "measured in terms of tourist arrivals" (Fourie & Santana-Gallego, 2011, p. 1369), because of the economic gains from new money injected by visitors into the economy. The advice of Fourie and Santana-Gallego (2011) to recognise spending by organizers and sponsors, not just visitor expenditure, appears to have been heeded, with Direct Spending second only to Tourism in Driver weighting, and Sponsor Activation weighted above event Security spend. Indeed, that Security was the lowest weighted element does not yet perhaps fully reflect that elevated security spending has become an enduring aspect of event hosting (Toohey & Taylor, 2012). Of note, however, was that the highly tangible event outcome of Infrastructure was only weighted third, behind direct spending by the event. That position does reflect a balance between the heavy investment for rarer large-scale events against the much wider range of events for which specific new infrastructure is not required; and consideration of the increasing use of temporary infrastructure for event hosting (Arup, 2017). The potential to increase local business activity is balanced between new spending and leakages from displacement, which like infrastructure, was seen as a mid-weighted outcome.

The rating of the Economic Drivers showed a stable opinion between the rounds and while there was a trend towards an increase in agreement, it was not significant. This resulted in a Type B classification denoting 'Stable opinion and stable moderate agreement'. However, a Type D – 'Stable opinion with increasing to moderate agreement' might be achieved if the sample base was large enough to reveal a significant closing in standard deviations.

Sport Drivers

As the second highest weighted Pillar and having a strong workstream cohort, the allocation of weights to the Drivers within the Sport workstream added an important dimension to the understanding of 'sport' within the context of sport event hosting. Despite the intense interest by both media and academia towards mega-events, the scale of the event was not seen as the most important driver of event impact. The primary position of Event Prestige is not just a reflection of mega-events but captures the value of other events in the opportunity for prize money and ranking points. Noting that areas interact to add value (Gratton et al., 2006b), prestige contributes through strong brand image associations for the host and attracts a premium in media value.

That Sport Reputation was rated second in weight and ahead of Sport Expansion and Sport Infrastructure was of particular interest. Attuned to "more often than not, gambling, cheating, match fixing, diving, doping, and so on make the front pages rather than the back" (Murray, 2012, p. 587), these weightings perhaps recognise that damage to reputation can do more to hamper the growth of the sport than the opportunities the events can provide.

If size were the only consideration, a low rating of Event Scale would not be expected. However, while Event scale is an important factor to leverage when an event is strong, if that event lacks outcome uncertainty, it "negatively affects public interests and hence broadcasting and match attendance" (Zheng et al., 2019, p. 46). Hence seeking events to host based simply on size instead of considering smaller but highly competitive events that can sustain ongoing attention and investment is consistent with Event Scale being afforded an overall lower weight.

As seen in the Economic workstream, all Sport Drivers demonstrated a consistent directional change and a reduced Standard Deviation over consecutive Delphi rounds. However, as the narrowing was not statistically significant, it was also classified as a Type B outcome, denoting 'Stable opinion and stable moderate agreement'.

Media Drivers

The globalised nature of sport and communications has raised the importance and value of media in the revenues generated from sport (Fujak et al., 2016). The notion of global audience reach and scale is even embedded into the classification of events, with the largest all described firstly as 'international or major spectator events' (Gratton et al., 2000), denoting the importance of the audience in the event profile and scale. Interesting, then, that it was not Audience Size that was given the highest weighting within the Media workstream. Instead, it was Media Output that was given the highest weighting of the drivers, with Audience Size second.

That ordering may at first seem incongruous given the high values being paid for media rights to reach a large sports audience. However implicit within those valuations is the expectation of high availability, high-quality and deep media content to be packaged, broadcast, streamed, posted, and hence consumed, commented, and shared. In that light,

the Media Driver weightings are ordered appropriately, even without consideration for the reliability of claimed audience size (Harris, 2007). The remaining three Drivers very narrowly ranged in weights and might be considered a balanced mix of elements needed to drive Media success. The strong alignment within the workstream on their weightings recognises Media Presence and Media Accessibility are needed to underpin content generation and audience reach. It also appears accepted that Sponsors Output matters equally in that they make their self-interested contribution to raising the event profile to recoup a return on their sponsorship investment (Cornwall, 2008, p. 31).

While the changes in mean scores between rounds were larger than for the Sport workstream, they remain within two points of the original weightings. Testing the mean scores and standard deviations found no statistically significant differences between the two rounds and hence the Media Drivers portion of the Delphi study is also classified as a Type B outcome denoting 'Stable opinion and stable moderate agreement'.

Social Drivers

The contribution of events to the development of social outcomes saw the Social Pillar weighting grouped alongside Media and Brand. Perhaps in an attempt to bring the Social perspective onto an equivalent basis with the Economic perspective (certainly implicit within the Triple Bottom Line perspective), the language used in assessing social outcomes often references the Social Capital being built. Indeed that social capital is a vital element in maintaining a healthy and vibrant civil society (Onyx & Leonard, 2000) was reflected in the Driver dimensions tested that included health, cohesion, engagement, and identity of communities.

The weightings of the Social Drivers do to some extent reflect the desire that events have an impact on building Social Capital in their host communities. The two highest weights ascribed to Community Cohesion and Social Engagement both reflect an expectation on events ahead of more directly applicable effects on Health and Active Living that might more seamlessly be presumed to flow from sporting events. The challenge of these weightings is twofold. The first is that the aspiration for social effect receives the highest weightings but the measurement of those social effects are not well-established, even though "social capital is a proxy for social cohesion" (Gibson et al., 2014, p. 113). The second is that while more directly measurable levels of sport participation and activity are collected in many developed markets, the causal effect of events in raising the latter remains missing.

The two lowest ratings of the Social drivers were the Goodwill/Pride from hosting an event and Volunteering. That the most measurable dimension of Volunteering should be afforded the lowest weighting seems to imply it is not as valued as an outcome, because it is less likely to be discounted for being difficult to measure. By contrast, Goodwill/Pride must be seen to be a more important outcome since it receives a higher weighting, despite it being less directly and consistently measurable. Those differences in the importance uncovered in the weightings despite differences in their reliable measurement reflect the observation that the political decision to provide enjoyment, pride, and excitement may exist even though "economists are still insufficiently capable of capturing this effect in their cost-benefit calculations." (de Nooij & van den Berg, 2018, p. 69).

While overall the consistent theme of stable group opinion continued in line with other Pillars, the Social pillar did show a significant narrowing of agreement around 'Community cohesion and quality'. As the top Driver in the pillar, this provided a concrete example of the consensus formation only seen as a trend (albeit consistent) in the other drivers. With

this significant change in the highest weight Driver, the Social Delphi rounds might qualify as prospective Type D – 'Stable opinion with increasing to moderate agreement'.

Brand Drivers

There was some discussion in the early stages of the EIF on whether the separation of Media and Brand was a semantic or a real distinction. The definitions underlying the Drivers showed Brand and Media to be quite different in focus and measures, and each was ascribed solid weightings in terms of the final EIF mean scores.

In understanding the assessment of event outcomes within the Brand workstream, the strongest weighting went to a single Driver, which was surprisingly not about the host, but instead the image building of the event. This may not be the news that prospective hosts wish to hear, but it is evidence of a rational judgement on events. Any event that cannot remain relevant and engage an audience cannot add value to its host, hence there is a degree of alignment of interests between the two key stakeholders 'owner' and 'host' in terms of ensuring the event maintains its 'brand health'. This is especially true for large scale events. Given the long lead times from being awarded the event to hosting it, they remain exposed to a poor host damaging the event brand before the city has their opportunity to host it. The viewpoint on the event brand attracting the highest weight was not universal within the workstream, as it also attracted the widest disparity in values.

Building from an event with a strong brand (and as noted in the Sport workstream, preferably high Prestige), to attract investment, prospective hosts seek to "enhance the city's image and generate revenue." (Sant & Mason, 2015, p. 42). These elements are reflected in the close weighting for prospective hosts to benefit across building their image, establishing a reputation as a credible host of events, and being seen to be attractive beyond immediate visitation (captured in the Economic pillar) through sustained positioning as an attractive location for work and investment.

Despite it being often cited as a key rationale given for event hosting, the lowest weighting was given to 'Host attention', reflecting the finding that "most of the awareness was oriented towards the attributes of the event itself" (Bob et al., 2019, p. 8). The principle that increasing competition for sport content "means broadcasters give blanket coverage at peak times for such events, enhancing the marketing benefits to the cities that stage them." (Gratton et al., 2006b, p. 44) is still true, but using events to promote and position a host city has become a highly congested and contested space, limiting the potential benefits to hosts from one-off events (Knott et al., 2015). Hence the inclusion of recurrent events under the EIF would provide an opportunity for a host to progressively build awareness and extract more value from 'Host Attention' than a one-off event of a similar scale may deliver.

Testing the mean scores and standard deviations of the Brand workstream found no statistically significant differences between the two rounds and hence the workstream was also classified as a Type B Delphi outcome denoting 'Stable opinion and stable moderate agreement'.

Environment Drivers

Despite having its own dimension in the TBL assessment model, the Environment Pillar received the lowest weightings and the fewest panel members who claimed expertise in the field. This limited the extent to which conclusions might be drawn from the weightings ascribed to the four Drivers within the Environment workstream.

The strongest weighting amongst the Environment drivers was ascribed to the Consumption Footprint of an event. If a driver of economic value is the arrival of visitors from outside of the host region, then seeking to reduce the consumption footprint of an event is at odds with other event outcomes (Chappelet, 2018). That does not reduce its importance as reflected in the weighting given by the workstream but requires work to overcome or offset the effects of event hosting. The next highest weighted drivers of Energy Efficiency and Resource Utilisation operate at either side of the consumption footprint. The early development work on assessing local event environmental impact focussed on "increased use of renewable energy on-site, waste minimisation and recycling, water conservation, and reducing noise and light pollution."(Collins & Cooper, 2016, p. 3). Those 'environmental cost avoidance' strategies are still valuable and increasingly built into or retrofitted into the sport infrastructure. Even then, the longer-term dimension of improvement in sustainable and efficient Building Design was not expected as an outcome for events that use existing infrastructure but should be built into longer-term plans. Hence while it receives a lower weighting as an outcome from an individual event, for most hosts its benefit (and cost) is spread across multiple events. Likewise, the use of existing built assets limits the opportunity to influence Resource Utilisation at a single event. However the mid-weight rating of importance perhaps reflects that it is curiously relevant to smaller-scale outdoor events that use the natural environment, and applies to mega-events. The smaller events can seek to maintain or preserve the existing landscape, and the major event brings the opportunity to regenerate a historical industrial development or contaminated land as a once in a generation benefit (Hellings et al., 2011). While these extremes of event scales offer examples that bookend the recognition of the resources used and improved in the places in which they are hosted, it is noteworthy that this element is considered of some importance within the environmental impact of events.

Noting the limited number of respondents, the Environment workstream is given a nominal Type B* Delphi assignment - Stable opinion and stable moderate agreement, but with the notation that it remains the result that would benefit most from a stronger panel base.

Summary of consensus classifications

A summary of the formation of opinions and the degree of consensus around those opinions from each of the Delphi studies conducted at both the Pillars and Drivers levels during the research shows initial classifications listed in the table below.

Table 96 - Summary of Delphi consensus outcomes

	Respondent group	Delphi typology	Description
Pillars	All panel	Type D	Stable opinion with increasing to moderate agreement
	Without outliers	Type C	Stable opinion with increasing to strong agreement.
Driver workstreams	Economic	Type B	Stable opinion and stable moderate agreement
	Sport	Type B	Stable opinion and stable moderate agreement
	Media	Type B	Stable opinion and stable moderate agreement

	Social	Type D	Stable opinion with increasing to moderate agreement
	Brand	Type B	Stable opinion and stable moderate agreement
	Environment	Type B*	Stable opinion and stable moderate agreement

*Denotes caution in fixing a conclusive type until a test is conducted with a larger sample size

The results in Table 96 suggest that consistent opinions can be formed when the definitions are clear and related to specific elements of event outcomes. Reconfirmation of those initial opinions through a second round attests to their stability. The notion of contestedness of the legacy concept lies not in recognition of the elements of event outcomes or their contribution per se. The source is one of the degrees of agreement, that is, the consensus around those weightings. In testing that determination, since 29 of the 30 Drivers tested all trended towards a narrowing of standard deviation (agreement), the overall result tends to support that the formation of a consensus is possible under the EIF. It is only under the very constrained circumstances of the Environment workstream that an increase in disagreement was observed in a single driver. That the constraint was the very small respondent base supports the proposition that a larger sample base might be a major constraining factor preventing a more definitive conclusion of consensus being drawn and a potential classification of Type C – 'Stable opinion with increasing to strong agreement' for the framework overall.

7.5 Event outcome measurements

In the development of the EIF, the following admonishment of Bob and Kassens-Noor (2012) from work in reviewing existing event assessment frameworks was noted: "despite these prominent categorisations of legacies, only a few researchers have suggested actual indicators to measure legacies, and those who did, were restricted to a specific subcategory of their selected field of study" (p. 13). This observation rang clear. The EIF needed to incorporate an extended list of indicators to ensure they addressed multiple dimensions of event outcomes.

Extending the ambition for the EIF, the design also sought to add explicit ratings for the indicators missing from prior work. Those ratings as noted were on two dimensions: the Importance [I] of each measure in assessing the potential pillar impact, and then the Reliability [R] with which each measure is consistently measured and available for assessing the pillar impact. In addition to differences in the perceptions of importance and reliability of indicators revealed within each pillar, the application of the ratings to each of the six pillars showed relative differences that would not have been seen had the model followed previous work and had been limited to a single specialty area.

Disparity in measures

The first of those insights is that some measures were considered more Reliable than they were rated as Important [R $\overline{x}$ > I $\overline{x}$]. The reverse is also true, that is some measures were considered to be Important but less Reliable. A combined construct of Impact Score [$\overline{x}$ I * $\overline{x}$ R] was used to rank the overall score for each measure. As a result, the top ranked measures were not always those that ranked highest on Importance but rather achieved well on both dimensions. This may be mathematically self-evident, but it surfaces some basic direct measures that might be overlooked in frameworks focussed on extended legacies and induced effects. For example, in the Economic workstream, simpler direct count measures

such as ticket sales and value of sales achieved higher Impact Scores than measures rated higher on Importance as compared to more difficult measures to capture accurately and consistently such as visitor spending or jobs created linked directly to the event.

A second effect that was possible to detect because the EIF captured measures across a range of Pillars was that whilst some Pillars showed the pattern of measures being rated as more Reliable than Important [R $\overline{x}$ > I $\overline{x}$], this was not true of all Pillars. Given that each set of measures was rated within the workstream by respondents with expertise in that area, it appears that the overall reliability of measures in some areas of event assessment was less than others. Indeed Economic, Sport, and Media had some measures showing evidence of greater reliability than importance [R $\overline{x}$ > I $\overline{x}$], but no measures rated in Social, Brand, or Environment were considered to be more Reliable than they were Important.

The critical role of reliability in building confidence and consistency in event assessment is discussed later, but it is noted that the findings provide critical evidence of the current state of measurement on the relevant Pillars that undermines their value to prospective hosts and in driving accountability of event outcomes. Reflecting on the relationship of those Pillars to the TBL model of event assessment, Economic and Media can provide some more surety in overall Economic results if the more reliable measures are selected, likewise for Sport under the TBL [Social] dimension. But there is work to be done on bringing the Brand, Social, and Environment measurements up to the same standard of reliability. This issue of developing more reliable measures is not new and was highlighted in relation to Brand measurement via Victoria's Major Event strategy, noting "both models also erroneously exclude the longer-term benefits of brand value [as] evidence is difficult to find …and we accept the point that more research is warranted." (Victorian Auditor General, 2007, p. 155).

