



THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e. g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

Capturing Engagement in Early Science Learning:
Triangulating Observational, Psychophysiological,
and Self-Report Measures

Alexia Michelle Revueltas Roux

The University of Edinburgh

Moray House School of Education and Sports

Doctor of Philosophy

2022



THE UNIVERSITY
of EDINBURGH

Alexia Michelle Revueltas Roux

*Capturing Engagement in Early Science Learning: Triangulating Observational,
Psychophysiological, and Self-Report Measures*

Doctor of Philosophy, January 2022

Final Word Count: 72,087

SUPERVISORS:

Professor Andrew Manches, Ph.D., *University of Edinburgh, United Kingdom*
Dr Josie Booth, Ph.D., *University of Edinburgh, United Kingdom*

EXAMINATION COMMITTEE:

Professor Paul Howard-Jones, Ph.D., *University of Bristol, United Kingdom*
Professor Judy Robertson, Ph.D., *University of Edinburgh, United Kingdom*

MENTOR:

Dr Zayba Ghazali-Mohammed, Ph.D., *University of Glasgow, United Kingdom*

Abstract

This thesis examined engagement in early science learning in the context of a science centre, particularly how, and to what extent, the process of engagement can be captured and measured in early years children. To achieve it, this thesis explored the potential of using a multimodal approach which captures the three components of engagement (cognitive, behavioural, and emotional) by triangulating simultaneous data from different tools, proposed theoretically by Azevedo (2015). Firstly, an observational study of children naturally interacting in a science centre showed that early-years children gravitate more and for longer to a hands-on exhibit compared to a planned-discovery exhibit. Secondly, a feasibility study showed it was feasible to use and triangulate the following tools in children from 3-7 years in a real-world context: engagement scales for behaviour coding using videorecording, a head-mounted camera, an electrodermal sensor (EDA), and a self-report questionnaire. Thirdly, the main triangulation study involved 28 children interacting with a sand exhibit at a Science Centre. Findings showed a relationship between children's cognitive-behavioural observations and their emotional arousal (EDA markers), but not with their self-report measures. Specifically, results showed that children's emotional arousal peaks were more likely to increase if they were doing cognitive-demanding behaviours such as strategic decision-making or looking away from the exhibit whilst searching for their parents. There was an increase in this effect the longer its duration lasted, but no effect was found for specific timepoints when a behaviour happened.. Children also engaged more when using new or previously used strategic behaviours rather than when adapting or immediately repeating them, as well as when persevering on a goal until fulfilled. A final study evaluated expert science practitioners' perception of engagement when they judged engagement as an outcome compared with as a process. A slider tool was developed to capture practitioners' continuous perception of the process of engagement while they watched a video of the interaction, as well as their overall perception of engagement by giving a single value. Their agreement amongst the three different videos used was also examined, and how much they

aligned with results from the previous triangulation study. Results showed no differences comparing between dynamic and discrete single scoring, however, the continuous dynamic score showed more nuances behind practitioner's rating of engagement throughout the interaction. These dynamic ratings also showed more agreement between the practitioners, and although practitioners' perceptions aligned to identify the video classified as the highest level of engagement, when levels were lower, evaluation of the level of engagement was challenging and practitioners did not agree on the intensity of the perceived engagement.

Overall, this body of research highlights how multiple sources of data can provide a richer picture of what could be understood as engagement, particularly when engagement is conceptualised as a continuous process, which may inform improvements in both facilitation and exhibit design through deeper understanding of engagement processes and tailoring them specifically to different age groups. However, the findings also highlight some of limitations some of the different tools used here have for specific situations. This research presents a theoretical advancement by conceptually examining engagement as a process as well as an outcome along with contributing a thorough examination of early years' informal science learning engagement and relevant tools to capture it. This is an area which has been greatly understudied compared to other age groups and contexts. This research can improve informal learning experiences in key developmental stages, particularly for populations with limited access to informal learning contexts.

Lay Summary

This thesis researched how early years children engage while they are at a science centre, particularly how the process of engagement can be captured and measured. To achieve this, the thesis explored the potential of a triangulation method proposed theoretically by Azevedo (2015) where different tools are used that can more effectively capture the individual three components of engagement (cognitive, behavioural and emotional), and then the resulting outputs from each tool triangulated. The tools proposed are engagement scales for coding behaviour using videorecording; a self-report questionnaire; a head-mounted camera to see children's perspective and an electrodermal sensor (EDA) that measures emotion from the body through the skin. Firstly, to understand and contextualise how children interacted and engaged typically at a science centre, a study observed them interacting with popular science exhibits. It was found that younger children would choose to engage with an open-ended exhibit more often and for longer periods when compared to an exhibit that has been designed to guide visitors' science discovery (planned-discovery exhibits). Then, a feasibility study showed that each of the tools proposed for the triangulation method was suitable to use with early years children between 3 and 7 years old in a real-world context and that the tools could all be used simultaneously. These two studies led to the main triangulation study which had 28 children interacting with a sand exhibit at a science centre. This study found a relationship between observations of children's cognition and behaviour and their emotional reactions from the EDA sensor; however, no relationship was detected between EDA sensor and information derived from a self-report questionnaire. Specifically, results showed that children's emotional activity was more likely to increase when they were doing cognitive-demanding behaviours, such as decision-making of strategies and when looking away from the exhibit while searching for their parents. The emotional activity would also increase with the duration of any of the coded behaviours. Children often persevered on the same self-set goal, and when using strategies they would often prefer using more new ones or re-using ones that they had previously used, than strategies that were adapted or immediately

repeated. Finally, to understand how expert practitioners evaluate engagement and how the triangulation method could be useful for practice, a final study evaluated practitioners' perception of engagement when this was judged continuously while watching a video of a child interacting with a science exhibit, compared to when they had to judge engagement overall by the end of the video by giving a single score to the intensity perceived. For this study a slider tool was developed that captures practitioners' judgement of engagement throughout the video they watched. Also, practitioners' agreement on their perception of how much engagement they were observing was compared between them, in both continuous and single scores. Their agreement on how they ranked the intensity of engagement was compared with the classification made of results obtained from the triangulation study showing three levels of engagement (low, moderate, and high). Although no significant differences were found between practitioners' ratings of engagement as an outcome or a process, a continuous rating showed more variation of their perception of children's engagement throughout the interaction. Practitioners also showed agreement on how they judged each of the videos continuously, although these did not match with the classification from the previous study. This means that all practitioners recognised a high level of engagement but there was disagreement on how they ranked the intensity of the lower levels. Overall, this body of work thoroughly examines early years' science engagement in a science centre and the methods used to capture it. This research contributes to advance the theory behind engagement by considering and measuring engagement as a process and not just an outcome. The findings from this thesis emphasize the advantage and importance of combining multiple sources of data which can lead to improve understanding of the processes behind engagement and how to adapt them specifically for different age groups. This can be used to improve the facilitation of science learning and exhibit design in the early years by allowing practitioners and researchers to track and modify behaviours and features that could enhance or hinder engagement. Improving engagement in informal learning experiences can be particularly helpful to enrich science learning for early years children with limited access to informal learning contexts.

Acknowledgements

Now that I am finishing this PhD I understand more why people often refer to it as a journey. It is not only the vast amount of knowledge you acquire about your topic, but it is also about how acquiring this knowledge impacts you. These past years have taught me more about myself, my strengths, and my limitations than any other period of my life. But I cannot take all the credit for making this PhD possible. This work would not be the same without the help and support of all the incredible people around me. I don't think there are enough words to express my gratitude, but I'll give it my best shot:

First of all, I am extremely grateful for my amazing team of supervisors. To Dr Andrew Manches, for the wonderful opportunity of joining the Move2Learn and always believing in my work even when I struggled to see it myself; your hope and optimism always kept me afloat. To my second supervisor Dr Josie Booth, for your unwavering support through all these years; you always had the right word at the right time, showing me academia can also be a nurturing environment. To my mentor and honorary supervisor, Dr Zayba Ghazali-Mohammed, for your inexhaustible guidance all these years, for always lending me a helpful eye, a kind ear and a good laugh. I am sincerely grateful to all of you for all the effort, time and hard work that has contributed to making this body of work a better research and of me, a better researcher. A very special thank you goes to Dr Katie Cebula, for your brief yet invaluable time as my temporary supervisor, particularly during such a tough time. Thank you all for always being supportive and encouraging, even when none of us were sure of the direction, it was great having you as partners in this exploration of new territories.

Secondly, I am so thankful for having been part of the Move2Learn team. I learned so much from all of your different areas of expertise and I deeply appreciate all the fun times we had as much as all the guidance and support you gave me, greatly influencing my work. I would like to particularly thank Dr Sharon MacNab, Susan Meikleham and the staff at the Glasgow Science Centre for helping me collect my data from their visitors, for helping me think and better understand how your science centre works, for all your valuable input when I was planning and

carrying out my study at the GSC and for always making me feel welcome and leaving your company with a reminder of why my research mattered. A special acknowledgement as well to all the children and parents that took part in my studies.

Thirdly, I do not have enough words for the fantastic support network I have around me. Starting with my family in Mexico, who have cheered on me and supported me from the very beginning, even though they had to endure me being an ocean away. You were always the light at the end of the tunnel showing me it was possible. To my family in Scotland for your constant support and motivation. And to my friends, the ones in Mexico, the ones in the UK, and the ones who I had the pleasure to make along the PhD, you were always interested in my work, encouraging me at the other end of a text, a call, or a door, no matter what I needed or when I needed it (in Mexico we'd say they were at the foot of the cannon, ready for action). Thank you all for always offering me all kinds of practical/methodological/ music/café/entertainment/life advice, and without whom I would have probably given up. However, a very special mention is for Euan Gardner who saved my analysis R code more than once, to Isabel Martinez Fornos for being my very first test subject and always willing to help me pilot my different tools, to Karim Rivera Lares and Ana Hibert Santana who endured me trying to explain my research design and analysis constantly until it made sense, and to Sarah Austin who helped me proofread my whole thesis in a record amount of time. And last but not least, to my partner David McClure for whom I do not have enough words to thank, particularly over the last year, who underwent the final toughest moments with me, with a loving smile and a strong embrace that helped me push through. To all of you, thank you for helping me achieve what I thought impossible.

Finally, I am very thankful to all the funders that made this research possible through the joint doctoral scholarship awarded by the Move2Learn project and the Mexican National Council for Science and Technology (CONACyT). The studentship from the Move2Learn project was awarded under a collaboration between the National Science Foundation (NSF), the Wellcome Trust, and the Economic and Social Research Council (ESRC) via a grant from the NSF (NSF grant no. 1646940), the Wellcome Trust with ESRC (Wellcome Trust grant no. 206205/Z/17/Z).

The doctoral scholarship awarded by the CONACyT was through their program “Becas al Extranjero” (Scholarships Overseas) (CVU:764909).

Contents

Abstract	iii
Lay Summary	v
Acknowledgements	vii
List of Tables	xiv
List of Figures	xv
Chapter 1 Introduction	1
Chapter 2 Literature Review	6
2.1 <i>Early Science Learning</i>	6
2.1.1 Importance of Early Science Learning	10
2.1.2 Contexts of Early Science Learning.....	15
2.2 <i>Engagement</i>	24
2.2.1 Engagement and Learning	24
2.2.2 Defining Engagement.....	28
2.2.3 Differences Between Engagement and Other Constructs.....	34
2.2.4 Components of Engagement.....	37
2.2.5 Continuum of Engagement.....	48
2.3 <i>Capturing Situational Engagement</i>	57
2.3.1 Current Tools Used to Capture Engagement as an Outcome	60
2.3.2 Current Tools Used to Capture Engagement as a Process	65
2.3.3 Tools That Capture Temporality of a Process.....	69
2.3.4 Potential of Triangulating Tools	78

2.4	<i>Overview of the Research</i>	88
2.4.1	Thesis Rationale.....	88
2.4.2	Current Research Project	89
2.4.3	Main Aim of the Research	90
2.4.4	Research Questions.....	91
Chapter 3 Children’s Natural Interaction with Exhibits and the Feasibility of Using a Multimodal Triangulation Approach		92
3.1	<i>Overview</i>	92
3.2	<i>Observing Engagement During Natural Interaction with Science Exhibits Study</i>	93
3.2.1	Method.....	96
3.2.2	Results	100
3.2.3	Discussion.....	107
3.3	<i>Feasibility of Method Study: Testing the Triangulation Multimodal Approach and Its Tools</i>	111
3.3.1	Method.....	117
3.3.2	Results	129
3.3.3	Discussion.....	132
Chapter 4 Understanding the Process of Individual Engagement in Children		135
4.1	<i>Overview</i>	135
4.2	<i>Capturing and Triangulating Multimodal Approach</i>	136
4.3	<i>Method</i>	138
4.3.1	Participants	138

4.3.2	Tools	138
4.3.3	Science Exhibit – Sand Patterns	139
4.3.4	Procedure.....	140
4.3.5	Data Processing.....	142
4.4	<i>Data Analysis</i>	154
4.5	<i>Results</i>	157
4.5.1	Descriptive Statistics	157
4.5.2	RQ 3.a - Is There a Relationship Between the Tools That Capture Engagement? 162	
4.5.3	RQ 3.b - Can Children’s Emotional Responses be Predicted by the Onset of Cognitive-Behavioural Responses?.....	167
4.6	<i>Discussion</i>	178
4.6.1	Relationship Between the Different Outcomes of Engagement After the Exhibit Interaction.....	180
4.6.2	Emotional Engagement Markers Predicted by Cognitive and Behavioural Responses throughout Exhibit interaction.	187
4.6.3	Reflection on the Tools Capturing Engagement.....	198
4.6.4	Limitations and Future Directions	203
4.7	<i>Conclusion</i>	206
Chapter 5 Science Practitioner’s Perceptions of Engagement		209
5.1	<i>Overview</i>	209
5.2	<i>Study of Practitioners’ Perceptions of Children’s Engagement</i>	210
5.3	<i>Method</i>	215
5.3.1	Participants	215

5.3.2	Tool- Slider for Tracking Engagement	215
5.3.3	Engagement Videos	217
5.3.4	Procedure.....	218
5.3.5	Data Processing.....	219
5.4	<i>Data Analysis</i>	221
5.5	<i>Results</i>	223
5.5.1	Differences in Dynamic vs Discrete Scoring	225
5.5.2	Participant’s Agreement Scoring Engagement	226
5.5.1	Practitioners’ Written Answers.....	229
5.6	<i>Discussion</i>	232
Chapter 6	General Discussion.....	238
6.1	<i>Overview</i>	238
6.2	<i>Capturing Engagement</i>	241
6.3	<i>Contributions</i>	248
6.3.1	Methodological	248
6.3.2	Theoretical	251
6.4	<i>Implications and Future Directions</i>	254
6.4.1	For Research	254
6.4.2	For Theory.....	256
6.4.3	For Practice.....	258
6.5	<i>Limitations</i>	260
6.6	<i>Conclusions</i>	261
	References	263

Appendix	I
<i>A - Engagement Observation Tools in Children.....</i>	<i>I</i>
<i>B - Rationale: Logic Model</i>	<i>IX</i>
<i>C - Exhibit Selection From Engaging Exhibit Criteria</i>	<i>XIII</i>
<i>D - Testing Feasibility of Engagement Scales and Fixed Video Camera</i>	<i>XIV</i>
<i>E - Testing Feasibility of Head-Mounted Camera</i>	<i>XXII</i>
<i>F - Testing Feasibility of Electrodermal Sensor</i>	<i>XXIX</i>
<i>G - Self-Report Questionnaire – Item Rationale</i>	<i>XXXVII</i>
<i>H – Behavioural Codes for Triangulation Feasibility</i>	<i>XXXIX</i>
<i>I -Self-Report Questionnaire.....</i>	<i>XLIII</i>
<i>J - Theory Behind the Codes for Engagement Codebook.....</i>	<i>XLVI</i>
<i>K - Engagement Codebook Items.....</i>	<i>L</i>
<i>L – Electrodermal Activity Decomposition per Participant.....</i>	<i>LVIII</i>
<i>M – Age Analysis</i>	<i>LXII</i>
<i>N- Additional Illustrative Examples</i>	<i>LXIV</i>

List of Tables

Table 3.1 <i>Frequency Distribution of Reasons for Leaving the Cycloid Collision Exhibit.....</i>	101
Table 3.2 <i>Frequency Distribution of Children’s Company in the Cycloid Collision Exhibit..</i>	101
Table 3.3 <i>Frequency Distribution of Reasons for Leaving the Sand Patterns Exhibit</i>	103
Table 3.4 <i>Frequency Distribution of Reasons for Leaving the Sand Patterns Exhibit</i>	105
Table 3.5 <i>How Do Tools Capture Components of Engagement</i>	112

Table 3.6	<i>Criteria for Selecting an Engaging Exhibit in GSC</i>	118
Table 3.7	<i>Behavioural Categories used from Engagement Scales</i>	126
Table 4.1	<i>Codebook items (Abbreviated version)</i>	146
Table 4.2	<i>Descriptive Information of Behavioural Variables</i>	159
Table 4.3	<i>EDA Descriptive Statistics per Participant</i>	161
Table 4.4	<i>Correlation Coefficients Behind Tool Triangulation</i>	163
Table 4.5	<i>Correlation Coefficients of Coded Behaviours after FDR adjustment</i>	164
Table 4.6	<i>NS-SCRs Descriptive Statistics Per Behaviour</i>	167
Table 4.7	<i>Model Comparison</i>	169
Table 4.8	<i>Model Results for Attention Off-task</i>	172
Table 4.9	<i>Model Results for Strategy</i>	173
Table 4.10	<i>Description of Interaction Times and Frequency Counts Within Behavioural Categories</i>	177
Table 5.1	<i>Measures Obtained from Experts' Observation</i>	220
Table 5.2	<i>Descriptive Results for Three Levels of Video-Coded Engagement</i>	223
Table 5.3	<i>Interrater Agreement and Consistency for Single Scores</i>	226
Table 5.4	<i>Interrater Agreement of Dynamic Scores Between Practitioners for Each Video</i>	228
Table 5.5	<i>Practitioner's Score Versus Score Assigned from Triangulation Model</i>	229
Table 5.6	<i>Practitioner's Conceptualization of Engagement</i>	230
Table 5.7	<i>Practitioner's Markers of Engagement</i>	231
Table 6.1	<i>Evaluation of Tools to Capture Engagement in Children</i>	243
Table 0.1	<i>Head-Mounted Camera Cognitive Markers</i>	XXVII
Table 0.2	<i>Markers for Electrodermal Activity</i>	XXXIV

List of Figures

Figure 2.1	<i>Continuum of Engagement</i>	51
-------------------	--------------------------------------	----

Figure 2.2 <i>Cycle of Engagement</i>	56
Figure 2.3 <i>Representation of Triangulation Model Proposed by Roger Azevedo</i>	83
Figure 3.1 <i>Cycloid Collision Exhibit in Glasgow Science Centre</i>	97
Figure 3.2 <i>Sand Patterns Exhibit in Glasgow Science Centre</i>	98
Figure 3.3 <i>Visualization of Holding Time per Participant in Cycloid Collision Exhibit</i>	102
Figure 3.4 <i>Visualization of Holding Time per Participant in Sand Patterns Exhibit</i>	104
Figure 3.5 <i>Distribution of Children’s Perceived Age</i>	106
Figure 3.6 <i>Distribution of Children’s Dwell Time</i>	106
Figure 3.7 <i>Proposed Design of the Triangulation of Multimodal Tools to Capture Engagement in This Thesis</i>	114
Figure 3.8 <i>Logic Model for Mapping of Tools with Research Aims</i>	115
Figure 3.9 <i>Designated Space Within Glasgow Science Centre</i>	117
Figure 3.10 <i>Exhibit Placement Inside Designated Space</i>	120
Figure 3.11 <i>Head-Mounted Camera</i>	121
Figure 3.12 <i>Empatica E4 Sensor Fitted on Child’s Hand</i>	122
Figure 3.13 <i>Example of Likert-Style Item with Smiley-O-Meter</i>	124
Figure 3.14 <i>Child Interaction with The Science Exhibit Using Equipment</i>	129
Figure 3.15 <i>Screenshot of ELAN Software Triangulating Trace Data</i>	130
Figure 4.1 <i>Exhibit Space in GSC For Study</i>	140
Figure 4.2 <i>Flowchart Showing Steps Followed by Participant from Recruitment Until End</i>	141
Figure 4.3 <i>Code Tree Showing Hierarchy of Codes from Engagement Codebook</i>	145
Figure 4.4 <i>Example of EDA Response After Using Convolution Deconvolution Method</i> ...	151
Figure 4.5 <i>Distribution of Total Behavioural Codes and Total Count Per Behavioural Category</i>	158
Figure 4.6 <i>Distribution of Behavioural Counts Per Participant</i>	160
Figure 4.7 <i>Positive Correlation plots between cognitive-behavioural markers</i>	165
Figure 4.8 <i>Negative Correlation plots between cognitive-behavioural markers</i>	166
Figure 4.9 <i>Distribution of NS-SCRs Count per Behavioural Variable</i>	168

Figure 4.10	<i>Nested Levels Behind Multilevel Structure</i>	168
Figure 4.11	<i>Visualisation of Model's Results</i>	170
Figure 4.12	<i>Plots of Model Predictions</i>	175
Figure 4.13	<i>Plots Showing Influence of Duration in Significant Behaviours</i>	176
Figure 4.14	<i>Boxplots showing Proportional Percentage Time Used in Different Strategies and Goals Sub-categories</i>	178
Figure 4.15	<i>Example Image of Child with Attention Off-Task Looking at Parent</i>	183
Figure 4.16	<i>Example Image of Child with Attention Off-Task Looking Away from Exhibit</i> ..	184
Figure 4.17	<i>Example Image of Child Progressing Through Series of Strategies</i>	191
Figure 4.18	<i>Example Image of Visuo-Tactile Attention Mismatch</i>	196
Figure 5.1	<i>Composition of Slider Tool</i>	216
Figure 5.2	<i>Examples of Practitioners Using the Slider Tool While Watching Interaction Videos</i>	218
Figure 5.3	<i>Visualization of Dynamic Scores in Each Engagement Level</i>	224
Figure 5.4	<i>Plots Showing Engagement Scores Depending on Type of Scoring</i>	225
Figure 5.5	<i>Plots Showing Practitioner's Dynamic and Discrete Scores</i>	227
Figure 0.1	<i>Head-Mounted Camera</i>	XXIII
Figure 0.2	<i>Example Grid Used on First-Perspective Video</i>	XXIV
Figure 0.3	<i>Example Snapshots From Head-Mounted Camera at National Museum of Scotland</i>	XXVI
Figure 0.4	<i>Comparison of Head-Mounted Camera Video With and Without Grid</i>	XXVII
Figure 0.5	<i>Empatica E4 Sensor</i>	XXXI
Figure 0.6	<i>Empatica E4 Sensor Fitted on Child's Hand</i>	XXXII
Figure 0.7	<i>Visualization of Data Obtained from E4 Empatica Sensor</i>	33

*To myself and anyone who comes next
may this serve as evidence
that even the darkest nights have a sunrise*

Chapter 1

Introduction

In our society, science plays an important role that goes beyond just understanding our current world. It helps by finding solutions to improve our world and contributes to the development of technologies that bring innovations to different communities around the globe. However, some people have limited engagement with science, maths, technology, or engineering (STEM) communication, content and activities, whereas they could also have a limited intention to pursue any science career paths as younger people, despite enjoying science. This may be due to a lack of access, resources, confidence, interest or due to a sense of not-belonging in structural science spaces (i.e. structural inequalities) (Godec et al., 2021). Science capital comprises the attitudes, resources, knowledge, experiences, behaviours and identity around STEM (L. Archer et al., 2015). It has been recently studied in order to understand the underlying reasons behind the different attitudes and approaches around science (DeWitt et al., 2016), as well as ways to enhance this capital in both communities and individuals. Having higher science capital can present socio-economic advantages at different stages of life, from schooling opportunities, academic attainment, teachers' and peers' perception, career paths, and job opportunities. Increasing not only the availability of science experiences but also the quality of these can help enrich the local and broader communities (L. Archer et al., 2015; DeWitt et al., 2013, 2016). Science centres and other informal learning environments could provide opportunities for developing science capital, whether it is by providing access to content and experts that are difficult to obtain within a home-learning environment or by providing safe learning spaces and activities that can encourage and enhance people's science experience. When it comes to studying and understanding science learning in informal contexts, engagement has been highlighted in

science centres' research as a necessary pre-cursors and facilitator for learning, because it contributes to making meaning of the learning situation by granting access simultaneous access both to the current situation and to previous experiences that could improve understanding (Barriault & Pearson, 2010). Thus, for improving science capital through science centres, engagement with the science learning experience is a key element.

Science centres in the world evolved from science and natural museums, starting as rare collections of objects and cabinets of curiosities with a focus on science and maths aimed more at technical visitors, until becoming spaces open to all people where they can test and explore interactively with science objects or scientific principles (Hauan & Kolstør, 2014). In 1951, the Science Museum in London started an interactive wing dedicated to children (Children's Gallery), providing interactive exhibit spaces for them to experience some of the basic science and technology concepts discovered so far with the aim of "stimulating interest and curiosity" (Science Museum, 2017). This demonstrates that science centres/museums from early on had considered the importance of including children from all backgrounds in the learning of science, as a concept and as a practice. Hence, understanding how to improve science learning experiences for children from an earlier age could provide better cognitive skills for the children by closing some of the inequality gap (Archer, 2015), while also positively impacting society by helping foster a community of individuals with an increased sense of curiosity and understanding of the world ready to improve current world conditions.

The UK currently has more than 60 publicly accessible informal science learning centres, with a range of science focus, all well-established and committed to the communities where they are located. According to the UK Association for Science and Discovery Centres (ASDC), these science learning centres/museums have amongst their goals to deliver to the public "unusual and exciting opportunities to discover, discuss, question and explore the latest science, engineering and research" (ASDC, 2021). The importance of these institutions

goes beyond promoting science since many of them also deliver science experiences programs beyond their locations to members of the community, reaching those who cannot easily access this science learning on their own account. This is one way in which science centres contribute to closing the attainment and inequality gap as it promotes early aspirations towards science career paths (DeWitt et al., 2013). Another way is through targeted development of science areas and activities for early years children, which is where it can provide longer-lasting effects. Due to their importance, more efforts have been made to bring together interdisciplinary researchers from different areas, like developmental psychology, learning sciences, early years education or human-computer interaction, to work alongside museum practitioners. This to better understand, update, improve and provide better learning experiences that start with the early years but also help enhance teenagers' and adults' science learning experiences to become a self-sustaining science learning cycle when these same adults engage in science learning with children.

The Move2Learn research project emerged from the need to understand how early years children learn, understand, think and communicate science through body-based experiences when they visit a science centre/museum. The Move2Learn project, was one of 5 Science Learning+ project funded by the Wellcome Trust, the National Science Fund (NSF) and the European Social Research Council (ESRC) and focused on embodied science learning of children by exploring how children explore, communicate and develop their science understanding through body-based interaction. This project also brings international collaboration and co-development from researchers from The University of Edinburgh, University College London, and the University of Illinois Urbana Champaign and practitioners from the US and the UK, such as the Glasgow Science Centre, Patricia Frost Museum of Science in Miami, London Science Museum, The Children's Museum of Indianapolis, Learning Through Landscapes and the Sciencecenter in Ithaca, New York.

