

Final Pre-publication text for the paper:

Fraser, J., Norlander, R. J., & Nock, K. (2023). Limiting claims about museums, STEM, and social issues. In J. Diamond & S. Rosenfeld (Eds.), *Amplifying informal science learning* (pp. 300–310). Routledge.

<https://www.taylorfrancis.com/chapters/edit/10.4324/9781003145387-34/limiting-claims-museums-stem-social-issues-john-fraser-rebecca-joy-norlander-kathryn-nock>

Limiting Claims about Museums, STEM, & Social Issues

John Fraser, Rebecca Joy Norlander, and Kathryn Nock

Introduction

Museum workers believe that museums are critical vectors for social change. The 2022 ICOM definition of museums made claimed that museums are necessary for fixing social wrongs, paths for cultural diplomacy, and venues for advancing a sustainable future. Unfortunately, there seems to be a scarcity of evidence to back up these social impact claims. An effort to synthesize research in the USA published in the first two decades of the 21st century sought to describe what can be considered common understanding in the museum field about how social issues and science, technology, engineering, and math (STEM) come together in museum practice.

Our study focused on the methods and data reporting: we examined where claims may overshoot what should be considered generalizable fact. To do that, we analyzed a subset of papers assembled through a configurative review of the Museums, STEM, and Social Issues domain in the USA.¹ The initial review described the topics and types of research related to our focal subject. Here, we focus on the choices made about the research methods. By selecting only those papers that assessed the intersection of STEM and social issues in museums, we were able to look across three primary sources of knowledge: peer-reviewed journals, grey literature from a national online repository, and dissertations or theses in the ProQuest database. We used these reports to understand whether there is sufficient evidence to make claims about the museum sector or museums as a class capable of supporting the many claims about their impacts. In this case, we focused only on museums' capacity to use STEM to engage audiences with social issues and acknowledge the exclusion of humanities content as a path for social change.

Background

The scholarly research into museum efforts make programs through community engagement has increased substantially since the turn of the millennium. These trends are evident in peer-reviewed journals and grant funded research. The topic was also in the forefront of the contentious debate over the definition of what constitutes a museum at the 2019 and 2022 International Council on Museums

congresses. The 2022 international definition demonstrated that there is a consensus that museums should be engaging communities with humanities issues of representation, human rights, and moral decision making. But this concept of community engagement around social issues may not pertain to all types of museums, including those that focus on the sciences, technology, engineering, or math (STEM).

We questioned the nature of what constitutes social science evidence in support of these larger claims about the museum enterprise and whether STEM content that might be useful lens for examining this contentious idea. We firmly believe that the museum enterprise can fulfill these aspirations with STEM content, but are concerned that much of the writing about what museums are capable of doing may be based on limited evidence and may even be a function of confirmation bias. Museums, as a type of intervention, cannot be easily subjected to long-term randomized controlled studies due to variation in visitor knowledge, museum types, and the nature of the learning that occurs, but that limitation does not preclude examination of any research that could suggest whether the form can have the social impacts it claims.

We took an appreciative position in our study, aiming to confirm claims of impact that could be attributed to museum form rather than a specific intervention. We value the hybrid knowledge that emerges when different epistemic traditions and methods examine their truths about a phenomenon. We explored how methodological choices afford comparison, contrast, additive, or explanatory value when brought together. Despite the breadth of disciplines that represent museum studies, we found the data to be sparse. This finding is notable, given that museum studies have been highly promoted for decades and there are more than 35,000 museums in the USA alone.²

In 2015, Fu and colleagues³ categorized STEM learning evaluation methods and found a remarkably narrow set of designs, suggesting the need for more empirical research designs that could quantify the impact, scale, and statistical models that would support generalizability. In parallel, Morrissey, and colleagues⁴ examined the online repository of grey literature to understand what can be said about STEM and social issues. Even with an expansive sample frame, they found that many of the projects tended to focus on temporal conditions during an experience, rather than the impact that might accrue over time. Together, the two studies demonstrated a lack of statistical rigor compounded by variation in cultural conditions, complexity, and specificity created that challenge any synthesis or claims about the form.

These two studies findings did suggest that museum visiting does have long-term impacts. The research demonstrates that museum experiences are memorable as significant life events,⁵ and have a mnemonic function in later memory reconstruction.⁶⁷ But, museums also face challenges with equal representation in exhibition display, staffing, and visitors^{89 10} and as both studies acknowledged, participant self-selection may exclude some voices.