In considering the balance in the weighting of the TBL model, should the default be towards those elements with strongly reliable measures such as Economic? The responsibility to ensure events are not an unreasonable burden on taxpayers and/or generate a return from investments tends to support the primacy of economic impacts, which is where the sector has previously focussed on impact assessments. Evidence that a more balanced approach to event measurement is important was found in the cross-workstream responses to the measure ratings. Challenged to rate the importance of measures across different Pillars, the panel overall did not simply default to rating Economic measures most strongly. Indeed, measures derived from three different Pillars were given the top three measure rating places, and four different Pillars were represented in the top five measure ratings. The overall balance in the cross-workstream ratings was important as it confirmed the weakness in the measurement scores within the Pillars is a 'reliability problem' for those areas, not a reflection on the importance of the areas themselves. It reconfirms that both dimensions are needed to be able to diagnose and improve the source of issues with measurement.

Internal consistency between Drivers and Measures

Capturing the importance of event dimensions on both the allocation of weightings and the ratings of measures allowed for testing the extent to which the two approaches may hold some internal consistency within the framework. The three tests undertaken involving correlating different constructs of measure ratings against the Driver weightings revealed two key insights that inform the ongoing development of event assessment frameworks.

The first is that using only the Importance ratings allocated to the measures dimensions gave the weakest correlation with the Driver weights allocated. Using the Impact Score from the 24 top rated measures improved on the 'Importance only' correlation, but it was the use of

Impact Score including all proxy scores that achieved the best correlation between the two dimensions. The improvement from $r^2 = 0.17$ for 'Importance only' to $r^2 = 0.50$ for the 'All Impact Scores' established that a solid internal consistency was achievable and established the basis on which refinement and future improvements might be made. The second, and from a generalizability perspective, the more important finding was that the results reiterated the value that Reliability brings in adding new information to the assessment of event outcomes. In light of the results of the internal correlation scores, a conclusion that the EIF would have been considerably weakened by only capturing the Importance ratings on the measures of event outcomes (without Reliability) appears well-founded.

Reliability, Impact Scores and event assessment improvement

Understanding how 'reliability' adds information to event outcome measures is not in looking at the accuracy of a single event but its effect in enabling the continuous collection across events. That notion of accuracy over the broad application of sources is also found in Generalizability Theory, where the "question of 'reliability' thus resolves into a question of the accuracy of generalization, or generalizability" (Shavelson et al., 1989, p. 922). Hence the description as being 'consistently and reliably available' in the rating scale seeks to understand if a measure can be usefully applied across events of different types and scales. To that end, the Impact Score, as the product of Importance and Reliability mean scores, adds an important dimension to the assessment of event outcomes. By operating at the measure level rather than the Pillar, it brings a focus on where specific gains might be made to improve event assessment and an opportunity to identify where a proxy measure of generalizability might be a better approach than the use of an unreliable direct score.

While shown to be helpful, the notion of a construct that combines importance and reliability dimensions is not new. Within engineering designs, Marginal Reliability Importance (MRI) measures the impact of any component on the system. The application of 'reliability importance' from engineering to event measures is of particular value in that it reveals that "improvements in reliability of components with the highest MRI cause the greatest increase in system reliability… this information can be used to determine which components should be improved first in order to make the largest improvement in system reliability" (Armstrong, 1995, p. 408). Given the unreliability of outcomes from just the economic impact projections noted by Gratton et al. (2006b), the ability to know which specific measures to target to rapidly improve the overall reliability of event assessment could substantially improve the ability to set and achieve realistic event outcomes. Hence improving reliability can be the fastest way to improve an overall system. In the case of the EIF, the system is a consistent approach to assessing event outcomes.

That same issue of uncertainty in economic impact projections of events mentioned above applies to one of the central tenets of Sport seeking support for public funding for event bids. The promise of increasing sport participation as a legacy of major events is a benefit often promulgated by supporters in event bids (Frawley & Van den Hoven, 2014). Despite the appeal of the message, post-hoc assessment of outcomes shows there is no evidence of an effect on participation levels in sport and physical activity (Homma & Masumoto, 2013). This is not that the effect might not exist but until "reliable and consistent participation data are available, the question of causality in the context of the wider sport development and participation system remains to be addressed" Veal et al. (2012, p. 155). Until a more reliable measurement basis can be formed, the value of the promised 'participation legacy' could take guidance from the weightings of the relevant Drivers in Sport (Sport Expansion) and Social (Health and Active Living). The lower to mid-level weightings given respectively suggest an objective funding valuation might avoid a heroic 'Inspire a

generation' designation and place a lower rating on any potential value arising from that dimension of events.

The second opportunity the Impact Score unlocks is to rank the generalizable measures against important but unreliable direct measures. For example, measures such as total television audience are seen to be an important measure of the scale of an event but also highly unreliable as "television ratings data is fraught by both methodological measurement weakness" (Fujak et al., 2016, p. 94) with claimed values often found to be grossly overinflated. For example, in the case of FIFA's 2006 World Cup, despite an audience of over one billion being claimed, an independent assessment concluded that "true audiences are between a quarter and a third of that size"(Harris, 2007). Hence while measures of television audience may improve, it could currently be classified as Importance = High, Reliability = Low. If a more reliable measure of an audience cannot be agreed, a more reliable basis for comparison of television viewership across events and time such as 'TV hours produced' (which uses a viewer's per hour of content relationship to assess its viability) may prove a more direct and reliable measure and a proxy for the scale of 'media coverage'. Likewise, 'media rights value' (where self-interest in not over-paying for audience size in a competitive market should lead to a discount for any uncertainty in audience size) could be another alternative proxy for Television Audience Size. A third alternative could be to use both in combination to raise overall reliability.

7.6 Event scale and Pillar weightings

While the Delphi study sought to allocate weights to the contribution of event impact on a generalizable basis, it is acknowledged that no single standardized model weightings will fit all events perfectly. Given the interest in the applicability of the EIF across different events, the question as to the extent that weightings might differ across events of different scales is pertinent. The standard model had shown stable opinions in the weightings of the Pillars and significant gains in consensus across the rounds when filtered for outliers. As noted in the research method, an optional question sought to elicit how those opinion weights vary across the different scales of events.

For consistency, the Sheffield Hallam SIRC Event Typology Table 6 was used to frame the scale of event options, which is reproduced below in Table 97 for ease of reference.

Table 97 - SIRC Typology for Pillar Weightings by scale

Type A	Type B	Type C	Type D	Type E
Irregular, one-off international spectator events, generating significant economic activity, and media interest (e.g. Olympics, FIFA World Cup)	Major spectator events, generating significant economic activity, media interest, and part of an annual domestic cycle (e.g. Open Golf, Wimbledon)	Irregular one-off major spectator/ competitor events generating limited economic activity (e.g. Grand Prix Athletics, World Championships)	Major competitive events, generating little economic activity, and part of an annual cycle of events (e.g. National Championships in most sports)	Minor competitor / spectator events, generating very little economic activity, no media interest and part of an annual domestic cycle of sport events. (e.g. local and regional events)

Despite the small respondent base that chose to complete the additional question, there were several significant differences detected in the Pillar weights ascribed to the impact of events of different scales. For the two largest scale events (Type A and Type B), there was no difference in mean weightings allocated to their Economic impact. However, for Type C

(one-off major events), Type D (national events), and Type E (domestic events), the weighting of economic outcomes was progressively reduced as the event moved down in scale. In the case of Economic impact, the weighting for Type E events was half that of Type C, with Media being given a similar weighting. Of the remaining EIF pillars, the Sport pillar reversed the previous Economic pattern, with Type C to Type E all reflecting a progressively strengthened Sport pillar, indicating sport impacts become a more important part of the event outcomes as scale shrinks. For Media and Social, only Type E showed a statistically different impact (lower) than Type A events. Brand was quite resilient and remained stable across events of all scales.

It is noteworthy that despite the description for Type E events as generating 'very little economic activity' and 'no media interest', the weightings ascribed to those pillars was not zero (or one as the minimum allowed score) but was reduced relative to the other pillars for that event. The results show a tailored distribution of weightings ascribed to the Pillars relative to what might be possible within that scale of an event, not simply applying a diluted version of the mega-event weightings. The ability of the EIF to be applied consistently and with varied weightings across events of different scales supports the premise that a framework can be designed to provide a consistent but comparative basis for the assessment of event outcomes.

7.7 Stakeholders and assessing event impact

While events are complex projects with a network of identifiable stakeholders, there are two key stakeholders classified as 'Dominant Stakeholders' due to holding a position of both Power and Legitimacy (Mitchell et al., 1997). In the context of event hosting, these stakeholders were identified as the Event Owner and the Event Host. Despite the shared position as Dominant Stakeholders, there is a potential for the misalignment of interests in the delivery of event-specific outcomes. Of those interests, the Event Owner holds "the rights to the name of an event…along with the rights to broadcast it, and to associate with the event as suppliers, sponsors, licensees, cities and destinations" (Getz et al., 2015, p. 92).

The Event Host who seeks the right to host - be an agent for - an event from the Event Owner (Principal), should look to maximise the value from their investment, which may include the use of public funds. However, Hosts can be restricted from doing so in practice as "highly formalised host agreements negatively impacted the host destinations' respective abilities to leverage the event." (Kelly, 2019, p. ii). From the Event Owner perspective, they seek hosts that ideally "act in the ISO's [International Sport Organisation] best interests, yet conflicting interests and information asymmetries give rise to agency problems" (Geeraert, 2016, p. 27).

The effect of making 'legacy' a requirement of event bids is an effort to "limit divergences from [owner] interest whilst delegating some decision-making authority to the agent" (Jensen & Meckling, 1976, p. 308). Hence, instead of allowing the Event Host to maximise their return, they instead cede some benefits of the principal [owner]. Indeed, the requirement to be able to claim extended legacies enhances the principal's property and raises its market value for future bidders. To that end, the incentives of owners and hosts, despite being both Dominant Stakeholders, are not fully aligned.

In the Delphi study, an assessment of the degree of alignment in Dominant Stakeholder interests was tested through the expert panel's ranking of their perceptions of Pillar outcome priorities. As noted in the results, the study found significant differences in the Event Host and Event Owner priorities on all six of the EIF Pillars. The key outcome priorities for Event Owners were distributed across Economic, Media, Brand, and Sport pillars. These

findings show the potential for significant misalignment even between the Dominant Stakeholders. Despite any rhetoric in communications, the expert panel perceived that Event Owners give little priority to the Social and Environmental outcomes, and instead Social was determined to be an Event Host matter, not the responsibility of the Event Owner. Hence if an event can deliver on these outcomes, the incentives are for the Principal to benefit from the Agent's work and for the Event Owner to claim these as an event outcome anyway. The latitude given in the descriptions of legacy to include all outcomes 'planned and unplanned' (Preuss, 2007) exacerbates the incentives of Event Owners to do so. Conversely, as the Event Host allocates resources to achieve its own outcomes priorities first and the Event Owner's priorities second, this adds further incentives to be even more proscriptive in the hosting agreement terms offered to future hosts.

Reflecting on the priorities ascribed to Event Owners compared to the TBL structure of event impact reports, it is noted that two of the three headline areas of the Triple Bottom Line - Social, and Environmental - are almost absent for the Event Owner and are seen to be reliant on the Host to achieve a more balanced outcome. It is important to note that consensus viewpoint by the panel does not ascribe a judgement as to whether that misalignment is acceptable, but rather that is believed to exist. The research, by making explicit the issue of balance of contributions towards positive event outcomes between dominant stakeholders, could encourage Owners to work more proactively with Hosts for Social and Environmental outcomes in future editions, in line with increasing ESG expectations of global organisations.

Given the divergent views on the Host and Owner priorities by the expert panel in the Delphi study, how strongly might the views be representative of the current state? The depth of experience and the range and responsibility of roles in the panel would underpin their value as an external perspective capable in judging a generalised Host/Owner relationship. The testing of the priorities also shows the broader application of the EIF pillars in providing a more detailed view of differences based on dimensions grounded in the literature. It would also be expected that individual and specific incentives of the Principal/Agent relationship would change with differences in event scale and across different event ownership models.

Reflection on using Impact and contestedness

As noted in the literature review, the term 'impact' has been pejoratively characterised as negative and criticised as focussed on shorter-term outcomes. In this research, the term 'impact' was specifically adopted and placed at the forefront in the naming of the EIF. Rather than concurring that the term is inherently negative or a limitation, it was foundational to the identification of the more generalizable near-term outcomes a structured and standardized assessment framework needs to allow comparison across a wide range of events in scale and type. The development of specific dimensions of impact and their sub-drivers within the EIF has been found to have consistency in the relative importance of those areas in event outcomes, with increasing consensus achieved across rounds.

In contrast to the deliberately constrained approach of 'impact', its counterpart 'legacy' has sought to expand its reach at the expense of specifics that might help resolve the elusive definition. Hence, as reflected in Preuss (2007), legacy is "all planned and unplanned, positive and negative, tangible and intangible structures created for and by a sport event that remain longer than the event itself" (p. 211). This provides event hosts and owners significant latitude in what might be claimed, and thus disagreement over the contents of legacy and its usage. Despite this lack of agreement on legacy's actual definition, it has

become a requirement to include a 'legacy plan' in many event bids. It is a concept that is agreed as valuable, but the nature of which cannot be defined, despite meeting the conditions of an Essentially Contested Concept (Gallie, 1956). The 'impact focussed' Delphi study of event outcomes afforded the opportunity to reflect on whether the findings might contribute to a reduction in the level of 'constestedness' around the concept of legacy.

Table 98 - Assessing Gallie's conditions from the Delphi study

	Gallie's ECC conditions	Event outcomes under 'impact' Delphi study
I	It must be appraisive in the sense that it signifies or accredits some kind of valued achievement.	Weights given to both Pillars and Drivers shows an explicit value afforded to each element of event outcomes
II	This achievement must be of an internally complex character, for all that it's worth is attributed to it as a whole.	The complexity of event hosting can be disaggregated and measured in individual dimensions. The worth of the event can be seen as a whole, but each area has its own value that is measured and valued for its contribution.
III	Any explanation of its worth must include reference to the respective contributions of its various parts or features	Specific and measurable dimensions allow reference to parts of events in contributions to the overall.
IV	Achievement must admit considerable modification in the light of changing circumstances; and such modification cannot be prescribed or predicted in advance.	Not prescriptive in design but reflective of outcomes. The value is ascribed to dimension. Changes may change weightings applied; the model can be recalculated in light of unexpected changes.
V	Each party recognizes that its own use of it is contested by those of other parties, and that each party must have at least some appreciation of the different criteria in the light of which the other parties claim to be applying the concept in question.	Experts do display a bias towards their own use but non-experts also ascribe some value to the same dimensions, recognising them to also hold some value. As parties are made aware of differences in views, the movement is not to contest others' usage but rather to move towards a consensus viewpoint.

As noted, all five conditions must be applicable for a concept to be described as Essentially Contested. Through the use of an 'impact' perspective, with a more direct focus on event outcomes and the weighting of explicit dimensions, the Delphi study adds quantified effects to challenging the applicability of two of Gallie's specific conditions, namely the second and fifth conditions, as seen in Table 98.

The second condition considers that the notion of worth can only be attributed to the whole. However, through breaking down the overall 'impact' of events on their host city into key dimensions and their sub-elements, each could be appraised for its individual contribution through the allocation of weights to represent their importance. This allows for the consideration of its worth to specific stakeholder audiences rather than only treating the outcome of the event in aggregate. The fifth condition is of interest in that the Delphi study did reveal bias within expert groups, showing that experts were applying the concept differently to non-experts. That, however, did not prevent the formation of a stable consensus opinion across the groups.