This doctoral thesis is also a part of the Move2Learn project, with a similar focus on early science learning in informal contexts, but with a different focus placed on children's engagement at a science centre, exploring and evaluating current methods that capture engagement which can lead to understanding engagement in this age period.

The current thesis starts Chapter 2 by discussing what early science learning is, its importance and the different contexts where this science learning happens, with emphasis on informal spaces such as science centres. This is followed by a section on the construct of engagement in learning, its different components, and a comparison to other constructs. In this section a conceptualisation of engagement is proposed as a temporal continuum that considers it both a process and an outcome when it comes to its relationship with learning. After, the different methods and tools used to capture engagement in different learning contexts are reviewed along with a discussion of research done around triangulating tools that capture different aspects of engagement simultaneously and the potential benefits this method could provide. This chapter finishes with an overview of the thesis research project, its aims, research questions and design.

Chapter 3 reports two studies to understand the current methods used to observe engagement in children interacting exhibits and the feasibility study of using a triangulation method with a single child. The first observational study focuses on how children engage naturally with two different types of science exhibits over the period of one hour, one was goal-oriented whilst the other one was open-ended, exploring the number of visitors each exhibit attracted as well as the attraction power measured by the amount of time spent in the exhibit. The second part on evaluating the feasibility of capturing and triangulating the outcomes from the different tools to capture engagement. The tools used were a combination of current tools used in research such as behavioural engagement coding scales, observation from video recordings and self-report questionnaires, and new tools that have

had a more limited use in a science learning context, including a head-mounted camera or an electrodermal activity sensor.

Chapter 4 reports the development of an engagement behavioural coding scale resulting from the previous chapter, used in the study presented in this chapter. The study in this chapter was conducted with a larger sample of children where the same tools evaluated for feasibility were used to capture, triangulate, and evaluate their engagement when they were individually interacting with a science exhibit for five minutes. This chapter discusses whether the triangulation of tools provides additional information on children's engagement, and critically examines whether emotional processes could be driven by cognitive and behavioural processes of engagement.

Chapter 5 explores expert science practitioners' judgement on children's engagement when asked to watch videos with different engagement levels which were identified from the prior multimodal triangulation approach. They were asked to rate engagement as an outcome (i.e. a discrete score at the end of each video), and as a process (i.e. dynamic scoring throughout the video). For the latter, a slider tool was developed to capture a constant score throughout the length of the interaction. This chapter explores practitioners' agreement on their perception of the intensity of engagement and how this matched the researcher. The chapter also interrogates the differences and implications of judging engagement as an outcome or a process and how this may contribute toward detecting different levels of engagement.

Lastly, Chapter 6 contains the general discussion of the thesis, addressing all the previous chapters and debating whether a multimodal triangulation approach is an effective method to capture young children's engagement when they are at a science centre, examining the limitations of the methods used, the implications and the general contributions of this research.

Literature Review

2.1 Early Science Learning

The *early years* refer 'generally' to the period in early childhood between birth and the age of 8 years (Farrell et al., 2015a). This period is characterised by acute developmental changes in the individual: biological, psychological, behavioural and sociocultural in their different spheres (Lerner et al., 2012). Children's brains and bodies experience fast growth, while in their social environment children are experiencing a myriad of new experiences, ways of communicating, cultural socialisation patterns, learning how to be with others and basically how to navigate the world that surrounds them.

Thanks to thorough research on different areas concerning the importance of early childhood, such as developmental, policy, social justice, education, among others (Farrell et al., 2015b), different international institutions that regulate the well-being of individuals, like the United Nations Educational Scientific and Cultural Organization (UNESCO) or the World Health Organization (WHO) agree with the period established between birth and 8 years for early years and the importance this period has for future development (UNESCO, n.d.; World Health Organization, 2020). Similarly, the Scottish government in their Early Years Framework (Scottish Government, 2009) uses the same period as a reference for early years, which has influenced policy regarding children and their well-being. One area of increased international importance within the early years is learning and education, since the WHO, UNICEF and the World Bank Group proposed in 2018 that having opportunities for

early learning is one of the 5 main components of nurturing care during the early years (World Health Organization et al., 2018). They describe early learning as:

“Early learning happens when children use their senses, move their bodies, hear and use language, experience different places, interact with people, and explore different objects. Opportunities for early learning present themselves in daily life” (WHO Team, 2020, p.7).

This definition enhances how learning in the early years is about experiencing the world, which education enhances by providing children from an early age with skills that can help them understand and make meaning of the world they are experiencing. This view is supported by John Dewey (1986) who mentioned that “all genuine education comes about through experience”(p.247). However, he also pointed out that “not all experiences are genuinely or equally educative [...] for some experiences are mis-educative” (p.247), meaning that some experiences can “arrest or distort the growth of further richer experiences” (p.247). Therefore learning in the early years should allow children to experience the world, by guiding and engaging them through education, to link previous experiences and prior knowledge to understand and make meaning of these new experiences. That is what Dewey meant by “education is a development within, by, and for experience” (p.249), in other words, it needs clarity about what the experience is to promote “the growth of future experiences”(p.247). Here is where engagement with learning and education plays a pivotal role, by bringing awareness to the child of the learning activity happening and encouraging them to take part in that experience. Hence, the experiences acquired in the early years are formative, and with the right guidance, these can be shaped and further developed to have an important role later in life. As an example of this during the preschool years, one of the largest projects to investigate longitudinally the effects of early education was the Effective Provision of Preschool Education (EPPE) Project. Its findings showed that having educational experience in a preschool setting improves children’s overall

development when compared to no-attendance (Sammons, 2010). Some of the effects remained throughout Key Stage 1 of Schooling (5–6-year-olds) setting a solid precedent of the benefits of early years education (Sylva et al 2004). Moreover, research has shown that having positive formative experiences in science at a young age influenced adults' decisions of going into a science career (Chakraverty et al., 2020; Dabney et al., 2013).

When talking about early science learning it is inevitable to think about what is meant by learning. Learning can take many shapes, so much so that a common definition has yet to be agreed upon by researchers and educators (for examples see de Houwer et al., 2013; Harel & Koichu, 2010; Lachman, 1997), as this can vary between fields that explore learning with different aims. However, this section will not address the different definitions of learning, rather it will explore different research that has been done around early science learning using the term similarly to Falk (2005): “a broad term that encompasses a wide range of changes in both the brain and body (...) that result in a more knowledgeable individual with an enhanced motivation and capacity to learn more in the future” (p.266). Falk uses this definition when discussing free-choice learning, which seems fitting in a context of the early years where, despite there being desired outcomes of their learning, it is approached and assumed that the child will choose and decide what they want to engage with during their learning, whether this is at home, a school setting or an informal-learning scenario. This is also reinforced by Barriault and Pearson (2010) who say that “learners are not passive recipients of knowledge or empty vessels to be filled. Learning is an active process, an interaction between ideas that learners currently hold and newly presented experiences” (p.92).

Moreover, when talking about early science learning, it is key to clarify and contextualise what is meant by science learning. After a thorough review of the development of scientific reasoning, Zimmerman (2000) stated that the term “science” is used to mean two things: (1) a specific body of conceptual knowledge that follows specific and rigorous

methods, and (2) the specific inductive and deductive methods that allow the development and acquisition of said body of knowledge. The most comprehensive view of science learning is presented by Klahr and Dunbar (1989 cited in Zimmerman, 2000) with their model called Scientific Discovery as Dual Search (SDDS), where they integrate domain-general skills with domain-specific knowledge. In other words, they tried to make a case where scientific thinking and learning is about more than just memorising content related to science such as facts, numbers, or science theories, i.e. domain-specific knowledge, rather it is about a process of understanding how the science content happens, focusing on developing cognitive skills related to scientific thinking such as critical thinking, observation, hypothesising, inquiry, curiosity, and several others, i.e. domain-general knowledge. Eshach and Fried (2005) make a parallelism between these two skills with the differences between conceptual and procedural knowledge, or in simpler terms “knowing that” and “knowing how” (p.316). They then later propose that these two simpler premises guide the rationale behind teaching science in the early years: *knowing that* emphasizes that science is about the real world and information about it should be learnt, whereas *knowing how* reflects that science develops reasoning skills that provide children with tools for exploration, and attuning both skills contribute towards their scientific inquiry (Eshach & Fried, 2005).

An essential component of science learning is engagement with the activity/content. A quality engagement during science learning can enhance scientific and learning skills acquired (Halliday et al., 2018; Tisdal & Perry, 2004). For example Halliday et al. (2018), evaluated engagement in preschool children in a laboratory while performing a set of tasks, where they found that the intensity of their observed behavioural engagement was predictive of their literacy and math performance in the classroom. At the Exploratorium in San Francisco, research has found that visitors who engage longer with interactive exhibits that foster inquiry and exploration are more likely to make connections between experience and prior knowledge (Tisdal & Perry, 2004) as well as show more learning gains about the topic-

content in a post-test evaluation (Dancstep et al., 2015). Thus, engagement with science learning activities might be likely to predict children's engagement in the future with both learning in general, as well as with science-related activities (Saçkes et al., 2011; Shwe Hadani et al., 2018). Eshach and Fried (2005), pointed out that "we see the world with the help of conceptions and ideas created by the human mind; they are like glasses that help us be aware of things to which we might otherwise be blind" (p.332). This is what science learning helps with, providing a lens with which to see and understand the surrounding world, but engagement would be the support that keeps the glasses on.

2.1.1 Importance of Early Science Learning

During the first years of life there is a fast brain growth period that helps facilitate learning skills and knowledge construction due to the increase of potential neural connections which, when established early and maintained, can help improve information processing and subsequently enhance more complex learning (Giedd et al., 1999; Gogtay et al., 2004; Knickmeyer et al., 2008). The current research and practice focus in early science learning encourages targeting basic skills related to developing scientific thinking in the child, such as observation, exploration, repetition of results, etc., which can then be used in multiple situations of daily life to explore and understand their world (Gopnik, 2012; McClure et al., 2017).

Eshach and Fried (2005) reviewed literature around science learning supporting why science should be taught in early childhood, finding six main reasons about the importance of starting early with science:

(1) *Children are natural observers and think and reflect about nature.* Children are already exploring their environment, paying attention and looking for cause-consequence in events they encounter. Evidenced by Buchanan & Sobel (2011), where they showed that

children since 3-4 years old can start making inferences with rationale on situations they are presented (e.g. when a light is connected, detect the correct source). This increases when their curiosity and observation are guided or pointed towards evidence that contributes to their understanding or solving new phenomena. For example, in a research study with children between 5-8 years old where they had to intervene and identify not only a single cause-consequence situation, but a causal system and or a causal chain that made shapes spin (or not) together (McCormack et al., 2016).

(2) Exposing young students to science (in a positive view) helps them develop positive attitudes towards science. Positive attitudes toward science, according to Osborne et al. (2003), are any positive feelings, beliefs and values held about science and scientists. A categorisation also made by Klopfer (1971 cited in Osborne et al. 2003) who establishes that these positive attitudes entail developing interest in science and science related activities or careers, the acceptance of scientific inquiry as a way of thought, as well as enjoyment of science learning experiences and adoption of scientific attitudes. Furthermore, research has shown that an early interest and positive attitude towards science can contribute to future engagement with science and science careers (DeWitt et al., 2013; Lindahl, 2007). However, this interest once lost is harder to recover, thus early exposition to positive experiences and attitudes towards science in children from guiding adults has the potential to make a long-lasting impression on them (DeWitt et al., 2013; McClure et al., 2017).

(3) Exposing children early to science events and phenomena can facilitate the understanding of these same topics when encountered later in life, particular in formal education. At this early stage, one fundamental way of exposing these scientific phenomena for meaning-making is through play in science inquiry (Vartiainen & Kumpulainen, 2020) which can also lead to imaginary situations and re-enactments that help develop scientific concepts after practising science inquiry. There is also evidence from the EPPE project, that specific skills such as maths, and literacy when encountered in preschool settings can

provide a significant cognitive benefit that can be maintained throughout some years in primary education (Sammons, 2010). Therefore, it could be expected that early science learning can provide similar results, specifically since multiple encounters are more likely to reduce the uncertainty of an encountered phenomenon and could increase the exploration process behind it, potentially leading to a better understanding. As evidence for this Tao et al. (2012) compared results from children of two different countries at the end of primary school, China and Australia. In the former, science education does not start until primary school whereas in the latter, science is taught from kindergarten. They found that when children came from a high SES status there were no significant differences found in how children understood science, however, when medium and low SES was considered, a protective factor was observed in children from Australia who have had an early encounter with science, having a higher average than the children from medium and low SES that had not had any science experience before primary school. Although contrary evidence has also been found where no particular effect was found in primary school after kindergarten's use of science resources (Saçkes et al., 2011) These mixed findings highlight that these results have to be critically assessed, since it also suggests the consideration of the difference in evaluation outcomes and tools used in the different education levels.

(4) Use of scientific informed language at an early age influences the future development and understanding of scientific concepts. Language shapes thinking and prior knowledge influences language (Eshach & Fried, 2005), which means that both experience and language are mutually impacted by each other. Introducing new concepts and ideas reshapes previous knowledge and provides future referents for understanding similar concepts (Weisleder & Fernald (2013). Scientific language does not necessarily need to be only verbal, since research has shown that visuospatial referents paired with science language, such as drawings and gestures while explaining a concept/idea help increase

language by providing a visual explanation of the words and increase understanding (Roth, 2002; Thomas Jha et al., 2020).

(5) *Children can understand scientific concepts and reason scientifically.* Most scientific topics, theories and phenomena can be stripped from technicalities and be explained in simple terms to facilitate basic understanding (as reviewed in Eshach & Fried, 2005), however, just because a concept can be simplified, it does not mean that it will stop being complex, abstract or inaccurate. In the same vein, Gopnik (2012) has reviewed scientific thinking in young children and argues that from a very early age children can follow and understand basic scientific concepts, such as hypothesis testing, probabilistic models, correlation of events, causality and reflection about events. As an example, Gopnik describes a study from Xu and Garcia (2008, cited in Gopnik 2010) where eight-month old infants are shown a box with a red and white ball, then the researcher randomly grabs a sample of some balls from the box, the study evaluated whether children could identify when a sample shown was truly random (i.e. matching the distribution of balls inside the box) or if the distribution failed to match the random selection of balls. Infants tended to look longer to the box with balls when the sample distribution did not match. Young children have also been tested to make direct hypotheses after showing them two sets of evidence, for example which food would be more likely to make a character sick (Zimmerman, 2000). However, a critique from Eshach and Fried (2005) is that most research around scientific thinking in early years is done in “nonscience” scenarios, testing their ability to gather evidence and connect hypothesis to it, but not grounding it in specific science concepts only provides an incomplete view of the processes behind children’s scientific reasoning. As a solution they propose that early years’ educators facilitate an early exposition to scientific situations where these scientific skills can be exploited.

6) *Science is an efficient means for developing scientific thinking.* The authors propose that scientific thinking skills are better developed when these are grounded in

explicit scientific topics, or domain-specific knowledge, since they consider that scientific content cannot be separated from the conceptual domain of scientific thinking (Eshach & Fried, 2005). Although they still acknowledge the importance of the procedural aspects, since these allow abstraction to everyday processes, they consider scientifically grounded activities to be more efficient, like the influence of lights on plants. The authors consider this grounding provides children with conditions that allow variation of a single variable at a time and focus their attention and reflection on the meaning of the control and experimental variables. It could be equated with a concrete and contextualised grounding of scientific thinking that could lead to a more tangible understanding of science, which can then be abstracted and generalised. This aligns with Bruner's proposal (1966) of conceptual development, where children first act on concrete objects, which allows them to form representations of the concrete objects and their operations leading to incorporating symbolic representations. This notion also touches on Piaget's proposal of constructionism (1970), where children construct their understanding of their environment by interacting with it. While Piaget proposes the starting age for this to be between 7-8 years old, other researchers have found that concrete thinking appears from as early as 6 years old (Porpodas, 1987; Samuel & Bryant, 1984). McNeil and Uttal (2009) point to a remark by Sarama and Clements (2009) in their research around manipulatives and computing skills: "What matters is not whether a learning material is concrete or abstract, but whether and how students gain insight into the meaning or purpose of what they are learning" (p.139). This last remark echoes the initial statement by Eshach and Fried (2005) regarding the importance of science learning to teach the *know-how of the world*. It is important to see that these principles are all interconnected with one another and therefore are steps that are equally important to the development of the older child and future adult.

Furthermore, early science learning can also have a positive impact on young children's personal growth, such as helping to develop a sense of self by learning and

exploring based on their own curiosity to find their own perceptions and attitudes (Kumpulainen et al., 2014). One example of this is shown by children's metacognition of their own thinking, realising it is they who think and discover (Bowers et al., 2015; Kuhn, 2000), which can then lead them to look for answers to their questions and think if they have found an answer to them. This also helps them develop and improve their sense of agency and confidence, by practising taking decisions based on their interests and realising they are capable of discovering and exploring their surroundings on their own (McClure et al., 2017). However, it is important at this early age to scaffold and support children's learning and exploratory behaviour, while still allowing room for them to be independent (Van Schijndel et al., 2010). This scaffolding approach has been known to be effective for learning for many years now, echoing Vygotsky's proposal of the Zone of Proximal Development (D. Wood, 2018). This is particularly important, because the role the adult could play in scaffolding new scientific concepts by linking it to any prior knowledge and previous experiences children have, can both facilitate and strengthen the understanding of the learner (Cabe Trundle, 2010).

2.1.2 Contexts of Early Science Learning

Early science learning can happen across multiple scenarios from a very early age (National Research Council, 2009). However, each scenario is likely to provide different benefits and different opportunities for children to engage with science at varying ages. Three main contexts to distinguish where early science learning happens are: (1) home-based spaces, such as baking, playing, gardening, doing carpentry, etc. (2) formal spaces, such as nurseries, preschools or kindergartens, and early elementary school and (3) informal spaces, such as museums, science centres, zoos, botanic gardens and outdoor centres. Although all three contexts have relevance for children's early science learning, this work focuses more in depth on informal contexts, particularly science centres, since these spaces

allow children a freedom to choose what to explore and learn while grounding this knowledge on domain-specific concepts of science.

2.1.2.1 Home-based spaces Science learning very often happens in the home. It can be encountered through multiple activities, these can be normal everyday activities such as baking, taking a bath, or making a cup of tea. It can also be encountered spontaneously, for example when a child asks ‘why do fish die outside water?’ (National Research Council, 2009). It is also the first context where the child tests their inquiry skills and exploratory abilities before any formal or informal contexts (Gopnik, 2012). Although, for this science learning to be evident, adults play a role in helping children making these connections between the everyday activities and the science learning behind them. Research shows that parents or adults who do or enjoy science are more likely to expose and project to the children a positive view of science (Dabney et al., 2013). There can also be science-focused activities in the house, for example, the use of a telescope, playing with science toys and using “science kits” to experiment with. All these different activities would contribute toward developing positive feelings about taking part in science activities (National Research Council, 2009; Shwe Hadani et al., 2018). Within the home environment there is availability for science related domain-specific content skills to be achieved, for example growing a plant from a seed, as well as access to situations to develop the domain-general skills that drive inquiry. Science learning can also be facilitated at home by different media content, for example when watching TV programs or YouTube videos (National Research Council, 2009), whether these are cartoons that teach about science related concepts like “Ask the StoryBots” or through shows that describe nature like BBC’s “Blue Planet” (Fisch et al., 2014). Other types of science-related media that children could encounter include science magazines or books for children, which explain science concepts or talk about historical figures that have had an impact on science. All of these examples enable an early everyday contact with science that explores the reasons behind the world as well as providing access

to information for children to construct their own knowledge and understanding and avoid context dependency (Fisch et al. 2014). These different examples of media give children a variety of exposure to explore a range of concepts through a myriad of examples and learning opportunities.

2.1.2.2 Formal Spaces The formal spaces for learning are those that have an established curriculum or that set particular learning goals to be achieved throughout a particular period of time, such as a school year. These tend to be spaces where learning is directed to pre-established content. Formal spaces have the advantage of being continuous spaces where science topics can be revisited at different moments. Some countries are establishing a defined curriculum for science learning in preschool such as life sciences, earth/space and physical science (Trundle & Saçkes, 2012).

An advantage of formal learning spaces is being able to provide a constant space that grants enough time to revisit content to strengthen the understanding of certain topics (Dejonckheere et al., 2016). This increases familiarity of the science theme with the children, facilitating them to make connections between the science concepts reviewed and their day to day lives. Evidence for this is presented by a research study carried out in Denmark with preschool children training in formal scientific reasoning (Dejonckheere et al., 2016). In this research, an inquiry-based intervention with one group of 4-6-year-olds where they did 15 investigative activities (e.g. sinking and floating) over seven weeks, this group was compared to a control group. All the activities from the intervention had three phases: introduction, exploration and trigger. In the trigger condition the researcher directed the child's attention to the phenomenon by asking probe questions. This last phase also had the intention of encouraging children to explore the relevant variables for the activity and to test causes and consequences. The research found that in the post-test activity intervention children had better guidance towards the relevant variables, they had more informative explorations and

were more prone to set-up experiments to look for additional information compared to the control group. They also found that these children were more inclined to explore only one variable at a time, reducing the number of uninformative explorations from their pre-test results. Their findings strengthen support for incorporating a more inquiry-based approach in the curriculum from the early years onwards. Schools have been encouraged to use inquiry-based science instruction to foster the development of exploratory and scientific skills, such as curiosity, testing hypotheses, and asking questions (Cabe Trundle, 2010; Saçkes et al., 2011). For example, talking about children's comprehension and attitudes towards a scientific phenomenon in biology such as the lifecycle of a monarch butterfly when this topic is reviewed with an inquiry-based approach (Samarapungavan et al., 2008).

Early years researchers and practitioners have advocated that science-related content taught in formal education at a younger age should try to be connected with children's previous knowledge or to things and experiences they might encounter again on their own (Saçkes et al., 2011; Tuckey, 1992; Zimmerman, 2000). An example of how to connect children's previous knowledge while taking advantage of children's playful nature and imagination to learn science is provided by Vartiainen and Kumpulainen's (2020) *Poetry Science*. Preschool children engaged in *scientific play* by combining a shared story-reading/telling, and interaction with a puppet that helps bring forward their everyday experiences related to the explored science concept through poems and graphs. These authors describe that children's scientific play arises by engaging in science talk in hypothetical scenarios, combining imaginary situations and problem-solving where children are actively doing science practices and "creating new knowledge" (Vartiainen & Kumpulainen, 2020).

In all the aforementioned studies, the teacher played a key role in facilitating and scaffolding the lessons and play for children to both understand and engage better with the science lesson. As a consequence, one disadvantage of these contexts is limiting free choice

within the science topics to be revised (Falk, Storksdieck, et al., 2007). According to a report on STEM Starts Early (McClure et al., 2017), an important constraint in this context is a systematic problem in training teachers in STEM-related content, which explains a reduced science content, and which impacts confidence – regarding science – of the teachers while teaching and the children’s who are learning it (Early Childhood STEM Working Group, 2017; Osborne et al., 2003).

2.1.2.3 Informal Spaces Informal spaces are those where learning happens outside or around a formal learning context and have different expectations with respect to learning. These contexts provide leisure time while supporting science learning by providing examples of real-world phenomena that can be associated with everyday experiences where visitors can develop new science interests and practice their science inquiry to make sense of their world (National Research Council, 2009). According to Falk (2005), people also visit informal spaces when they are searching for new interesting experiences that satisfy their intellectual curiosity. Informal science learning spaces are children’s museums, science centres, zoos, botanical gardens, aquariums, and libraries (National Research Council, 2009) covering a wide range of scientific topics. These spaces are designed environments that first started being used to understand learning experiences outside the classroom from school field trips (Rennie et al., 2003). A defining difference between informal and formal spaces is that informal spaces allow children to encounter and experience things that they would not normally have access to in their daily spaces, for example, wild animals, already established props that demonstrate a physics experiment, or encounter a demonstration of how a black hole works (Schwan et al., 2014). All of these different informal learning science spaces provide visitors with different pathways to learning about science, for example, a science centre is designed for voluntary interactivity with its science content through exhibits (Heath

& Vom Lehn, 2008), allowing visitors to directly experience and explore specific scientific phenomena. These experiences can also prompt further science-related activities that can be carried out at home or explored in formal contexts (Schwan et al., 2014). Informal spaces facilitate access to public understanding of science, research and science literacy (Schwan et al., 2014). Thus, these spaces promote scientific engagement in the early years allowing young children to experience and explore a myriad of scientific content from a young age.

From all these informal science learning spaces, science centres are spaces exclusively created and designed for visitors to experience science, unlike other spaces like zoos or botanical gardens where they also have the function of preservation of living specimens (Schwan et al., 2014). Science centres share pedagogical aims regarding science learning with formal schooling, whilst the latter are more explicitly delivered through teaching and directed facilitation and the former are enabled through free exploration (Faria et al., 2015). Thus, science centres provide an opportunity to understand how visitors explore, learn and understand science content where they have the freedom to choose what they want to learn about but without the pressure of providing any results or any other prejudices that could be experienced in formal contexts.

Science centres are designed to be interactive exploratory spaces where visitors can experience different science concepts, promoting scientific skills and disseminating science content delivered in unexpected and surprising events (Allen, 2004; Gutwill, 2008; National Research Council, 2009; Tisdal & Perry, 2004). One of the biggest attributes of science centres is their interactive and hands-on exhibits and spaces (Rennie & McClafferty, 1996). Interactive exhibits are exhibits that generate an action-reaction loop between the visitor and the exhibit, where the visitor carries out an action in the exhibit and as a consequence the exhibit not only reacts to the visitor but facilitates opportunities for further responding to reaction from the exhibit. For example a button that when pressed displays a light, but there are other items that can be used to interact with the light to understand the phenomenon of

reflection. Contrary to reactive exhibits where there is no invitation for additional action from the visitor, for example an exhibit where a button is pushed and starts a whirlpool in a tube of water (Allen & Gutwill, 2004; Rennie & McClafferty, 1996). Although some authors warn that interactivity should be used with moderation to maintain a good balance between the interactivity loop with the exhibit and the social interaction and meaning construction around the exhibit (Allen & Gutwill, 2004; Heath & Vom Lehn, 2008). For example Heath and Vom Lehn (2008) argue that some computer-based exhibits that rely on single users with an audience can hinder co-participation, whilst also providing a false sense of choice where only a limited number of possible actions are offered which could constrain visitors' engagement. Allen and Gutwill (2004) describe some of the common pitfalls of interactive exhibits design, for example having too many options of equal salience or secondary features that conceal the main features of the exhibit, which can undermine the main educational contribution from the exhibit; while having multiple users that can interfere with one another or disrupt the phenomenon displayed can also curb engagement and cause frustration. Therefore a sweet spot is needed when designing interactive science exhibits to maximise the cognitive and socio-emotional benefits that these spaces can provide.