Knowledge Production

Methods are a purposeful choice, driven by research aims rather than any need to justify the museum sector's work. Research and evaluation operate at the intersection of understanding specific

conditions for monitoring or improvement, and the efforts to generalize knowledge based on aggregate learning. When research or evaluation is undertaken to support improvement, design, change, or manage risk, they fulfill their goals. But when considered as a body of knowledge, we must acknowledge that methods preference some truths while suppressing awareness of others.

Museum studies, like many other fields, rely extensively on case studies, comparative assessments, and non-randomized studies. Using a quantitative approach can explain how common something is across a population, or to what extent something is occurring. Qualitative approaches give us a better understanding of what happens and why, often reflecting the deeper and unique lived experience of individuals, but frequently without the ability to make broad claims about representativeness.

Methodologists advancing mixed methods theory, the use of qualitative and quantitative studies in combination, reveal a series of crises that challenge claims. They have demonstrated a crisis of validity because of a dichotomization of samples as either qualitative or quantitative, leading to reductive assessment that can be misleading. They describe these challenges as self-limiting crises of representation, legitimation, integration, and epistemic politics¹¹.

Museum studies is also a victim of non-representativeness in attendance, exacerbated by opt-out bias, and the challenge of integration when qualitative and quantitative methodologies are used in isolation, the relative affluence of museum visitors, and historical defaults to heteronormativity.¹²¹³¹⁴ These conditions lead to challenges when researchers focus on a preconceived normative museum visitor when attempting to address a community level social issue.

Common methodologies & sample sizes in the data

We identified 127 separate studies that reported sample size in 71 papers (Figure 30.1). We consider a sample size to be small if the margin of error in statistical tests would not generalize beyond 5,000 people, or effectively $N = 300$. We found that median sample sizes in peer-review were $N=116$, theses and dissertations $N=28.6$, and grey literature hovering near $N=82$, the minimum required for a two-tailed test. We conclude that studies focusing on STEM and social issues in museums relied on non-representative, convenience sampling with responses from samples too small for quantitative analysis, or sample sizes that prevent validity testing. There were a few notable exceptions where large sample sizes allowed robust statistical analysis of the findings which we explore in the Synthesis section below. Across our review, the majority garnered at least 30 responses, considered the upper limit of the minimum range recommended for experimental testing¹⁵ or grounded theory¹⁶ but below the minimums recommended by methodologists for correlational one-tailed testing (64 participants), two-tailed testing (82 participants) or causal-comparative study (51 participants). That is to say, the topic of STEM and social issues was most commonly represented by qualitative research, even when researchers were reporting quantitative data.