A practical impact of reducing contestedness is a better shared understanding of implicit expectations of event outcomes and the resources needed to achieve them. By making explicit differences in stakeholder expectations and perspectives, the host can recognise how changes to the hosting environment and plans will impact stakeholders differently. Understanding their perspective enables better engagement with stakeholders in formulating and delivering a successful response that aligns with their priorities and expectations. As the results of the Delphi study indicate, when awareness of others' use of the concept is made explicit and available, the movement is towards a shared viewpoint rather than reinforcing any perceived contestedness. Hence the concern that constestedness may be a persistent condition endemic to the assessment of events is challenged. A shift to a more direct impact perspective may also contribute to a reduction in reliance on the 'catch-all' definitions associated with the use of 'legacy' in time.

7.8 Key characteristics of the framework and its benefits and application

A negative characterisation of 'impact' as focussed on shorter-term outcomes fails to recognise that a shortened timeframe allows for some benefits in event assessment not available under the extended timeframe of the legacy perspective. The first is that a shorter horizon allows for a greater level of certainty about the outcome and that the outcome is directly attributable to the event and not related to other confounding factors. The second is that under the impact timeframe, annual and recurrent events may be included and assessed regularly, enabling the development of standardised information to inform decisions on future events.

Timeframe

In 2015, the report on legacy outcomes of the London 2012 Games finished with the researchers concluding "as with any long-term project that is intended to be a catalyst for long term change and transformation, the analysis of three years into legacy that this report presents is only the beginning" (UEL, 2015, p. 185). For the very rare events that qualify for mega-event status, the extended timeframe in which three years post-event is 'the beginning' may be appropriate. For most events, and especially those that are held annually, for the outcomes from any edition to be assessed, the timeframe must be short enough to not be confounded with previous or future editions. A benefit of that shortened timeframe is the greater certainty of both the attribution to the event and the achievement of the outcome. The decades-long research of the Good Judgement Project found the ability to accurately forecast outcomes declines the more extended the time horizon, with the reliability of predictions beyond a year declining rapidly outside the exceptional group of so-named 'superforecasters' (Tetlock & Gardner, 2016).

In capital markets, that uncertainty as the time horizon extends is factored into Net Present Value (NPV) and Discounted Cashflow (DCF) methods. They place greater value on near-term cash flows due to the effects of time on the value of money and future risks. While a shorter horizon produces an overall lower value by constraining the claimed future periods, it also avoids being discounted for future uncertainties. A 'legacy focus' extends the timescale over which the benefits are accrued and the value is calculated, with the Brisbane 2032 assessment using "a 20 year time period including the 10 years leading up to the Games, the Games event, and the 10 years post the Games (i.e. 2022 to 2042)" (KPMG, 2021, p. ii). It is not that those extended horizons should not be used. Where there are significant infrastructure investments for mega-events, the extended horizon of legacy frameworks is entirely appropriate in matching costs to the asset's useable life. What is challenged is that the extended timeline of the 'legacy' model should not be the default

perspective, but rather the exception. This is closer to the reality of how frequently events of that scale might be hosted.

Standardised information

A consequence of including more events by constraining the timeframe is that more outcomes might also become available from which to assess events, as well as their potential for prospective hosts. With the broader adoption of a standardised framework, under which a stable consensus was achieved during the Delphi study, a greater availability of standardised data could available across events. The value of standardising information is reflected the challenge that despite the significant investments required "it is not possible to compare successive mega-events in the absence of reliable and transferable reporting and auditing systems" (Hayes & Karamichas, 2012, p. 13).

In the absence of reliable and consistent information on events, some prospective hosts have looked to the use of contingent valuations to understand the level of support residents have for an event. The method requires the placement of a value on the item under examination as it seeks to "reveal the monetary trade-off each person would make concerning the value of goods or services." (Carson, 2012, p. 28). In effect, it is a method for understanding the value of an event without an active market for either price discovery or the availability of reliable information. It is not possible to conduct a contingent valuation on each event a city might consider hosting, but how much should a host bid for an event? Siegfried and Zimbalist (2006) caution that where an event places a "public obligation for debt service, [any] possible cost overruns [would] generate a substantial budgetary problem" (p. 423). The potential for drawing on public funding is exacerbated when "less than reliable economic benefit calculations cast a dark shadow on the return on investment claims made by both sport event organizers and government backers" (Lee & Taylor, 2005, p. 596).

Encouragingly, however, increased timeliness and availability of standardised event data does not need to be perfect before benefits can be realised by prospective hosts. Art, used by Gallie (1956) as an exemplar of Essential Contested Concept, is not prevented from enjoying an open flourishing global art market despite that 'contestedness'. In economics, a market "is considered to be efficient if prices which prevail at any time are found to be an unbiased representation of all currently available information." (Louargand & McDaniel, 1991, p. 54). Noting the reference to 'unbiased representation' echoes the explicit identification of the biases under the EIF. Increasing the availability of information helps the market improve its efficiency over time. "Even though the basic institutions of the art market may have evolved as early as the 18th century, it appears to have steadily tended towards greater efficiency…due to the increase in quality and quantity of public information about auction prices"(Goetzmann William, 1995, p. 34). Hence a market becomes more efficient in increments as new information becomes available. Given the level of investment in hosting events – especially with the use of public funds – more consistent and frequent information about event outcomes could be achieved by consistent application of the standardised EIF, which should enable greater efficiency in the pricing of events.

Adaptability

Designing, developing, and testing the degree of consensus around explicit weights on the pillars of the EIF allows for those weightings to be reweighted and recalibrated over time within the existing structure, as stakeholder needs or societal perceptions of value change or better measures become available. To some degree, this process is already underway with "researchers hav[ing] started to shift their attention towards more intangible assets or returns

of events" (Agha & Taks, 2015, p. 200) including social and environmental impacts. Even those whose profession is based on tangible assets in building new infrastructure are looking to an event industry based more on the use of temporary venues and multi-purpose venues (Arup, 2017). A transition away from 'hard assets' towards intangible value should not be surprising. As demand for events increases and costs from rights holders rise, shifting the focus to sources of intangible value that are location-specific and not accrued to the rights holder avoids the transfer of value from the host to the event owner. That shift towards intangible value hinges on more consistent use and application of measures to avoid the fragility in intangible methodologies (Lee & Taylor, 2005). The adaptability of the EIF design is inherent in that it captures both tangible and intangible dimensions, and allows for reweighting the value of events across the balance of the dimensions both at the Pillar and Driver level. As methodologies for capturing the intangible value of events improves, they can be reflected in an explicit shifting of consensus over time.

Extended applications of a standardised framework

While the timeframe, standardised information and adaptability have specific technical benefits, the broader application of the EIF and how it could be applied to benefit different stakeholder groups has been considered. Making explicit differences in perceived value of events and their impact allows hosts to accommodate those differences and plan for outcomes under different scenarios and different biases.

Uncovering hidden value

A consequence of timely and consistent information to improve the event market efficiency may be to uncover hidden value in events. While examples of over-budget and costly events may dominate the narrative, being able to assess events on a standardised basis could assist prospective hosts in finding events that have been undervalued relative to their potential for impact. Targeting and securing those events may achieve a better return on investment for any taxpayer funds used in supporting the sporting event sector. That potential for benefit is not wholly directed to the hosts, as event owners may also find they can raise the value of previously under-developed or under-recognised event properties for which they hold the rights. Similarly, the desire for brands to associate themselves with sport and sporting events has made sponsorship a key source of event funding. The ability to assess event outcomes could enable potential sponsors to more critically identify events of different scales that offer value for their available sponsorship investments. It may also assist them to uncover events or sports that are gaining in momentum and hence secure longer-term rights before the value rises.

Rational expectations

The extended timeframe and 'catch-all' nature of the legacy-dominated view of events can inflate expectations of an event's potential impact disproportionately to its scale. Attempting to adopt the open-ended legacy approach to event outcomes within a single host location holds the potential for different events to become confounded and inseparable within a portfolio. The establishment and availability of benchmarked data from previous editions of the same or similar events could provide a rational basis of what any single event might deliver, so expectations within government and political agenda can be established on an objective basis. That establishment of a credible baseline may avoid under-resourcing events based on overly optimistic expectations, or the idea that any single event might be a panacea to a host's existing issues.

Portfolio alignment

Reflecting on the management of a host's event portfolio, hosts seeking to expand or build an event portfolio might also benefit from a standardised framework such as the EIF, which captures events across impact dimensions, especially those intangible elements.

Instead of defaulting to using the scale as the simplest comparator across events, prospective bidders could combine scale with the 'fit' of the event, by considering more broadly the profile of the events within their current portfolio. That impetus to deepen alignment between the host and the event is also reflected in the shifting narrative around event bidding and planning. As noted by Taks et al. (2015) "whereas legacy planning focuses on the event and the outcomes it might render for the community, event leverage focuses on the community and the ways that it can integrate each event into its marketing and management strategies"(p. 1). This development towards a deeper integration of events within the host's capabilities and strategies will likely create stronger alignment between event and host and the potential to realise the expected benefits and outcomes sought.

However, the potential of an event may still go underdeveloped unless the process can also recognise and highlight elements not being addressed. The extended breadth of the EIF dimensions and its focus on comparability could assist in identifying unmet 'fit' across the portfolio within the current event strategy and the host's longer-term development goals.

7.9 Limitations and further developments

Research choice limitations

As noted in the research methodology, the selection of a quantitative Delphi study in the form of a 'judgement function' could result in the omission of more subjective variables or elements that may be difficult to measure. The exclusion of those elements does not ascribe them to be of any lesser value but is a limitation within the initial design of the framework and research method selected. Seeking additional input from other qualitative methods of research on event outcomes, and inclusion of those dimensions where possible, would enhance future iterations of this research.

Technical limitations and developments

The testing of the internal consistency within the EIF between the weightings allocated to Drivers compared to the rating of Measures could be more rigorously assessed if each potential Measure was explicitly related to a specific Driver, rather than assigned post-hoc. There was insufficient room in the initial Delphi study to have respondents contribute to those allocations in addition to all the other elements tested without risking panel retention or data quality. The assignment of Measures to Drivers could be conducted as a separate companion study and applied post-hoc to the current study results. Alternatively, using the now established list of Measures based on Impact Score, a shorter list of Measures could now be included in a rerun of the existing Delphi study.

Concerning Pillar and Driver weightings, instead of respondents being able to ascribe a potential value from a minimum of 1 to a maximum of 95, the weightings could be constrained within a narrower range. This would have two effects. The first would be to reduce the degree of variation in an initial Delphi round by reducing the potential for extreme values, that is outliers, and being able to detect consensus in Drivers even within a similar panel size. The second effect would be more subtle but would signal to respondents within each area of expertise to be wary of their biases by limiting how much their own field can contribute to the overall weighting. Based on the data collected for Maximum and

Minimum values collected during this research, a reduction of the highest possible value to 50 and raising the lowest value to five would not compromise the range of weightings expressed for all but the most extreme outliers.

In the development of the conceptual design of the EIF, it was acknowledged that it may be desirable to add new measures and dimensions over time. Anticipating this requirement, the nested structure allows for it to adapt easily to include new measures into the framework. The adoption of a more advanced version of the BWS used in the Delphi to establish the relative strength of measures across workstreams could be developed. Specialist 'incomplete block design' techniques (such as a Bandit MaxDiff) might allow for the ranking of an extended set of Measures as a single full list rather than being compartmentalised under the relevant workstream. Not all respondents would be required to rate all measures, but the ranking of the full list would be established across the panel. New metrics could be added and tested more frequently under this approach, and it could be designed to enable identification of any cross-Pillar or 'multiplier' effects that key measures may have, which is not possible under the current workstream focussed approach.

Future developments

Noting the disparity in weightings provided across audience groups is in part what confirms the contested nature of legacy. But likewise, the consistent core weightings within the consensus range shows that it is possible to move towards a 'less contested concept'. While the study included an external view of how Event Host and Event Owners may prioritise the impacts they seek from events, those audiences were not directly tested in those roles. An important future development of the EIF would be to research the views of Event Hosts and Event Owners.

In addition, the resident or public view on what they value in terms of event outcomes, particularly from events of different scales, was not tested as the Delphi study was limited to an informed and expert panel only. Historically, the testing of prospective event value within the public has been researched within economic and social dimensions through contingent valuations. When conducted around events or prospective event bids, the aim is to determine a value (preparedness to pay for an event) that residents place on hosting that event. Extending a contingent valuation to include a task of assigning weightings across the six pillars as undertaken in the EIF would enable not just a value being placed on the event, but a more detailed view of which specific outcomes are being seen as the most and least valued by residents. Targeting events that deliver on the higher valued areas are therefore more likely to be met with support from residents as they perceive taxpayer funds being used in ways that align closer to their priorities.

7.10 Planning for future uncertainty

The past two years have seen significant disruption and uncertainty in the global event market, with events of all sizes impacted and having to be postponed, cancelled, or modified. The development of scenarios enables planners to think across and anticipate a range of possible outcomes when dealing with conditions of turbulence; uncertainty; novelty, and ambiguity (Ramirez & Wilkinson, 2016).

By providing a well-defined set of impact dimensions, the EIF could be used to provide a structured approach to building and analysing different hosting scenarios, the likelihood of those occurring, and how hosts could adapt or respond. Crucially, the discipline of exploring alternate scenarios acknowledges that the process of delivering an event is dynamic and not linear, and that outcomes promised in bid documents are in fact uncertain

and not conditional. Additionally, in the building of scenarios, event hosts need to work with stakeholders in a structured way to pre-test a range of conditions from ideal to adverse. They can explore who is impacted by the challenges under each condition and gain their buy-in towards achieving and delivering on agreed (modified) outcomes. That process itself will surface a diversity of views that might otherwise go either assumed or unacknowledged.

Applied well, they should provide "consistent and coherent descriptions of alternative hypothetical futures …which can serve as a basis for action". (Notten, 2005, p. 1605). At present, the use of the term 'unplanned legacies' is used to cover unforeseen events such as security breaches in Munich 1972, political protests as occurred in Mexico in 1968, or boycotts as per Moscow 1980 and Los Angeles 1984 (Gold & Gold, 2010). It is also used to cover highly foreseeable outcomes such as classifying the 'gentrification' of the London 2012 redevelopment area resulting in pricing out long-term residents (Dickson et al., 2011, p. 298) as an 'unplanned negative legacy'. It may have been unplanned, but it was certainly not unforeseeable.

At the core of scenario planning is the question of the likelihood of a range of potential outcomes (positive and negative) and that "in order to develop meaningful scenarios, creative input concerning possible future trends and developments is crucial" (Nowack et al., 2011, p. 1603). It requires the planner to be 'thinking the unthinkable'. While the case of a global pandemic may have been outside the range of 'creative inputs' considered, event organisers could have considered disrupted global travel under numerous scenarios and planned for their response in advance. Instead, the use of 'scenarios' about events is typically bounded within single disciplines of event outcomes (economic, social, media, etc.). The term is also used in economic modelling of events when describing a potential range of forecasts (low, central, and high 'scenarios') (Li et al., 2011). Other applications consider environmental impact footprints of alternate visitor transportation scenarios (Collins & Flynn, 2008), but again is constrained to individual dimensions of event outcomes.