Although science museums and science centres started out as object-based spaces that displayed objects and instruments that were used to carry out science, it has now changed, becoming spaces where science ideas are demonstrated (McManus, 1992). According to Sue Allen (2004), a museum researcher at the Exploratorium in San Francisco, these informal learning spaces have the responsibility of being entertaining places that allow visitors freedom in following personal interests, while also being expected to be a place where some form of learning would be achieved following a visit, whether it is science-themed, affective or about introspecting after experiencing a particular visit. Paris (1997) also points to the motivation that visitors have towards topics that may be personally relevant. He considers four factors that motivate a visitor to engage with informal science learning: (1)

construct personal meaning, (2) challenges, (3) control and (4) positive consequences. One of the main characteristics of an informal learning space is the free-choice in learning that match their interests and allows them to explore it at will (Falk, 2005; Falk & Dierking, 1998) This freedom of choice grants children certain autonomy and agency in what topics they decide to engage and discover, and for how long they want to explore a learning space. However, the child never interacts alone, since according to Falk and Dierking's (2000) Visitor Experience Framework there are three relevant contexts at simultaneous play that influence a visitor's learning: (1) physical context – which considers the physical space, the architecture and design of space, the apprehendability and the exhibit organization. (2) sociocultural context – takes into account the social interaction happening around the exhibit or museum, that is the group the visitor comes with, or the groups present at an exhibit when they arrive, any relevant conversations or interactions arising from the exhibit and any relevant social facilitation. (3) personal context – refers to the individual's prior knowledge, interest, motivations, expectations and preconceptions. Rennie and McClafferty (1996) used this Visitor Experience Framework as a base to design a successful visit to a science centre, since it explains why visitors learn and enjoy themselves, in both the cognitive sense and the affective sense, while also being influenced by the surrounding socio-affective situations.

Research around science learning in science centres has addressed a diversity of topics, mainly focused on different learning gains from the different exhibits by observation and survey analysis from the visitors, for example, both Ben-Eliyahu et al. (2018); and Shaby et al. (2017) followed a group of visiting children from a school as they wandered around a designated section of the science centre. Another commonly researched subject in science centres is the comparison of different exhibits to detect relevant features for increasing engagement and learning gains with them (Dancstep née Dancu et al., 2015; Gutwill, 2008). While others have focused on exhibit features, such as labels and facilitation of the science content (Denham et al., 2012; Gutwill & Dancstep (née Dancu), 2017; Serrell, 2020) or to find

better measures to compare exhibit performance across a variety of different exhibits, for example using the square footage and holding time of an exhibit (Serrell, 1997b, 2020). However, this research usually focuses on older children or adults, who are the main target audiences of these spaces. Likely, one of the main reasons early years in science learning has not been as explored has to do with the intrinsic challenges of researching informal science learning, which according to Andre et.al (2017) tend to have many confounding factors attributable both to the age of participants and to factors that influence them. Additionally, challenges of working with children in the early years should be considered, such as planning and designing activities to also take into account children's shorter attention spans and longer reaction times when it comes to decision-making (Akshoomoff, 2002), as well as ensuring children have enough verbal skills to understand, participate or provide any feedback in research studies (Yarosh et al., 2011). Although one relevant project that is targeting this limitation is the Move2Learn project, which is aimed at exploring how preschool-age children could use their body movements and gestures to learn about science in science museums/centres and how to use these to communicate their science understanding better than when using verbal vocabulary, (M2L Team, n.d.-a; Thomas Jha et al., 2020).

Therefore, the literature in informal science learning presents a gap when it comes to science learning in the early years. This gap is exacerbated when targeting specific processes for science-learning, such as engagement. Addressing both of these gaps could provide an increased understanding of how science learning develops in different contexts. This is particularly important because behaviour is not the same in a school setting as in a science centre, focusing on engagement in an informal context will have differences and it is important to understand how this context is promoting and enhancing early science learning through engagement. Barriault and Pearson (2010) emphasize this when highlighting what should be the focus during research with informal science learning: "assessing the learning

outcomes in science centres needs to not only focus on cognitive gains, but must take into account the conditions, the processes and the engagement that lead to learning”. For example, understanding how engagement with a science exhibit starts and develops is a good starting point to further understand its potential role when it comes to free-choice learning. That is, how does the situated engagement could help facilitate the learning process by giving and maintaining a space (mentally and physically) to further develop relationships between what is known and what is being discovered throughout a particular science learning experience.

However, the definition behind engagement and the research about its relationship with learning should be addressed first.

2.2 Engagement

2.2.1 Engagement and Learning

Engagement has been highlighted as a needed and inherent element in any learning process, regardless of the different learning scenarios (Azevedo, 2015; Boekaerts, 2016; Falk & Dierking, 2000; Shaby et al., 2017; Sinatra et al., 2015; E. Wood & Wolf, 2008). Boekaerts describes engagement when referring to students as “the basic processing operation that describes *how* students react to and interact with the learning materials, and with people present in the learning environment to enhance domain-specific knowledge and skills” (2016, p.81). Although taken from a formal learning environment, this description could be translated to visitors in informal learning scenarios, such as science centres, since learners have to interact with science exhibits (i.e. the learning materials). While learners interact, they would also experience other people in a particular learning environment

inherent to the science centre, an environment that will also depend on the type of science exhibit being used.

Learning science in informal contexts, such as a science centre, is shaped by a free-choice and self-paced environment that is voluntarily driven by the visitor's interest and motivation to learn in a contextually rich environment (Falk & Dierking, 1998). It also considers that learning can be socially constructed from the interactions between the visitor and its social environment (Falk & Dierking, 2000). Learning has also been related to play and exploration in children, showing that play leads to exploration, which leads to developing scientific skills (Zimmerman, 2000). Having multiple choices to explore offers more possibilities for visitors to ask questions and explore the answers, which according to Perry (1993) will lead to visitors learning something. However, she argues that the obtained learning will not always be something new, other times learning might be in the shape of increased awareness, a new interest, or improved psychomotor, social or affective skills.

In science centres, engagement has been seen as the gateway to learning or a pre-requisite needed for learning (McClure et al., 2017; Serrell, 2020). Engagement is placed as the process that follows after a visitor's interest and curiosity have been captured, the combination of these two would then trigger the learning process (more is discussed in following sections). The National Research Council in its review of Learning Science in Informal Environments mentions that "engagement creates an opportunity for learners to experience a range of positive feelings and to find meaning in relation to what they are learning" (p. 58). In other words, engagement is perceived as an individual decision that then drives a visitor to act in response to their interests and motivations, in turn influencing their learning experiences and outcomes after interacting (Renninger & Bachrach, 2015). This conceptualization of engagement is so recurrent that interactive science exhibits are designed and built to create engagement with the assumption that this would facilitate learning (Gutwill, 2008; Shaby et al., 2017). For example, Gutwill (2008) mentions that

counter-intuitive exhibits, those that present discrepant event between what is common knowledge and what is being observed, creates an immediate hook for the visitor's attention and creates in the visitor the need to resolve and understand that phenomenon, ideally then leading to an inquiry process that accommodates new information or restructures prior knowledge. Whilst Gutwill (2008) argues that these exhibits are not always successful in fostering longer-term inquiry because the phenomenon shown is only possible in unique circumstances, which reduces exploration and hinders any connection with prior knowledge. Contrarily, Gutwill (2008) argues for open-ended exhibits that foster active prolonged engagement as better suited to encourage playful exploration that leads to learning with a positive affect. Although research has shown that some exhibits designed specifically to be engaging and enjoyed, like immersive exhibits, will not always be related to greater knowledge gains (Dancstep née Dancu et al., 2015), whereas exhibits that foster active prolonged engagement, like open-ended exhibits, facilitated visitor's exploration and intellectual inquiry about the topic of the exhibit leading to greater learning gains about relevant concept compared to immersive exhibits demonstrating the same concept.

Research in a science centre by Shaby et al. (2017) showed that learning can also be observed through the visitor's behaviour and their verbal responses to the science exhibits. They looked at 1800 children between 10-and-12 years old from school visits to a science museum, and after examining 200 children in 9 different exhibits, found that children made connections of relevant information to their prior knowledge. They also found that familiarity and apprehendability (ease of use) with the exhibit lead to faster use and longer engagement time, also increasing social group engagement was observed to lead to increased understanding of the topic being observed.

One explanation behind the relationship between engagement and learning might be the link between exhibit content and a visitor's prior experiences, which has been observed when parents or science facilitators elaborate on a topic by asking children open-ended

questions to help them reflect critically on the phenomenon (Benjamin et al., 2010). Thus, when conversation links children's own experience with the science content they are experiencing, their conceptual engagement (i.e. engagement relevant to the content that is being experienced) flourishes (Callanan et al., 2017; Shwe Hadani et al., 2018).

Sustained attention has also been related to effective learning since it involves a cognitive process where a person is actively and uninterruptedly interacting with a task (Erickson et al., 2015; Sarter et al., 2001). Thus, the relationship between engagement and learning could be because sustained attention has been studied as a proxy or equivalent for engagement (Akshoomoff, 2002; López et al., 2005). As evidence, Serrell (1997) has used time spent in an exhibit to measure engagement, which generates more opportunities to explore and gain a deeper understanding of prior knowledge or discover new information.

Many reasons have been discussed how engagement is relevant for the many shapes learning can take in informal scenarios, however, there is a key challenge in improving our understanding of this relationship: the interchangeable use of some terms with engagement (i.e. motivation, interaction, attention, self-determination, curiosity, interest, etc.), which causes a lack of consistency in the definition and methods used to measure and capture engagement (Christenson & Reschly, 2012). Even more so, the fact that the term engagement is commonly used in a myriad of learning fields (e.g. digital engagement, student engagement, visitor engagement, civic engagement, customer engagement, user engagement, etc.), but meaning slightly different things in each of them further complicates the problem. With each field/researcher adopting a slightly different use of the term engagement, the meaning, the specific associated actions, and the tools for measuring it will vary. This negatively impacts the generalisation of results or the development of a more unified theory that could contribute to discerning the role engagement plays with other cognitive and social processes.

Given the importance highlighted of science learning in the early years, it is valuable to gain deeper understanding of how engagement works in this key developmental stage. This could also lead to explore the relationship between early years engagement and science learning. To achieve this, it seems necessary to bridge some of the operational and ontological definitions, that translates into better measurements, methods and theories used and provide a comprehensive framework.

2.2.2 Defining Engagement

A lack of an agreed ontological definition or theory around engagement causes the term to be used inconsistently by making it mean the same and different things, which impacts in a lack of clarity when it comes to capturing, measuring, and comparing results around engagement (Fredricks et al., 2004). Monique Boekaerts specifically reported this in her commentary on a special issue of the Learning and Instruction Journal called *Student Engagement and Learning: Theoretical and Methodological Advances* (2016). She found that most of the papers included in the special issue did not provide a working definition even though they were all focused on student engagement, and due to a missing theory of engagement, authors borrow and explain engagement through different psychological theories and frameworks. There was also barely any consensus on the boundaries around the construct of engagement and not enough conceptual clarity to reach an agreement. Although she did find that all the researchers among its descriptions of engagement, agreed that “engagement is malleable and not stable across learning situations and subjects” (p.77). Unfortunately, this only adds to the illusion of agreement, as well as to what Christenson and Reschly (2012) called a *jingle/jangle* fallacy, where two things with the same name mean two different concepts or two concepts with a different name, might be meaning the same thing.

It is necessary to understand how the term engagement is used in ordinary language to appreciate how some connections and assumptions start being made to the word and its

related constructs (Sheppard, 2011). Understanding how the definition of a word is currently being used in a conventional conversation by referring back to their lexical definitions (e.g. dictionary definitions) provides information about the different contexts and or meanings associated with it, as well as the different ways in which it can be applied (Michie et al., 2019; Swedberg, 2020). Engage is derived from the Old French word *engager*, which means “to pledge” or “in bind of something”; one of its associated meanings is “to attract and occupy the attention of” (Cresswell, 2009; Hoad, 1996). Sheppard (2011) traces the use in the ordinary language of the word *engagement* and some of its derivatives showing two main strands of use and meaning for the term *engagement*: one points to a voluntary commitment to something that the individual has perceived as having intrinsic value, thus the use reflects some individual agency and responsibility when it comes to making a choice or a promise, also called substantive. While the other points to activities, items or experiences that are designed to persuade or attract the individual because of their extrinsic value, thus here, the responsibility and agency do not emanate from the individual but rather it is conferred to them and is also referred to as procedural. Hence the importance of having an operational definition when it comes to research, since it provides clarity and avoids ambiguity (Boring, 1945). Furthermore, definitions can play an active role when it comes to empirical research because they are “used as a prescription for what to study” (Swedberg, 2020, p.437). Boekaerts (2016) also argued that definitions of engagement are tied to the method used to capture it, as well as to the availability of these instruments, while J. S. Eccles (2016), one of the pioneers in the field of student engagement since the 90s, reminds researchers in engagement that any data obtained by specific measures is “inherently limited by the range of items assessed and the population selected to answer them” (p. 72). In other words, a clear and concise definition would shape and provide limits to the research construct, it would dictate what can be or not measured in order to fulfil the research needs.

Eccles also presents a commentary in the same Special Issue (2016), where she presents two metaphors to describe what she considers the two main problems with the definition and conceptualisation of engagement. The first, “rubber meets the road” where a car’s rubber wheels represent the different components of engagement and the road symbolises the context in which engagement is happening, both the specific activity and environment. According to this “engagement is assumed to be a momentary, emergent property derived from all of the ways in which a person could engage in the moment in an activity or a contextual setting”. This metaphor highlights the importance of the context of study when defining engagement, it emphasizes engagement as situated and dependent on its defining terms as much as the context where is elicited. Whereas the second metaphor references “the blind men trying to represent an elephant”, where each of the men grabs different unique parts of the elephant but is unable to perceive the elephant’s totality. Eccles references the blind men as the different research fields that work with engagement where she questions the production of the different quantitative and qualitative measures of engagement as to whether these can capture the emergent property of engagement discussed in the previous metaphor. Thus, the second metaphor echoes the challenge of having multiple measures of engagement but no unifying view of engagement.

Recognising there is no agreed definition for any construct or term is important, since it acknowledges how unlikely it is to be able to include all the phenomena associated with it. Nonetheless, in these cases an operational definition could work better, or what Swedberg calls a “stipulative” definition, in which the author decides what the definition will be, but also is very clear about it, allowing them to bring clarity and precision to the meaning, as well as providing room for a critical assessment of previous definitions and possible improvement and refining of the term (2020). In his words: “a stipulative definition is well suited for the reconceptualization of a concept” (Swedberg, 2020, p.435). Consequently, having a clear operational definition of engagement also provides a space to explore what other elements

constitute a construct, or how it could be interacting with other possible constructs or processes happening in the individual and the environment.

Several attempts have been made to review the concept of engagement and come out with a conceptual framework to give a definition that works with learning (Finn & Zimmer, 2012; Fredricks et al., 2004; Murphy et al., 1996; O'Brien & Toms, 2008; Perski et al., 2017; Turner, 2010; Whitton & Moseley, 2014a). So far, the most accepted and replicated definition used, at least in the field of education, is the one from the review paper of Fredricks, Blumenfeld and Paris (2004) which sees engagement as a “meta-construct” that implies engagement as commitment (to a varying degree) that “results from an interaction of the individual with the context and is responsive to variation in environments” (p.61). They also propose that engagement is a multifaceted construct that has a behavioural, a cognitive and an emotional component, which work together to provide a richer understanding of the nature of engagement and how learners behave, feel and think about something. However, even this key paper eludes having a formal definition of what they consider engagement since they recognise that a definition of engagement may more likely involve terms related to either one of the components they propose or would tie it to a specific context. For example, Seifert and Metz (2017) describe engagement as the degree of attention, curiosity, interest, optimism and passion that students show when they are learning or being taught. In this example, the definition could align with the multi-faceted construct proposed above by categorising its different elements into one of the proposed facets of engagement: attention and curiosity as cognitive engagement and optimism as emotional engagement. Thus, this vague conceptualization would be a strength that has allowed research in the field of engagement to move forward, even with its limitations, it still leaves a space for a clear definition of engagement. Other research papers on engagement tend to provide an operational definition, if at all, that is only aimed towards the particular context they are focused on (e.g. student engagement, user experience and engagement, digital engagement) which helps

clarify the meaning the authors have for engagement and the way in which the results can be used. The latter facilitates future readers to decide if the definition is applicable to their research in either formal or informal learning scenarios.

Furthermore, Elizabeth Wood and Barbara Wolf (2008) examine the concept of engagement across museums and science centres, where they describe how science centres and museums have looked at the Learning Sciences for agreement on a definition of engagement, since engagement becomes, what Sawyer (2005) calls “a measure for a hybrid form of cognitive and affective learning” but also as an institutional measure for impact. E. Wood and Wolf argue the need for museum practitioners to agree on what engagement looked like in their particular context since this would lead to aligning the design and development of experiences with the aim of the institution. E. Wood and Wolf also regret that engagement has been seen more as a goal that needs to be achieved, rather than an accompanying process of learning. This conceptualisation also complicates a full understanding of a visitor’s behaviour when entering a museum space. However, from different research carried out in museums and science centres, it could be noted that they have defined and conceptualised engagement as a form of active participation and interaction throughout their visit, considering both the museum space and the accompanying group of an individual (Baniyamin & Rashid, 2016; Ben-Eliyahu et al., 2018; Falk & Dierking, 2000; E. Wood & Wolf, 2008). Furthermore, it is relevant to the museums and other informal learning institutions studying their visitors, both their behaviour and their learning, to understand visitors’ exploration patterns, and what the individual’s process of meaning-making is while connecting previous knowledge to current information encountered (Andre et al., 2017; Fu et al., 2016; Serrell, 2020; E. Wood & Wolf, 2008). Informal learning contexts have also acknowledged the importance of different factors that surround engagement, like their cognition, their interaction with the environment and the emotions that result from both

of these, thus putting in question the one-dimensionality of engagement (Tisdal & Perry, 2004; E. Wood & Wolf, 2008).

Engagement, just as situational interest, cannot be taken out of the context of where it is situated (Eccles, 2016; E. Wood & Wolf, 2008), especially since the result of engagement will drive the person to strengthen their interests on a topic, or develop new interests which would lead them to reengage with in the future (Renninger & Bachrach, 2015). It can also have the opposite effect, where a lack of engagement can cause the individual to disengage and lose interest in a topic completely (Finn & Zimmer, 2012). Thus, engagement can be seen as a construct that encompasses the environment and the individual at a specific moment in time, as well as the combination of processes that are keeping the individual in that activity. Therefore, there seems to be agreement in the literature about the benefits of adopting a multicomponent view of engagement that unifies not only the outcome of engagement but also the process it goes through. This would lead researchers to use more appropriate measures that can align with the definition proposed for engagement, but also to the use of measures that can be extracted and adapted for different learning contexts without losing track of the particularity of engagement within the specific context that is being researched.

For the purposes of this thesis, engagement is defined as a multicomponent construct that will be characterised according to where it is situated and what the intended learning outcome is. In this research, engagement is situated within the context of a science centre environment and is defined as a process that results from an active interplay of interest, curiosity, sustained attention, situated thinking and emotional attachment that allows an individual to ground themselves physically, mentally and emotionally during the time and space of an activity.

2.2.3 Differences Between Engagement and Other Constructs

Multiple constructs throughout the literature have been used in close relation to engagement, where it is not always being clear whether these are considered within the definition of engagement, if they are being used as interchangeable synonyms, or which role these other constructs could play in relation to engagement. (Heath & Vom Lehn, 2008). An example of this would be the term *interaction* and *interactivity*, closely related to engagement, but conceptualised as a facilitator of engagement by promoting direct action on the direct environment (Adams et al., 2004). This has been mainly established by research in museum exhibits and technology-based learning that frames interaction as the direct observable action of engagement since it is assumed to be the intent behind this (Falk et al., 2004; Rennie et al., 1993). Thus, the concept of interactivity tends to reflect an observable behaviour between a person and an object or activity they are currently involved with (Rennie & McClafferty, 1993), whereas engagement takes into account underlying cognitive and affective processes that are elicited due to that interaction. This is reflected on the definition of Adams et. al (2004) of interactivity as “a range of experiences that fully engage visitors personally, physically and emotionally” (p. 157-158), thus also acknowledging the multicomponent nature of engagement.

Motivation is another term that is constantly mentioned when researching engagement (Eccles & Wang, 2012; Fredricks et al., 2016; Pham & Avnet, 2009; Weston et al., 2015), particularly when looking at engagement as a goal achieved over a period of time. Sometimes the term motivation is mentioned as related to engagement (Ben-Eliyahu et al., 2018; Eccles, 2016; Eccles & Wang, 2012; Rennie & McClafferty, 1996; Renninger & Bachrach, 2015), and other times it is used to mean engagement by incorporating it as one of its direct components (Pekrun & Linnenbrink-Garcia, 2012) or as an equivalent (Martin, 2009). This lack of congruency can become problematic when it comes to researching either of these, which highlights the importance of differentiating it as a construct separate from

engagement. Motivation has been related to learning in three main theories: the self-determination theory (Deci & Ryan, 1985) where they address if a person's capacity to fulfil their goals or desired outcomes comes from a reward from the environment (i.e. extrinsic), or from within (i.e. intrinsic). The task-value theory looks specifically at student motivation as a result of the individual's beliefs, values and goals behind a specific behaviour (Eccles & Wigfield, 2002). In self-efficacy theory, motivation is seen as the driver of an individual's own beliefs about successful achievement of their outcomes (Bandura, 1986). Overall, in the literature motivation is related to an individual's inner experience, their beliefs, feelings, and thoughts concerning a situation (Meyer & Turner, 2006). Even more, motivation and engagement could be both influenced by similar factors such as feelings of competence, autonomy and opportunity to act. A reason for this is provided by Fredricks & McColskey (2012, p.3) after analysing results from Skinner et al. (2009), where they say: "engagement tends to be thought of in terms of action, or as the behavioural, emotional, and cognitive manifestations of motivation" (p.3, 2012). According to Christenson and Reschly, (2012), motivation is necessary but not sufficient for starting engagement. While for Ben-Eliyahu and colleagues (2018), in their research of the relationship between motivation and engagement, motivation is different from engagement processes, these are produced by motivation, for which they conclude that motivation "instigates engagement" (p.89). In conclusion, motivation could be an emotional aspect that leads to engagement with an activity and engagement involves the emotional aspects during an activity.

Interest is also closely related to engagement, which has been defined as the likelihood of investing energy in one (specific) set of stimuli rather than other (Csikszentmihalyi & Hermanson, 1995), which is different from *curiosity*, which guides attention towards novel stimuli (Oudeyer et al., 2016). Some authors have presented interest as a driver of engagement (Mahatmya, Lohman, Matjasko, & Farb, 2012), whereas others proffer engagement as a consequence of interest (Hidi & Renninger, 2006). Interest has also

been defined as “the psychological state of engaging or the predisposition to reengage with particular classes of objects, events, or ideas over time” (Hidi & Renninger, 2006). Thus, engagement seems to be the physical or active manifestation of interest development as it is the process needed to create attachment and predisposition to certain events contrary to others. Similarly, it is important to distinguish between *interest* and *interests*: interest has more to do with curiosity, novelty and that something might seem appealing enough to attract the attention of an individual in a particular situation, also known as situational interest (Flowerday & Shell, 2015; Hidi & Renninger, 2006). On the other hand, interests would be when something becomes rewarding and it is more likely for the person to reengage or revisit it more than once, leading to incorporating it into the attitudes and personality of the individual. Influencing the person’s interests will help to guide and narrow attentional focus towards selecting something that is known to be more attractive for them. Moreover, interest when it is situational or topical has been found to influence, directly and indirectly, learning and engagement (Flowerday & Shell, 2015). Therefore, interest can be both a precursor and an outcome of engagement.

Given the complexity of the interplay between engagement, its precursors and its possible outcomes, it is key to remember that clarity is needed when it comes to operational definitions since these shape the limits around a construct as well as guide the development of any future tools to capture each of these constructs (Eccles, 2016). Richard Swedberg (2020) warns that when different definitions of the same term are used, they risk falling into the *illusion of agreement*, that is, only superficial agreement comes out of using the same word but giving it a different meaning. This could be one of the potential reasons for the perceived “messiness” behind studying the concept of engagement as mentioned by Fredricks et al.(2004) and why it is so important to provide its operational definition.

2.2.4 Components of Engagement

Having established that engagement is a multifaceted and complex construct (Ben-Eliyahu et al., 2018) in any learning scenario, evidence suggests that there are both domain-general skills – applicable to any learning context, and domain-specific skills – related to engagement in a particular learning context (Sinatra et al, 2015). Specifically, the 3 main components that are simultaneously present or active simultaneously are *behavioural*, *cognitive*, and *emotional* and were first proposed by Fredrick et. al (2004) in a review of student engagement. To understand the underlying nature of these components, first it will be addressed what is meant by cognition, emotion and its relationship with behaviour.

Cognition is the combination of inner processes through which humans can experience and interact with the world by acquiring, storing and transforming information from these experiences (Eysenck & Brysbaert, 2018). In other words, it refers to perceptual processes (using the senses), that can guide action and reaction with stimuli from the environment through sensory-motor processes (i.e. behaviour) and creates a representation of the surrounding information, which when the information is relevant for the organism, it is encoded and stored in memory for future use and reference. Another aspect of cognition is being able to access and retrieve the encoded information, while also using it when previous or novel experiences are encountered (Glass, 2016). Cognition also involves higher-order processes, some that may be unique to humans or primates, such as language or executive functions which allow them to think, problem-solve, learn and reason (Eysenck & Brysbaert, 2018). These higher-order processes give individuals more voluntary control and regulation over their thoughts and actions. Thus, due to cognition individuals can navigate and explore their environment, they can choose and self-establish goals and plan how best to achieve them through their interaction and exploration of their environment,

Emotion research has divided the definition of emotion into two main aspects: the emotional state and the conscious emotional experience (Tsuchiya & Adolphs, 2007). According to Damasio (2001) “the term ‘emotion’ should be rightfully used to designate a collection of responses triggered from parts of the brain to the body, and from parts of the brain to other parts of the brain, using both neural and humoral routes”(p.103), and these emotions can originate from internal or external events (Scherer, 2001). Emotional states are the end result of this array of processes and are envisioned as changes within the body (e.g. viscera, brain) (Damasio, 2001). For Tsuchiya and Adolphs (2007), this emotional state also encompasses behavioural and cognitive processes along with the physiological ones, which also lead to neural responses for an appraisal and preparation of an ‘action-response’ behaviour (e.g. fight/flight), although these latter may not always be conscious (Scherer, 2001). On the other hand, the conscious experience of this emotional state can also be referred to as ‘feeling’ (Damasio, 2001; Tsuchiya & Adolphs, 2007). According to Damasio, this is a complex ‘mental state’ that includes a cognitive representation of the psychophysiological changes that were occurring, the alterations in cognitive processing, as well as the associated context, thoughts and/or previous experiences surrounding the emotional state.

When it comes to differentiating between emotions there are also multiple perspectives that could be considered: some divide the emotions according to different dimensions (e.g. pleasantness-unpleasantness), others argue that there are discrete and unique different patterns that characterise specific emotions (e.g. anger, joy, sadness), while others propose that emotions are not specific, but rather constructed from core affects (positive/negative), relying on what is ‘known’ about the experienced emotion/affect which will have a varied pattern according to the situated context in which this emotion arises (Barrett, 2011; Scherer, 2001). The main neural areas associated with emotion are the cingulate cortex, the amygdalae, the hypothalamus, the prefrontal cortex (ventromedial and

orbital regions), somatosensory cortices (primary and secondary), the insular cortex, hippocampus and other areas in the brainstem and spinal cord in charge of afferent/efferent signalling to the peripheral system (Damasio, 2001; LeDoux & Phelps, 2008; Rolls, 2019; Tsuchiya & Adolphs, 2007).