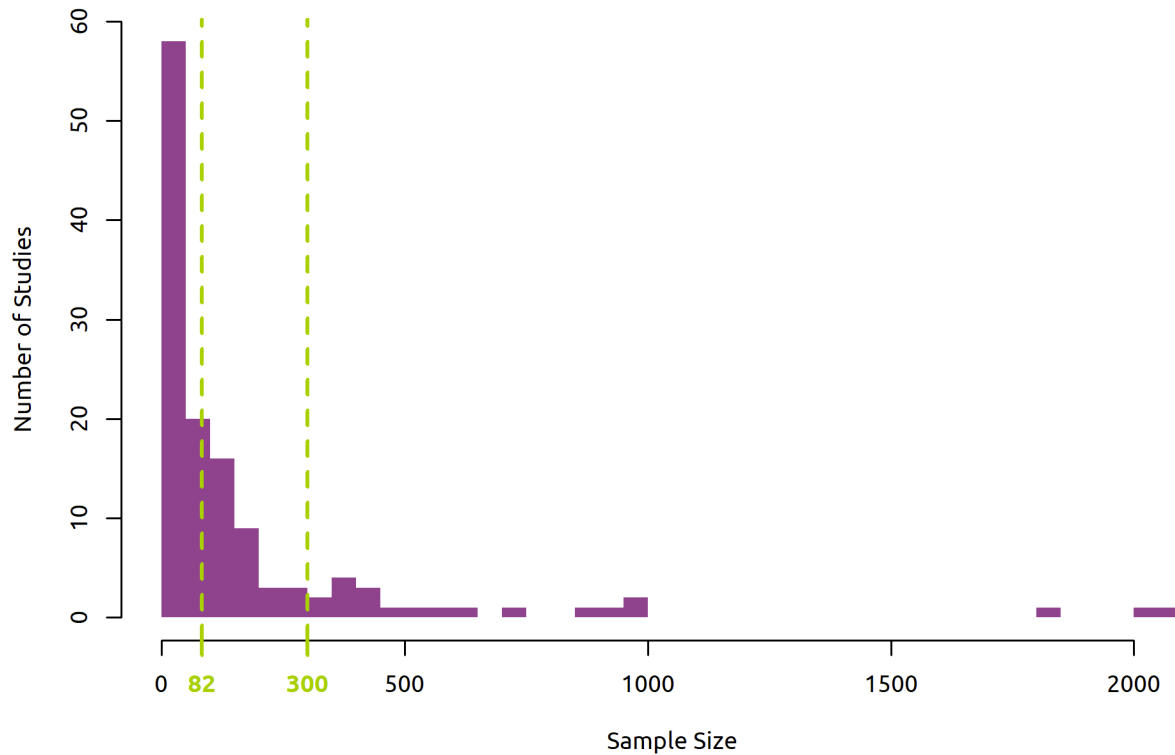


Figure 30.1 *Sample sizes compared to the number of studies relating to social issues and STEM in museums, with break-points at 82 participants suitable for two-tailed tests of significance, and 300 participants where margin of error may be large enough to make our inventory.*

Sampling Strategies

Studies that use truly randomized sampling strategies enable researchers to more easily conduct meta-analysis because they can be combined to provide a more generalizable understanding of a phenomenon. Studies in our dataset tended to focus on surveys, many with small convenience samples collected at a single institution or a small set of institutions. Criterion or convenience sampling seemed to be the most widely used with visitor audiences. A great many articles did not describe a sampling strategy, so our inferred the methodology from how the data were described. While criterion and convenience were the most frequently employed, we also found explorations of variation, including stratified, representative, purposive, extreme case, critical case, snowball, maximum variation, or the use of key informants.

In a few notable cases, deliberate criterion selection or invitation of a participant or participant group based on a specified characteristic offered insights into some social issues. These choices may offer explanatory value when interpreting findings in larger quantitative studies because the criterion specifically illustrates minority concerns that might be concealed in quantitative studies. Unfortunately, without a companion quantitative representative study, these data remain limited in their utility to the museum phenomenon itself.

Analytic Approaches

Some papers in our dataset were clearly grounded in methodological theory, including detail about the steps involved and their decision-making. This clarity was more common in evaluations, theses and dissertations, while peer review was notably lacking detail. The most frequently reported techniques were descriptive statistical measures, such as counts, percentages, and averages.

Analyses tended to be either quantitative or qualitative, with few references to triangulation between types of data.. The qualitative analysis consisted of coding with predefined codes, pattern emergence, *post hoc* code books, inductive coding, or thematic coding. In most cases, projects involving more than one coder included data did report inter-coder reliability and the use of well-known software. Only 31 papers had samples that were large enough to be used in a synthesis study, and few offered benchmarking useful for connecting to larger studies.

We also uncovered some serious concerns in the studies involving STEM and social issues that may extend into the larger museum studies field. We observed studies that employ data summaries and tools without describing how that analytic approach was appropriate to the data. We found calculations of mean and standard deviation or statistical comparison of paired ratings for sample sizes that fall below recommended minimums, or reporting proportions without *N* values, making it hard to know whether the results were statistically significant. We commend the work by Fife¹⁷ and others who offer strategies that attempt to resolve the majority of statistical traps researchers may fall into, but note that our data included quite a few examples that replicate misconceptions or lack of training in applying statistical methods for small samples.

Synthesis

Even though we uncovered methodological issues, we observed a handful of projects that stand as methodological models, and may offer some generalized understanding of the affordances of museums. These models illustrate some of the latent potentials for museums to contribute to public understanding, attitudes, and behavior. They also offer some strategies for addressing methodological critiques revolving around representation, legitimation, integration, and politics.

Representation: Museums, viruses, & vaccines

Papers by Diamond and various colleagues¹⁸¹⁹ described a three-phase sequential mixed methods research strategy to identify the root causes of public debate or confusion about viruses and vaccines, and to test interventions to address these gaps. The project exemplifies an effective approach to addressing representation because the team gathered data that reflected beliefs and then compared specific populations.