The latent acceptance of 'planned and unplanned' outcomes dimensions as central to the definition of event legacy is an indictment of the lack of critical foresight being applied to the range of possible outcomes across multiple impact areas. A more disciplined separation of 'foreseeable consequences' from the 'unexpected', would reduce the likelihood of 'unplanned' & 'negative' outcomes arising. The failure to make alternate scenarios explicit gives rise to 'uncomfortable knowledge' (Stewart & Rayner, 2015), where changes to emerging outcomes go unrecognised or unacknowledged by event stakeholders. A key part of that acknowledgment is to understand that event outcomes are the result of complex stakeholder interactions, rather than a utopian legacy 'bestowed' upon a host. This reframes the planning process to deepen understanding of the 'wicked' nature of legacy development and delivery (Byers et al., 2021).

While the development and application of the Event Impact Framework does not in itself reframe the process of legacy planning, its contribution in establishing a base of standardised data across a wide range of event outcomes held in both favourable and challenging circumstances may provide a baseline from which to 'think the unthinkable' and forecast for their effects. Without applying this more disciplined and integrated approach to planning for future uncertainties of event hosting, event hosts will likely continue to report on the emergence of 'unplanned legacies' rather than their ability to anticipate and adapt around the planned legacies or outcomes.

8 Conclusion

Assessing sporting events impacts within a standardized comparative framework

This research sought to identify how sporting events of different types might be assessed for their impact on their host location within a standardized comparative framework. The foundation of the framework was grounded in the research literature from which key dimensions in event assessment across event types were identified.

The resulting six core areas (pillars) and thirty sub-elements (drivers) were targeted to capture event outcomes from across disciplines to build a comprehensive view missing from other grounded models. The value of the areas and sub-areas was established through assimilating the views of a global panel of experts to test the applicability of the framework to events. A key advance on other frameworks this research has delivered is the combination of breadth of areas considered and presented for impact, with the specificity at the sub-element and individual measure level. The deliberative approach to specifying the outcomes in terms of actual measures forced the consideration of the extent to which each of the 200+ measures might apply to a wide range of event types.

In addition, three key aspects were identified as enabling the framework to be more widely applicable than existing frameworks, which are typically only applied to mega-events or large-scale events. The first was to constrain the timeframe to the year in which the event was hosted, enabling annually recurrent events to be included within the framework without their outcomes being confounded across years. The second was to remove the need or expectation for an event to leave an enduring or structural change on its host, allowing the inclusion of smaller events that reuse existing assets. The third element was to adopt a zero-based approach to event outcomes. This means the framework should not be tailored if an outcome area does not apply to an event. Instead, it should remain standardised, with no (zero) outcome recorded for that area. In combination, this has enabled a framework design that is more widely applicable to multiple events of different scales and types hence avoiding the creation of tailored and non-comparable versions.

Identifying the key areas of impact

The framework established in this study comprised a multi-level structure consisting of pillars, drivers, and measures. The structure allowed for the standardising of the areas of impact to be assessed but still provided for future adaptions if measures of event outcomes change or emerge.

The six pillars established - economic, sport, media, social, brand, and environment - captured the breadth of dimensions that are relevant to event hosts and stakeholders in the bidding and delivery of a wide range of events. The use of the multi-level structure enabled each of the six areas to capture a wider range of outcomes (through the drivers contained within each) than a single level framework would allow. The approach allowed well-established areas of impact such as tourism, infrastructure, and audience size to be represented, but not dominate or exclude other areas, as would have been the case had they been represented at the pillar level. A seventh area of political or ‘soft-power’ was reviewed but excluded due to the subjective nature of the outcomes and the lack of applicability to events below the global level (and which already dominate existing frameworks).

While developed independently, the resulting alignment of the six pillar areas to the triple bottom line model of event assessment confirms the breadth of scope fits with the existing

focus of the contribution of events. However, it also brings two additional aspects. The first is the ability to surface any potential shift in weight from tangible to intangible value in the economic area of the triple bottom line by capturing economic, media, and brand dimensions of value separately. The second is to contribute explicit weightings to three areas of the triple bottom line that might otherwise be assumed to be equivalent.

Establishing weightings for different areas of event impact

The results of the Delphi study established a stable consensus on the weightings of each area of event outcomes. That stable opinion was evident for both the six pillars and at the more detailed level for the 30 drivers included within the pillars. Indeed, the ability of the framework to provide a stable opinion on the weightings from only a single survey round, and confirmed by a further round, suggested that the level of specificity in the framework design and definitions of areas of impact assisted with obtaining more consistent agreement.

Rather than the dominance of a single dimension, the weightings at the top level of the framework established a balanced profile across the six pillars, grouped into three tiers of similar weights. The pillars of Economic and Sport outcomes occupied the top tier of the outcomes from events, with the second tier comprised of Media, Social and Brand outcomes. The lowest tier contained the final pillar of Environment. The weightings of the pillars and the grouping in the tiers were consistent across the panel and within subgroups of the panel, establishing a solid basis on which those weights could be confidently applied within a standardised framework.

Important insights arose in testing the allocation of weightings that have implications for hosts assessing whether to initiate or retain events. The largest disparity in the allocation of weightings was from the bias with which respondents allocated greater weight to their own areas of expertise. Whilst not surprising, it was only in being able to test and surface this explicitly that cities can recognise the need to ensure a diversity of views within their assessment panels and working groups. The second strong area of bias was from academic researchers, who are often sought independent perspectives from hosts and event owners. A higher level of collaboration between academics and industry is needed to ensure academic research remains rigorous but more closely grounded in resolving the issues and challenges the industry faces.

In keeping with the aim of a specific and standard approach, weightings for each of the drivers was established within their relevant pillar. Reflecting the results of the pillar weights, a strongly stable opinion was formed for the importance of the 30 drivers by the panel in the first round of weightings and confirmed by the second round.

As an approach for comparison across events, the testing of the allocation of pillar weights to events of different scales suggested that the pillars could be successfully applied and that the weightings reflected the significant differences in the event prospective outcomes expected based on the typology descriptions. The research process has thus tested and established an explicit and stable weighting for two levels of event outcomes grounded in a framework that was not previously available.

Identification of measures for assessment of event outcomes

The absence of a comprehensive set of specific measures linked to areas of event assessment was noted in the literature. Drawing from the individual areas in the academic literature and industry practice, the list of potential measures collated was refined to those applicable across events of different scales and types. The rating of measures on both their importance and their reliability established the need to capture both dimensions in the

selection of measures of event outcomes. The use of 'Impact Score' as a combined score of a measure's contribution to event assessment established the basis for selecting measures that are highly relevant and directly measurable but might be otherwise overlooked if only considering the single dimension of perceived importance. The combined construct adds rigour to an area of event measurement that was under-represented in the literature on event framework.

The fact that all measures were rated as more important than reliable for the Brand and Social pillars confirms the need to build greater consistency and accuracy in the measurement of these areas.

Benefits of a comparative framework to host cities

Recognising that very few cities will ever host a mega-event, the value of a standardised framework such as that established in this study is in the ability to assess events of differing scales and types for those that best fit the host's long term development plans and capabilities. This research has already made explicit the consensus weightings and evidence of biases in the perceived importance of areas of event outcomes. Host cities can reflect on those two insights and consider the target weightings they have placed on their existing portfolio of events, and how that might differ from the allocation of weightings found in this study. The process of reviewing and surfacing different perspectives within a diverse panel can make explicit hidden assumptions that might hamper resourcing to achieve desired outcomes or create a misalignment in expectations between stakeholders.

Perhaps the most practical and valuable outcome of a standardised framework that enables comparison of events across editions and hosts is the grounding of expectations in evidence, thus avoiding overcommitting of public resources for poor returns. By creating greater efficiency in the market for events, hosts may be able to discover events that are undervalued but fit their capacities, and in doing so, better use public funds and existing assets through a more targeted portfolio.

9 Reference List

AASB. (2018). *Australian Accounting Standards Board*. Australian Accounting Standards Board. Retrieved September, 2018 from Australian Accounting Standards

ABS. (2016). *Australian Census* http://www.abs.gov.au/websitedbs/D3310114.nsf/Home/Census

Acquier, A., Daudigeos, T., & Pinkse, J. (2017, 2017/12/01/). Promises and paradoxes of the sharing economy: An organizing framework. *Technological Forecasting and Social Change, 125*, 1-10. https://doi.org/https://doi.org/10.1016/j.techfore.2017.07.006

Aengenheyster, S., Cuhls, K., Gerhold, L., Heiskanen-Schüttler, M., Huck, J., & Muszynska, M. (2017, 05/01/May 2017). Real-Time Delphi in practice — A comparative analysis of existing software-based tools. *Technological Forecasting & Social Change, 118*, 15-27. https://doi.org/10.1016/j.techfore.2017.01.023

Agha, N., Fairley, S., & Gibson, H. (2012, 2012/02/01/). Considering legacy as a multi-dimensional construct: The legacy of the Olympic Games. *Sport Management Review, 15*(1), 125-139. https://doi.org/https://doi.org/10.1016/j.smr.2011.08.004

Agha, N., & Rascher, D. A. (2016, 2016/05/09). An explanation of economic impact: why positive impacts can exist for smaller sports. *Sport, Business and Management: An International Journal, 6*(2), 182-204. https://doi.org/10.1108/SBM-07-2013-0020

Agha, N., & Taks, M. (2015). A Theoretical Comparison of the Economic Impact of Large and Small Events. *International Journal of Sport Finance, 10*(3), 199-216.

Allen, D. (2013). 'The successes and challenges of hosting the 2010 FIFA World Cup': the case of Cape Town, South Africa. *Soccer & Society, 14*(3), 404-415. https://doi.org/10.1080/14660970.2013.801268

Allen, D., Knott, B., & Swart, K. (2013). 'Africa's Tournament'? The Branding Legacy of the 2010 FIFA World Cup™. *The International Journal of the History of Sport, 30*(16), 1994-2006. https://doi.org/10.1080/09523367.2013.861131

Allen, J., O'Toole, W., Harris, R., & McDonnell, I. (2008). *Festival and Special Event Management*. Wiley.

Allen, J., O'Toole, W., Harris, R., & McDonnell, I. (2010). *Festival and Special Event Management, 5th Edition*. Wiley. http://ebookcentral.proquest.com/lib/vu/detail.action?docID=5017568

Allmers, S., & Maennig, W. (2009). Economic impacts of the FIFA Soccer World Cups in France 1998, Germany 2006, and outlook for South Africa 2010. *Eastern Economic Journal, 35*, 500-519. https://doi.org/DOI: 10.1057/eej.2009.30

Alm, J., Solberg, H. A., Storm, R. K., & Jakobsen, T. G. (2016). Hosting major sports events: the challenge of taming white elephants. *Leisure Studies, 35*(5), 564-582.

Ammon, R., Southall, R. M., & Nagel, M. S. (2016). *Sport facility management: Organizing events and mitigating risks* (3rd ed.). Fitness Information Technology.

Arcodia, C., & Whitford, M. (2006). Festival Attendance and the Development of Social Capital. *Journal of Convention & Event Tourism, 8*(2), 1-18. https://doi.org/10.1300/J452v08n02_01

Armstrong, M. J. (1995). Joint reliability-importance of components. *IEEE Transactions on Reliability, 44*(3), 408-412.

Arup. (2017). *Cities Alive - Rethinking legacy for host cities*. Arup. www.arup.com/citiesalive

ASOIF. (2014, 05 Dec 2011). Now Tokyo 2020 Says It Will Slash Budget after Spending $150m on 2016 Bid. http://www.asoif.com/News/News_Article.aspx?ID=1467

Baade, R. A., & Matheson, V. A. (2004). The Quest for the Cup: Assessing the Economic Impact of the World Cup. *Regional Studies, 38*(4), 343-354. https://doi.org/10.1080/03434002000213888

Bajc, V. (2007). Introduction:Debating Surveillance in the Age of Security. *American Behavioral Scientist, 50*(12), 1567-1591. https://doi.org/10.1177/0002764207305148

Bakardjieva, M., Svensson, J., & Skoric, M. (2012). Digital citizenship and activism: Questions of power and participation online.

Barajas, A., Coates, D., & Sanchez-Fernandez, P. (2016). Beyond retrospective assessment. Sport event economic impact studies as a management tool for informing event organization. *European Research on Management and Business Economics, 22*(3), 124-130. https://doi.org/10.1016/j.iedee.2015.05.001

Barney, R. K. (2003). *The Olympic legacy of wealth: A double edged sword* (M. d. Moragas Spà, Ed. Vol. The legacy of the Olympic Games, 1984 - 2000 : International Symposium, Lausanne). International Olympic Committee.

Bek, D., Merendino, A., Swart, K., & Timms, J. (2018). Creating an enduring developmental legacy from FIFA2010: the Football Foundation of South Africa (FFSA). *European Sport Management Quarterly, 19*(4), 437-455. https://doi.org/10.1080/16184742.2018.1535608

Bertella, G. (2014). Designing small-scale sport events in the countryside. *International Journal of Event and Festival Management*.

Bob, H., Henry, W., Adam, J., Tim, B., Xiaoyan, X., Juan Luis Paramio, S., Masayuki, Y., & Inge, D. (2019). Inducing Destination Images Among International Audiences: The Differing Effects of Promoting Sport Events on the Destination Image of a City Around the World. *Journal of Sport Management*, 1-12. https://doi.org/10.1123/jsm.2018-0101

Bob, U., & Kassens-Noor, E. V. A. (2012). An indicator framework to assess the legacy impacts of the 2010 FIFA World Cup. *African Journal for Physical, Health Education, Recreation & Dance*, 12-21.

Bob, U., & Swart, K. (2010). Sport events and social legacies.

Borders, W. (1974, Jan. 19, 1974). Montreal's Olympic Market: The World. *The New York Times*. https://www.nytimes.com/1974/01/19/archives/montreals-olympic-market-the-world-the-world-montreals-olympic.html

Boykoff, J., & Mascarenhas, G. (2016, 2016/04/02). The Olympics, Sustainability, and Greenwashing: The Rio 2016 Summer Games. *Capitalism Nature Socialism, 27*(2), 1-11. https://doi.org/10.1080/10455752.2016.1179473

Bradley, R. A., Thomas, D., Freeman, R. E., Michael, C. J., Ronald, K. M., & Donna, J. W. (2008). Dialogue: Toward Superior Stakeholder Theory. *Business Ethics Quarterly*(2), 153.

Brittain, I., Bocarro, J., Byers, T., & Swart, K. (2017). *Legacies and Mega Events : Fact or Fairy Tales?* Milton : Taylor and Francis, 2017.
1st ed.

Brownlee, E., Greenwell, T. C., & Moorman, A. (2018). An Experimental Approach to Assessing the Effectiveness of Official Sponsor Designations in an Ambush Marketing Scenario. *Sport MArketing Quarterly, 27*(3).

Burns, J., Hatch, J., & Mules, T. (1986). *The Adelaide Grand Prix: The Impact of a Special Event* (19881850907).

Business Research. (2021). *Sports Market 2021*. T. B. R. Company. https://www.thebusinessresearchcompany.com/report/sports-market

Byers, T., Hayday, E., & Pappous, A. (2021). A new conceptualization of mega sports event legacy delivery: Wicked problems and critical realist solution. *Sport Management Review, 23*(2), 171-182. https://doi.org/10.1016/j.smr.2019.04.001

Byrne, C. (2014). Relationship of convenience? The diplomatic interplay between the Commonwealth Games Federation and the Commonwealth Games host city. *Sport in Society, 17*(9), 1204-1219.