Ultimately, the role of emotion is to help an organism survive by helping them do a fast appraisal of whether to approach or avoid certain stimuli, whether that is food, locations, or other living beings (Damasio, 2001). In humans, emotions can also contribute to the decision-making process since these can be paired during learning situations (e.g. context situation when an emotion arose, the outcome of a particular decision, or available/selected options when solving a problem), therefore when a similar situation arises, emotional experiences or emotional states are recalled (conscious or non-consciously) along with the cognitive memory and can help guide the decision-making process when reason is not enough (Antoine Bechara et al., 2000). This emotional 'memory' that is elicited and intuitively used while making decisions is called the Somatic Marker Hypothesis and was proposed by Antoine Bechara, Hanna Damasio and Antonio Damasio (Antoine Bechara et al., 2000; Antoine Bechara & Damasio, 2005). These authors propose that when humans are reasoning or having to make decisions they are supported by overt and covert processes, such as somatosensory perception, attention, working memory and emotions, and ultimately depend on having access to different types of knowledge which can be innate or acquired (e.g. actions and outcomes, contextual situation, facts, bioregulatory processes and resulting body states, and linkages between all the above). However, the access to this knowledge is not as a direct recall, but rather as a 'dispositional knowledge storage' (i.e. implicit), through the ventromedial Prefrontal Cortex, which is an area related "to holding links between facts that compose a given situation, and the emotion previously paired within an individual's contingent experience" (p.296) (Antoine Bechara et al., 2000). What the authors mean as dispositional knowledge, is not necessarily recalling the specific fact or emotional state, but

rather like a short-hand to being able to re-activate that linkage between an emotion or bio-regulatory process (i.e. physiological arousal such as increased heart-rate, facial expressions, skin sudoration), which is also commonly known as 'gut-feeling'(Antoine Bechara et al., 2000; Antoine Bechara & Damasio, 2005; Damasio, 2001). In other words, emotions could be reflected through facial and bodily expressions as much as by psycho-physiological processes that have been associated with a specific event.

Cognition, emotion and behaviour cannot exist without one another or without the respective environment in which an individual is situated. Since cognition and emotion are a set of inner processes through which there is an interaction between the individual and their environment, it is through the physical actions of behaviour that these interactions with their surroundings are facilitated. Each action/reaction with the environment simultaneously generates an emotional response to evaluate the recent experience and a combination of inner cognitive processes to judge what the best next action is needed in response. This entails identifying the experience, whether this has been encountered before, evaluating if the previous encounter fits the current situation (this evaluation involves stored memories along with their corresponding associated emotional experiences) and whether a next action or plan of actions is needed as a response or whether to continue to previously established goals.

A recent theory in cognitive sciences that envisions the body as an 'acting' entity which is not separate from the brain and its cognitive/emotional processes is Embodied Cognition (Borghi & Cimatti, 2010). This theory proposes that cognition is situated, grounded and constrained by the different bodies and environments each individual inhabits, where some cognitive processing is off-loaded and stored in the interaction between the body and the environment (Anderson, 2003; Clark, 1999; Pulvermüller, 2013; Wilson, 2002). It is supported by research evidence from psychology and neuroscience studies on the role of the body and the self on cognition, such as how language around objects/emotions is associated

to motor actions and activated simultaneously in the brain even when there is no direct movement involved (Borghi & Cimatti, 2010; Calvo-Merino et al., 2005; Clark, 1999; Novack & Goldin-Meadow, 2015; Wakefield et al., 2019; Wilson, 2002). Recent research has also looked at embodied learning and cognition in how children learn, understand and communicate science and maths-related learning (Broaders et al., 2007; Ghazali-Mohammed et al., 2020; Goldin-Meadow, 2018; M2L Team, n.d.-b; Shapiro & Stolz, 2019; Thomas Jha et al., 2020). Embodied Cognition could be a helpful framework to understand why engagement has been studied and conceptualised as a multicomponent process and as a domain-general process, particularly since engagement has also faced the decomposition of its different parts (the body, the mind and the environment as equivalents) while agreeing that they work simultaneously. It has also been acknowledged that each of these components can act as a unique domain-general process (e.g. cognition or emotion). Not only that, but it also highlights the importance of considering the situated context of any learning activity when researching and understanding engagement, taking into account that each individual will have a different lived experience of a learning activity both because of their unique social and bodily situations. Moreover, it also emphasizes that a multi-component perspective can provide a broader picture that facilitates understanding of the relationships between the different components that lead to a combined domain-general skill.

In the literature, engagement has been studied and described as a multi-dimensional or multicomponent construct, where its main separate components for analysis are behavioural, engagement, cognitive engagement, and emotional engagement. The following descriptions are a synthesis of how each of these components have been used and described according to the literature (Bailey et al., 2017; Deater-Deckard et al., 2013; Fredricks et al., 2004; Halliday et al., 2018; Silpasuwanchai et al., 2016).

Behavioural engagement reflects the physical interaction, manipulation of physical/virtual environments or active participation during an activity, and it could be defined

as any type of action or movement related to the activity that is taking place. This is the type of engagement that is most reported and measured because it can be observed and evaluated from an external point of view (i.e. the researcher/practitioner).

Cognitive engagement manifests itself through situated thinking, that is, conscious and purposeful thought processing relevant to the task at hand, mainly higher-order cognitive functions which are in charge of regulating behaviour such as focused attention, staying on task (inhibitory control and working memory), thinking about the task (problem-solving) and how to reach the task goals (planning).

Emotional engagement often refers to the feelings and emotional expressions observed or reported throughout the activity. These can be positive (enjoyment, happiness, excitement) or negative (frustration, anxiety, sadness) and would generally work as enhancements or impediments to an effective engagement. Emotional engagement has also been studied as arousal, or a physiological activation, using sensors for heart rate and electrodermal activity,

The first component, behaviour, is the easiest to measure since it relies on observable actions from an external perspective. The other two components, cognitive and emotional, involve more internal, subjective processes that are much harder to observe directly and evaluated without the individual's insight, whether this is captured by answering questions in between pauses throughout the experience or reporting it afterwards. Thus, the behaviour observed is the output or the reflection of covert cognitive processes that are happening, both at the same time as the behaviour and as a result of said behaviour. Simultaneously, the interaction is being emotionally appraised as positive or negative, which would then determine if the engagement was to be continued, diminished or terminated. Therefore, an individual could be reported as engaged by an external observer, but as Renninger & Bachrach (2015) point out, they "may not be aware of their own engagement",

until being prompted about the experience. The opposite can also be true where someone reports feeling engaged, but an observer does not perceive it.

Despite the three components of engagement being described as separate processes, the three are not isolated processes that happen independently or that only arise one at a time. On the contrary, they are interdependent and can act simultaneously within the individual (Ben-Eliyahu et al., 2018). However, given the methodological challenges of observing three interwoven processes, it is typical that studies of engagement measure each of the components separately and according to their own nature (Fredricks et al., 2004).

Some authors have adapted the three main components depending on their research aims, creating operational definitions according to the field of study in which the term engagement is being observed, including adding or removing components to these three. For example, Appleton et al. (2008) focused on school engagement and added an academic component, defining the 4 components as follows: *academic* (time on task, credits earned, homework completion), *behavioural* (attendance, participation, preparation for class-school), *cognitive* (value/relevance, self-regulation, goal-setting) and *affective* (belonging, identification with school). In this case, the academic component would be better classified as a domain-specific component, since it is pertinent and applicable only in long-term schooling, although it could also be argued that some of the elements within academic engagement (i.e. time-on task) could be placed in a different categorised (i.e. behavioural engagement) as it has been widely used in other contexts, like science centres (Serrell, 2020).

Another case where the components have been subdivided and combined is proposed by Pekrun and Linnenbrink-García (2012), where importance is placed on the role of “academic emotions” and their function to students. In this model, they considered emotions and emotional engagement (i.e. enjoyment of learning) as the precedents for their proposed components of engagement. They added two new components, *motivational* (i.e.

intrinsic and extrinsic) and *social-behavioural* (i.e. social on-task behaviour) while using a combination of *cognitive* (i.e. attention and memory), *behavioural* (i.e. effort and persistence), and *cognitive-behavioural* (strategy-use and self-regulation). The authors recognise their five components overlap with the three main components identified by Fredricks et al (2004), yet they were interested in clarifying which emotions relate to engagement and the role these play. Unfortunately, Pekrun and Linnenbrink-García's proposed classification encounters a limitation regarding components named after the interaction or overlap of two of the main components proposed, which then fails to recognise that these components are at a constant interplay rather than as separate processes as found by Ben-Eliyahu et al.(2018). On the other hand, their *social-behavioural* component provides a window into a context where the individual is interacting and how this drives engagement. This is a similar direction that Fredricks et al. (2011, 2012) also adopted when they proposed *social engagement* as a relevant component when studying engagement. Or like Wang et al.(2016) who added a social component that encompasses any type of social interaction that was relevant to the task –whether these were instructional or affective reactions. Although if a strict criterion is applied in this component, as discussed by Eccles (2016), the social factions could be categorised simply as behavioural and affective engagement. Researchers in science centres had also highlighted the importance of considering the social context and social engagement as key elements to understanding engagement (Falk & Dierking, 2000; Rennie & McClafferty, 1996; Tisdal & Perry, 2004). Thus, a social component is proposed as a separate component, and many have found it a useful addition, however, since it considers the context or influences of the context to engagement, it would be arguably better to not include it as a distinct separate aspect, but like a relevant influence on engagement.

A different position to the multifaceted construct of Fredricks et al. (2004) exists. Whitton and Moseley (2014) reviewed research around the topic of engagement and also noted that engagement had multiple meanings and different components proposed in the

context of educational research depending on the desired outcome, including formal contexts of learning and gaming activities. The authors propose a perspective rethinking engagement by using new categories that instead adapt to the function engagement can play, proposing two higher-order categories: Superficial and Deep Engagement. Superficial engagement relates to behaviour and extrinsic motivation only, where participation and attention are considered relevant. The second category, Deep engagement, is subdivided into four uses of engagement: enthrallment; feeling; belonging and being. Although the authors proposed this separation to facilitate the meaningful measurement of the construct, the end result arguably complicates definitions more by touching upon overlapping parts of other already defined constructs, such as motivation. Whilst this model skips what Christenson and Reschly (2012) denominate the jingle-jangle fallacy (i.e. giving multiple meanings to the same concept, or give the same meaning to different concepts) by categorising what already exists in the literature under a different name, it risks expanding the concept of engagement enough that it becomes less meaningful. Furthermore, the authors do not discuss if all these categories are present simultaneously, as well as being unclear on how best to capture them. Additionally, it does not provide guidance on how these different uses of the term engagement could transition from one into another, particularly since a gradient from superficial to deep is proposed. However, one key suggestion that adds value to the multicomponent concept of engagement from this model is the proposal of intensity in engagement, going from a superficial engagement into a deep engagement. Conceptualising a dimension of intensity within the component of engagement would also be related to current conceptions that propose engagement as part of a process (Azevedo, 2015; E. Wood & Wolf, 2008) which can transition through different stages of intensity. This suggests that work measuring engagement should attempt to quantify engagement in terms of the level of intensity.

Given the debate found regarding engagement and its multicomponent theory, Ben-Eliyahu and colleagues (2018) took on the task of determining if engagement was a unidimensional or a multidimensional construct, and if the latter was correct, whether these facets were as clearly separable as the concepts seemed to infer, or if these facets could also be interacting with one another. These authors studied children's engagement whilst taking part in science activities in two different learning scenarios, schools and science museums. The authors were particularly interested in how the different dimensions of engagement could "empirically co-occur", using theory both from Fredrick's (i.e. components) and Pekrun's (i.e. affective states around engagement) approach. Thus, they propose that "engagement processes (affective, behavioural, and cognitive) are active processes that emanate from within the learning situation as a result of the learner-activity interaction" (p.89). Moreover, since the aim of the study was to examine whether these facets acted separately, they took into account a measure for separateness of the components which was having evidence from one component despite having the opposite evidence for the other two components. For example, if participants finished the task despite feeling indifferent, bored or cognitively disengaged, cognition and emotion would be separate from behaviour. In sum, the authors tested with a Confirmatory Factor Analysis 3 different models of how engagement could be working: 1) engagement as a unidimensional construct; 2) engagement as a multidimensional construct using correlational models- this had two variations, three individual dimensions, or two dimensions combined, with cognition and behaviour acting as one factor and emotional engagement separate; 3) engagement as context-independent higher-order construct, engagement is a latent factor of the specified dimensions (as before, the same two variations were established). Additionally, in order to capture engagement more precisely, the authors left out any type of concept that could be related to it, but would not define it, such as motivation, self-regulated learning or deep-shallow processing, and instead measured motivation before and after the activity took place. Their research showed that engagement is both a unidimensional and a multifaceted

construct and comprises behavioural, cognitive and affective states that need to be considered when measuring and capturing engagement, at least in relation to learning tasks or situated activities. Their analysis showed that behaviour and cognition were merged as one single factor, which they provide multiple explanations: a) could be a result of a coupling of the demands of the science task that children were asked to evaluate, mainly because in science tasks, thinking and doing often appear simultaneously, or b) it could be because cognitive processes are deduced from observation, or c) a developmental confound of how they conceptualise their thinking and doing that is refined as children grow older. It could also be an artifact of the measuring tools they used to evaluate the term (i.e. self-report survey), as the items were not balanced between cognitive and behavioural, having 4 items of the first and 8 of the second one. Withal, Ben-Eliyahu et al, make the case that these results could be attributed to the fact they were focusing on engagement in science activities, and the topic of the activity could be influencing the participants' inclination towards the topic. Finally, the authors concluded that "engagement is more than the sum of its parts" (p. 97), emphasizing the importance of contextualising research around engagement. Whether this is to measure engagement as an outcome, as a precursor or as a process, this differentiation could result in a more adequate selection of measuring tools, which in turn would also clarify how the facets of engagement could be working as one.

Thus, after the review of the different views and conceptions of engagement in the literature, engagement could be understood as the active interplay of an individual's sustained attention-interaction (what individuals do), situated thinking (what individuals think), and emotional attachment (what individuals feel) while interacting with something during a specific time frame. As mentioned earlier, engagement has been construed as a meta-construct (Ben-Eliyahu et. al, 2018), that needs all components to be considered when capturing and measuring engagement to provide a full representation of how engagement operates. For example, engagement is not only a feeling or an emotion, because an emotion

on its own is a psychophysiological change associated with the situation that is eliciting it and it will impact behaviour and cognition in return (Adolphs et al., 2003), whereas emotional engagement considers that this emotion is as much a part of the experience as it is a potential factor behind engagement being sustained or stopped. Thus, engagement is not only the emotion itself but what resulting actions come out of it. Following the previous example, engagement cannot be sustained attention only, since attention is already a construct on its own defined as the process by which a subset of information is selected from a broader available source in order to focus and enhance the processing and integration of it (Ward, 2008). Since both cognitive and behavioural engagement need sustained attention to filter relevant information from an intricate situation in order to have a follow-up behaviour/thought process. Hence, engagement is a process that will be impacted by the temporality and the granularity of the context that the researcher is examining.

2.2.5 Continuum of Engagement

Among the different definitions provided for engagement, two things are recurrent: engagement is perceived either as a *process* or as an *outcome* (Whitton & Moseley, 2014a; E. Wood & Wolf, 2008). That is, in some definitions, engagement is perceived as the ultimate long-term goal that enables the individual to return to something constantly or over a long period of time, for example, a successful engagement outcome is being able to prevent school drop-out (Mahatmya, Lohman, Matjasko, & Feldman Farb, 2012), or a long-lasting positive impact from a visit to a science centre (E. Wood & Wolf, 2008). On the other hand, engagement is also studied as a *process* that happens alongside learning, in these cases it tends to be studied only during the length of a certain activity of interest (Ben-Eliyahu et al., 2018; Boekaerts, 2016). This clear difference seems to be related to the two main uses found in the ordinary use of the engagement definition by Sheppard (i.e. substantiative and

procedural). However, these could also be pointing to a temporality axis that follows engagement during different moments in an interaction.

This would not be the first time proposed that different continuums seem to exist within the construct of engagement (Azevedo, 2015; Fredricks et al., 2004; Shaby et al., 2017; Sinatra et al., 2015; Wiebe et al., 2014). Sinatra et al (2015) propose a continuum that goes from a microlevel that considers only what happens to the individual, to a macrolevel that encompasses a group of learners. This continuum highlights an important factor about engagement: the environmental context of the subject of study and its interaction with it matter. Another continuum widely reported is based on the intensity of the observed engagement, ranging from engagement to dis-engagement, which follows whether learners were completely immersed in the activity at hand – superficial/deep-engagement (Whitton & Moseley, 2014a; Worsley, 2018) or whether they directed their attention and behaviour away from the task – disengagement (Shaby et al., 2017; Worsley, 2018).

Considering intensity as a continuum contributes to acknowledging that engagement is a more nuanced construct with more variability throughout its appearance. An example of its applications is the Visitor Engagement Framework (Barriault & Pearson, 2010) which was designed to capture engagement during science learning interactions and proposes a transition in engagement behaviours going from initial contact, until they reach a breakthrough by engaging deeper and making meaningful connections. However, this intensity should be placed within the same restrictions of being tied to a context and that intensity would be also applicable to its multi-components. Moreover, although a continuum of intensity allows the tracking of a progression of the process, it is not enough to show or capture the temporality associated with engagement, whether this is an immediate associated process or a long-term outcome.

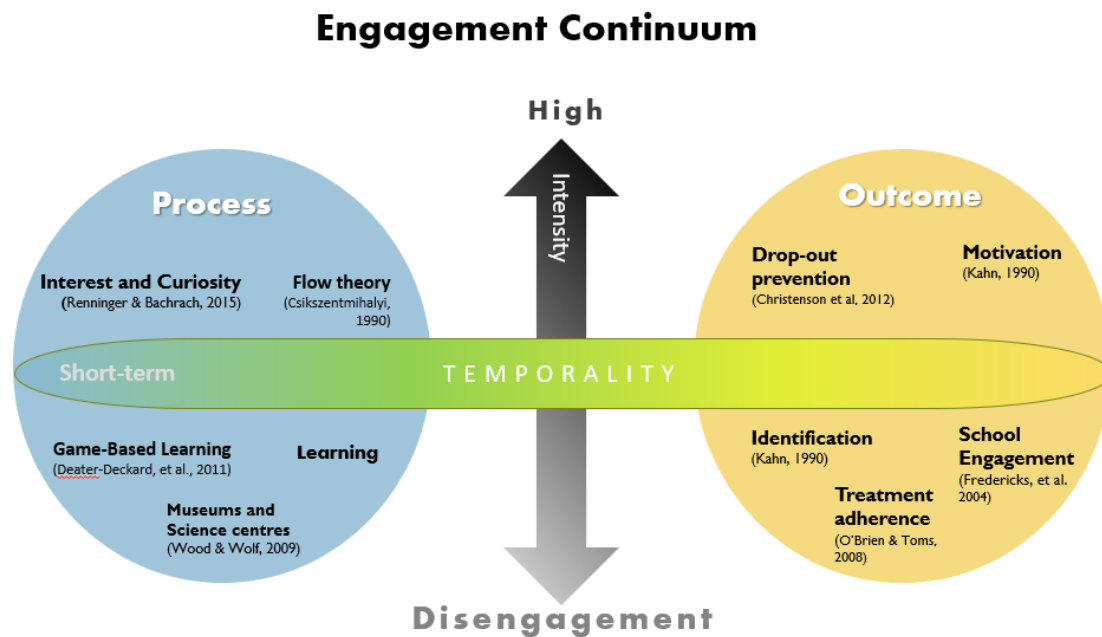
From the literature discussed there is a need to differentiate between learning engagement being considered in the short or long-term of an activity, as well as need to align

if engagement is being considered as a process happening during an interaction, or as an outcome resulting from the interaction (J. S. Eccles, 2016; E. Wood & Wolf, 2008). Since engagement has been researched as a process associated with learning, it has become relevant to understand how engagement is initiated and sustained throughout an interaction (Azevedo, 2015; Ben-Eliyahu et al., 2018).

A particular framework is proposed here, where engagement is seen as a continuum with two main axes: one based on the temporality of engagement and another based on intensity. The conceptualization of both continuums is shown in Figure 2.1, where the temporal axis is shown horizontally while the intensity axis is shown vertically. The first incorporates both engagement as a process and engagement as an outcome and how these can be situated within a temporal window. For example a short-term window, where engagement is considered during the interaction with a science exhibit, or a long-term window where engagement is considered a factor involved in visitors' return to a science centre. The accompanying axis moving vertically throughout the temporal continuum would be the intensity plane, which allows the variability in the process of engagement to be followed while also accounting for the temporality of engagement used. The next sections address and discuss each of these axes separately. Figure 2.1 also illustrates where some of the research on engagement could be placed within the temporality axis.

Figure 2.1

Continuum of Engagement



2.2.5.1 *Temporality Axis*

From the literature, engagement is constantly researched under two-different time frames: short-term (Csikszentmihalyi & Hermanson, 1995; O'Brien & Toms, 2008; Pekrun & Linnenbrink-Garcia, 2012; Shaby et al., 2017; Vile Junod et al., 2006; E. Wood & Wolf, 2008) and long term (Appleton et al., 2008; Finn & Zimmer, 2012; Fredericks et al., 2004; O'Brien et al., 2018), showing that engagement could be placed within a temporal continuum. Within the short-term window, engagement is situated and would finish along with the task at hand, that is, it would require action-response for the current task, and it would be bound to the demands of the context where the task is taking place (E. Wood & Wolf, 2008). Fredericks (2004) referred to this short-term end of the continuum as procedural engagement, and it was defined as “[the type of engagement when] trying to complete the task requirement lasting as long as the task itself” (p.67). This definition emphasizes that a short-term perspective on engagement does not necessarily mean there is a prescribed short duration, but rather that engagement cannot be expected to

last beyond the task itself. Likewise, Pekrun & Linnenbrink-Garcia's (2012) approach contains areas on their model that tap into the "short-term" duration of engagement. Since their focus is aimed at the emotions that arise while a student is taking part in a learning activity, they propose engagement could appear at any point throughout an activity with a variable duration. Moreover, Ben-Eliyahu and colleagues (2018) focused on the short-term dimension of the temporality continuum by focusing on process-oriented engagement during an activity, and defining it as "the intensity of productive involvement with an activity, involvement including focus, participation and persistence on a task"(p.87). Along a very similar line, D'Mello and Graesser (2012) working with young children defined engagement as "a state where the concentration is intense, attention is focused and involvement is complete" (p.146). Again, in these two cases, the duration of engagement seems to move in parallel with the duration of the task, while this use of the term engagement accentuates that within a short-term frame of engagement, the aim will more likely be related to examining the process of engagement. The notion of a short-term engagement, also relates to the importance of looking at the process of engagement which captures the nuances that initiate, drive and sustain engagement. An example has been studied during engagement with different digital platforms, describing engagement as part of the experience a user goes through while interacting with a digital system (O'Brien & Toms, 2008). Seeing that other fields outside education also have conceptualised engagement as a process, provides good evidence of this potentially underlying temporal continuum of engagement that is not context-dependent. E. Wood and Wolf (2008), museum researchers, argue that a short-term window is particularly well suited to study the relationship between engagement and learning in science centres. Particularly when analysing engagement in a science museum there is a need to look at the process of engagement that arises during learning rather than just focusing on the end result of engagement. Namely, this research field pays particular interest to the possible triggers and drivers of engagement in the short-term when visitors are interacting with a particular science exhibit, because it is recognised within this context that

visitors have brief episodic learning opportunities, compared to the continuous opportunity to explore engagement in learning in formal education (Falk, 2005; National Research Council, 2009). Thus, a short-term end of engagement can help the field of engagement by examining the onset and features of the process of engagement.

A *long-term* frame of research within engagement sits on the other end of the temporal continuum. This temporal window is mostly associated with attitudes and self-identification over time with an institution (Finn & Zimmer, 2012; Fredricks et al., 2011; Kumpulainen et al., 2014; Mahatmya, Lohman, Matjasko, & Feldman Farb, 2012), a topic (Bathgate et al., 2014; Mantzicopoulos et al., 2008), or a game (Whitton, 2011). In other words, *long-term* engagement is conceived as an *outcome* of an experience, which is why it would usually represent commitment with an activity or institution because of a perceived intrinsic value in it. For example what prompts an individual to return to said activity/context, like a school, or what influences a constant or prolonged use of a digital platform or game. Moreover, engagement is captured and studied after the situation has finished and the individual has been asked to reflect about what their thoughts and feelings towards the situation would imply for any future interaction with it again. As an example of this, the handbook of student engagement (Christenson & Reschly, 2012) is particularly aimed at preventing school dropout, defining engagement as the attendance, participation, identification and sense of belonging with the school. Furthermore, this perspective is most pervasive in the field of student engagement, where most research aims are directed to understand what keeps students in school, what drives them to participate and take an interest in the different activities offered and what the possible reasons behind students leaving school, whether this is high school or higher education. Another example where the long-term form of engagement has been observed is in museums or science centres, however the focus is around visitor engagement and the reasons that would drive those visitors to return to the museum many more times after (Baniyamin & Rashid, 2016; Schwan

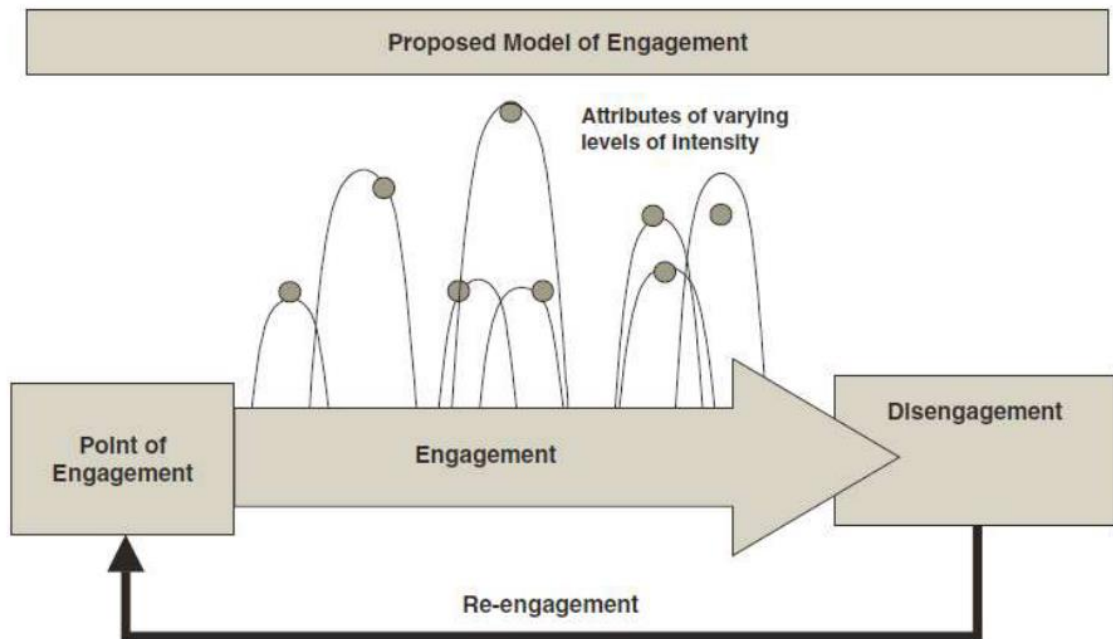
et al., 2014). Thus, in these cases, the attention is placed on what visitors experience throughout their whole visit to the museum, which events, exhibits or areas in the museum would they leave, modify or remove. Here, it becomes clear that when engagement is conceived as an outcome for impact it will imply the development of behaviours, attitudes, and emotions towards an activity over time. Since engagement is part of a continuum, it can be the case that it works both as a process and as an outcome. Perski and colleagues (2017) give an applied example of this double use of engagement when researching engagement with digital behaviour change interventions (e.g. fitness apps) and how to maintain people using them over a long period of time. They defined engagement as “a dynamic process that is expected to vary both within and across individuals over time” (p.258). This means engagement could start by being a short-term situated engagement (e.g. a science exhibit), that is constantly reinforced/re-engaged over different situations in the same context (e.g. interacting with different exhibits in a day), moving in the temporality axis towards a long-term outcome (e.g. positive attitudes towards science after engagement in a science centre visit). Thus, if a process of engagement is successfully started then it becomes an outcome of the activity short-term, if this was then restarted or repeated long enough over the same activity or multiple times over different activities, engagement could become a resulting attitude towards the initial task or the context where the task was situated.