They presented a foundational descriptive quantitative study conducted through a set of questions added to a biannual state survey of residents to establish deficits and a rationale for their intervention

with specific target audiences. The second phase employed stratified, purposeful, and semi-structured interviews of key informants, followed by a third phase randomized comparison test of their intervention with their target audiences.

These authors demonstrated that museums can identify a social issue that requires STEM reasoning to understand and make decisions about, critical audiences for that intervention, and testing of specific tools to address that deficit. Their approach led to an intervention useful for a general population, and avoided the self-selection crisis.

We conclude that museums are well-equipped to create learning experiences that can build essential STEM literacies to help a population navigate a social issue. Their methods address the challenge of representativeness by expanding their source data beyond visitor populations. By sampling in controlled settings and using key informants, they were able to overcome potential bias found in relying on visitors alone for data. These studies also demonstrated that museums represent a node in the social process of learning, that complements other settings and materials like formal school, comic books, and public media.

Legitimation: Participatory research, reflexivity, & controversy

Ostman, Zirulnik, and McCullough Cosgrove²⁰ undertook one of the few studies that anchored itself in the general movement to address societal issues at museums. Their study of museum's use of controversial topics affords an examination of the legitimation challenge -- specifically, single methods studies that attempt to represent the diversity of audiences. They used the apparent conflicts that emerge between religious theologies and science enterprise as their prime focus, with the intention of synthesizing claims about museum social experiences and how STEM content can align with audiences' needs. More importantly, they recognized that analyzing a sample of projects around a controversial topic would offer broad theory detection for museum practice.

They noted that despite the promise of museums to provide integrative experiences that bring STEM to social issues, the literature was without compelling evidence beyond small topical studies that we affirmed in our own study. Therefore, they employed participatory research methods, a strategy that acknowledges researchers own experience and knowledge as central to theory detection. Their constructivist approach acknowledged the presence of cultural histories, potential biases, and values of the museum staff as part of the knowledge construction, and the importance a reflexive process to theory construction.

Mixed-methods research theorists assert that the legitimation crisis results from any research paradigm taking precedence over another as a path to truth. Through five unique interventions, they developed a quantitative dataset triangulated to qualitative reporting from participants, program leaders, and the research team. Their instruments focused on the construct level they labeled engagement, rather than any focus on specific content learned. This approach enabled them to compare across subject matter to build generalizable knowledge that could speak more generally to a phenomenon. Although the final dataset ($N = 315$) offered a limited degree of certainty and was weighted toward two of the five settings, the model remains instructive. It revealed a set of critical practices that support using social

issues as a motivational tool for engaging with STEM content in a museum setting. In particular, they identify four principles for museum experience, which represent foundational constructs for future research that we paraphrase as:

Neutral space: through creating a space for users; personal reflection, idea sharing ideas, and encouraging open-minded attention to the perspectives of others.

Presenting visitor's contrasting perspectives to aid affective visitor experiences.

Creating space for personal reflection and documentation to enhance learning outcomes;
and

Providing opportunities for engagement beyond the experience.

The affordances of setting would likely vary based on the content area being pursued. But more importantly, they addressed the legitimization concern by describing their findings as tentative and calling for future exploration to challenge their theory in service of advancing generalizable knowledge.

From a methodological perspective, Ostman, Zirulnik, and Cosgrove demonstrated that STEM and social issues in the museum sector will likely continue to be in the theory development phase, relying on more qualitative approaches, and noting that theory requires sufficient agreement on constructs across topics and types before it can be pursued through large scale quantitative study. Their work offered a useful critique of the qualitative bent in museum studies and over-reliance on evaluation reports reporting frequencies and percent change for small populations irrespective of significance.

Given both the case study implication of Ostman and colleagues' work and our findings, we raise concern about the disconnect about a post-positivist preference for numbers that may not have meaning²¹ The acknowledgement of the early phase of this field, suggests that the interpretivist and critical theory paradigms may continue to be more useful until constructs related to social impact are broad enough to explore the principle of social impact that museums aspire to achieve.

Integration: Triangulating between knowledge worlds

A challenge with literature synthesis lies in the possibility of triangulating, comparing, expanding or consolidating quantitative data when the source material is small and purpose driven. Leblang and Osche²² evaluated a multi-site film, lecture, and discussion program, employing parallel measures to understand how the intervention impacted individuals in different roles and settings. They gathered data from the public, students, and teachers at five sites, supplementing this work with online qualitative studies, and interviews with the program leaders.