Cahill, D., Wee, M., & Yang, J. W. (2017). Media sentiment and trading strategies of different types of traders. *Pacific-Basin Finance Journal, 44*, 160-172. https://doi.org/10.1016/j.pacfin.2017.07.001

Campbell, S. (2021, 04/01/). Regenerative-led design driving sustainability. *Australasian Leisure Management, 142*(142). https://doi.org/10.3316/informit.647495769848902

Carneiro, M. J., Eusébio, C., & Caldeira, A. (2018). The influence of social contact in residents' perceptions of the tourism impact on their quality of life: A structural equation model. *Journal of Quality Assurance in Hospitality & Tourism, 19*(1), 1-30.

Carson, R. T. (2012). Contingent Valuation: A Practical Alternative when Prices Aren't Available. *Journal of Economic Perspectives, 26*(4), 27-42. https://doi.org/10.1257/jep.26.4.27

Cashman, R. (2003). *The legacy of the Olympic Games: 1984-2000: International Symposium, Lausanne, 14th, 15th and 16th November 2002*. International Olympic Committee.

Casswell, S., & Maxwell, A. (2005). Regulation of Alchohol Marketing: A Global View. *Journal of Public Health Policy, 26*(3), 343-358.

CGF. (2010). *Legacy of Delhi 2010*. Commonwealth Games Federation. Retrieved 19 October 2019 from http://d2010.thecgf.com/legacy_delhi_2010

CGF. (2011). *2018 Commonwealth Games Candidature, Report of the CGF Evaluation Commission*. https://thecgf.com/sites/default/files/2018-03/2018_CGF_Evaluation_Commission_Report.pdf

Chan, G. R. Y. C. (2008, 2008/01/01). Aligning collections budget with program priorities. *Library Collections, Acquisitions, & Technical Services, 32*(1), 46-52. https://doi.org/10.1080/14649055.2008.10766192

Chappelet, J.-L. (2012). Mega sporting event legacies: A multifaceted concept. *Papeles de Europa, 25*, 76.

Chappelet, J.-L. (2018). Challenges and opportunities for past and future Olympic cities [Working Paper]. *Université de Lausanne*. https://serval.unil.ch/resource/serval:BIB_29F28F9E5192.P001/REF

Chavanat, N., Martinent, G., & Ferrand, A. (2009). Sponsor and Sponsees Interactions: Effects on Consumers' Perceptions of Brand Image, Brand Attachment, and Purchasing Intention. *Journal of Sport Management, 23*, 644-670.

Chen, S. (2013). *An evaluation of the sub-regional legacy/impacts of the London 2012 Olympic Games in a non-hosting sub-region : a case study of Leicestershire* Loughborough University]. edsble.

Chen, Y., Qu, L. E. I., & Spaans, M. (2013). Framing the Long-Term Impact of Mega-Event Strategies on the Development of Olympic Host Cities. *Planning Practice and Research, 28*(3), 340-359. https://doi.org/10.1080/02697459.2013.750072

Chhabra, D., & Gursoy, D. (2009). Life dynamism explorations on perceived quality of life and social exchange paradigms in casino settings. *Leisure Sciences, 31*(2), 136-157.

Chiu, W., Won, D., & Kim, S. (2019). Extended Model of Sport Spectator Goal-directed Behavior: The Role of Event Prestige in Nonmajor Sport Events. *Event Management, 23*(1), 119-133. https://doi.org/10.3727/152599518x15378845225393

Choi, N., & Majumdar, S. (2014, 2014/05/01/). Social entrepreneurship as an essentially contested concept: Opening a new avenue for systematic future research. *Journal of Business Venturing, 29*(3), 363-376. https://doi.org/https://doi.org/10.1016/j.jbusvent.2013.05.001

Clarkson, M. E. (1995). A stakeholder framework for analyzing and evaluating corporate social performance. *Academy of Management Review, 20*(1), 92-117. https://doi.org/10.5465/AMR.1995.9503271994

Climate Council. (2021). *Brisbane Olympics 2032: the first climate positive Olympics*. Retrieved 9th March 2021 from https://www.climatecouncil.org.au/climate-positive-brisbane-olympics-2032/

Coates, D., & Humphreys, B. R. (1999). The growth effects of sport franchises, stadia, and arenas. *Journal of Policy Analysis and Management, 18*(4), 601-624.

Collier, D., Daniel Hidalgo, F., & Olivia Maciuceanu, A. (2006, 2006/10/01). Essentially contested concepts: Debates and applications. *Journal of Political Ideologies, 11*(3), 211-246. https://doi.org/10.1080/13569310600923782

Collins, A., & Cooper, C. (2016). Measuring and managing the environmental impact of festivals: the contribution of the Ecological Footprint. *Journal of Sustainable Tourism, 25*(1), 148-162. https://doi.org/10.1080/09669582.2016.1189922

Collins, A., & Flynn, A. (2008). Measuring the environmental sustainability of a major sporting event, A case study of the FA Cup Final. *Tourism Economics, 14*(4), 751-768.

Collins, A., Flynn, A., Munday, M., & Roberts, A. (2007). Assessing the Environmental Consequences of Major Sporting Events: The 2003/04 FA Cup Final. *Urban Studies, 44*(3), 457-476.

Comrey, A. L. (1950). A proposed method for absolute ratio scaling. *Psychometrika, 15*(3), 317-325.

Cornelissen, S. (2008, 2008/07/01). Scripting the nation: sport, mega-events, foreign policy and state-building in post-apartheid South Africa. *Sport in Society, 11*(4), 481-493. https://doi.org/10.1080/17430430802019458

Cornelissen, S. (2011). Mega Event Securitisation in a Third World Setting. *Urban Studies, 48*(15), 3221-3240. https://doi.org/10.1177/0042098011422392

Cornelissen, S., Bob, U., & Swart, K. (2011). Towards redefining the concept of legacy in relation to sport mega-events: Insights from the 2010 FIFA World Cup. *Development Southern Africa, 28*(3), 307-318. https://doi.org/10.1080/0376835x.2011.595990

Cornwall, T. B. (2008). State of Art and Science in Sponsorship-Linked Marketing. *Journal of Advertising, 37*(3), 41-55. https://doi.org/10.2753/JOA0091-3367370304

Crompton, J. L. (2006). Economic Impact Studies: Instruments for Political Shenanigans? *Journal of Travel Research, 45*(1), 67-82. https://doi.org/10.1177/0047287506288870

Czinkota, M. R., & Ronkainen, I. A. (2005). A forecast of globalization, international business and trade: report from a Delphi study. *Journal of World Business, 40*(2), 111-123.

Daniels, M. J., Norman, W. C., & Henry, M. S. (2004). Estimating Income Effects of a Sport Tourism Event. *Annals of Tourism Research, 31*(1), 180-199. https://doi.org/10.1016/j.annals.2003.10.002

Davies, L., Coleman, R., & Ramchandani, G. (2013). Evaluating event economic impact: rigour versus reality? *International Journal of Event and Festival Management, 4*(1), 31-42. https://doi.org/10.1108/17582951311307494

de Nooij, M., & van den Berg, M. (2018). The bidding paradox: Why politicians favor hosting mega sports events despite the bleak economic prospects. *Journal of Sport & Social Issues, 42*(1), 68-92. https://doi.org/10.1177/0193723517748552

Death, C. (2011). 'Greening' the 2010 FIFA World Cup: Environmental Sustainability and the Mega-Event in South Africa. *Journal of Environmental Policy & Planning, 13*(2), 99-117. https://doi.org/10.1080/1523908x.2011.572656

Delamere, T. A., Wankel, L. M., & Hinch, T. D. (2001). Development Of A Scale To Measure Resident Attitudes Toward The Social Impacts Of Community Festivals, Part I: Item Generation And Purification Of The Measure. *7*, 11.

Deneau, L. (2016). *Implementation of the Mayor's Advisory Panel recommendations for future "mega" international event bidding and hosting in the Toronto region* (AFS 22897).

Dickson, T. J., Benson, A. M., & Blackman, D. A. (2011, 2011/11/01). Developing a framework for evaluating Olympic and Paralympic legacies. *Journal of Sport & Tourism, 16*(4), 285-302. https://doi.org/10.1080/14775085.2011.635014

Dinnie, K. (2004). Country-of-origin 1965-2004: A literature review. *Journal of Customer Behaviour, 3*(2), 165-213.

Donaldson, T., & Preston, L. E. (1995). The Stakeholder Theory of the Corporation: Concepts, Evidence, and Implications. *The Academy of Management Review, 20*(1), 65-91. https://doi.org/10.2307/258887

Donlan, L., & Crowther, P. (2014, 2014/07/04). Leveraging sponsorship to achieve consumer relationship objectives through the creation of 'marketing spaces': An exploratory study. *Journal of Marketing Communications, 20*(4), 291-306. https://doi.org/10.1080/13527266.2012.684068

Doyle, J. P., Filo, K., McDonald, H., & Funk, D. C. (2013, 8/1/August 2013). Exploring sport brand double jeopardy: The link between team market share and attitudinal loyalty. *Sport Management Review, 16*, 285-297. https://doi.org/10.1016/j.smr.2012.11.001

Drummond, R., & Cronje, J. (2018). Building a white elephant? The case of the Cape Town Stadium. *International Journal of Sport Policy and Politics, 11*(1), 57-78. https://doi.org/10.1080/19406940.2018.1508053

Dubinsky, Y. (2019, September 01). From soft power to sports diplomacy: a theoretical and conceptual discussion. *Place Branding and Public Diplomacy, 15*(3), 156-164. https://doi.org/10.1057/s41254-019-00116-8

Dudek, F. J., & Baker, K. E. (1956). The Constant-Sum Method Applied to Scaling Subjective Dimensions. *The American Journal of Psychology, 69*(4), 616-624. https://doi.org/10.2307/1419084

Dwyer, L., Forsyth, P., & Spurr, R. (2005). Estimating the Impacts of Special Events on an Economy. *Journal of Travel Research, 43*, 351-359. https://doi.org/DOI: 10.1177/0047287505274648

Dwyer, L., Mellor, R., Mistilis, N., & Mules, T. (2000). A framework for assessing 'tangible' and 'intangible' impacts of events and conventions. *Event Management, 6*(3), 175-189.

Edwards, T., & Lazzara, C. J. (2019). Sector Effects in the S&P 500® The Role of Sectors in Risk, Pricing, and Active Returns. *Indexology Magazine*. https://www.spindices.com/indexology/sectors/sector-effects-in-the-sp-500

Eick, V. (2010). A neoliberal sports event? FIFA from the Estadio Nacional to the fan mile. *City*(3), 278. https://doi.org/10.1080/13604813.2010.482275

Essex, S. J., & Chalkley, B. S. (2003a). *The infrastructural Legacy of the Summer and Winter Olympic Games: A Comparative Analysis* (M. d. Moragas Spà, Ed. Vol. The legacy of the Olympic Games, 1984 - 2000 : International Symposium, Lausanne). International Olympic Committee.

Essex, S. J., & Chalkley, B. S. (2003b). Urban transformation from hosting the Olympic Games University lecture on the Olympics. University lecture on the Olympics [online article]. *Barcelona : Centre d'Estudis Olímpics (UAB). International Chair in Olympism (IOC-UAB)*. <http://olympicstudies.uab.es/lectures/web/pdf/essex.pdf.>

Evnine, S. J. (2014, 01//). Essentially Contested Concepts and Semantic Externalism. *Journal of the Philosophy of History, 8*(1), 118-140. https://doi.org/10.1163/18722636-12341268

Faden, R. R., Beauchamp, T. L., & King, N. M. P. (1986). *A history and theory of informed consent*. New York : Oxford University Press, 1986.

Fairley, S., Lovegrove, H., & Brown, M. (2016). Leveraging events to ensure enduring benefits: The legacy strategy of the 2015 AFC Asian Cup. *Sport Management Review, 19*(4), 466-474. https://doi.org/10.1016/j.smr.2016.01.003

Fama, E. F. (1970). Efficient Capital Markets: A Review Of Theory And Empirical Work. *Journal of Finance, 25*(2), 383-417.

FGR. (2021). *A truly sustainable football club*. Forest Green Rovers Football Club. https://www.fgr.co.uk/another-way

Finkelstein, S., & Hambrick, D. C. (1990). Top-management-team tenure and organizational outcomes: The moderating role of managerial discretion. *Administrative science quarterly*, 484-503.

Finn, A., & Louviere, J. J. (1992). Determining the appropriate response to evidence of public concern: the case of food safety. *Journal of Public Policy & Marketing, 11*(2), 12-25.

Fisher, R. A., & Yates, F. (1963). Statistical tables for Biological, Agricultural and Medical Research.

Florian, M. (2000). Das Ladenburger „TeleDelphi ": Nutzung des Internets für eine Expertenbefragung. In *Die Delphi-Technik in den Sozialwissenschaften* (pp. 195-215). Springer.

Florida, R. (2003). Cities and the creative class. *City & community, 2*(1), 3-19.

Flynn, T. N., & Marley, A. A. (2014). Best-worst scaling: theory and methods. In *Handbook of choice modelling*. Edward Elgar Publishing.

Flyvbjerg, B., Stewart, A., & Budzier, A. (2016, 07/14/). The Oxford Olympics Study 2016: Cost and Cost Overrun at the Games.

Fourie, J., & Santana-Gallego, M. (2011). The impact of mega-sport events on tourist arrivals. *Tourism Management, 32*(6), 1364-1370. https://doi.org/DOI: 10.1016/j.tourman.2011.01.011

Frawley, S., & Van den Hoven, P. (2014). Football participation legacy and Australia's qualification for the 2006 Football World Cup. *Soccer & Society*, 1-11. https://doi.org/10.1080/14660970.2014.882817

Fredline, E. (2005, 2005/06/01). Host and Guest Relations and Sport Tourism. *Sport in Society, 8*(2), 263-279. https://doi.org/10.1080/17430430500087328

Fredline, L., Jago, L., & Deery, M. (2003). The development of a generic scale to measure the social impacts of events. *Event Management, 8*(1), 23-37. https://doi.org/10.3727/152599503108751676 (The 'Events and Place Making' conference)

Freeman, R. E., & McVea, J. (2001). A Stakeholder Approach to Strategic Management. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.263511

Freeman, R. E., Phillips, R., & Sisodia, R. (2020). Tensions in stakeholder theory. *Business & Society, 59*(2), 213-231.

Fujak, H., Frawley, S., Lock, D., & Adair, D. (2020). Consumer behaviour toward a new league and teams: television audiences as a measure of market acceptance. *European Sport Management Quarterly*, 1-21. https://doi.org/10.1080/16184742.2020.1770310

Fujak, H., Frawley, S., & Morgan, A. (2016). Broadcasting sport mega-events. In *Managing Sport Mega-Events* (pp. 103-118). Routledge.

Gallie, W. B. (1956). Essentially Contested Concepts. *Proceedings of the Aristotelian Society, Vol. 56*, 167. http://www.jstor.org/stable/4544562

Gastwirth, J. L., Gel, Y. R., & Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. *Statistical Science, 24*(3), 343-360.

Geeraert, A. (2016). Theorizing the governance of sport mega-events: a Principal–Agent perspective. In *Managing sport mega-events* (pp. 38-50). Routledge.

Gehman, J., Glaser, V. L., Eisenhardt, K. M., Gioia, D., Langley, A., & Corley, K. G. (2017). Finding Theory–Method Fit: A Comparison of Three Qualitative Approaches to Theory Building. *Journal of Management Inquiry, 27*(3), 284-300. https://doi.org/https://doi-org.wallaby.vu.edu.au:4433/10.1177/1056492617706029

Geland, A. (2003). Local Economic Impacts The British Open. *Annals of Tourism Research, 30*(2), 406-425. https://doi.org/doi:10.1016/S0160-7383(02)00098-1

George, P. (2009). Sport in Disrepute. *Australian and New Zealand Sports Law Journal, 4*(1), 24-54. https://doi.org/10.3316/informit.395653493310639

Getz, D. (1991). *Festivals, special events, and tourism*. Van Nostrand Reinhold.