2.2.5.2 Intensity Axis Engagement has also been defined according to intensity, from high to low (Chapman et al., 1999; Shaby et al., 2017; Wiebe et al., 2014) , or categorically – engagement vs disengagement (Christenson & Reschly, 2012; Skinner, Kindermann, & Furrer, 2009; Skinner, Furrer, Marchand, & Kindermann, 2008; Worsley, 2018). However, as Christenson and Reschly (2012) noted, these two conceptions of intensity levels of engagement are not mutually exclusive and can both exist as part of a continuum since their engagement can change over time. Related to this, O’ Brien and Toms

(2008) proposed a model which underpins the cycle that situational engagement goes through (Figure 2.2). The model shows four stages: *Point of engagement (1)*, the moment or object that attracts attention, it starts with an experience which is novel or aesthetically attractive. This transitions to the *period of engagement (2)*, which can be sustained as long as the person can maintain their attention and interest in the interaction, normally characterized by a positive emotion. This particular period would be characterized by the different behaviours done as part of the established engagement with varying degrees of intensity. Then, *Disengagement (3)* could happen, mainly the sudden stop of the interaction, which can be caused by many reasons, whether it is external (e.g. distractions in the environment) or internal (e.g. boredom). Finally, the last stage *re-engagement (4)*, brings an opportunity to restart the cycle, this will be dependent on the quality of the engagement elicited and the reasons for disengaging. A person could go through an engagement cycle several times during a single interaction or session. This model represents a cyclical rather than a linear process of intensity in engagement. Although this model was developed for digital platforms, it can be applied to different contexts where engagement takes place since none of its elements are context-dependent. Its use can be exemplified for engagement in science centres, particularly since these allow the individual to self-pace their visit having constant stimuli competing for the visitor's attention, which means many potential points for engagement and driving them to interact or not with particular exhibits. It is possible that the competing stimuli managed to capture the visitor's attention momentarily and then they would re-engage with their current exhibit, or it could prompt them to disengage with the current exhibit and establish a new point of engagement.

Figure 2.2

Cycle of Engagement



Note. Image reproduced with permission from O'Brien and Toms (2008)

Some authors in the literature warn against using an intensity dimension when studying and conceptualising engagement since it risks oversimplifying the construct of engagement (Azevedo, 2015; Ben-Eliyahu et al., 2018). However, if it is conceptualised as a gradient, or even better a cycle, it provides more room to explain possible underlying nuances of engagement, as evidenced by Worsley (2018) who argues in favour of tracking dis-engagement which can facilitate automatization of algorithms that track processes of engagement. Even more so, when both dimensions are combined, or additional continuums are incorporated, as the granularity continuum proposed by Sinatra et. al (2015), there is an increase in the potential to better explain the complexity behind the multifaceted construct of

engagement by providing a space to situate engagement under different circumstances and account for more confounding variables.

Throughout this chapter, engagement has been shown as more than a pre-requisite for learning, but as a process that accompanies and influences learning. Understanding how this relationship is established requires a thorough examination of the multiple components that constitute engagement while also considering the intrinsic factors associated with each distinctive learning context. When it comes to science learning in science centres, it is needed to focus on the short-term end of engagement, the situated engagement that arises from interacting with a particular science exhibit. More specifically, the cycle of engagement should be examined, capturing how individuals go from a starting point in their engagement when first entering an exhibit, what happens during this established engagement and what drives them to disengage and/or re-engage with the current exhibit. In order to start addressing this complexity in a science centre, there also needs to be clarity on how different tools capture, and what do they measure as engagement.

2.3 Capturing Situational Engagement

The previous section indicated two key factors to better define, capture, and measure engagement when designing research about it: (a) what is the context under which engagement will be studied and (b) what is the underlying goal behind the activity or process that is being studied. Context matters because it highlights the demands placed on the individual's engagement, specifically, the nature of each context and situating the environmental experience (Eccles, 2016). It also helps understand what the expectations and permitted behaviours are when being in that context, the limitations and challenges each context brings, both physically and behaviourally, and the different possible interrelations happening between the individual and the different agents in that environment (Falk &

Dierking, 2000). For example, researching engagement in a science centre context will differ from other contexts with similar attributes, like a botanic garden or an aquarium, where learning and interactivity look different, even within the different types of exhibits. In the case of learning in science centres, the expectations and demands placed upon children's engagement in science learning will not be the same when comparing engagement in a school than in a science centre. On the other hand, if examining fun and interactivity, expectations of children's engagement behaviour in the science centre would be different when compared to children being in an interactive playground (Poppe et al., 2014), where fun would be at the forefront rather than learning. Research has highlighted the cognitive and affective benefits of a particular context, such as a museum context, by analysing the differences in appreciation and memory recall of an art exhibition when people experience artwork in a real museum in comparison to a computer-simulated virtual version of the same museum (Brieber et al., 2015). They found that when people interacted with real art rather than simulated, their cognitive and affective processes relevant for art appreciation were enhanced, as was their long-term memory encoding for it. These findings recognise the context in which an experience occurs is always relevant even if the activity is the same.

Once the influence of context has been recognised, the specific goal of the activity needs to be considered when researching engagement. Doing this helps delimit what *can* or *cannot* be achieved as part of the interaction and the research. For similar reasons around interactivity, Adams et al. (2004) emphasize the importance of "clarity of purpose and intentionality" (p.160) behind a specific exhibit. Identifying the aim of the activity helps recognise the affordances and limitations each activity allows, which places more realistic expectations on the actions, processes and outcomes resulting from the interaction. For example, Toni Dancstep and colleagues (2015) explored the differences in visitor experiences between immersive exhibits, those where the exhibited phenomenon surrounds the participant, and hands-on exhibits, those where the phenomenon exists only in front of

the user. Their findings show that although immersive exhibits were reported as being more enjoyable and fun, they did not reflect as much time spent or intellectual engagement with the content from the exhibit, this was the same whether the visitors were adults or children. Thus, attention should be placed on *how* the child is interacting with the exhibit and evaluate if the exhibit learning goal is being met. If the child only interacted and tapped a screen or free-played with the features from a table-top exhibit, without following any of the intended learning goals of the science exhibit, it could be wrongly categorised as the child being engaged with the content; although the child might have been engaged, elicited engagement would be focused on elements that did not align with the goal of the activity. For example, a study by Rennie and McClafferty (1993) showed that when school children used more inquiry play compared to fantasy play, their likelihood to learn some of the scientific principles from the interactive exhibits increased. Nonetheless, some exhibits are intentionally designed to allow free play or exploratory behaviours with its elements. Gutwill (2008) found that open-ended exhibits attract and hold visitors for longer when compared to counter-intuitive exhibits that show and direct interaction, because it allows visitors to plan and set their own learning goals within a specific content, for example creating ice sculptures under a screen. Their findings help exemplify how the learning goal of the activity needs to be accounted for.

As mentioned in the previous section, one last thing to consider is whether the interest and capacity of the research are on engagement as an outcome (once the interaction has finished), a process (during the interaction) or both. Making this distinction clear can help identify and ask testable research questions, grounded on effective tools that align better with the expectations behind engagement and with the context and underlying goal of the activity. The research presented here aims at exploring the process of engagement and bridge them with the resulting outcome of engagement. Since this thesis focuses on science centres and early years, the following section discusses tools that are currently being used for capturing engagement in a science centre context and/or in the early

years. It is also discussed how other tools not currently used for researching either this context or this population could potentially benefit this field.

2.3.1 Current Tools Used to Capture Engagement as an Outcome

A scoping of different research carried out in science centres indicates that the focus on engagement has been as the outcome of an activity, using tools that provide a score, or a collection of answers after an interaction or visit is over. One of the easiest and most commonly used methods, according to Scharf, (2018 in Serrell, 2020) is timing and tracking used by 48% of the different types of museums, which means that engagement is measured as the interaction time a visitor spends in a science exhibit/museum area (Serrell, 2020; Yalowitz & Bronnenkant, 2009). Serrell is one of the main proponents and researchers using this method (Serrell, 1997b, 2020), where she states that this practical method allows the collection of the simplest of data which is widely compatible and not costly, facilitating the comparison between exhibits. To achieve this, she created an index that considers the size of the exhibit with the time spent, to control for the dimensions of the exhibit. Although it still lacks the capacity to provide insight into the individual's cognitive, emotional or behavioural engagement, this measure provides a good screening tool to scan how science centres' visitors' engagement is being distributed and allocated between the different areas and science exhibits. However, Serrell, acknowledges that timing by itself cannot provide any information beyond tracking the length of the engagement (Serrell, 1997a), encouraging the use of additional measurable behaviours to record what is happening *throughout* the activity, such as reading the labels or talking about the exhibits. While this method encompasses engagement as a single outcome, additional research has complemented it with interviews to the visitors to capture their self-reported emotional engagement or their subjective experience (Tisdal & Perry, 2004), which according to Renninger and Bachrach (2015) presents an advantage on capturing a cognitive and emotional element that can be related with the observed timed engagement.

Surveys and questionnaires are tools that also measure engagement as an outcome when done at the end of the activity (Read & MacFarlane, 2006). The surveys are either self-reported or reported by any accompanying adults in the case of younger children who may not be old enough to take part, whilst the interviews would normally be carried out by the researcher. Although these can touch upon the visitor's experience *during* their interaction/activity, like the Experience Sampling Method (D'Mello & Graesser, 2012; Fredricks et al., 2004; Fredricks & McColskey, 2012) where constant reminders in a device make an individual pause their interaction to report about their self-perception of engagement, these tools are used more in formal schooling or gaming contexts than in science centres. Unfortunately, a fundamental drawback of this approach is the interruption of the activity, engagement is disrupted and any process happening before the interruption is finished, turning any engagement captured into an outcome. Measuring engagement via interruptions not only disrupts the activity, but also affects the phenomena being measured, providing two or more different experiences of engagement.

Young age can also be a limitation when it comes to filling in surveys or questionnaires, as participants could not have the reading and comprehension skill level needed for answering a questionnaire or a self-report (Zaman et al., 2013). Tools, like the fun Toolkit (Kano et al., 2010; Read, 2008; Read et al., 2002; Read & MacFarlane, 2006), exist to address this limitation and aim to capture self-reported reflection of children's experiences using visual representations such as "smiley" faces/emojis or thumbs-up/down for depicting different emotions. Despite criticism made to the use of emotions that may not be related to possible emotions elicited by the experience being observed (Zaman et al., 2013), using age-appropriate prompts help access children's ideas to verbalise and share some of their feedback from taking part in an activity. Luckily, additional representation imagery has been created, for example as part of the fun toolkit, sorting tangible examples (photos, objects) from the activity into boxes from "most/worst enjoyed" have shown to help children enact

their preferences by using smaller categories (Kano et al., 2010; Read, 2008). Nonetheless, surveys and questionnaires still have disadvantages that arise when trying to capture any subjective experience and have been criticised before for capturing limited or biased information (Azevedo, 2015). Amongst the main criticisms and limitations are: 1) *Agreeableness/ prone to bias/suggestibility* – a tendency to agree when asked open questions or to try to answer what they think the research wants to hear, which is further exacerbated when working with children (Zaman et al., 2013). 2) *Averaging experience* - these tools rely on the individual's capacity to reflect on their own engagement experience, and on many occasions, provide an overall score to summarise that experience which provides limited information subjected to what the individual remembers. Some authors have also mentioned that it depends on the subject's capacity to be aware of their own engagement, and its extent, while this is happening (Renninger & Bachrach, 2015).

Relevant to children's science engagement in science centre, only one survey/questionnaire was found in the literature: the *Activity Engagement Survey* (Ben-Eliyahu et al., 2018), which considers the cognitive, behavioural and affective domains, but reported from the first-person's perspective and used to compare children's science experiences when at a school or a science museum visit. This survey was always used after each science activity or interaction with the science exhibit had just finished, reinforcing the engagement outcome. While the Likert scale used to answer this survey was easy to understand (YES! – yes -no- NO!), the items were aimed at children in upper primary or older, which relies on both a reading comprehension, and a higher ability to self-reflect and abstract the meaning behind each question asked. Another engagement survey that appeared as relevant to capture engagement was found in the user experience field, the *User Engagement Scale* (O'Brien et al., 2018), developed for evaluating digital platforms (e.g. video gaming, online shopping, web searching and educational webcast), which interestingly attempts to track the process of engagement by capturing points of

engagement, disengagement and reengagement based on answers from participants. Two more surveys were found in the realm of formal education: the *Student Engagement Instrument* (Appleton, Christenson, Kim & Reschly, 2006) aimed at preventing school dropout during high school, and although it considers the multi-components of engagement, it also adds an additional component that tracks academic engagement, which turns it into a very specialised survey that targets the long-term outcome of engagement. The other survey, *Motivation and Engagement Scale* (Martin, 2009), combines the evaluation of processes they define as adaptive and maladaptive behaviours and cognitions. This survey targets a broad population, from elementary level to university level, but it defines engagement differently from the multi-component framework as well as adding elements of the concept of motivation. Although as discussed previously, Renninger and Bachrach,(2015) have argued that motivation is a different construct from engagement, where engagement can be perceived as the observable action and motivation can be seen as the drive behind it.

From all the tools presented, most of them were created for a different context, formal education or use of digital platforms, and for a different age group, upper primary and high school. This makes it difficult to adapt to a science centre context since the use and purpose of each of these tools are vastly different as shown in the aims of each one. Even for the surveys that include learning, research has shown that the context where an activity is taking place matters, for example children's preference for engaging with science learning activities is different depending on whether the activity is taught at school at home or in a science museum (Bathgate et al., 2014). Furthermore, the behaviour in a formal learning setting is different in nature to the one found in free-choice learning environment (Falk, Dierking, et al., 2007). Firstly because the time available for activities in the formal contexts is longer than the average time spent in a science centre exhibit, whilst a digital platform provides longer interaction times and higher opportunities for re-visiting the digital platform, compared to episodic visits to a science centre.

Moreover, formal education contexts allow longitudinal progress to be tracked about engagement in a way that is almost impossible for a science centre to do with its visitors, since learning in these contexts is designed to be sporadic and episodic rather than recurring (National Research Council, 2009). Also, the different definitions of engagement used in each of the surveys described muddles the comparison between them, while they can align in some areas, some others consider and combine additional concepts, like motivation or mastery of skills as part of engagement which only complicates the research and definition of the concept of engagement.

Finally, the language comprehension required for most of these tools is aimed at older participants, from upper primary children to adults, who are able to comprehend and self-reflect on the different statements presented (Appendix A shows the detailed items asked). Even more so, all the surveys presented here are still susceptible to evoke biases such as those surrounding expectations from the researchers, regardless of the age in which they are presented (B. C. K. Choi & Pak, 2005). These tools reflect the importance of considering the factors mentioned throughout this chapter around the importance of the context, the goal of the activity, and the audience with which research will be carried.

Contrary to surveys, interviews are used to explore and better understand the reasons and variance found in participants' engagement, which can be done with a pre-set structure, semi-structured or open to the participant's answers. Interviews provide more insight into the individual's self-reported thoughts and feelings, a good complement to any behavioural or survey analysis (Fredricks & McColskey, 2012). For example, Heath and Vom Lehn (2008) interviewed children after participating in a collaborative activity in a museum, or Silpasuwanchai et al. (2016) used semi-structured interviews with video-games participants to know more about their experience while playing videogames. However, interviews show very similar limitations to surveys and questionnaires, where participants are prone to bias or might answer with what they think the interviewer wants to hear. Age could also influence the

quality of the data, since interviews still rely on younger participants' verbal and emotional capacity to express their thoughts, actions and feelings, although interviews could potentially bypass that limitation by addressing and adapting the questions according to each participant (Zaman et al., 2013). In any of these methods, evaluating subjective perceptions of engagement with science learning, topic-content is an additional important factor that is often overlooked, but should also be considered when comparing visitors' experience (Bathgate et al., 2014).

The three methods presented here for capturing engagement as an outcome have all presented limitations in how much information they can provide on what happens *during* an interaction with an activity. The next section addresses tools that can be complimentary by capturing different engagement-relevant processes throughout an activity.

2.3.2 Current Tools Used to Capture Engagement as a Process

In a science centre, tools have attempted to unobtrusively capture the **process** of engagement, providing richer details from beginning to end of an interaction. One of the main methods used is observation analysis which can be done live or by observing video recordings (Fu et al., 2016). Participant observation consists of watching a specific group or activity of interest and analysing their interaction (Jorgensen, 2011). These observations can be structured (systematic protocol), or unstructured (free observation) (Grack Nelson & Cohn, 2015). Observations can be carried out by expert practitioners of the context or by researchers who may have experience with observation methods, but not too much familiarity with the science centre context (Jorgensen, 2011). When it comes to tracking the process of engagement, structured observations use scales and coding schemes developed to capture different levels of engagement or different behaviours associated with engagement.

There have been a few measures developed to capture engagement, either in children or related to science learning. Two of these measures were developed specifically to evaluate visitors in science centres: the *Visitor Engagement Framework (VEF)* (Barriault & Pearson, 2010) and the *Exploratory Behaviour Scale (EBS)* (Van Schijndel et al., 2010). The VEF is catered to all visitors of science centres and captures visitors' initiation behaviours, transition behaviours and breakthrough behaviours relevant to engagement with an interaction. The EBS is not specifically designed for engagement, however, it is designed to evaluate children's behaviour while exploring their environment and how to support parental scaffolding, which can be a good indicator of engagement. This scale follows children's behaviour from passive contact with the environment to active manipulation and to exploratory behaviour. Both of these coding schemes track the process of intensity in engagement by containing three categories that increase in their intensity, through which engagement could be considered as transitioning from one level to another. Observers keep track of these behaviours through a tally count of the number of times these happened in live observations. However, neither of these coding frameworks capture the multi-dimensionality of engagement, only its intensity.

Other engagement scales found in the literature were aimed at the early years, like the *Learning Engagement Coding Scheme (LECS)* (Halliday et al., 2018), which is a scale used to evaluate preschool-age learning engagement of children while solving a reading task and a tangram-building task. This scale contains seven indicators of engagement and the rating of each of these depends on the behaviour descriptors of each point in each item, which overall range from no indication of behaviour to high indication of behaviour. The tasks were done in a controlled laboratory environment which has different constraints compared to any formal or informal learning environment.

The *Child Behaviour Rating Scale (CBRS)* (Mahoney & Wheeden, 1999) was designed to examine the effects of different styles of teacher interaction on engagement of

preschool children with special learning needs. It has seven categories that capture engagement when the child is working alone and with a teacher, which are rated using a 5-point Likert Scale (very low to very high).

A third scale found was the *Engagement States Observational Coding System* (eSOCS) (Deater-Deckard et al., 2013), which was an instrument designed to detect developmental differences in engagement from preschool to high school learners by observing and quantifying various developmental aspects of cognitive, behavioural, and affective engagement based on Fredricks et al. (2004). This scale captures 13 behaviours that capture independent and dyadic interaction and is rated in a Likert scale of perceived frequency (Present all the time to Never present).

The last scale found that worked with early years children was the *Engagement Quality Observation System (E-Qual III)* (Aguiar & McWilliam, 2013; McWilliam & De Kruif, 1998; Pinto et al., 2004), which was developed to track engagement in child-care classrooms while 2-year old children free played. They capture engagement with nine engagement levels that differentiate between the quality of engagement and the focus of the child's behaviour. In a more recent paper (Aguiar & McWilliam, 2013), the resulting engagement classification has been divided into sophisticated engagement (which only includes levels 1-4), and non-engagement.

All of the scales presented here, besides being appropriate for the early years, also consider the cognitive, emotional and behavioural aspects of engagement proposed by Fredricks et al. (2004) in the rationale used for the items by its authors. They have also been used better in video recordings to capture detailed information and to code with more certainty. However, most of these tools are not as effective for tracking engagement *during* an interaction, that is, they can capture behaviours that happened throughout an interaction, but it is harder to assess what was the order of such behaviours, except for the *E-Qual III* where one of its strengths is the screening every 15s during the 15minutes of interaction.

Moreover, these tools share a similar limitation when it comes to using them in science centres, which is they have been developed for use in formal contexts (or a laboratory like the LECS) with specific goals and items. Although, some of these tools could be adapted to use in science centres, like eSOCS, since a context has not been defined, adjustments would also have to be made to add the temporality of the events to track *when* during an interaction these events happen and what might be driving it.

All the observation coding scales provided vary in their use, where some have been used in live-contexts, and others have been used watching video recordings. The engagement scales shown here are an example of the influence of different definitions of engagement and how it can impact the tool used for capturing it. Although some scales have been used in contexts of science learning in museums and/or with young children, they all share a similar limitation with the observation reports because they rely on someone to observe and code, giving an inferred assumption into the cognitive and emotional components of engagement from the perspective of the observer. Moreover, the majority of the engagement codes and scales do not present any information regarding their validity as reliable measures that capture engagement, and while they can be used to descriptively compare between results, it is uncertain whether they are a valid tool to use to capture engagement. This again refers back to considering what the interest behind researching engagement is and what the capacity and need for it is. For example, if the aim is to provide an overarching sense of what behaviours happen throughout the interaction, or if the aim is to understand how engagement is being elicited in a particular scenario and how to capture and improve the beneficial objects to be able to generalise to similar activities.

Furthermore, observation studies when done live are susceptible to misinformation due to human limitations in attention and memory, whereas video recordings present an advantage over them when there is a need to capture more details on the process of engagement, allowing the researcher to go back to the event as much as needed and to

keep track of the moment in time where the different engagement markers happened (vom Lehn et al., 2002). Although video-coding tends to be more time-consuming than live-observation due to the level of detail required for it (Grack Nelson & Cohn, 2015), depending on the aim of the research, the amount of additional information might be worth the additional time needed for interpretation. Observation studies also present only one side of the picture (i.e. information limited to dimensionality of the videoframe), and like with any type of observation, when a videorecorder or a person observing is in sight it could impact and change the behaviour where the learners could modify or diminish their behaviour (Montgomery, 2014; Sparrman, 2005).

2.3.3 Tools That Capture Temporality of a Process

The advancement of technology has made accessible tools that are better at capturing fine-grain details of physiological processes that are not easily observed. These processes tend to be covert from both the individual and the observer (Cacioppo et al., 2000). Tools that can capture, measure and keep track of these processes help understand the role that any psychophysiological change could be having in observable behaviours of interest. While there has also been research that claims that engagement elicits cognitive and emotional neurobiological markers (Cowley & Ravaja, 2014; Kamzanova et al., 2011; Mercier & Charland, 2013), offering a new promise of methods to measure engagement. Since all of the engagement scales and observational methods above depend only on what can be observed from the individual, interest has risen around tools that could capture and measure covert or unobservable inner processes related to the process of engagement *during an interaction*, both on a cognitive and an emotional level (Azevedo, 2015; D'Mello et al., 2017; Emerson et al., 2020; Lee-Cultura et al., 2021; Malmberg et al., 2019). The last decade has seen an increase in interdisciplinary research between fields like neuropsychology or psychophysiology with learning contexts using tools that provide a window into these inner processes in real time (Cain & Lee, 2016; Cruz-Garza et al., 2017;

Gashi et al., 2019; Hedman, 2008; Leister et al., 2016, 2019; Malmberg et al., 2019; Marín-Morales et al., 2019; Mercier & Charland, 2013; Wass et al., 2016, 2015). The added benefit of working with these areas of research is the premise that unobservable physiological changes happen constantly, depending on the situation being encountered or experienced, changes that the individual may not even be aware that have happened (Mcguigan & Schoonover, 1973). The main tools that offer the possibility to track unobservable changes on the individual and that have been used in different learning contexts are:

(1) Electroencephalography (EEG) and Near-infrared spectroscopy (NIRS) capture brain activity in real-time and can be adjusted to look for activity patterns and brain areas that are related to learning (Ansari et al., 2012; Johnson & de Haan, 2011). The former tracks electric signals, while the latter records the oxygen saturation of the brain's blood flow. These tools normally require a lab set-up as they have multiple wires connecting different brain areas from the individual with the recording machine, which tend to need fast data processors to have better precision in capturing any brain activity. Additionally, both these tools require a rubber cap to insert the electrodes that capture brain activity, making the process lengthier and invasive, since some devices also need gel on sections of the scalp to improve data quality. For EEG, studies have developed an *Engagement Index* based on a decrement in task vigilance, taken from measures in fronto-parietal regions and use the proportion of faster electric waves indicating high brain activity (β) in comparison to slow electric waves that might indicate lower activity or relaxation ($\theta+\alpha$) (Mercier & Charland, 2013). This has been used to study adults' engagement while playing a video-game in the lab (Cowley & Ravaja, 2014), or how adults manipulate their workload manipulation while engaging with a demanding and stressful task in a lab, like air traffic control (Kamzanova et al., 2011), or to optimise work environment to increase motivation and productivity (Berka et al., 2007). Wireless mobile options both for NIRS and EEG have already been developed

and used in learning contexts like art museums, for example Cruz-Garza et al. (2017), evaluated different mobile EEG devices, one vs four electrodes, while participants visited an art exhibit in a public museum, whereas Marín-Morales et al. (2019) used a mobile EEG to compare experiences between exploring an art museum live or through an immersive virtual reality device. Research like this opens up room for exploring and understanding real-world learning environments under a different light that considers both external and internal relevant factors. More compact versions of EEG, either wired or mobile, have been used to evaluate learning and engagement in classrooms with adults and older children (Halderman et al., 2021; Khedher et al., 2019), but research is still scarce with young children. To the date this review was written, no studies could be found using EEG in early years while at a science centre, which presents novel unexplored avenues for future research in live contexts, such as science centres, and with new populations contexts as the early years. Having research using this type of tool could provide an additional perspective of unobservable neuropsychological changes, as well as clarify how certain cognitive processes children might experience while engaged in a science learning activity

(2) Eye-tracking / Head-Mounted Cameras capture visual attention by tracking the eye-movement patterns used to analyse and interact with the environment. Visual attention provides information about a scene that might be experienced, however, it can only focus on a small portion of that scene at a time (Rayner, 2009). Smith and her colleagues (2015) propose that three perspectives exist for a person's visual environment: a third-person view, a first-person view and fixations within the first-person view. Although some methods rely on observation of the participant's eye gaze in the wider context, according to Smith et al. (2015) this would still fall within the third-person view, since it would be captured from outside of the person. Conversely, for understanding the first-person view the key tool used for tracking visual attention is eye-tracking which provides a method of tracking the different eye fixations, both in direction, length and duration. Eye-movements have been associated with

volitional cognition, planning and problem-solving (Land et al., 1999; Ohlendorf et al., 2010; Rayner, 2009), which are also related to processes needed for establishing cognitive engagement (Deater-Deckard et al., 2013). An eye-tracker is used to capture them, normally placed on top of a screen where an infrared light is directed at the pupil centre and reflected on the cornea, it then records the movement of the cornea's light reflection; this is called Pupil Centre Corneal Reflection (PCCR). To do this, the participant needs to stand still and place their head on a specific head-rest to prevent them from constantly moving their head. This technique is harmless and unobtrusive. Two main types of eye-trackers exist: screen-based and mobile/wearable eye-trackers. Screen-based eye-trackers have high temporal resolution and can be used with any task that is performed while watching a screen. This type of eye-tracker can be used with all ages, from infants to elders. Eye-tracking has been used diversely, mainly to explore how individuals read (Rayner, 2009; Rayner et al., 2006; Schotter et al., 2014), how dynamic or static images are analysed (Jarodzka et al., 2010; Roberts et al., 2013; T. J. Smith & Mital, 2013), and how gaze plays a role in social cognition (Chita-Tegmark, 2016; Foulsham et al., 2010). On the other hand, mobile eye-trackers are used in live contexts, allowing the exploration of visual attention in a third-dimension environment (Krogh-Jespersen et al., 2020). For example, in museum research, mobile eye-trackers were used with older children to record their interactions with different art exhibits and how their examining gaze patterns of artwork was different from an adult's, highlighting that children and adults do not focus on the same aspects of an image (Walker et al., 2017). This type of tool brings the potential to understand where children's attention is located as they explore a science centre visit or how they react to an engaging exhibit, such as what features they choose to focus (Eghbal-Azar & Widlok, 2013). Evidence in a case study with a boy using a mobile eye tracker during a visit to a science museum, showed that is possible to capture patterns of use with the exhibits and how he engaged with others during the visitors (Jung et al., 2018). However, mobile eye-trackers have lower quality and temporal resolution

than their static counterparts since these have to compensate with the head movement of the individual by having a lower recording frequency (Eghbal-Azar & Widlok, 2013). Also, as evidenced by Jung et al. (2018), they can be costly, and more affordable options present technical difficulties with battery duration and remote connection between its elements.