Their methods demonstrated that visitors tend to be equipped with knowledge about the social issue they will encounter, and use museum assets to strengthen these users understanding. They demonstrated that 80% of program participants expanded their knowledge of a topic, and most claimed they could apply their learning to different contexts and concerns. The integrative approach revealed a tension important for all questions of how STEM content and contentious social issues might be well-

suited to museum. Furthermore, their results demonstrate that small samples can be used to explore the larger thematic questions of museum affordances and opportunities if used in mixed methods approaches.

Discussion

We found that the motivation for exploring informal STEM learning about social issues at museums remains an emerging field, relying heavily on qualitative inquiry. Unfortunately, there appears to be little cross-referencing between sources, creating a risk of under-representation of the full breadth of evidence that can demonstrate the value of museums bringing STEM to social issues, or where foundational precedent may be found.

The lack of cross-referencing creates a challenge for researchers, evaluators, and research consumers. For researchers, it suggests that each study exists in autonomy and undermines the scientific principle of building on prior knowledge through the use of frameworks and precedent research. For evaluators treating each project *de novo* is expensive, time-consuming, and may be less cost-effective than building on prior knowledge or triangulating data to precedent. For evaluation consumers, it suggests that synthesis and benchmarking before commissioning new studies can offer a more focused use of resources.

We acknowledge that this study was possible because self-published evaluation reports generously offered the most detailed information on methodology and more frequently illustrated the complexity of a mixed-methods study, while the peer-reviewed literature tended to focus on a single question or subset of the data without the value that comes from triangulation between methods. This review suggests that peer review journal editors should encourage the publication of methods to support the advancement of synthesis and reduced redundancy.

Thesis and dissertations were more likely to acknowledge the positionality of the researcher, while peer-reviewed literature adhered more closely to a post-positivist paradigm that may not fully represent the impact of an intervention. More concerning, the noted absence of thesis and dissertation research from the references in the evaluation and peer-reviewed literature suggests a blind spot that could fill gaps in knowledge.

We were surprised by the widespread lack of validated and published instruments for research and evaluation. The tendency to construct targeted instruments for each study without reference to the literature or triangulation with other datasets related to social issues appears to be a deficit in the field. It is impossible to assess broad or generalizable findings.

We were similarly concerned about a sampling fallacy that seems to pervade descriptions of museum populations without caveats or benchmarking. It is well-known that museum visitors cannot be considered representative of a general population. Without refusal rate reporting and remediation strategies for non-participants, “random” sampling may produce a false positive affirming the bias of

museum professionals that an intervention is achieving desired goals. Although many articles acknowledge these challenges, most reports tended to claim representativeness of the sample to their attendees based on specious indicators like sex or age, a practice that can confound explorations of impact because aggregating data containing errors in source data can report on a majority of the majority, exacerbating exclusion.

Conclusion

This exploration of methods used for the study of STEM and social issues in museums reveals a overweighting toward case studies that should limit claims about the role of museums' ability to link STEM content with social issues. This challenge can lead to a misrepresentation of the full depth and value of the museum sector as it supports its service populations' engagement with social issues. Our results offer a cautionary lesson that museums should limit their claims to impact until there is sufficient evidence to stand behind. We recommend that more effort be given to anchoring new case studies to precedent, using validated scales to help build enough data to support future meta-analysis, and paying attention to minoritized voices to ensure that we do not continue to preference the majority of the majority. Through minimal additional effort in reporting, it may be possible to demonstrate the aspirations of the museum sector and the social purposes at the heart of the ICOM 2022 definition.

Acknowledgements

This manuscript was based upon work supported by the National Science Foundation under Grant No. DRL-1906556. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Notes

¹ Morrissey, K., Fraser, J., Norlander, R., Nock, K., Ball, T., & Flinner, K. (2021). A Research Synthesis: Addressing Societal Challenges through STEM. Knology Publication #NSF.022.499.01a. Knology.

² Institute of Museum and Library Services (2014). *Government doubles official estimate: There are 35,000 active museums in the U.S.* Author. <https://www.imls.gov/news/government-doubles-official-estimate-there-are-35000-active-museums-us#:~:text=The%20Institute%20of%20Museum%20and,of%2017%2C500%20from%20the%201990s>.