Getz, D. (2003). Bidding on Events: Identifying Event Selection Criteria and Critical Success Factors. *Journal of Convention & Exhibition Management, 5*(2), 274. https://doi.org/http://dx.doi.org/10.1300/J143v05n02_01

Getz, D., MacDonald, D., & Parent, M. M. (2015). The sport event owners' perspective. In *Routledge handbook of sports event management* (pp. 91-108). Routledge.

Giannarou, L., & Zervas, E. (2014). Using Delphi technique to build consensus in practice. *Int. Journal of Business Science and Applied Management, 9*(2).

Gibson, H. J., Walker, M., Thapa, B., Kaplanidou, K., Geldenhuys, S., & Coetzee, W. (2014). Psychic income and social capital among host nation residents: A pre–post analysis of the 2010 FIFA World Cup in South Africa. *Tourism Management, 44*, 113-122. https://doi.org/10.1016/j.tourman.2013.12.013

Gioia, D. A., & Chittipeddi, K. (1991, 01 / 01 /). Sensemaking and sensegiving in strategic change initiation. *Strategic Management Journal, 12*(6), 433-448. https://doi.org/10.1002/smj.4250120604

Girgin, F. A., & Tasci, O. E. (2019). Mega-Event Organization Considering Safety, Security and Resilience. *TeMA-Journal of Land Use, Mobility and Environment, 12*(3), 249-264.

Girginov, V. (2012). Governance of the London 2012 Olympic Games legacy. *INTERNATIONAL REVIEW FOR THE SOCIOLOGY OF SPORT, 47*(5), 543-558. https://doi.org/10.1177/1012690211413966

Girginov, V., & Hills, L. (2008). A Sustainable Sports Legacy: Creating a Link between the London Olympics and Sports Participation. *International Journal of the History of Sport, 25*(14), 2091-2116.

Glaser, B., & Strauss, A. (1967). Grounded theory: The discovery of grounded theory. *Sociology the journal of the British sociological association, 12*, 27-49.

Goetzmann William, N. (1995). The Informational Efficiency of the Art Market. *Managerial Finance, 21*(6), 25-34. https://doi.org/10.1108/eb018522

Gold, J. R., & Gold, M. M. (2010). *Olympic Cities : City Agendas, Planning, and the World's Games, 1896 – 2016*. https://doi.org/10.4324/9780203840740

Gold, J. R., & Gold, M. M. (2013). "Bring It under the Legacy Umbrella": Olympic Host Cities and the Changing Fortunes of the Sustainability Agenda. *Sustainability, 5*, 326-3542. https://doi.org/doi:10.3390/su5083526

Gordon, T., & Pease, A. (2006). RT Delphi: An efficient, "round-less" almost real time Delphi method. *Technological Forecasting and Social Change, 73*(4), 321-333. https://doi.org/10.1016/j.techfore.2005.09.005

Gough, C. (2021). *Average regular season attendance of selected sports leagues worldwide in 2019*. Statista. Retrieved 29 November 2021 from https://www.statista.com/statistics/1119911/average-sports-attendance-by-league/

Graeff, B., Monteiro Gutierrez, D., Sardá, T., Bretherton, P., & Bettine, M. (2019). Capable, splendorous and unequal: international media portrayals of Brazil during the 2014 World Cup. *Third World Quarterly, 40*(4), 796-814.

Grant Thornton. (2010). *Updated economic impact of the 2010 FIFA World Cup*.

Gratton, C., Dobson, N., & Shibli, S. (2000). The economic importance of major sports events: a case-study of six events. *Managing Leisure, 5*(1), 17-28.

Gratton, C., & Preuss, H. (2008). Maximizing Olympic Impacts by Building Up Legacies. *The International Journal of the History of Sport, 25*(14), 1922-1938. https://doi.org/http://dx.doi.org/10.1080/09523360802439023

Gratton, C., Shibli, S., & Coleman, R. (2006a). The economic impact of major sports events: a review of ten events in the UK. *Sociological Review, 54*(2), 41-58. https://doi.org/DOI: 10.1111/j.1467-954X.2006.00652.x

Gratton, C., Shibli, S., & Coleman, R. (2006b). The economic impact of major sports events: a review of ten events in the UK. *Sociological Review, 54*, 41-58. https://doi.org/10.1111/j.1467-954X.2006.00652.x

Greatorex, J., & Dexter, T. (2000). An accessible analytical approach for investigating what happens between the rounds of a Delphi study. *Journal of Advanced Nursing, 32*(4), 1016-1024. https://doi.org/https://doi.org/10.1046/j.1365-2648.2000.t01-1-01569.x

Grix, J. (2012). 'Image' leveraging and sports mega-events: Germany and the 2006 FIFA World Cup. *Journal of Sport & Tourism, 17*(4), 289-312. https://doi.org/10.1080/14775085.2012.760934

Grix, J., & Houlihan, B. (2014). Sports Mega-Events as Part of a Nation's Soft Power Strategy: The Cases of Germany (2006) and the UK (2012). *British Journal of Politics & International Relations, 16*(4), 572-596. https://doi.org/10.1111/1467-856X.12017

Grupp, H., Blind, K., & Cuhls, K. (2000). Analyse von Meinungsdisparitäten in der Technikbewertung mit der Delphi-Methode. In *Die Delphi-Technik in den Sozialwissenschaften* (pp. 43-65). Springer.

Häder, M., & Häder, S. (2000). *Die Delphi-Technik in den Sozialwissenschaften*. Springer.

Halliwell, M., & Freeman, M. B. (2014). *Social Media Influence On Viewer Engagement* 25th Australasian Conference on Information Systems, 8th -10th Dec 2014, Auckland, New Zealand. https://openrepository.aut.ac.nz/bitstream/handle/10292/8050/acis20140_submission_262.pdf;sequence=1

Harris, N. (2007, 1st March 2007). Why Fifa's claim of one billion TV viewers was a quarter right. *The Independent*. https://www.independent.co.uk/sport/football/news/why-fifa-s-claim-of-one-billion-tv-viewers-was-a-quarter-right-5332287.html

Hasson, F., Keeney, S., & McKenna, H. (2000, Oct). Research guidelines for the Delphi survey technique. *J Adv Nurs, 32*(4), 1008-1015. https://doi.org/10.1046/j.1365-2648.2000.t01-1-01567.x

Hayes, G., & Karamichas, J. (2012). Introduction: Sports Mega-Events, Sustainable Development and Civil Societies. In G. Hayes & J. Karamichas (Eds.), *Olympic Games, Mega-Events and Civil Societies: Globalization, Environment, Resistance* (pp. 1-27). Palgrave Macmillan UK. https://doi.org/10.1057/9780230359185_1

Hellings, J., Lass, M., Apted, J., & Mead, I. (2011). Delivering London 2012: geotechnical enabling works. *Civil Engineering, 164*, 5-10. https://doi.org/10.1680/cien.2011.164.6.5

Helmer, O., & Rescher, N. (1959). On The Epistemology Of The Inexact Sciences. *Management Science, 6*(1), 25-52.

Heslop, L. A., Nadeau, J., O'Reilly, N., & Armenakyan, A. (2013). Mega-event and Country Co-branding: Image Shifts, Transfers and Reputational Impacts. *Corporate Reputation Review, 16*(1), 7-33. https://doi.org/10.1057/crr.2012.23

Hiller, H. H. (2000). Mega-events, Urban Boosterism and Growth Strategies: An Analysis of the Objectives and Legitimations of the Cape Town 2004 Olympic Bid. *International Journal of Urban and Regional Research, 24*(2), 449-458. https://doi.org/10.1111/1468-2427.00256

Holey, E. A., Feeley, J. L., Dixon, J., & Whittaker, V. J. (2007). An exploration of the use of simple statistics to measure consensus and stability in Delphi studies. *BMC medical research methodology, 7*(1), 52-61.

Homma, K., & Masumoto, N. (2013, 2013/06/01). A Theoretical Approach for the Olympic Legacy Study Focusing on Sustainable Sport Legacy. *The International Journal of the History of Sport, 30*(12), 1455-1471. https://doi.org/10.1080/09523367.2013.825251

Hood, M. (2016). The Purse Is Not Enough. *Journal of Sports Economics, 7*(3), 289-308. https://doi.org/10.1177/1527002504271356

Horne, J. (2007). The Four 'Knowns' of Sports Mega-Events. *Leisure Studies, 26*(1), 81-96. https://doi.org/10.1080/02614360500504628

Horne, J., & Manzenreiter, W. (2006). An introduction to the sociology of sports mega-events1. *The Sociological Review, 54*, 1-24. https://doi.org/10.1111/j.1467-954X.2006.00650.x

Hotchkiss, J. L., Moore, R. E., & Zobay, S. M. (2003). Impact of the 1996 summer olympic games on employment and wages in Georgia. *Southern Economic Journal, 69*(3), 691-704.

Huber, G. P., & Delbecq, A. (1972). Guidelines for combining the judgments of individual members in decision conferences. *Academy of Management Journal, 15*(2), 161-174.

Humphreys, B. R., & Prokopowicz, S. (2007). Assessing the impact of sports mega-events in transition economies: EURO 2012 in Poland and Ukraine. *International Journal of Sport Management and Marketing, 2*(5-6), 496-509. https://doi.org/DOI: 10.1504/IJSMM.2007.013963

IEG. (2016). Average Activation-To-Fee Ratio Passes Two-To-One Mark For First Time. Retrieved 16/02/2021, from https://www.sponsorship.com/iegsr/2016/12/19/Average-Activation-To-Fee-Ratio-Passes-Two-To-One-.aspx

IEG. (2018). *What sponsors want & where dollars will go in 2018.* http://www.sponsorship.com/IEG/ les/f3/f3cfac41-2983-49be-8df6-

IOC. (2006). *What is the Olympic Games Global Impact Study?* Internaitonal Olympic Committee. Retrieved 12/1/2016 from http://www.olympic.org/Documents/Reports/EN/en_report_1077.pdf

IOC. (2015). Olympic Games: legacies and impacts. Retrieved 2/2/2016, from https://stillmed.olympic.org/AssetsDocs/OSC%20Section/pdf/bibliography_on_the_olympic_games_legacy_and_impacts_.pdf

IOC. (2019). *Olympic Marketing Fact File 2019 Edition*. I. O. Committee. https://stillmed.olympic.org/media/Document%20Library/OlympicOrg/Documents/IOC-Marketing-and-Broadcasting-General-Files/Olympic-Marketing-Fact-File-2019.pdf

IOC. (2020). *Olympic Charter*. https://olympics.com/ioc/olympic-charter

IOC. (2021a). *Future Host Questionnaire – Olympic Games*. https://stillmed.olympics.com/media/Document%20Library/OlympicOrg/Games/Future-Olympic-Hosts/Future-Host-Questionnaire-Olympic-Games.pdf

IOC. (2021b). *Olympic Agenda 2020 : Closing report*. https://stillmed.olympics.com/media/Document%20Library/OlympicOrg/IOC/What-We-Do/Olympic-agenda/Olympic-Agenda-2020-Closing-report.pdf

ISO. (2012). *Event sustainability management systems*. https://www.iso.org/iso-20121-sustainable-events.html

Israel, M., & Hay, I. (2008). Informed Consent. https://doi.org/10.4135/9781412963909.n214

Jackson, S. J. (2013). The contested terrain of sport diplomacy in a globalizing world. *International Area Studies Review, 16*(3), 274-284. https://doi.org/10.1177/2233865913498867

Jago, L., Dwyer, L., Lipman, G., Lill, D. v., & Vorster, S. (2010). Optimising the potential of mega-events: an overview. *International Journal of Event and Festival Management, 1*(3), 220-237. https://doi.org/10.1108/17852951011078023

Jasmand, S., & Maenning, W. (2008). Regional income and employment effects of the 1972 Munich Summer Olympic Games. *Regional Studies, 42*(7), 991-1002. https://doi.org/DOI: 10.1080/00343400701654095

Jensen, M. C. (2002). Value Maximization, Stakeholder Theory, and the Corporate Objective Function. *Business Ethics Quarterly, 12*(2), 235-256. https://doi.org/10.2307/3857812

Jensen, M. C., & Meckling, W. H. (1976, 1976/10/01/). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*(4), 305-360. https://doi.org/https://doi.org/10.1016/0304-405X(76)90026-X

Jolly, R. (2013). *Commonwealth Games: friendly rivalry* (Parliament of Australia, Issue.

Jones, C. (2008). Assessing the impact of a major sporting event: the role of environmental accounting. *Tourism Economics, 14*(2), 343-360.

Jutbring, H. (2014). Encoding destination messages in media coverage of an international event: A case study of the European athletics indoor championships. *Journal of Destination Marketing & Management, 3*(1), 29-36. https://doi.org/10.1016/j.jdmm.2013.12.004

Kahneman, D., & Sibony, O. (2021). Sounding the alarm on system noise. *McKinsey Quarterly*(3), 1-7.

Kaplanidou, K., Karadakis, K., Gibson, H., Thapa, B., Walker, M., Geldenhuys, S., & Coetzee, W. (2013). Quality of Life, Event Impacts, and Mega-Event Support among South African Residents before and after the 2010 FIFA World Cup. *Journal of Travel Research, 52*(5), 631-645. https://doi.org/10.1177/0047287513478501

Karadakis, K., & Kaplanidou, K. (2012). Legacy perceptions among host and non-host Olympic Games residents: a longitudinal study of the 2010 Vancouver Olympic Games. *European Sport Management Quarterly, 12*(3), 243-264. https://doi.org/10.1080/16184742.2012.680067

Kasimati, E., & Dawson, P. (2009). Assessing the impact of the 2004 Olympic Games on the Greek economy: A small macroeconometric model. *Economic Modelling, 26*(1), 139-146. https://doi.org/DOI: 10.1016/j.econmod.2008.06.006

Kassens-Noor, E., Wilson, M., Müller, S., Maharaj, B., & Huntoon, L. (2015, 2015/11/02). Towards a mega-event legacy framework. *Leisure Studies, 34*(6), 665-671. https://doi.org/10.1080/02614367.2015.1035316

Kavetsos, G., & Szymanksi, S. (2009). From the Olympics to the grassroots: What will London 2012 mean for sport funding and participation in Britain? *Public policy research*.

Kelly, D. (2019). Essays on Event Leverage: The strategic use of event leverage in host destinations.

Kelly, D. M., Fairley, S., & O'Brien, D. (2019). It was never ours: Formalised event hosting rights and leverage. *Tourism Management, 73*, 123-133. https://doi.org/10.1016/j.tourman.2019.01.016

Kidd, B. (2003). The global sporting legacy of the Olympic movement. Legacy of the Olympic Games 1984–2000, International Symposium,

Kidd, B. (2008). A new social movement: Sport for development and peace. *Sport in Society, 11*(4), 370-380. https://doi.org/10.1080/17430430802019268

Kim, S., Liu, J. D., & Love, A. (2015). Event prestige as a mediator between sport involvement and intent to continue participation. *Perceptual & Motor Skills, 121*(2), 371-389.

Kim, Y., Lee, H.-W., Magnusen, M. J., & Kim, M. (2015). Factors influencing sponsorship effectiveness: A meta-analytic review and research synthesis. *Journal of Sport Management, 29*(4), 408-425.