While this type of eye-tracker offers more mobility and ease of use, they are not yet designed to take young children's smaller features into account without being obtrusive. A viable solution that accounts for mobile first-person perspective in naturalistic settings is a head-mounted camera (Aslin, 2009; Schmitow et al., 2013; L. B. Smith et al., 2015; Yoshida & Smith, 2008). Moreover, both eye-tracking tools and head-mounted cameras provide a better idea of how the scene is being experienced from the participant's first-person perspective (Schmitow et al., 2013; Smith et al., 2015). Thus a head-camera becomes a proxy for the field of vision where the focus of the attention could be tracked by following the centre of the scene being captured, since moving the centre of the image might implicate that the person is turning towards any new focus of interest nearby. Yoshida and Smith (2008) were some of the first to evaluate the use of a head-mounted camera on infants between 18 and 24 months old, where they found a coupling between eye gaze and head direction when playing with toys. They also found that since children tend to place objects of interest closer to them, dominating the field view, it facilitates capture by the head-mounted camera. Schmitow et al. (2013) studied attention related to head-direction when 6- and 12-months old infants look at an object of interest. They found that head and eyes are tightly coupled together, and that object direction is followed systematically by the head in the horizontal plane. L. B. Smith et al.(2015) discussed the contributions of head-mounted cameras to study visual attention in young children in a review paper, where they presented evidence that when the head and eyes align in watching an object of interest, quality in visual processing increases, arguing that head cameras with a head-centred view can provide optimal markers for children's attention. However, because it is acknowledged that

misalignment between eyes and head could be possible or hard to track in vertical planes, it is suggested to have another camera that can capture the child's behaviour and track where their attention was focused. In informal science centre environments, having a camera that records the person's point of view has been evaluated by Burris (2017) who placed an action camera on the participant's chest. However, after her findings revealed a limited understanding of the participant's field of vision she recommends using a head-mounted camera to gather better evidence of where exactly a participant might be focusing their attention, since a chest-focused camera does not account for occasions where a visitor might be turning their head.

A head-mounted camera has been suggested as a more appropriate tool for recording both looking behaviour and the potential visual environment around, particularly when the activity is related to free-play, where it is more useful to gather information from the surroundings of the child by letting them direct their attention and explore (Schmitow et al., 2013). Therefore using this method can contribute to capturing the process of engagement by providing the child's first-person view of their interaction pattern, along with the third-person view captured by a fixed video recorder, which can be triangulated to show possible triggers within the exhibit or outside the child's point of view that could be influencing their engagement.

(3) *Electrodermal Activity Sensor (EDA)* - Electrodermal activity, also referred to as galvanic skin response, has long been used to measure psychophysiological changes due to emotional arousal often in response to stimuli (Critchley, 2002). An EDA sensor measures the skin conductance of an individual's finger, palm or wrist's surface, commonly due to an activation of the eccrine sweat glands. The sympathetic nervous system is entirely responsible for the eccrine sweat glands, making electrodermal activity a good marker of sympathetic activity (Critchley, 2002). Sympathetic activity, when aroused, has been associated with higher sensorial vigilance and a heightened readiness to react and move

which also regulates the levels of attention and emotion displayed by an individual (Critchley, 2002). The sympathetic branch of the Autonomous Nervous System is also influenced by the Central Nervous System via the hypothalamus and the limbic system. The former is associated with thermoregulation, and the latter with emotional arousal and motivation along with activation of frontal cortical areas, responsible for attention, inhibition and decision-making processes (Boucsein, 2012). Thus, the neural mechanisms show that EDA is associated with similar pathways that are present in emotion experience and expression.

EDA is divided into two types of activity, a tonic activity which is slowly varying and shows the overall level of activity, and a phasic activity which is fast-varying activity representing a short response or reaction that can reflect stimulus-specific responses or spontaneous responses not specific to anything that can be traced (Benedek & Kaernbach, 2010a; Boucsein, 2012). Cognitive processes can also be accompanied by emotional arousal, thus further linking cognition and emotion (Boucsein, 2012) as evidenced by the theory of the somatic marker (A. Bechara et al., 2005). Skin Conductivity was used as a measure to track somatic markers before and after decisions were being made using the Iowa Gambling task (which evaluates positive and negative outcomes and the associated rewards/punishments); their results showed that participants started to develop anticipatory skin conductivity responses to what they perceived 'risky decisions' which would guide them more towards positive outcomes decisions even before being fully aware of their reasons for choosing some cards above others. This evidence has been replicated in different research (Li et al., 2010; Webb et al., 2014) including research with children (Cassotti et al., 2014) and it would indicate that skin conductivity is a sensitive measure for tracking psychophysiological arousal (i.e. understood as emotional states) that can be related to attention, memory and decision-making processes. In other words, skin conductivity has been shown to reflect emotion, but it is still advised to use a combination of behavioural and psychophysiological measures for interpretability (Buelow & Suhr, 2009). EDA sensors can be found in two main

formats: wired sensors and wearable devices. Wearable EDA sensors have started to be included in research around learning environments since these allow the individual to move freely in their environment while taking part in a learning activity (Garbarino et al., 2014; C. McCarthy et al., 2016). Having the capacity to measure the psychophysiological markers for emotional arousal in specific contexts where participants can move around (Cain & Lee, 2016; Hedman, 2008; Hernandez et al., 2014) presents an advantage when it comes to explore and better understand that context. Unfortunately, there is little literature around ambulatory devices with early years children (Hedman, 2008; Malmberg et al., 2019; Sridhar et al., 2018),

EDA sensors have been used in research in museums with adults exploring a world-war exhibit and how their emotional arousal was influenced by their visit (Mitas et al., 2020). Regarding children's engagement, research using skin conductivity has been done in different scenarios: while they free-played with Lego pieces (Hedman, 2008), when they were interacting socially with adults (Hernandez et al., 2014), when preschool children had to respond to performance tests (Sridhar et al., 2018), to evaluate children's programming in an after-school course (Cain & Lee, 2016) or when playing with an augmented reality learning game (Lee-Cultura et al., 2021). In most of these studies, EDA sensors were used in combination with some of the other tools presented here to capture more than one aspect of an individual's engagement (Cain & Lee, 2016; Fowles et al., 2000; Hernandez et al., 2014; Malmberg et al., 2019) and to help discern and triangulate what caused the emotional arousal. All the different applications presented here where mobile EDA has been used speak to the potential this tool has for providing a glimpse into the unobservable emotional processes an individual goes through and might be using as a 'gut-feeling' reference for deciding how to engage while interacting with a learning activity. Acknowledging that a combination of all the psychophysiological markers (e.g. Heart rate, pupil dilation, SCR)

provides a broader picture of the individual's emotional arousal (Garbarino et al., 2014), the use of each of these will depend on the researcher's aims.

Using any of these new technologies offers the potential to reduce an observer's bias by providing more information about the participant's inner experience with an activity. By capturing data continuously throughout an activity, they can offer quantifiable evidence of how a process operates in time. Wass et. al (2015, 2016) explored the co-variation of task-related peripheral measures of psychophysiological arousal in infants while they watched different types of videos. Admittedly this research is neither mobile, nor with early years children, however, they provide evidence of how different measures interact over time, how they can be temporarily asynchronized yet be caused by a similar event, which is done by taking into account the lag of the response time depending on the physiological marker. Noticeably, the measures and tools that provide psychophysiological feedback cannot be analysed individually or without any context; they need contextual information to ground and direct what is being captured and observed to understand and make sense of the data obtained (Sridhar et al., 2018). Many of the studies presented here use a combination of psychophysiological tools with observation or self-report studies to track relevant events to explain the obtained measures (Ding et al., 2018; Harley et al., 2015; Malmberg et al., 2019). While other studies also combine different psychophysiological tools to capture different aspects of an inner experience, for example to capture mind-wandering, a combination of eye-tracking and EDA has been used (Brishtel et al., 2020), or combining EEG and EDA to detect engagement while learning using a game (Cowley & Ravaja, 2014). (Wass et al., 2015, 2016) Moreover, mobile and wearable technology bring an additional benefit for researching engagement in a science centre since they have been developed for use in specific contexts beyond a laboratory set-up (Mercier & Charland, 2013). Higher quality data gives more in-depth information about a specific part of a process, whilst mobile tools allow participants and researchers to be immersed in the live learning environment. However, often this means increasing noise or lower data resolution in exchange for free movement from the

individual (Cruz-Garza et al., 2017; Krigolson et al., 2017; Malmberg et al., 2019). Despite this, Azevedo (2015) proposes that if the aim is to understand the experience within a specific context, a combination of mobile tools can be key to capturing multiple points of data to build a more complete perspective of engagement. In a review of methods used in science museum learning, Fu and colleagues (2016) make a call to incorporate more methods that collect direct measures of the outcomes of the informal learning experience. However, because many of these tools are still not widely accessible, and even though they are feasible to use in the context, they can still be costly and require time allocation and expertise to analyse them. Nonetheless, the research evidence that has already been presented using these mobile tools in museum contexts (Burris, 2017; Cruz-Garza et al., 2017; Eghbal-Azar & Widlok, 2013; Kress, 2016; Marín-Morales et al., 2019; Nubani et al., 2018) presents an optimistic view also shared by practitioners in the museum field like the Smithsonian museum (Massachusetts et al., 2019) where they have hired a neuroscientist to examine how art perception and personal reflection could vary depending on certain curated prompts. Therefore, more research in these contexts should be pushing to incorporate the use of psychophysiological tools but also considering younger populations, like the early years, which has the potential to explore developmental links between observable behavioural processes and underlying cognitive and affective processes. Moreover, an approach that allows researchers to combine different tools that capture the process of engagement, taking advantage of each of the tools presented here while also compensating its different limitations and disadvantages with more traditional research methods could potentially present a more complete picture of engagement as a whole in the contexts of science centres.

2.3.4 Potential of Triangulating Tools

While gathering best practices from researchers in the field of engagement to examine how the methodology for capturing engagement has improved over the years, some

authors have noticed that when tools are used individually they might reflect an incomplete representation of engagement (Azevedo, 2015; Boekaerts, 2016; Fredricks & McColskey, 2012; Sinatra et al., 2015). One of the most multifactorial perspectives used in science centres has been proposed by Falk and Dierking (2000) in their three contexts of the Visitor Experience Mode (personal, social and physical), emphasizing the importance of observing “natural” data of the interaction to arrive at better conclusions. While this approach can help encompass what a museum experience entails by providing a context to what personal, social and physical factors might influence a child’s engagement with a science activity, it does not fully address how these factors interact together to guide the process of engagement and science learning. One attempt to explain this relationship between engagement and learning in science centres was done by Barriault and Pearson (2010) after years of designing their Visitor Engagement Framework to characterise “observable behaviours and activities related to engagement that are indicative of learning” (p.94). Unfortunately, this approach captures a narrow view of all the different components of engagement, missing out on some of the richer information about the process of engagement or the unobservable changes underlying engagement. While they do consider that engagement might go through a process by acknowledging different stages (i.e. Initiation, transition, breakthrough), they limit their own instrument by measuring engagement as a binomial outcome of a specific activity (e.g. engaged / not engaged). Other research has also addressed a multicomponent view of engagement countering this binary state by defining different behavioural markers for different components of engagement. For example, *Going Ape!* (Tisdal & Perry, 2004) was a project at the Exploratorium looking for exhibit attributes that elicited active prolonged engagement during an interaction to facilitate “deeper cognitive experiences” (p.1). They analysed engagement from what they refer to as perspectives: emotional engagement, physical engagement, social engagement, intellectual engagement and timing of engagement. Each of these perspectives captured different factors from a visitor’s interaction, facilitating the comparison between exhibits because this

classification would explicitly point to different areas of engagement where one exhibit could do better than another. The last perspective, 'timing', would also incorporate some temporality into the comparison. Although this project was able to detect and examine better exhibits that fostered active engagement by using different perspectives, the authors also report that the process of intellectual engagement was not yet understood. Here is where combining cognitive science with visitors' studies can extend understanding of relevant cognitive processes, like the ones considered in this study for intellectual engagement, which the authors described as "various ways in which visitors engaged with their minds. Includes connections visitors make to existing knowledge during their interactions, the conceptual understanding and the questions they have" (p.7). New tools that claim to detect physiological and psychological markers of an experienced emotion could also enrich the emotional perspective by considering how to capture an emotional marker of an experience of engagement. However, care should still be placed on not relying solely on this marker, and if possible, observing or tracking, all the possible reasons behind the obtained emotional markers when making assumptions from it. This would lead to clarifying or discarding any particular factors that could be associated with the observed physiological marker.

There is also a new debate as to whether a more automated approach that benefits from current computer programming technology would be useful for capturing engagement by reducing biases and increasing time-cost efficiency (Azevedo, 2015; D'Mello et al., 2017). This kind of approach called an advanced, analytic and automated (AAA) approach, proposed by D'Mello and colleagues (2017), suggests taking advantage of advanced technological methods, like machine-learning, to automate the analysis of the events and discriminate between different parts of the process of engagement. Their aim is to reduce human involvement and increase temporal resolution by using tools, coupled with the context, which acquire data directly from a participant (i.e. person-oriented), in real-time. This is presented as more objective since it is claimed that data captured from these tools can

limit biases associated with self/observer reports and prevent data loss due to human fatigue and momentary lapses in attention/memory. The authors ground their method by reviewing 15 case studies that have used cost-effective, non-intrusive and user-friendly person-oriented tools, such as eye-tracking, electrodermal activity sensors, facial expressions and log files from any digital learning environment. While all these tools provide a window into both, cognitive and emotional, covert processes that the individual might not be aware of that are happening, these still need to be accounted for and considered as part of the overall situated behaviour of the individual.

The automation of tools that capture and analyse engagement presents an innovative approach that could benefit the field of engagement, particularly since it considers the learning context first and uses multiple data points. Although they can potentially help uncover hidden patterns in the process of engagement and set engagement research one step closer to generalising context-dependent events, these automated models should be used with caution, since any automation still relies on a human-based definition, human programming, supervising and interpreting, which are all prone to err. Furthermore, D'Mello et al., (2017) recognise that the automation process they propose, howbeit beneficial for specific fixed learning scenarios (i.e. screen-based), still presents limitations when working on live-scenarios since most of the tools and methods proposed rely more on digital reliable environments. Being that a science centre can be a messy, busy context (Chamberlain et al., 2012), it would capture a higher degree of noisy signals from these tools that could be wrongfully misinterpreted.

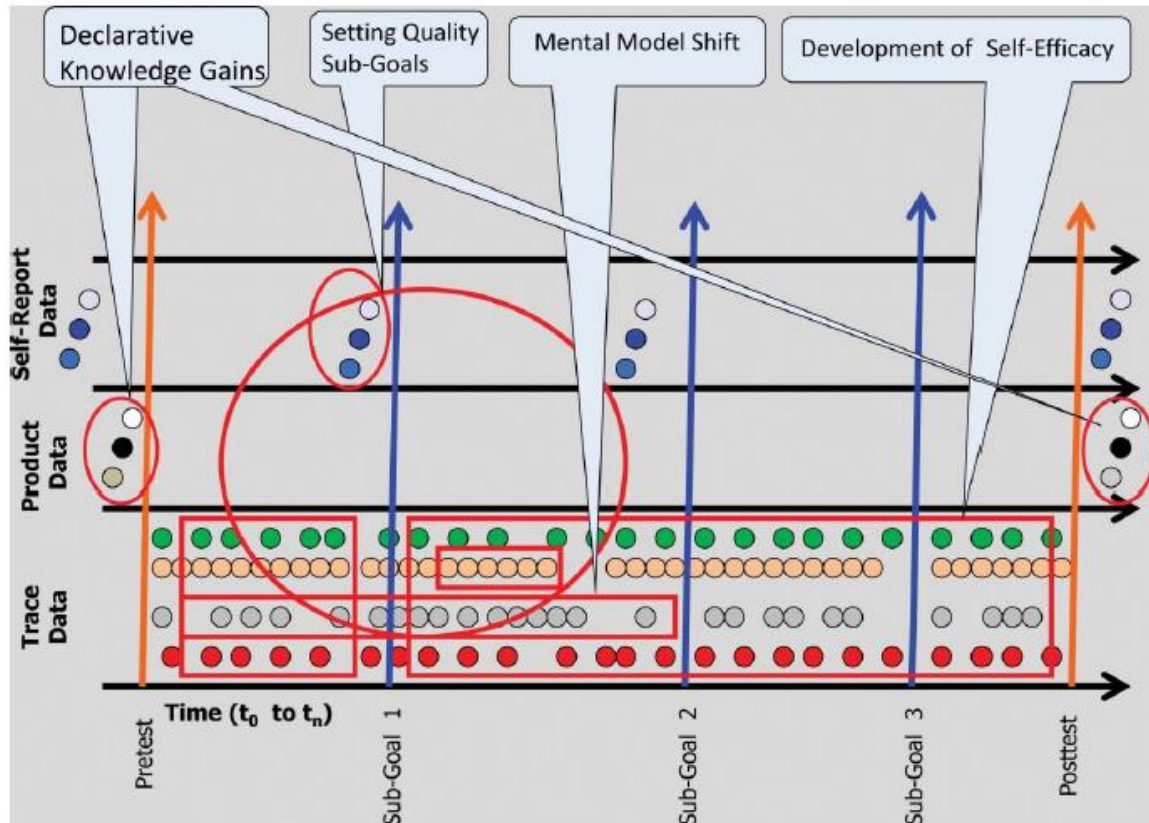
A balance is needed therefore between a method that depends upon single tools and one that uses a fully automated perspective. The solution could be a method that considers multiple data entry points from different tools, reliant on both external and internal signals of an individual, capturing different aspects of the process of engagement. A good example is described by Roger Azevedo (2015) in his commentary on the current state of research on

student engagement. Azevedo calls for “an overarching model of engagement that can be used to generate hypothesis and make assumptions regarding the role, timing, triggers, duration and quality of specific processes, mechanisms and constructs of engagement” (p. 88). He proposes extending and integrating different methods that capture engagement in order to answer the *how, why, when, what* and *by whom* behind engagement and learning in different learning scenarios. According to Azevedo, answering these questions provides a frame of reference that allows a researcher to identify, capture, measure and understand how engagement is characterised depending on the context where it is being elicited. Azevedo (2015) puts forward a research strategy that triangulates processes, outcomes and self-reported experiences of engagement using complex quantitative statistical, computing and tracing methods, like machine learning and data mining, and combines them with psychophysiological tools that capture traces of emotional and cognitive data covert from an individual using EDA and eye-tracking respectively, finishing with an additional layer of observational coding to tie everything together. As part of this strategy, Azevedo marks different points in time to account for and track specific temporal fluctuations that depend on the nature of each tool, whereas to test the activity learning outcomes and self-reported experiences of engagement, he proposes doing them before and after the interaction. This proposed triangulation means using different tools that “measure engagement” through different perspectives in different time-points, which might provide different pieces of the whole picture of engagement, helping characterise what could, or could not be, engagement in a particular setting.

Figure 2.3 shows Azevedo’s hypothetical representation of how different types of data could paint a more complete vision of engagement, with each point representing an individual stream of data (e.g. observational measures or eye-tracking results) and the contribution each of these provides. The trace data highlights the processes *during* the learner’s interaction and is captured between the pre and post-tests. This figure also shows

Figure 2.3

Representation of Triangulation Model Proposed by Roger Azevedo



Note. Image reproduced with permission from Azevedo (2015).

how the trace data could show a development of patterns over time when compared to the initial patterns observed, which he attributes to a shift in the mental model being used that might be leading to the development of self-efficacy. The figure implies that this could be driven by sub-goals within the main activity that lead to the mental model shift. Since Azevedo works with student engagement in self-regulated learning and metacognition (Azevedo, 2018; Azevedo & Cromley, 2004; Taub et al., 2018), the speech bubbles are very likely aimed at observing and capturing the variables of interest for this particular framework. However, the underlying triangulation strategy of different tools capturing the process of engagement lends itself to be adapted to other scenarios with other populations without being dependent on a specific theoretical framework. It allows flexibility of use as long as the

researcher remains aware of the specific information each of the tools used for tracing the process of engagement can provide, as well as what are the variables of interest being tracked for each situated context.

With any research strategy, there will be challenges and limitations associated with it. Azevedo acknowledges some of the main methodological challenges of this approach would be not knowing exactly how many data channels would be necessary to allow reliable inferences to be made from the results of the data, or if the statistical methods to analyse the data would be too complex and therefore not as accessible or understandable for other researchers. Additional limitations of this strategy would also be failing to consider and acknowledge the context and learning goal where the desired engagement is being observed as well as having tools that might not be appropriate or limited to capture a phenomenon outside the lab, and measures that risk being too rigid within less controlled scenarios, like a live interaction in a science centre or a school. Despite these, Azevedo believes that “focusing on process data will lead to advances in models, theory, methods, analytical techniques and instructional recommendations for learning contexts that effectively engage students” (p.93, 2015). While the latter paints an exciting picture to move the research field of engagement forward, it is critical for any researcher to stop and consider the limitations of using tools that capture some form of data and then claiming to have observed the “real” version of a construct, rather than only a layer of that construct. This follows the core propositions behind the philosophy of critical realism, where there is a difference between what *is observed or captured* of an event (i.e. empirical level of reality), an *event that happens in the world regardless of it being observed/captured* (i.e. actual level of reality) and what *causal mechanism are behind the events that are observed/captured* (i.e. real level of reality) (Fletcher, 2017). In other words, considering a critical realism perspective contributes by reminding the researcher that whenever engagement happens in the world (i.e. ontology) – naturally existing and operating independently of our knowledge of it, cannot be reduced to

what is known about engagement (i.e. epistemology) – and it should be considered with its situated socio-historic context used to construct this knowledge of what engagement is (M. Archer et al., 2016; Bhaskar & Hartwig, 2016; Price & Martin, 2018). Instead, there can be more than one possible explanation of what is being observed/captured, and as researchers, it is important to judge which possible theoretical explanation could be closer to the “real” reality by finding and understanding possible causal mechanisms behind what is observed. Hence, an interdisciplinary approach becomes more convenient, since it can observe and capture multiple layers of the same *real* construct, while also considering the specificity around the nature of the event, that is, its situated and unique context (Bhaskar & Hartwig, 2016). Thus, translating this back to Azevedo’s proposed research strategy, it is necessary to add a multi-perspective lens to research around engagement to improve its understanding, but whichever obtained outcome should not be used as a decontextualised generalisation of how the construct of engagement could be operating. Contrarily, having a broader knowledge of how engagement works in different situated contexts can lead to better understand some of its core mechanisms and some of its underlying relationship with learning.

Since Azevedo’s triangulation strategy (2015) was proposed, some research in learning environments has started applying their versions of a multimodal multichannel strategy with different populations. For example, research by Harley et al. (2015) used a combination of self-reported experienced emotions, electrodermal activity and automatic facial expression recognition to understand adult learners’ emotions while interacting with a multi-agent computerised learning environment where they had to learn about complex science topics. Malmberg et al. (2019) also used multichannel data combining electrodermal activity, automatic facial expression recognition and observational data to examine high-schoolers interactions during collaborative learning and its different phases while designing a healthy meal for an athlete. These two research examples use both observation and facial

expression recognition to ground the psychophysiological changes observed in an EDA sensor and provide context to understand what is driving emotional arousal.

A simplified way for this method has also been tested by Cain and Lee (2016) with two 10-12-year-old children taking part in a mechatronic-programming workshop, where children were only given a wearable wrist EDA sensor and a wearable camera (to record from a first-person perspective of the children's experience). This research aimed to understand the psychophysiological changes children went through while learning common programming skills. Interestingly this research defined the use of a wearable camera as an observation tool that also allows users to track where children direct their visual attention, since the camera moves along with the child, lending further support that head-mounted cameras can be a feasible (and more affordable) tool that can provide similar information to an eye-tracker when the temporal resolution and the high detail of eye-scanning pattern of a scene can be spared (Burris, 2017). Finally, the first ones to use a triangulation of measures with a preschool population were Sridhar et al. (2018) who were interested in understanding children's cognitive load as a relevant influence in their cognitive-affective states during learning experiences. Thus, they examined the feasibility of triangulating measures from self-reports, performance scores, coded observations from researchers and psychophysiological markers from wearable sensors (i.e. EDA and HRV). Here they explored whether physiological data combined with behavioural analysis could disclose any new information about the state of the participant's mental effort beyond observable performance scores. The authors worked with 18 kindergarten children between 4 and 7 years old in a quiet school classroom, asking them to wear an E4 sensor while they were asked to carry out tasks from neuropsychological batteries evaluating different cognitive processes like executive functions and overall skills. Their findings show that it was feasible to obtain direct psychophysiological and behavioural measures of cognitive load in young children, and while cognitive states were easier followed by observational measures, they found that these were always

accompanied by emotional markers. Their results showed that affective states determined how children coped with a task, but most importantly it provides evidence that by triangulating their observational measures with psychophysiological markers they obtained more information than using these markers alone, especially when both affective markers have similar measures. The authors also conclude that the future direction for understanding learning lies in researching the phenomenon in-situ.