³ Fu, A. C., Peterson, L., Kannan, A., Shavelson, R. J., & Kurpius, A. (2015). A framework for summative evaluation in informal science education. *Visitor Studies*, 18(1), 17-38.

⁴ Morrissey, K., Petrie, K., Canning, K., Windleharth, T.W., Montano, P. (2014) *Museums & Social Issues: A Research Synthesis of an Emerging Trend*. Available at: <http://www.informalscience.org/museums-social-issues-research-synthesis-emerging-trend>

⁵ Fraser, J., (2009). *An examination of environmental collective identity development across three life stages: The contribution of social public experiences at zoos*. (Doctoral dissertation, Antioch University New England, 2009). Dissertation Abstracts International, http://www.ohiolink.edu/etd/view.cgi?acc_num=antioch1244223241

Fraser, J., & Sickler, J. (2009). *Why zoos and aquariums matter: Handbook of research key findings and results from national audience surveys*. Silver Spring, MD: Association of Zoos and Aquariums.

⁶ E.g. Anderson, D., Shimizu, H., & Iwasaki, S. (2021). Recollections of Who We Were: Nostalgic Retrospective Perceptions of Japanese Society Following a Visit to a Shōwa-era Museum. *Curator: The Museum Journal*, 64(1), 17-40. and others by these authors.

⁷ Dean, D. (2013). Museums as sites for historical understanding, peace, and social justice: Views from Canada. *Peace and Conflict: Journal of Peace Psychology*, 19(4), 325.

⁸ National Research Council. (2009). *Learning science in informal environments: People, places, and pursuits*. National Academies Press.

⁹ Isselhardt, T., & Cross, L. (2020). You love them, but you don't know them: Recognizing and welcoming lived experiences. *Curator: The Museum Journal*, 63(4), 571-578.

¹⁰ Martinez, N. (2020). Increasing museum capacities for serving non-white audiences. *Curator: The Museum Journal*, 63(4), 585-590.

¹¹ Collins, K. M., Onwuegbuzie, A. J. & Jiao, Q. G. (2007). A mixed methods investigation fo mixed methods sampling deisgns in social and health science research. *Journal of Mixed Methods Research* 1(3), 267-294.

¹² Coffee, K. (2008). Cultural inclusion, exclusion, and the formative roles of museums. *Museum Management and Curatorship*, 23(3), 261-279.

¹³ DiMaggio, P. (1996). Are art-museum visitors different from other people? The relationship between attendance and social and political attitudes in the United States. *Poetics*, 24(2-4), 161-180.

¹⁴ Dawson, E. (2014). "Not designed for us": How science museums and science centers socially exclude low-income, minority ethnic groups. *Science Education*, 98(6), 981-1008.

¹⁵ Onwuegbuzie, A. J., Jiao, Q. G., & Bostick, S. L. (2004). *Library anxiety: Theory, research, and applications*. Scarecrow Press.

¹⁶ Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education.

¹⁷ Fife, D. (2020). The eight steps of Data analysis: A graphical framework to promote sound statistical analysis. *Perspectives on Psychological Science*, 15(4) 1054 –1075.

¹⁸ Diamond, J., Jee, B., Matuk, C., McQuillan, J., Spiegel, A. N., & Uttal, D. (2015). Museum monsters and victorious viruses: improving public understanding of emerging biomedical research. *Curator: The Museum Journal*, 58(3), 299-311.

¹⁹ Diamond, J., McQuillan, J., Spiegel, A.N., Hill, P.W., Smith, R., West, J., & Wood, C. (2016). Viruses, vaccines, and the public. *Museums & Social Issues*, 11(1), 9-16.

²⁰ Ostman, R., Zirulnik, M.L., Cosgrove, J.M. (2019). Storytelling, science, and religion: Promoting reflection and conversation about societal issues. *Curator: The Museum Journal*. 62(2):117-134.

²¹ Treagust, D. F., Won, M., & Duit, R. (2014). Paradigms in science education research. In N. G. Lederman * S. K. Abell, *Handbook of Research on Science Education*. Volume II. Routledge, pp. 3-17.

²² Leblang, J. & Osche, E. (2011). *Ice Planet Earth: Summative Evaluation Report*. Lesley University - Program Evaluation & Research Group.