Knott, B., Fyall, A., & Jones, I. (2013). The Nation-Branding Legacy of the 2010 FIFA World Cup for South Africa. *Journal of Hospitality Marketing & Management, 22*(6), 569-595. https://doi.org/10.1080/19368623.2012.663155

Knott, B., Fyall, A., & Jones, I. (2015, 3//). The nation branding opportunities provided by a sport mega-event: South Africa and the 2010 FIFA World Cup. *Journal of Destination Marketing & Management, 4*(1), 46-56. https://doi.org/http://dx.doi.org/10.1016/j.jdmm.2014.09.001

Koehler, D. J., Brenner, L., & Griffin, D. (2002). The calibration of expert judgment: Heuristics and biases beyond the laboratory. *Heuristics and biases: The psychology of intuitive judgment*, 686-715.

Koenigstorfer, J., Bocarro, J. N., Byers, T., Edwards, M. B., Jones, G. J., & Preuss, H. (2019). Mapping research on legacy of mega sporting events: structural changes, consequences, and stakeholder evaluations in empirical studies. *Leisure Studies, 38*(6), 729-745. https://doi.org/10.1080/02614367.2019.1662830

Korhonen, J., Nuur, C., Feldmann, A., & Birkie, S. E. (2018, 2018/02/20/). Circular economy as an essentially contested concept. *Journal of Cleaner Production, 175*, 544-552. https://doi.org/https://doi.org/10.1016/j.jclepro.2017.12.111

KPMG. (2006). Economic Impact Study of the Melbourne 2006 Commonwealth Games. Retrieved October 2006, from https://opus.lib.uts.edu.au/handle/10453/19802

KPMG. (2021). *Brisbane 2032 Olympic and Paralympic Games : Prelimiinary economic, social and environmental analysis*. https://www.premiers.qld.gov.au/publications/categories/reports/assets/2032-qld-games-economic-analysis-summary-report-final.pdf

Kushner, S. (2005). Confidentiality. *The SAGE encyclopedia of qualitative research methods*, 75-76.

Lacey, R., Sneath, J. Z., Finney, Z. R., & Close, A. G. (2007). The Impact of Repeat Attendance on Event Sponsorship Effects. *Journal of Marketing Communications, 13*(4), 243-255. https://doi.org/10.1080/13527260701250752

Lechner, E., & Solberg, H. A. (2021). The competition for government funding of major sports events: why do some applicants pass the needle's eye? *International Journal of Sport Policy and Politics, 13*(1), 13-27. https://doi.org/10.1080/19406940.2020.1859586

Lee, C.-K., & Taylor, T. (2005). Critical reflections on the economic impact assessment of a mega-event: the case of 2002 FIFA World Cup. *Tourism Management, 26*(4), 595-603.

LEED. (2021). *WHAT IS LEED?* Retrieved 23 July 2021 from http://leed.usgbc.org/leed.html

Leopkey, B., & Parent, M. M. (2012a). The (Neo) institutionalization of legacy and its sustainable governance within the Olympic Movement. *European Sport Management Quarterly, 12*(5), 437-455. https://doi.org/https://doi.org/10.1080/16184742.2012.693116

Leopkey, B., & Parent, M. M. (2012b). Olympic Games Legacy: From General Benefits to Sustainable Long-Term Legacy. *The International Journal of the History of Sport, 29*(6), 924-943. https://doi.org/10.1080/09523367.2011.623006

Li, D. (2017). Expertise versus Bias in Evaluation: Evidence from the NIH. *American Economic Journal: Applied Economics, 9*(2), 60-92. https://doi.org/10.1257/app.20150421

Li, M., & Yang, J. (2019). Effects of CEO duality and tenure on innovation. *Journal of Strategy and Management*.

Li, S., Blake, A., & Cooper, C. (2011). Modelling the economic impact of international tourism on the Chinese economy: a CGE analysis of the Beijing 2008 Olympics. *Tourism Economics, 17*(2), 279-303.

Linstone, H. A., & Turoff, M. (2002). The Delphi Method - Techniques and Applications. *Delphi Book*.

Linstone, H. A., & Turoff, M. (2011). Delphi: A brief look backward and forward. *Technological Forecasting and Social Change, 78*(9), 1712-1719. https://doi.org/10.1016/j.techfore.2010.09.011

LOCOG. (2004). *Summer Olympic Games Official Report London 2012*. https://stillmed.olympic.org/Documents/Reports/Official%20Past%20Games%20Reports/Summer/2012/ENG/2012-RO-S-London_V1_I_eng.pdf,

Logan, I. (2018, 25th May 2018). Why host the Youth Olympic Games? *Sportcal*. https://www.sportcal.com/Insight/Opinion/118027

London Assembly. (2007). A lasting legacy for London? Assessing the legacy of the Olympic Games and Paralympic Games. *London Research Institute (LERI), University of East London, for the London Assembly of the Greater London Authority.* https://www.london.gov.uk/about-us/london-assembly/london-assembly-publications/lasting-legacy-london-assessing-legacy-olympic

Louargand, M. A., & McDaniel, J. (1991). Price efficiency in the art auction market. *Journal of Cultural Economics, 15*(2), 53-65.

Louviere, J. J., & Islam, T. (2008). A comparison of importance weights and willingness-to-pay measures derived from choice-based conjoint, constant sum scales and best–worst scaling. *Journal of Business Research, 61*(9), 903-911. https://doi.org/10.1016/j.jbusres.2006.11.010

Mair, J., & Whitford, M. (2013a, 2013/03/15). An exploration of events research: event topics, themes and emerging trends. *International Journal of Event and Festival Management, 4*(1), 6-30. https://doi.org/10.1108/17582951311307485

Mair, J., & Whitford, M. (2013b, 01//). An exploration of events research: event topics, themes and emerging trends. *International Journal of Event & Festival Management, 4*(1), 6. http://0-search.ebscohost.com.library.vu.edu.au/login.aspx?direct=true&db=edb&AN=86132660&site=eds-live

Majumdar, B. (2012). The 'accidental legacy' of Commonwealth Games 2010. *South Asian History and Culture, 3*(1), 126-132. https://doi.org/10.1080/19472498.2012.639526

Malfas, M., Theodoraki, E., & Houlihan, B. (2004). Impacts of the Olympic Games as mega-events. *Proceedings of the Institution of Civil Engineers: Municipal Engineer, 157*(3), 209-220. https://doi.org/DOI: 10.1680/muen.157.3.209.49461

Mallen, C., Stevens, J., Adams, L., & McRoberts, S. (2010). The assessment of the environmental performance of an international multi-sport event. *European Sport Management Quarterly, 10*(1), 97-122. https://doi.org/10.1080/16184740903460488

Maloney, M. T., & McCormick, R. E. (2000, 01 / 01 /). The Response of Workers to Wages in Tournaments: Evidence From Foot Races. *Journal of Sports Economics, 1*(2), 99-123. https://doi.org/10.1177/152700250000100201

Marcouiller, D. W. (1997, 1997/02/01). Toward Integrative Tourism Planning in Rural America. *Journal of Planning Literature, 11*(3), 337-357. https://doi.org/10.1177/088541229701100306

Marden, J. I. (2019). *Analyzing and modeling rank data*. Chapman and Hall/CRC.

Mason, D. S., & Slack, T. (2005, 2005/03/01). Agency Theory and the Study of Sport Organizations. *Sport in Society, 8*(1), 48-64. https://doi.org/10.1080/1743043052000316614

Matheson, V. (2006). Mega-Events: The effect of the world's biggest sporting events on local, regional, and national economies.

Matheson, V. A., & Baade, R. (2004). An economic slam dunk or march madness? Assessing the economic impact of the NCAA basketball tournament. *The economics of college sports*, 111-133.

May, V. (1995). Environmental implications of the 1992 Winter Olympic Games. *Tourism Management, 16*(4), 269-275.

Mazodier, M., & Quester, P. (2014). The role of sponsorship fit for changing brand affect: A latent growth modeling approach. *International Journal of Research in Marketing, 31*(1), 16-29. https://doi.org/10.1016/j.ijresmar.2013.08.004

Mazodier, M., Quester, P., & Chandon, J.-L. (2012). Unmasking the ambushers: conceptual framework and empirical evidence. *European Journal of Marketing, 46*(1/2), 192-214. https://doi.org/10.1108/03090561211189284

MBIE. (2013). Economic evaluation outcomes: Major Events Development Fund. Retrieved 14/1/2016, from http://www.majorevents.govt.nz/pdf-library/news/MEDF-evaluation-report.pdf

McCartney, G., Butler, R., & Bennett, M. (2008, 2008/11/01). A Strategic Use of the Communication Mix in the Destination Image-Formation Process. *Journal of Travel Research, 47*(2), 183-196. https://doi.org/10.1177/0047287508321201

McGillivray, D. (2013). Digital cultures, acceleration and mega sporting event narratives. *Leisure Studies, 33*(1), 96-109. https://doi.org/10.1080/02614367.2013.841747

McGillivray, D., & Frew, M. (2014). From Fan Parks to Live Sites: Mega events and the territorialisation of urban space. *Urban Studies, 52*(14), 2649-2663. https://doi.org/10.1177/0042098014550456

McLuhan, M. (1964). *Understanding media : the extensions of man* ([1st ed.] ed.). McGraw-Hill.

Metfessel, M. (1947, 1947/10/01). A Proposal for Quantitative Reporting of Comparative Judgments. *The Journal of Psychology, 24*(2), 229-235. https://doi.org/10.1080/00223980.1947.9917350

Miles, S. (2012). Stakeholder: Essentially Contested or Just Confused? *Journal of Business Ethics*(3), 285.

Mill, J. S. (1859). *Utilitarianism (1863)*.

Mitchell, R. K., Agle, B. R., & Wood, D. J. (1997). Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts. *The Academy of Management Review, 22*(4), 853-886. https://doi.org/10.2307/259247

Monclús, F. J. (2000). Barcelona's planning strategies: from 'Paris of the South' to the 'Capital of West Mediterranean'. *GeoJournal, 51*(1), 57-63. http://www.jstor.org/stable/41147497 .

Motoarca, I.-R. (2015). Kinds of Fair Play and Regulation Enforcement: Toward a Better Sports Ethic. *Journal of the Philosophy of Sport, 42*(1), 121-136.

Mules, T., & Dwyer, L. (2005). Public Sector Support for Sport Tourism Events: The Role of Cost-benefit Analysis. *Sport in Society, 8*(2), 338-355. https://doi.org/10.1080/17430430500087864

Murphy, M., Black, N., Lamping, D., McKee, C., Sanderson, C., Askham, J., & Marteau, T. (1998). Consensus development methods, and their use in clinical guideline development. *Health technology assessment (Winchester, England), 2*(3), i-88.

Murray, S. (2012, 2012/09/01). The Two Halves of Sports-Diplomacy. *Diplomacy & Statecraft, 23*(3), 576-592. https://doi.org/10.1080/09592296.2012.706544

Nickell, D., & Johnston, W. J. (2019). An attitudinal approach to determining Sponsorship ROI. *Marketing Intelligence & Planning*.

Norway Government. (2019). *Norway Government Sports Policy*. Department of Civil Society and Sports. Retrieved 4 April 2021 from https://www.regjeringen.no/en/topics/culture-sports-and-non-profit-work/innsiktsartikler/sports-policy/id2001187/

Notten, P. v. (2005). *Writing on the wall scenario development in times of discontinuity*.

Nowack, M., Endrikat, J., & Guenther, E. (2011). Review of Delphi-based scenario studies: Quality and design considerations. *Technological Forecasting and Social Change, 78*(9), 1603-1615. https://doi.org/10.1016/j.techfore.2011.03.006

Nye, J. (2017). Soft power: the origins and political progress of a concept. *Palgrave Communications, 3*(1), 1-3.

Ocean Tomo. (2015, March 5, 2015). *Annual Study of Intangible Asset Market Value from Ocean Tomo, LLC* http://www.oceantomo.com/2015/03/04/2015-intangible-asset-market-value-study/

Ofcom. (2019). *Listed Events: Identifying services that are free-to-view and widely available*. https://www.ofcom.org.uk/__data/assets/pdf_file/0020/155090/listed-events-statement.pdf

Ogden, R. (2008). Anonymity. *The SAGE encyclopedia of qualitative research methods*, 17-18.

Okoye, A. (2009, 11//). Theorising Corporate Social Responsibility as an Essentially Contested Concept: Is a Definition Necessary? *Journal of Business Ethics, 89*(4), 613-627. https://doi.org/10.1007/s10551-008-0021-9

Oldenboom, E. R. (2008). The impact of the broadcasting of sports events on the image and awareness of host cities abroad. *Sports geography*, 2-13.

Onyx, J., & Leonard, R. (2000). *Rural renewal and social capital: The case of Sweden and Australia*. University of Technology, Sydney, Centre for Australian Community ….

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS*. Routledge.

Parent, M. M., Eskerud, L., & Hanstad, D. V. (2012). Brand creation in international recurring sports events. *Sport Management Review, 15*(2), 145-159. https://doi.org/10.1016/j.smr.2011.08.005

Peeters, T., Matheson, V., & Szymanski, S. (2014). Tourism and the 2010 World Cup: Lessons for Developing Countries. *Journal of African Economies, 23*(2), 290-320.

Pellegrino, G., & Hancock, H. (2010). *A lasting legacy : How major sporting events can drive positive change for host communities and economies*. https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Public-Sector/dttl-ps-alastinglegacy-08082013.pdf

Powell, C. (2003, Feb). The Delphi technique: myths and realities. *Journal of Advanced Nursing, 41*(4), 376-382. https://www.ncbi.nlm.nih.gov/pubmed/12581103

Pranic, L., Petric, L., & Cetinic, L. (2012). Host population perceptions of the social impacts of sport tourism events in transition countries: Evidence from Croatia. *International Journal of Event and Festival Management, 3*(3), 236-256. https://doi.org/10.1108/17582951211262684

Preuss, H. (2004). *The economics of staging the Olympics : a comparison of the Games, 1972-2008*. Edward Elgar.

Preuss, H. (2005). The Economic Impact of Visitors at Major Multi-sport Events. *European Sport Management Quarterly, 5*(3), 281-301. https://doi.org/http://dx.doi.org/10.1080/16184740500190710

Preuss, H. (2007). The Conceptualisation and Measurement of Mega Sport Event Legacies. *Journal of Sport & Tourism, 12*(3-4), 207-228. https://doi.org/10.1080/14775080701736957

Preuss, H. (2013). The contribution of the FIFA world cup and the Olympic games to green economy. *Sustainability (Switzerland), 5*(8), 3581-3600. http://www.scopus.com/inward/record.url?eid=2-s2.0-84886243276&partnerID=40&md5=403ee678c09e12cb97db91cbb183e354

Preuss, H. (2015). A framework for identifying the legacies of a mega sport event. *Leisure Studies, 34*(6), 643-664.

Preuss, H. (2018). Event legacy framework and measurement. *International Journal of Sport Policy and Politics, 11*(1), 103-118. https://doi.org/10.1080/19406940.2018.1490336

Putnam, R. D. (2000). Bowling alone: America's declining social capital. In *Culture and politics* (pp. 223-234). Springer.

PWC. (2005). *Olympic Games Impact Study*. http://gamesmonitor.org.uk/files/PWC%20OlympicGamesImpactStudy.pdf

Ramchandani, G. (2014). *Economic, sport development and elite performance consequences of sports events* Sheffield Hallam University]. edsble.

Ramchandani, G. M., & Coleman, R. J. (2012). Testing the accuracy of event economic impact forecasts. *International Journal of Event and Festival Management, 3*(2), 188-200. https://doi.org/10.1108/17582951211229726

Ramirez, R. (1999). Value Co-Production: Intellectual Origins and Implications for Practice and Research. *Strategic Management Journal, 20*(1), 49-65. http://www.jstor.org/stable/3094231

Ramirez, R., & Wilkinson, A. (2016). *Strategic reframing: The Oxford scenario planning approach*. Oxford University Press.