Therefore, given the lack of evidence regarding engagement in science learning during the early years, it is important to engage not only with current technology but with a solid methodology which draws from the different examples shown here. Doing this can help address challenges when measuring aspects of engagement that are less easy to observe, such as its cognitive and emotional component, whilst tools such as the head-mounted camera add an additional mean to access another perspective of participation. That is why this thesis proposes using an adapted version of Azevedo's research strategy. Triangulating these tools presents a unique opportunity to capture in-depth, real-time, in situ data on engagement in informal learning contexts and within the early years. Within this research project, triangulation of tools refers to the use of different sources of data that have been designed to capture different cognitive, emotional and psychophysiological processes with different temporalities that have been related to the phenomena of engagement by different authors. As part of this triangulation strategy, the different data perspectives will be temporally aligned and it will be explored where their outcomes align and where there could be a divergence between them to actively understand the phenomenon of engagement. Using this approach could provide more detailed information about whether there is an influence of different cognitive, emotional or social factors during a short learning interaction.

2.4 Overview of the Research

2.4.1 Thesis Rationale

The early years are a key developmental stage for children, where it has been shown that improving learning experiences during these years can lead to significantly improved outcomes in later life. These early learning experiences combined with the cognitive benefits obtained from science learning, can enhance children's skills and help them to understand and navigate their world more accurately. As an essential part of science learning, engagement drives the simultaneously occurring cognitive and emotional processes that accompany and strengthen a learning experience, which can have long-lasting effects that develop into personal interests and attitudes that would allow future re-engagement with science.

Informal science learning experiences, such as science centres, give children the freedom to explore their own interests by having a vast display of science content, technologies and facilitation that cannot always be encountered in formal education contexts. Thus, exploring a method that captures and measures engagement more broadly, such as a multimodal triangulation approach, has the potential not only to help understand how younger children engage when it comes to science learning at an informal learning context such as a science centre, but also to extend this understanding to other learning contexts, both formal and informal, and to other population groups.

Focusing on improving engagement of a learning experience can also bring understanding about how particular features within an informal learning environment are playing a role in children's science learning, such as curiosity, previous knowledge, emotions, etc. This can benefit exhibit designers and science practitioners to purposefully increase engagement in diverse areas of a science centre and consequently elevating the cultural and science capital impact these learning spaces have on children. Moreover,

informal science learning spaces make efforts to invite children with little science capital, which has been related to lower levels of engagement with science (DeWitt et al., 2016), thus improving their engagement is of particular importance, since positive engagement from the early years can help foster future interest, engagement and positive attitudes towards science (DeWitt et al., 2013).

Current methods used when researching engagement in science centres tend to be limited and provide a brief answer of children's engagement after they have taken part in an interaction, mainly because the observation alternative can be more time and labour-intensive. Having an approach that combines multiple points of data while being adaptable to different learning contexts has the potential to be the base to future research methodological advances that are less costly, financially, in time and in previous knowledge needed to operate them.

The potential of triangulating tools for researching engagement in different contexts, such as a science centre is relevant, not only for developing a theoretical and practical understanding of the nature of engagement, but also to help learning practitioners make a more informed decision when it comes to selecting tools to capture different aspects of engagement. That means, providing them with the strengths and limitations of using single and combined tools to research engagement, which can be beneficial for designing learning experiences, science exhibits, or improving user experience, and ultimately help maximise science learning opportunities.

2.4.2 Current Research Project

This thesis seeks to explore how young children engage with a science exhibit by testing a multimodal triangulation method with tools that capture the three different components of engagement (behavioural, cognitive and emotional). In doing so, this work provides a more in-depth perspective of the learning experiences. The proposed triangulation

method also evaluates whether the process of engagement can be captured in an informal learning context with young children and the implications this could bring to engagement as a theory, since this perspective on the dynamism of engagement has yet to be examined in this context.

Given the scarce literature explicitly measuring early years' engagement in a science centre and the relatively new use of triangulating information via a multimodal approach, an exploratory approach was needed. According to Stebbins (2001), exploratory research "needs first hand understanding of the human acts being observed" (p.5), which then allows "the production of inductively derived generalizations about the situation under study"(p.5). In this case, this thesis concatenates a series of studies to explore a theoretically proposed research method for engagement, with the aim of developing and contributing possible generalizations about engagement in the science centre context and the early years population. These studies also seek to evaluate the feasibility and outcomes obtained from using a multimodal tool triangulation approach.

2.4.3 Main Aim of the Research

The main aim of this doctoral research was to examine engagement in early science learning in the context of a science centre to explore how and to what extent the process of engagement could be captured in young children aged 3 to 7 while they were interacting in a science exhibit. This was explored by using a multimodal-multichannel approach that combines and triangulates measurements from new and current tools that capture engagement simultaneously and evaluates how the tools that are currently being used in this context measure engagement. By doing this, the project also aimed to detect if certain events and variables might be characterising engagement, along with those that could trigger an emotional arousal and how these would impact children's engagement. The specific rationale for this research is explained in detail in the logic model in Appendix B.

2.4.4 Research Questions

Main Research Question: How could a multimodal triangulation approach using current and new tools capture engagement for early science learning in children aged 3-7 at a science centre exhibit?

To answer this question, this research project used the following 5 sub-questions:

- 1) How do young children interact naturally with different exhibits at a science centre (chapter 3)?
- 2) Is a multimodal triangulation approach feasible for children in the early years within a science centre (chapter 3)?
- 3) What is the relationship between the different outputs of the tools that measure different components of engagement (chapter 4)?
 - a) What is the relationship between the different variables used to capture and measure engagement?
 - b) Can children's emotional responses be predicted by the onset of cognitive-behavioural responses?
- 4) Do science practitioners' perceptions of engagement align with results from the multimodal approach (chapter 5)?
 - a) Is there a difference in practitioners' perceived engagement as a dynamic process vs engagement as a discrete outcome?
 - b) How much agreement is there between practitioners' perceived ratings of engagement?

Chapter 3

Children's Natural Interaction with Exhibits and the Feasibility of Using a Multimodal Triangulation Approach

3.1 Overview

In order to better understand the complexity of engagement in a particular learning experience, it is first important to gain a better sense of how children interact naturally in the science centre context - how long they choose to engage with different designs and the possible factors that may lead them to move on. Measuring engagement with the tools that are most used in science centres can provide a sense of the breadth or depth of what practitioners and museum researchers collect and interpret as engagement. Exploring these methods can point towards which underlying factors influencing engagement could be observed what is beyond the tools' capacity and provide guidance around how best to adapt and integrate tools that can capture and measure engagement from a multi-modal perspective, which are both accessible for children and an in-the-wild context.

Hence, this chapter examines RQ1 "How do young children interact naturally with popular science exhibits at a science centre?" which was answered by observing children interacting in a science centre with different types of exhibits. The chapter starts by discussing the different types of exhibits that can be found within a science centre, and their

different qualities, benefits, limitations, and affordances. This study helped to contextualise how young children frequented the chosen exhibits, their total interaction time and some of the reasons children leave the exhibit. The methods used also provided a perspective of a typical observation approach used in practice and what can be learned from it.

The potential to explore and understand the process of engagement in a science centre requires the use of new tools, which has been limited by pragmatic factors of the context and population of interest, that is, working with early years children in a chaotic space, an age period and context often neglected in research. Thus, the chapter also examines RQ2 “**Is a multimodal triangulation approach feasible for children in the early years at a science centre?**” to assess the feasibility of combining current and new tools in the desired context and population. This section discusses how the triangulation approach of different tools would capture the different components of engagement in children within a science centre exhibit and discusses the importance of having a feasibility study before carrying out a bigger data collection. The chapter ends by assessing and discussing how feasible it was to combine all the tools and their outputs to triangulate information for a single participant while they interacted with a selected science exhibit for five minutes.

3.2 Observing Engagement During Natural Interaction with Science Exhibits Study

A science centre has diverse types of interactive science exhibits with different aims and learning goals. One of the main differences in types of exhibits is how these are presented: for example, some are table-top based, where a table contains on top the elements required for the participant to interact and which exists only in front of the visitor (Dancstep née Dancu et al., 2015), others may be more immersive, where the visitor is surrounded by the exhibit and all the features are contained in the space. These two types of

exhibits can also be categorised into open-ended or planned-discovery. The former has been proposed as a space that encourages free exploration of a topic, encouraging the visitor to make its own experiments and build new knowledge using prior experiences and present interactions with the exhibit (Humphrey & Gutwill, 2005; Tisdal & Perry, 2004). As part of research for the Exploratorium in San Francisco, Gutwill (2008) found that attractive features of open-ended exhibits often have “beautiful aesthetics, opportunities for creation and intriguing representations that lead to active prolonged engagement” (p.192). In contrast, planned-discovery exhibits are goal-directed, that is they guide visitors’ interactions towards a desired outcome that has already been pre-set and it is up to visitors to ‘discover’ and interpret the content of the exhibit. Amongst the planned-discovery exhibits, exists a popular type of exhibit where counterintuitive phenomena are displayed. These were one of the initial types of interactive exhibits in science centres and science museums as they started to represent ideas and concepts more than they did objects (Rennie & McClafferty, 1996). Counterintuitive exhibits show visitors a discrepant event that conflicts with a “common knowledge” belief (Gutwill, 2008) making them surprising and attractive for a brief period of time, allowing visitors to discover new phenomena. Although the author argues that these types of exhibits can hinder the learning process by interrupting the inquiry process in a space where it cannot be further investigated due to time and feature limitation. Particularly since the underlying scientific explanation can not only be hard to explain, but hard to replicate given the narrow circumstances that might allow such discrepant event to happen.

According to museum researchers, interactive exhibits have been found to promote engagement with the exhibit and the content, facilitate understanding and recall of the exhibit and its contents along with positive affective memories of the interaction (Allen & Gutwill, 2004; Dancstep née Dancu et al., 2015; National Research Council, 2009). However, an increase in interactivity can also lead to detrimental results in visitors’ learning and engagement. Amongst the most common pitfalls of interactive exhibits where learning can be

hindered is the exhibit design interfering with the inquiry process (Allen & Gutwill, 2004). This may be due to having equal salience across all the exhibit features or due to interfering visitors – such issues, obscure the identification of the critical phenomenon shown (Allen & Gutwill, 2004). Furthermore, some computer-based interactive exhibits have also been criticized for diminishing the importance of the social component of visitor groups and interactions, by favouring individual interactions with a screen whilst the rest of the group observes, thus reducing topic discussion (Heath & Vom Lehn, 2008). These arguments highlight the diverse range of different interactive science exhibits in science centres and their possible influence on engagement.

To explore and understand how visitors navigate these different types of exhibits, systematic unobtrusive observations are typically carried out. The first step commonly used by museum researchers is to observe without being detected how real visitors naturally interact and engage with the space (Dancstep née Dancu & Sindorf, 2018; Tisdal, 2004; Tisdal & Perry, 2004). One of the quintessential methods to track an exhibit's engagement is by measuring holding time (Dancstep née Dancu et al., 2015; Gutwill, 2008; Serrell, 1997b, 2020; Tisdal & Perry, 2004; Yalowitz & Bronnenkant, 2009), whether that is timing users' interaction with a particular exhibit or tracking their route through a science centre. Holding time has been typically measured from the moment where a visitor plants both feet facing an exhibit, with their eyes or gaze directed towards the exhibit for at least two to three seconds. It is more than just an affordable and practical technique; it is not time-consuming and easy to use, thereby helping set comparative benchmarks both for old/new exhibits and also between similar types of exhibits. It has been found that exhibits categorised as 'promoters of active prolonged engagement' can have an average holding time of 132 seconds though interaction times can range from 1 min to > 23 min. Whereas the average interaction time with exhibits that have not been categorised as 'promoters' is of around 51 seconds and is normally no longer than 3 min (Tisdal & Perry, 2004). While it has been acknowledged that

holding time in itself is not enough of a tool to understand visitors' underlying reasons for engaging or choosing to end their engagement (Serrell, 1997a; Shettel, 1997), it provides a starting point from which to identify, track and record interaction or behaviour trends to explore further with more in-depth tools. The current study addresses RQ1: **How do young children interact typically with science exhibits at a science centre?** which aimed to provide a situational context of two types of science exhibits, a planned- discovery and an open-ended one, by comparing how long young children interact with them over an hour during regular visiting hours of the science centre.

3.2.1 Method

3.2.1.1 Participants Eighty-one children between 3 and 10 years old who visited Glasgow Science Centre for a day were observed in the present study. As no personal details were collected from participants, visitors to the exhibit were considered for observation if they: a) had entered the space of the exhibit and/or started interacting with it; b) appeared to be Primary age or under (under 12 years old).

Ethics – Approval for the observation study was granted by the Moray House School of Education and Sports ethics committee. The Glasgow Science Centre placed existing exhibit observation signs signalling that observation would be happening around the chosen exhibits. This is standard procedure in the Glasgow Science Centre when observation research for the science exhibits is being carried out because it ensures visitors are aware someone might be looking at them and taking notes, while also giving visitors the option to opt-out of that exhibit if they are not comfortable with the observation or to ask questions for any additional clarification about the research. Moreover, the observation signs were also placed to raise awareness of educational research happening at the GSC, to highlight that the researcher is working on a specific task and cannot provide help with locations and other

exhibits as well as to remind other staff in GSC that the researcher is working in the area, and they could support both researcher and visitors.

3.2.1.2 Exhibits To examine the interaction between exhibits, two exhibits were chosen that differed in their designed approach: a planned-discovery exhibit called *Cycloid Collision* and an open-ended exhibit called *Sand Patterns*. These particular exhibits were suggested by expert practitioners at the Glasgow Science Centre as popular science exhibits where young children had often spent large amounts of time or had reported them as favourites outside the designated area for early years children.

a. Cycloid Collision

Figure 3.1

Cycloid Collision Exhibit in Glasgow Science Centre



Description of Exhibit: A rectangle-shaped table that can potentially fit four people around it but only two people at the ends could interact at a time. It contains a U-shape metal track and two metal balls which can be placed inside both ends of the track. This exhibit

shared the table with another exhibit that allowed experimenting with speed and curves in smaller flat tracks (Figure 3.1). The aim of the exhibit is to explore what happens to an object's speed when two objects collide with the same speed from opposite directions.

Context of the Exhibit: The exhibit is situated on the first floor at the main entrance to all the exhibits. Visitors have to walk through this area to make their way around the science centre.

b. Sand Patterns

Figure 3.2

Sand Patterns Exhibit in Glasgow Science Centre



Description of Exhibit: A sand table that can accommodate between four and six children around it with two similar stations that allow multiple interactions. It contains two elevated discs and two big red buttons which make the discs spin when pressed. While at the table, children could interact with a range of objects of different sizes and colours:

beakers, graduated cylinders, scrapers and funnels (Figure 3.2). The aim of the exhibit is to explore the properties of sand with different objects and moving platforms.

Context of the Exhibit: The exhibit is situated on the second floor of the science centre, in one of the corridors that connects two different areas on the floor.

3.2.1.3 Procedure. Each of the exhibits was observed for 60 minutes. The researcher observed from a distance of at least 1 m away to limit children's awareness of the observation. The exhibits were observed in order of typical visitor trajectory: *Cycloid Collision* first and *Sand Patterns* next. In each of the exhibits, data was collected for children's holding time, as well as their perceived gender, an estimation of their age, who they were accompanied by (if anyone), and, if notable, the reasons for leaving the exhibit (e.g. left with parents). Their holding time was registered from the moment children first looked at and approached the science exhibit until they left the science exhibit space (i.e. ~50cm around the exhibit). Time was recorded in minutes, for analysis it was rounded up to the nearest minute. If children returned to the exhibit, this was recorded later in the time visualization but counted towards the same visiting time.

3.2.1.4 Data Analysis Firstly, descriptive statistics are presented for each exhibit. Tests that evaluate a normal distribution of the data were used with both samples' holding time data distribution to determine whether tests should be parametric or non-parametric. To compare the holding power of each of the exhibits, a non-parametric test (Mann-Whitney) was used to compare the average holding time of each exhibit along with the corresponding effect size.

3.2.2 Results

3.2.2.1 Cycloid Collision – Planned-Discovery Exhibit. Forty-five participants (24 female) visited this exhibit during the 60 minutes observation. The average holding time was of 1.9 minutes (SD = 0.81). Visualization of the visitors' holding time distribution can be observed in Figure 3.3, where each square represents a minute, and the colour represents the estimated age group of the visitor. Most children left the exhibit without any notable reason, as shown in Table 3.1. However, there was variation in the observed reason for children leaving, with children leaving to directly engage with another exhibit, or because parents left and/or asked them to leave.

During the 60 minutes of observation, children experimented briefly with the metal balls, testing movements and different ideas with the balls from both ends of the track. However, for this exhibit children did not stay to inquire long. The majority of the participants tested the collision of the exhibit's metal balls more than once before moving to the exhibit on the other side of the table. In some cases, children kept the metal balls that were part of this exhibit to use in the adjacent exhibit since both exhibits used the same metal balls. This impacted the length of stay whenever children could not find both balls to carry on the full experiment and fulfil the goal of the exhibit. Additionally, when it came to accompanying visitors of children few adults were present in this exhibit (29%), but children visited accompanied by other children as shown in Table 3.2.

Table 3.1*Frequency Distribution of Reasons for Leaving the Cycloid Collision Exhibit*

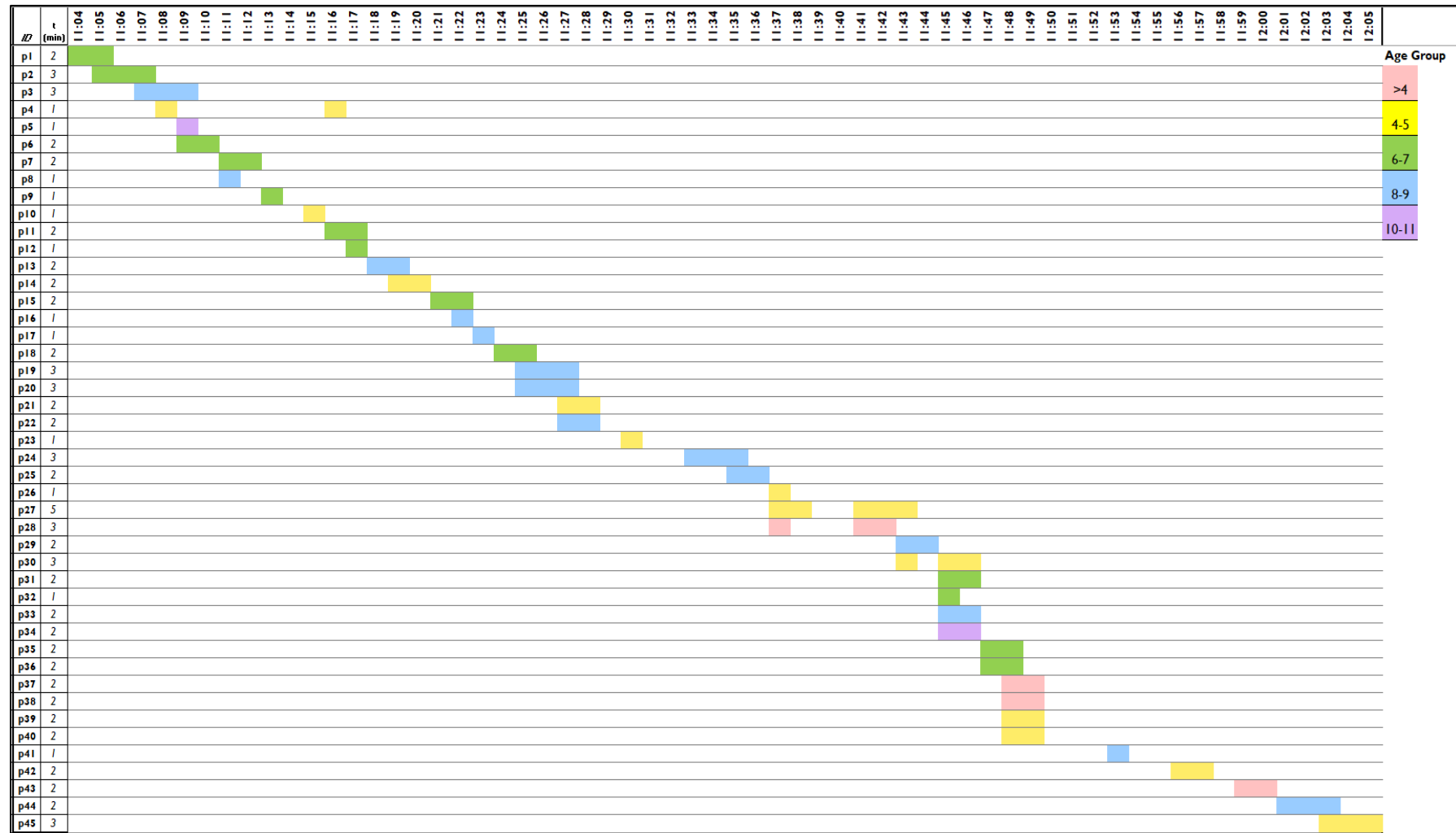
<i>Exhibit</i>	<i>Reason to Leave</i>	<i>n</i>	<i>%</i>
Goal	Unknown	30	66.7
Goal	Leaves with parent	6	13.3
Goal	Change exhibit	5	11.1
Goal	Parent left exhibit	3	6.7
Goal	Taken away	1	2.2

Table 3.2*Frequency Distribution of Children's Company in the Cycloid Collision Exhibit*

<i>Exhibit</i>	<i>Accompanied</i>	<i>n</i>	<i>%</i>
Goal	No	30	66.7
Goal	Mother	8	17.8
Goal	Father	3	6.7
Goal	Another child	2	4.44
Goal	Grandmother	1	2.2
Goal	Grandfather	1	2.2

Figure 3.3

Visualization of Holding Time per Participant in Cycloid Collision Exhibit



Note. Sixty minutes of observation of planned-discovery exhibit Cycloid Collision. Each square represents 1 min of interaction. Each colour represents one of the different age groups.

3.2.2.2 Sand Patterns – Open-Ended Exhibit

Thirty-six participants (16 female) were observed interacting with this exhibit throughout the 60 minutes of observation. Their average holding time was of 4.67 minutes (SD = 3.40). Figure 3.4 shows the distribution of visitors' holding time throughout the observation in order of appearance, where each square represents a minute and the colour of the estimated age group of the visitor. It can also be observed in Figure 3.4 that this exhibit had more children re-engage with the exhibit.

For this exhibit, most children also left without any notable reason. The second most common reason was parents leading them away, whilst some children changed exhibits intentionally (Table 3.3). During the observation, children would use the different objects available at the exhibit to experiment with the sand, often staying in the same place where they started the interaction. When children were accompanied by an adult, adults would either watch from a distance or try to get involved with the children by pointing to different available objects. However, most children did not appear to be accompanied (Table 3.4).

Table 3.3

Frequency Distribution of Reasons for Leaving the Sand Patterns Exhibit

<i>Exhibit</i>	<i>Reason to Leave</i>	<i>n</i>	<i>%</i>
Open-end	Unknown	26	72.2
Open-end	Taken away	6	16.6
Open-end	Change exhibit	2	5.6
Open-end	Leaves with parent	1	2.8
Open-end	Walks away by himself	1	2.8

Table 3.4*Frequency Distribution of Reasons for Leaving the Sand Patterns Exhibit*

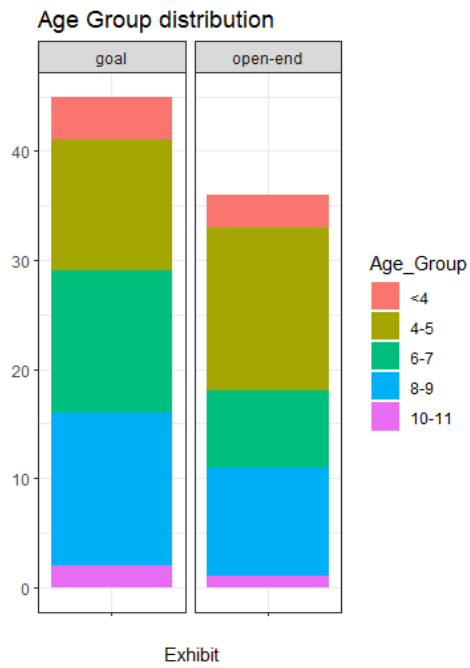
<i>Exhibit</i>	<i>Accompanied</i>	<i>n</i>	<i>%</i>
Open-end	No	28	77.8
Open-end	Mother	3	8.3
Open-end	Father	1	2.8
Open-end	Grandfather	1	2.8
Open-end	Mother and sister	1	2.8
Open-end	Family	1	2.8
Open-end	Older sister	1	2.8

3.2.2.3 Comparing Both Exhibits

The planned- discovery exhibit, *Cycloid Collision*, showed more attraction power – the number of visitors interacting – by having 45 children over the period of an hour compared to the 36 children that visited the open-ended exhibit. Figure 3.5 shows the difference in perceived Age group between both exhibits, where the planned-discovery exhibit showed a higher participation of 8–9-year-old (31.1%), whilst the highest age group for the open-ended exhibit was the 4–5-year-olds (41.7%). Both exhibits had approximately the same percentage of visiting children under 4 years old (goal = 8.9%, open = 8.3%) and had the lowest visit percentage of the eldest group 10–11-year-old (goal = 4.4%, open = 2.8%). Contrarily, the open-ended exhibit showed more retaining power and higher ranges in holding time (Figure 3.6). After failing to pass a normal distribution test ($W = 0.85$, $p < 0.0001$) a Mann-Whitney test was used to compare the distribution of visitors' holding time between both exhibits, which showed that holding time for the open-ended exhibit ($M = 4.67$ min) was significantly higher than for the planned- discovery exhibit ($M = 1.98$ min), $W = 363.5$, $p < .001$, $r = -.48$.

Figure 3.5

Distribution of Children's Perceived Age

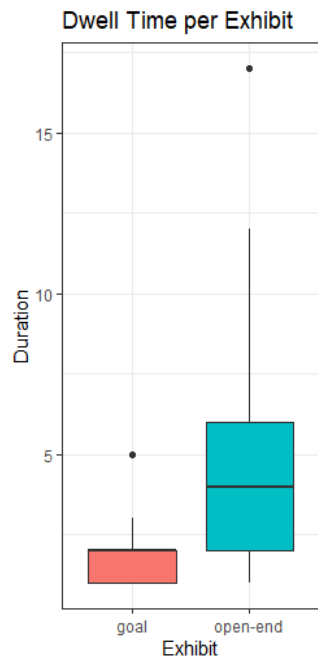


Note. Bar plot with stacked distribution per exhibit.

Y axis represents total count of children

Figure 3.6

Distribution of Children's Dwell Time



Note. Average duration time spent in each exhibit.

Y axis is measured in minutes.

3.2.3 Discussion

This study aimed to first investigate children's natural behavioural engagement in the context of exhibit interaction in a science centre, which would be the main focus of the next steps in this research, to further understand how this context operates daily. The observation was done with two different types of exhibits over the period of an hour using holding time as an indicator of engagement.