Ransom, I. (2018). Games: Gold Coast businesses fume over Commonwealth 'disaster'. *Reuters*. https://www.reuters.com/article/us-games-commonwealth-business-idUSKBN1HK0JD

Reddin, E. (2017). Cultural Citizenship, Social Utility, and Positive Network Externalities: The Role of Anti-Siphoning Legislation. *PLATFORM: Journal of Media & Communication, 8*(2).

Reis, A. C., de Sousa-Mast, F. R., & Gurgel, L. A. (2013). Rio 2016 and the sport participation legacies. *Leisure Studies, 33*(5), 437-453. https://doi.org/10.1080/02614367.2012.758304

Ritchie, B. (1984). Assessing the Impact of Hallmark Events: Conceptual and Research Issues. *Journal of Travel Research, 23*(2), 1-11.

Rocha, C. M. (2020). Temporal variations in the relationship between legacies and support: A longitudinal case study in Rio 2016 Olympic Games. *Journal of Sport Management, 34*(2), 130-146.

Rogerson, R. J. (2016). Re-defining temporal notions of event legacy: lessons from Glasgow's Commonwealth Games. *Annals of Leisure Research, 19*(4), 497-518. https://doi.org/10.1080/11745398.2016.1151367

Rose, J. (2002). *Something for Nothing? The Legacy of Ephemeral Events* Events and Place Making, Event Research Conference, UTS : Business, Sydney.

Rowe, D. (2020). Subjecting pandemic sport to a sociological procedure. *Journal of sociology, 56*(4), 704-713.

Rowe, D., & Baker, S. A. (2012). Live sites in an age of media reproduction: mega events and transcontinental experience in public space. *Global Media Journal - Australian edition*.

Rowe, D., & Miller, T. (2021). Communication and Sport. In B. Michael (Ed.), *14 MediaSport: over production and global consumption:* (pp. 255-274). De Gruyter Mouton. https://doi.org/doi:10.1515/9783110660883-014

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting, 15*, 353-375.

Saayman, M., & Saayman, A. (2012). The economic impact of the Comrades Marathon. *International Journal of Event and Festival Management, 3*(3), 220-235. https://doi.org/10.1108/17582951211262675

Sailer, M. O. (2015). *Package 'crossdes'*. CRAN.RSTUDIO. Retrieved 12 August 2015 from https://cran.rstudio.com/web/packages/crossdes/crossdes.pdf

Sallent, O., Palau, R., & Guia, J. (2011). Exploring the legacy of sport events on sport tourism networks. *European Sport Management Quarterly, 11*(4), 397-421. https://doi.org/10.1080/16184742.2011.599208

Sammut, R., Griscti, O., & Norman, I. J. (2021). Strategies to improve response rates to web surveys: a literature review. *International Journal of Nursing Studies*, 104058.

Sant, S.-L., & Mason, D. S. (2015). Framing Event Legacy in a Prospective Host City: Managing Vancouver's Olympic Bid. *Journal of Sport Management, 29*(1), 42-56. https://doi.org/10.1123/jsm.2013-0294

Schirato, T. (2013). *Sports discourse*. A&C Black.

Schubert, M. (2014, 2014/10/07). Potential agency problems in European club football? The case of UEFA Financial Fair Play. *Sport, Business and Management: An International Journal, 4*(4), 336-350. https://doi.org/10.1108/SBM-02-2014-0006

Schulenkorf, N., & Edwards, D. (2012). Maximizing positive social impacts: Strategies for sustaining and leveraging the benefits of intercommunity sport events in divided societies. *Journal of Sport Management, 26*(5), 379-390.

Serrano-Durá, J., Molina, P., & Martínez-Baena, A. (2021). Systematic review of research on fair play and sporting competition. *Sport, Education & Society, 26*(6), 648-662.

Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. *American Psychologist, 44*(6), 922.

Shen, P. (2000). The P/E ratio and stock market performance. *Economic review-Federal reserve bank of Kansas city, 85*(4), 23-36.

Sheskin, D. J. (2003). *Handbook of parametric and nonparametric statistical procedures*. Chapman and Hall/CRC.

Siegfried, J., & Zimbalist, A. (2006). The Economic Impact of Sports Facilities, Teams and Mega-Events. *The Australian Economic Review, 39*(4), 420-426.

Small, K. (2007). Social Dimensions Of Community Festivals: An Application Of Factor Analysis In The Development Of The Social Impact Perception (Sip) Scale. *Event Management, 11*, 45-55.

Smith, A., & Fox, T. (2007). From 'Event-led' to 'Event-themed' Regeneration: The 2002 Commonwealth Games Legacy Programme. *Urban Studies (Routledge), 44*(5/6), 1125-1143. https://doi.org/10.1080/00420980701256039

Solberg, H. A., & Preuss, H. (2007). Major Sport Events and Long-Term Tourism Impacts. *Journal of Sport Management, 21*(2), 213-234. https://doi.org/https://doi.org/10.1123/jsm.21.2.213

Stefani, R. (2020, 9-11 November 2020). Understanding how sports like Sport Climbing and Break Dancing were added to the Olympics: Could eSports be next? 15th ANZIAM Mathsport, Wellington, NZ.

Stensholt, J., & Korporaal, G. (2021, 08/12/ 2021 Aug 12). World Cup bid ready to kick off. *The Australian,* 2. https://www.proquest.com/newspapers/world-cup-bid-ready-kick-off/docview/2560149165/se-2?accountid=14844

Stewart, A., & Rayner, S. (2015, 2016/04/02). Planning mega-event legacies: uncomfortable knowledge for host cities. *Planning Perspectives, 31*(2), 157-179. https://doi.org/10.1080/02665433.2015.1043933

Stiglitz, J. E., & Weiss, A. (1981). Credit Rationing in Markets with Imperfect Information. *The American Economic Review, 71*(3), 393-410. http://www.jstor.org/stable/1802787

Strauss, H. J., & Zeigler, L. H. (1975). The Delphi Technique and Its Uses In Social Science Research. *The Journal of Creative Behavior, 9*(4), 253-259. http://dx.doi.org/10.1002/j.2162-6057.1975.tb00574.x

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor.

Swart, K., Linley, M., & Muresherwa, G. (2017). Sport mega-events and the media. In *Legacies and Mega Events* (pp. 235-243). Routledge.

Taks, M., Chalip, L., & Green, B. C. (2015). Impacts and strategic outcomes from non-mega sport events for local communities. *European Sport Management Quarterly, 15*(1), 1-6.

Taks, M., Green, B. C., Misener, L., & Chalip, L. (2014). Evaluating sport development outcomes: the case of a medium-sized international sport event. *European Sport Management Quarterly, 14*(3), 213-237. https://doi.org/10.1080/16184742.2014.882370

Taylor, M., & Edmondson, I. (2007). Major sporting events—planning for legacy. *Proceedings of the ICE - Municipal Engineer, 160*(4), 171-176. https://doi.org/10.1680/muen.2007.160.4.171

Tetlock, P. E., & Gardner, D. (2016). *Superforecasting: The art and science of prediction*. Random House.

Thomson, A., Cuskelly, G., Toohey, K., Kennelly, M., Burton, P., & Fredline, L. (2019, 2018/07/06/). Sport event legacy: A systematic quantitative review of literature. *Sport Management Review, 22*(3), 295-321. https://doi.org/10.1016/j.smr.2018.06.011

Thomson, A., Schlenker, K., & Schulenkorf, N. (2013). Conceptualizing Sport Event Legacy. *Event Management, 17*(2), 111-122. https://doi.org/10.3727/152599513x13668224082260

Toohey, K., & Taylor, T. (2012). Surveillance and securitization: A forgotten Sydney Olympic legacy. https://doi.org/10.1177/1012690211433456

Turner, M., & McCrory, P. (2003). Social drug policies for sport. *British Journal of Sports Medicine, 37*(5), 378. https://doi.org/10.1136/bjsm.37.5.378

Turner, P. (2007). The Impact of Technology on the Supply of Sport Broadcasting. *European Sport Management Quarterly, 7*(4), 337-360. https://doi.org/10.1080/16184740701717055

Tversky, A., & Kahneman, D. (1973, 1973/09/01/). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207-232. https://doi.org/https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

UBC. (2013). *Olympic Games Impact (OGI) Study for the 2010 Olympic and Paralympic Winter Games - Post-Games Report*. U. o. B. Columbia. http://cfss.sites.olt.ubc.ca/files/2011/10/OGI-UBC-Post-Games-Report-2013-10-23.pdf

UEL. (2015). *Olympic Games Impact Study – London 2012 Post-Games Report*. https://www.uel.ac.uk/sites/default/files/10083.pdf

UKSport. (2015). *eventIMPACTS*. United Kingdom Sports Council. http://www.eventimpacts.com/#

UN. (2018). *Creating the Greenest Football Club in the World - Forest Green Rovers*. United Nations - Climate Change. Retrieved 22 July 2021 from https://unfccc.int/climate-action/momentum-for-change/climate-neutral-now/creating-the-greenest-football-club-in-the-world-forest-green-rovers

USGBC. (2016). LEED Silver Nationals Park is a model for green sports. https://www.usgbc.org/articles/leed-silver-nationals-park-model-green-sports

Veal, A. J., Toohey, K., & Frawley, S. (2012). The sport participation legacy of the Sydney 2000 Olympic Games and other international sporting events hosted in Australia. *Journal of Policy Research in Tourism, Leisure and Events, 4*(2), 155-184. https://doi.org/10.1080/19407963.2012.662619

Veltri, F. R., Miller, J. J., & Harris, A. (2009). Club Sport National Tournament: Economic Impact of a Small Event on a Mid-Size Community. *Recreational Sports Journal, 33*(2), 119-128. https://doi.org/10.1123/rsj.33.2.119

Victorian Auditor General. (2007). *State Investment in Major Events*. https://www.audit.vic.gov.au/sites/default/files/20070523-Investment-in-Major-Events.pdf

von der Gracht, H. A. (2012). Consensus measurement in Delphi studies. *Technological Forecasting and Social Change, 79*(8), 1525-1536. https://doi.org/10.1016/j.techfore.2012.04.013

Waldron, T. (2016, 16th October 2016). Sports Governing Bodies Must Take Responsibility For Human Rights. *Huffington Post*. https://www.huffpost.com/entry/martina-navratilova-human-rights-ioc-fifa_n_57fffba8e4b0162c043b09d6

Welty Peachey, J., Cunningham, G. B., Lyras, A., Cohen, A., & Bruening, J. (2015). The Influence of a Sport-for-Peace Event on Prejudice and Change Agent Self-Efficacy. *Journal of Sport Management, 29*(3), 229-244. https://doi.org/10.1123/jsm.2013-0251

Westerbeek, H. (2010). *"Driving away white elephants"— an essay on strategically planning the legacy of the Commonwealth Games*.

Westerbeek, H., & Linley, M. (2012, Summer2012). Sponsorship and branding: Research paper Building city brands through sport events: Theoretical and empirical perspectives. *Journal of Brand Strategy, 1*(2), 193-205.

Westmattelmann, D., Sprenger, M., Hokamp, S., & Schewe, G. (2021). Money matters: The impact of prize money on doping behaviour. *Sport Management Review, 23*(4), 688-703. https://doi.org/10.1016/j.smr.2019.09.005

Wilson, C. G., Shipley, T. F., & Davatzes, A. K. (2020). Evidence of vulnerability to decision bias in expert field scientists. *Applied Cognitive Psychology, 34*(5), 1217-1223.

Wilson, R. (2006, 2006/01/01). The economic impact of local sport events: significant, limited or otherwise? A case study of four swimming events. *Managing Leisure, 11*(1), 57-70. https://doi.org/10.1080/13606710500445718

Wilson, S. (2013). *Thinking differently about leadership: a critical history of the form and formation of leadership studies* Victoria University of Wellington]. http://hdl.handle.net/10063/2898

World Athletics. (2017). *Nitro Athletics get the seal of approval from IAAF President Sebastian Coe.* World Athletics. Retrieved 14 May 2021 from https://spikes.worldathletics.org/post/nitro-athletics-approved-by-sebastian-coe

Zagnoli, P., & Radicchi, E. (2009). Do Major Sports Events Enhance Tourism Destinations? *Physical Culture and Sport. Studies and Research, 47*(1). https://doi.org/10.2478/v10141-009-0031-z

Zavyalova, A., Pfarrer, M. D., Reger, R. K., & Hubbard, T. D. (2016). Reputation as a benefit and a burden? How stakeholders' organizational identification affects the role of reputation following a negative event. *Academy of Management Journal, 59*(1), 253-276.

Zeng, G., Go, F., & Kolmer, C. (2011). The impact of international TV media coverage of the Beijing Olympics 2008 on China's media image formation: a media content analysis perspective. *International Journal of Sports Marketing & Sponsorship*, 319-336.

Zhang, J. J., Kim, E., Mastromartino, B., Qian, T. Y., & Nauright, J. (2018). The sport industry in growing economies: critical issues and challenges. *International Journal of Sports Marketing and Sponsorship, 19*(2), 110-126. https://doi.org/10.1108/ijsms-03-2018-0023

Zhao, R. (2012). Research on Evaluation Index System of Media Performance in Sports Events. 746-749. https://doi.org/10.1109/iccis.2012.243

Zheng, J., Dickson, G., Oh, T., & De Bosscher, V. (2019). Competitive balance and medal distributions at the Summer Olympic Games 1992–2016: overall and gender-specific analyses. *Managing Sport and Leisure, 24*(1-3), 45-58. https://doi.org/10.1080/23750472.2019.1583076

Zyphur, M. J., & Pierides, D. C. (2019). Making Quantitative Research Work: From Positivist Dogma to Actual Social Scientific Inquiry. *Journal of Business Ethics, 167*(1), 49-62. https://doi.org/10.1007/s10551-019-04189-6

Appendices

Appendix 1 - Pillar definitions provided to respondents

Summary pillar definitions made available when ascribing weightings during Delphi study.

Economic	Social	Media	Environment	Sport	Brand
Economic scale of event through direct economic contributions	Social scale of the event in terms of advancement of social development	Media scale of the event defined by media presence and output	Environmental scale of event in terms of its environmental footprint	Sporting scale of the event in its contribution to development of the sport	Branding scale of event in terms of development of host and event image
Including: • event attendance and event-led tourism • direct event spending (operational, capital, security), infrastructure and asset development, • increased economic activity for existing businesses • spending by sponsors on activation programmes (not sponsorships rights value)	Including: • belonging ('feel-good' and public engagement and support) • outreach programmes educational, cultural, human development • community involvement in hosting (volunteering) • promotion of health, well-being & active lifestyle • public infrastructure & 'quality of life' legacies e.g. green space, health, safety and security	Including: • volume of coverage, including press, broadcast, online and social media • number of media organisations, production and technical staff; • media prominence & access ('free to air vs. pay) and timing (live vs. delayed &/or highlights) • media engagement, reach and consumption across media channels	Including: • Efficiency in energy, water, CO^2, land and resource utilisation • mitigation of impact of increased consumption and travel during event • positive legacies of best practice and improved technologies	Including: • numbers and breadth of of participants • quality of performances during the event • support & enforcement of 'clean sport' • positive sport infrastructure and talent pathway legacy • development of the sport and interest in future editions of the event	Including: • profile and presence of the host and event in media • familiarity, associations and perceptions of host destination & the event, • use of iconic elements, landmarks and people for enduring image • expression of host's unique culture or place

Appendix 2 - Delphi study surveys: Rounds 1 – 3.