A planned- discovery exhibit, where children directly tested the collision and speed of solid objects, was found to initially attract more children over the 60 minutes of observation when compared to the open-ended exhibit. The attraction power observed for this exhibit could be related to the location of this exhibit, which is at the entrance of the first floor where the science centre visit starts and is one of the first stops of the visiting path. Research on museums' layouts has shown that exhibits that are in spaces where more people are visible or moving through tend to be more attractive (Y. K. Choi, 1999; Wineman & Peponis, 2010). Simultaneously, since the open-ended exhibit was on the second floor in an area next to the lifts and in between corridors, the location could also explain a relatively high number of visitors. When it comes to holding power, the Cycloid Collision exhibit had an average holding power of almost 2 minutes, which matches the reported average time spent on interactive science exhibits (Dancstep née Dancu et al., 2015; Tisdal & Perry, 2004). Although the average time could also position this exhibit as one that encourages active prolonged engagement according to the time measures observed by the Exploratorium (average of 132s), the maximum range of time spent by children under twelve years old was never more than 3 min which matches more the behaviour observed for those exhibits that do not promote active prolonged engagement. This holding time would equate to a brief (>1-3min) or very brief (<1 min) engagement or holding power according to a category of engagement level that is relative to the time spent in a science exhibit developed by the *Going APE!* team at the Exploratorium (Tisdal & Perry, 2004). Nonetheless, Serrell (1997b) also argues that when an exhibit's message is "clear and quickly and easily accessible"

(p.121) the time needed for understanding and using the exhibit will not be as high, which matches the evidence observed of children testing the collision of the metal balls no more than two or three times before moving on to a different exhibit. The impact of the order of exhibits should also be considered, where Cycloid Collision was one of the first stops after starting their visit, which could explain why children's engagement was higher than the average detected previously in research, for example if they are more excited because it was one of their first experiences in the science centre that day. Although this engagement was not held as long as with the Sand Patterns exhibit, probably because they still wanted to explore the rest of the science centre. This would have also kept children moving more towards other exhibits, therefore allowing more visitors to interact with it. Another variable that should be considered is that the exhibit only included two objects to experiment reducing the affordances of the exhibit. This was exacerbated when children took the objects from that exhibit to use in a nearby exhibit, hindering the exhibit's usability and potential to foster longer experimentation.

Conversely, the open-ended exhibit, although it had fewer visitors, had a higher holding time and re-engagement events once children started interacting. The average holding time was around 4min, which is higher than the average 132s found for exhibits that foster active prolonged engagement (Tisdal & Perry, 2004). Since the range of holding time in this exhibit ranges from 1-18 min, according to the time categories aligning with engagement levels proposed by Tisdal and Perry (2004), this exhibit would be encouraging not only very brief and brief engagement, but also substantial (>3-5min), extended (>5m – 8 min) and very extended (>8 min -23min). These higher levels of observed engagement were evidenced in the exhibit when children sometimes had to be physically removed from the exhibit by parents and other accompanying adults in order to leave. These results are not surprising for an open-ended exhibit, since this type of exhibit has been found as related to active prolonged engagement which allows for more exploration and creation, increasing the number of possible affordances to be done (Gutwill, 2008; Sandifer, 2003; Tisdal, 2004;

Tisdal & Perry, 2004). Another possibility for the increased holding time of this exhibit could be related to the location of the exhibit on the second floor, since research has found that the longer the duration is, the more likely for visitors to feel comfortable and confident with the environment (Nubani et al., 2018). This could lead to an increase in exploration time of some of the exhibits, where research has also found that the further along a visitation path goes visitors show more agency and choice (Wineman & Peponis, 2010), which an exhibit on the second floor might be likely to facilitate engagement and exploration longer.

Other factors which may have greatly influenced the data include the positioning of exhibits, the type of day (e.g. holiday, weekend) or signage. Additionally, since the science exhibits were chosen from GSC experts' perceptions, this study does not attempt to point to the open-ended exhibit as the most engaging exhibit in the science centre, rather than highlighting the differences in the observed interaction and engagement between an open-ended and planned discovery exhibit. In other words, multiple contextual factors are likely to influence children's decisions of *if*, *how*, and *how long* they engage with particular exhibits. Nevertheless, comparing interaction of multiple children between two exhibits in this study does reveal key factors such as the type of design (open/directed) and the influence of peers/adults.

This study also used a slightly different approach to reported timing and tracking data, where the focus was placed on the exhibits' holding power over a set time (i.e. 60 minutes), rather than the holding time of a set number of visitors using the exhibit (Dancstep née Dancu et al., 2015; Serrell, 1997b, 2020; Tisdal & Perry, 2004). Although it is reassuring to have found timing results similar to those reported previously on comparing active prolonged engagement exhibits with those that are not, showing that timing and tracking were used correctly. However, the findings also highlight how using just time can fall short when providing more detailed information regarding what is fostering said engagement. Thus, this method only manages to partially capture one of the aspects of engagement, the behavioural. Even if efforts were made to capture markers encompassing the other aspects,

cognitive and emotional, for example, using facial expressions or verbalizations that indicate out-loud thinking of children, it would be challenging to capture all of them in real-time or with any nuanced depth. If only this timing and tracking approach were to be used to detect whether children engaged more cognitively or emotionally with the exhibits, it would be relying mainly on the observer's subjective assumptions of what longer time with an exhibit meant. Whereas, detailing the specific interaction patterns of behaviour during a live observation would require a more in-depth analysis. An additional limitation present in this study relates to the observation notes and details captured through the data collection since attempting to record all possible information from their visits, along with arrival and departure times to the exhibit, can become problematic when several children are interacting with the exhibit and some relevant data can be missed (e.g. target visitors not recorded). It is also highly challenging to capture any specific interaction processes of each visitor or accurately track the different steps that could lead to engagement beyond the superficial information. Therefore, future studies that start examining engagement in-depth of children would benefit from focusing on a single participant or controlling as much as possible some of the confounding factors within the context, such as preventing being taken away in the middle of an "engaging" interaction.

In conclusion, this study highlights that engagement in young children could look different depending on the type of exhibit, leaving holding time as a measure that allows screening an exhibit's potential in attraction and holding power. Despite possibly resulting in engagement as an outcome of total interaction time, timing and tracking remain an incomplete measure to capture the process engagement since it lacks the capacity to explore children's behaviour during that holding time and gives no indication of emotional or cognitive engagement. Therefore, when observing engagement in young children at a science centre it is key to move beyond superficial measures as proxy for engagement and instead incorporate a multifactor perspective that considers engagement with its cognitive, emotional and behavioural components to better understand what could be influencing both

their process and outcome of engagement with each science exhibit (Allen, 2004; Azevedo, 2015; Rennie et al., 2003; Rennie & McClafferty, 1996; Tisdal & Perry, 2004).

3.3 Feasibility of Method Study: Testing the Triangulation Multimodal Approach and Its Tools

Azevedo (2015) called for researchers in engagement to incorporate new tools to capture engagement better, these tools have not been widely used with early years children outside a laboratory context. Also, whilst his method was tested with adult learners interacting with video game-based science learning on a screen (Taub et al., 2017), in self-regulated learning (Taub et al., 2018) or with young children's mental effort (Sridhar et al., 2018), it has yet to be tested as feasible for engagement in a real-world context and with a population of early years children. Feasibility studies are a type of study done before any large-scale intervention studies with the purpose of answering whether the study can be done (Tickle-Degnen, 2013). These types of studies handle any uncertainty that might be had before any main study, that is, they evaluate if the needed parameters, population, context or tools being considered are appropriate, sustainable and relevant for the main study and whether it is possible to use them in a full-scale study (Bowen et al., 2009; Lancaster, 2015; Whitehead et al., 2014). Feasibility studies are a necessary part of research regardless of the positive or negative outcomes found, and regardless of the field of study, because they help ensure that the research being conducted is based on robust methods (Lancaster, 2015). This way any possible outcomes would not be due to strange variables related to any parameters in the research design that were not accounted for (Tickle-Degnen, 2013).

Since this thesis is proposing the use of a method that has not been tested with individual tools nor with the specific early years in a science centre context, **RQ2 Is a**

multimodal triangulation approach feasible for children in the early years within a science centre? was needed to examine to what extent can the selected tools be triangulated and aligned in time whilst being used in a single early years child while they interact at a science exhibit. This triangulation method was based on Azevedo's work (2015) which proposes capturing different components of engagement through different tools for a dynamic analysis of the process of engagement throughout the time of the activity. Each tool has the function of capturing a different aspect of the components of engagement – cognitive, behavioural and emotional – and thus complementing each other. Cognitive and behavioural components were both measured by using engagement coding schemes for the interaction used with the video recording of the interaction. Video recording used a fixed-camera that provides a third-person point of view, and a head-mounted camera that provides a first-person perspective and which serves as a proxy to track visual attention. Emotional engagement would be measured by having the child use a wearable electrodermal sensor that captures their psychophysiological response (i.e. emotional arousal or emotional state) to the interaction. Table 3.5 shows how each tool could completely (double tick), partially (single tick) or fail (cross) to capture the three components of engagement based on Azevedo's table of components (2015).

Table 3.5
How Do Tools Capture Components of Engagement

<i>Component</i>	<i>Cognitive</i>	<i>Behavioural</i>	<i>Emotional</i>
<i>Tool</i>			
Engagement Codebook	✓	✓	✓
Video-Camera	✓	✓ ✓	✓
Head-Camera	✓	✓ ✓	✓
EDA	x	x	✓ ✓
Self-Report / Interview	✓		✓
Expert's Observation	✓	✓	✓

Therefore, this study tests the feasibility of the triangulation method of some of the tools proposed by Azevedo (2015) to capture the process and outcome of engagement, as well as the feasibility of using those tools with early years children interacting in a science centre context. The triangulation of a multimodal approach is designed to look as shown in Figure 3.7, where children will be asked questions before and interact with a science exhibit for 5 minutes, which is above the average (~2-3 min) in a normal science exhibit interaction, even if they foster active prolonged engagement (Humphrey & Gutwill, 2005; Tisdal & Perry, 2004). Also, after results from the previous study where children could interfere with each other during the interaction, it was of interest to first focus on children's individual interactions to reduce confounding variables.

To further understand how all these tools can interrelate with one another as well as with the objectives of the research, a logic model was developed (Figure 3.8) explaining the rationale behind each of the tools, the activities and outputs obtained from the tools and what the possible outcomes in short, intermediate and long-term could be and what they entail.

Figure 3.7

Proposed Design of the Triangulation of Multimodal Tools to Capture Engagement in This Thesis

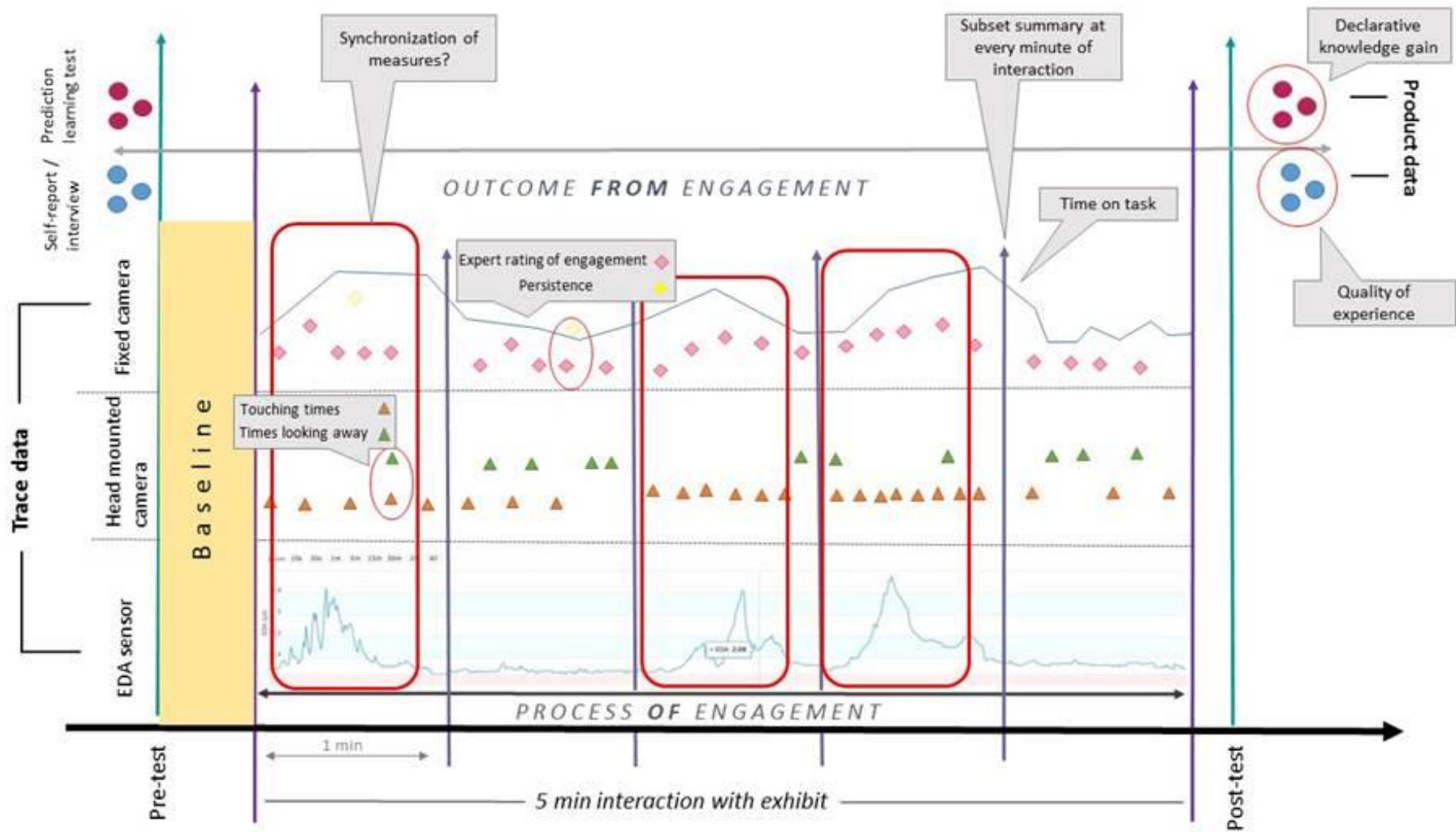
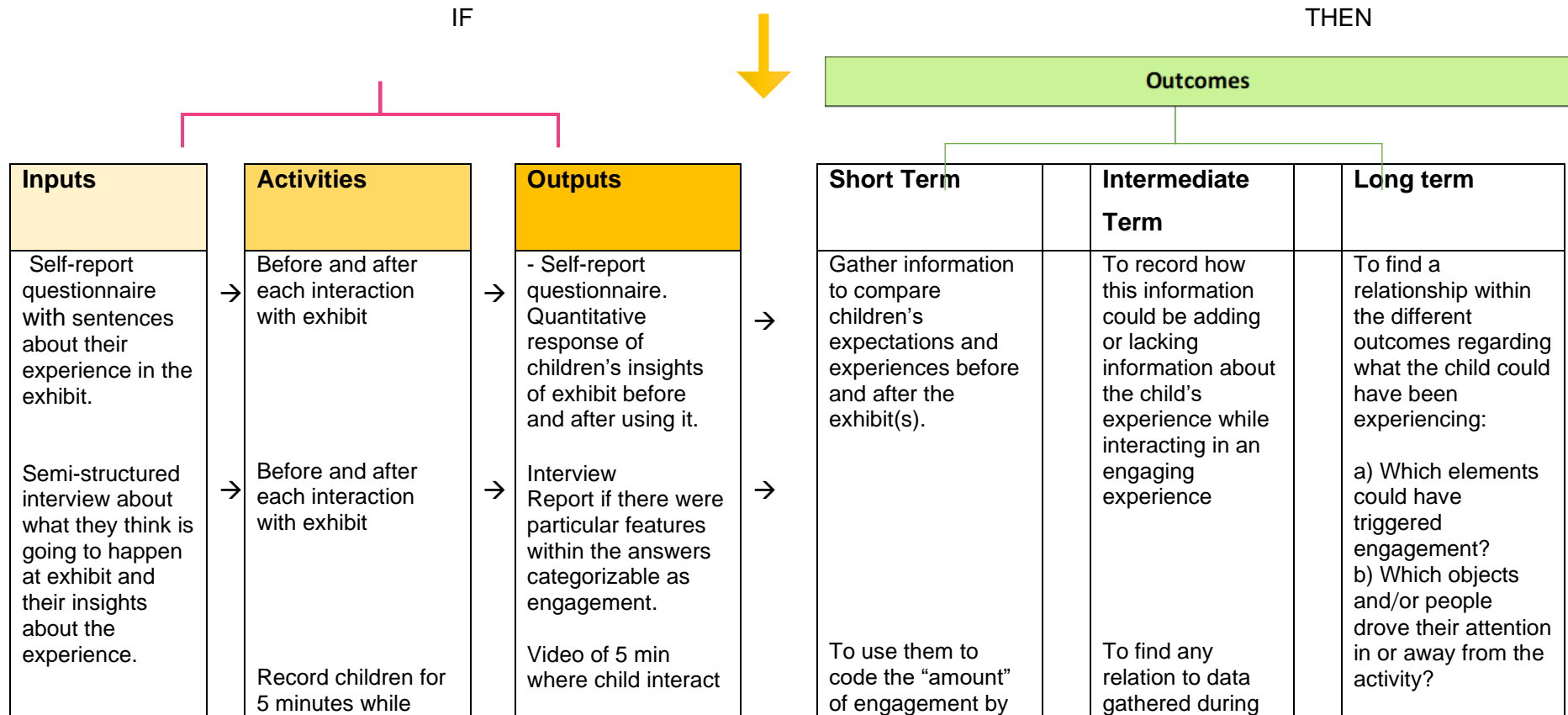
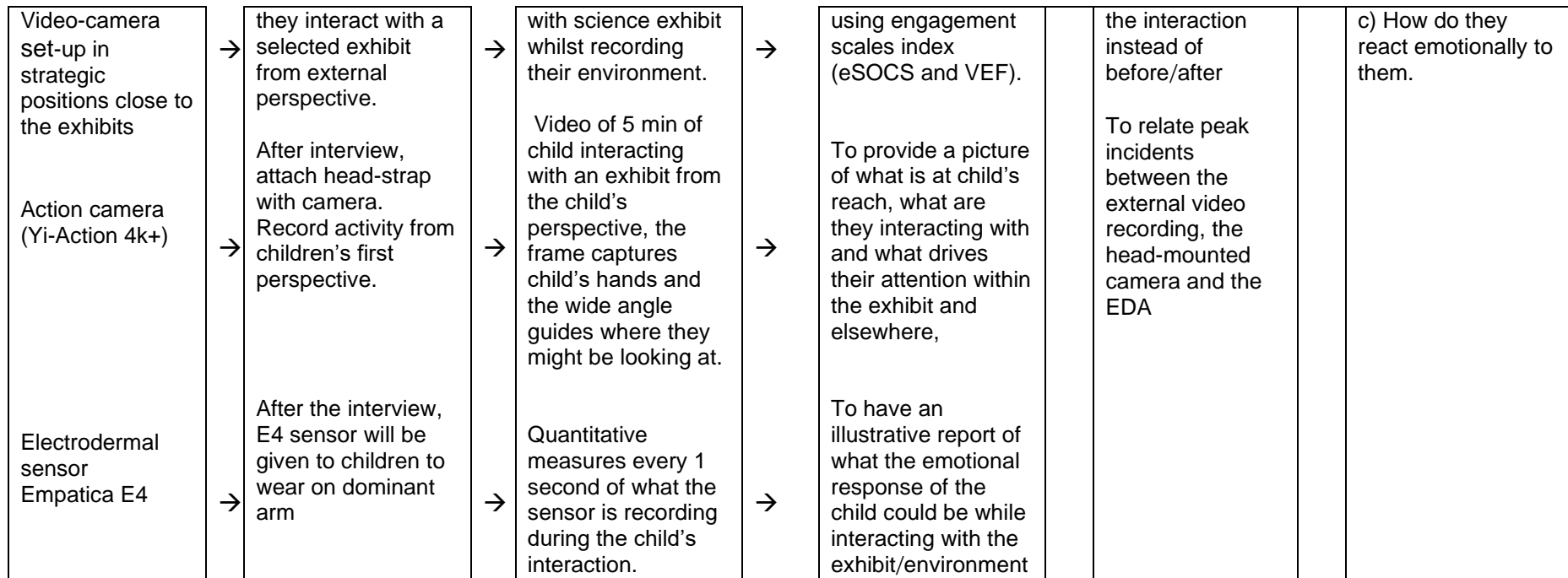


Figure 3.8

Logic Model for Mapping of Tools with Research Aims

Find the relationship between the tools that measure engagement in younger children during an engaging experience: Most of the tools used currently in real-world context are single-observational methods and psychological reports, while research has suggested a combined approach of tools that measure psychophysiological and cognitive responses with observation tools could lead to a better understanding of how engagement is working.





- External factors that could influence:**
- Child's attention and interest to interact for 5 min
 - Child does not find exhibit interesting or does not understand
 - Too many children that don't let child interact
 - Facilitator explaining
 - Someone standing in front of the camera covering the view while researcher is closer to child
 - Camera/sensor stops working
 - Noise that avoids hearing the child.

3.3.1 Method

3.3.1.1 Participants One child (7 years old boy) was recruited through convenience sampling from Glasgow City from a child of staff within the University of Edinburgh.

Ethics- The study was approved by the Research Committee Ethics of Moray House School of Education and Sport, University of Edinburgh. Information was provided about the project when inviting the child and their parents to participate, including showing them the tools and a photo of the exhibit in child-friendly terms. Written and verbal assent and consent were sought from both parents and the child, paying careful attention to any verbal or non-verbal signs of inhibition, reluctance or discomfort from the child.

3.3.1.2 Location This study was done in a designated space within the Glasgow Science Centre called Project Lab (Figure 3.9). This is a small room for temporal exhibitions located at the back of the first floor, thus providing an environment similar to that surrounding the science exhibits, while at the same time allowing a semi-controlled environment. The exhibit was placed within this Project Lab space.

Figure 3.9

Designated Space Within Glasgow Science Centre



3.3.1.3 Exhibits

A science exhibit that allowed a variation of levels of engagement was needed to better track the process of engagement. The criteria used for selecting a single engaging science exhibit was based on research done on designing engagement at the Exploratorium in San Francisco (Allen, 2004; Dancstep née Dancu & Sindorf, 2018; Tisdal & Perry, 2004), which can be found in Table 3.6 along with the rationale for each criteria point. The chosen science exhibit was selected from a shortlist provided by the Glasgow Science Centre which contained exhibits reported by children as favourites, or where children under 7 often engaged large amounts of time outside the designated Big Explorer area (i.e. area for early years children). Each science exhibit from the shortlist was rated to determine how many “engaging” criteria points they fulfilled. Each point had three levels to determine whether a science exhibit fully met the criteria, met a criterion with caveats or did not meet the criteria¹.

Table 3.6

Criteria for Selecting an Engaging Exhibit in GSC

Criteria	Points to Consider	Rationale
<i>Engagement Levels (high/low)</i>	Varying engagement through interaction	Allows the capture of different levels of engagement throughout interaction
<i>Location</i>	Relatively isolated, or easy to change location	Isolated - To reduce possibility of exposing children’s personal privacy by recording them accidentally if they were not part of the research and came into the exhibit space. Easy to change location – Allows transportation to the Project Lab room.
<i>Type of exhibit</i>	Open-ended	It should allow children to explore and inquiry the exhibit freely.

¹ See Appendix C for full details of exhibit selection according to rating of criteria points.

<i>Theme</i>	Preferably that relates to everyday life.	Relating to everyday life encourages more exploration and facilitates meaning making by connecting prior knowledge with new discoveries. This can be helped by containing familiar objects.
<i>Interaction</i>	Hands-on exhibit	An exhibit that allows for a reciprocal cycle of action-reaction between child and exhibit.
<i>Labelling / Instructions</i>	Clear explanation of exhibit <i>theme</i> . Either aims or topic	This could guide children's free exploration.
<i>Inquiry allowed</i>	An exhibit that allows questions like: "What's going on?" and "So what?"	Allowing for inquiry could provide varying levels of engagement depending on moment of inquiry process.
<i>Attraction Power (Look and Feel)</i>	Visited by children 4-7 years old. Playful, Whimsical or humorous.	To appeal to interests of this age group without the theme being too complicated to understand
<i>Exclusion criteria.</i>	Not in the big explorer zone (under 7s)	Big Explorer Zone is designed specifically for early years, however, it resembles more a soft play environment. Additionally, the space is very constricted, which conflicts with the placement of an external camera and isolation of the space.
<i>Capacity</i>	Allow for more than one person	It should allow for more than one person, but not restrictive to cooperative playing. That is, enough space for child to play - could be different stations; but not limiting the interaction due to a lack of participants.

Figure 3.10

Exhibit Placement Inside Designated Space



Note. Chosen exhibit Sand Patterns placed inside Project Lab. Video camera is also placed inside showing full set-up of video-camera.

3.3.1.4 Tools For assessing the feasibility of the triangulation method a

combination of different tools was simultaneously used.

- 1) Video Camera - To enhance the triangulation of events between the third-person and first-person perspective a video camera was set up in a strategic position close to the science exhibit that captured all of the exhibit, this would record both the child interacting and any possible events outside the exhibit. A wireless microphone was also connected to the video recorder to record sound from an external camera, further enabling evaluation of the feasibility of using a microphone along with the rest of the equipment (Figure 3.10).

2) ELAN – Open-software for video-coding developed by the Max Planck Institute. (*ELAN - Linguistic Annotator*, n.d.)

3) Action camera (Yi4k+) and head-mount² - An action camera (Yi Action 4k+[®]) was mounted onto the child's head using an adjustable head-band strap. Figure 3.11 illustrates what the camera looks like and how it fits a participant. The head-mount was used without any hat underneath as feasibility findings pointed to interference from the hat. The camera was connected wirelessly to an iPad that displayed the same image as the action camera.

Figure 3.11

Head-Mounted Camera



Note. Head-strap elastic band with camera (*left*). Example of child wearing it (*right*).

4) Electrodermal sensor (Empatica E4)³ - Wearable wrist-band device (Figure 3.12) – This device contains four sensors to track 5 different psychophysiological markers: galvanic skin response sensor to measure electrodermal activity, an accelerometer to measure movement, a photoplethysmography to measure heart-rate variability and blood-volume pulse, infrared thermophile to capture peripheral

² Details on feasibility of using HMC with young children year-olds while exploring a science museum can be found in Appendix E

³ Details on feasibility of using EDA with an early years child while exploring a science museum can be found in Appendix F

skin temperature and movement using an accelerometer. The device also contains a button in the centre that when pressed recorded it as an event marker. The E4 sensor can also connect to a mobile device via Bluetooth and stream the real-time data acquisition to the screen. *Ledalab* (Benedek & Kaernbach, 2016), was used to process the raw data. MATLAB- based open-source software for skin conductance analysis that allows importing, cleaning and processing of the electrodermal raw activity using a Convolution-Deconvolution method that separates the phasic from the tonic data (Benedek & Kaernbach, 2010b, 2016).

Figure 3.12

Empatica E4 Sensor Fitted on Child's Hand



5) *Engagement Observation Scales* - Two coding frameworks for engagement were combined to evaluate whether this combination allowed the tracking of behavioural and cognitive engagement⁴: (1) *Engagement States Observation Coding Scheme* (eSOCS- Deckard et al, 2011), which contains 13 items used individually, whether a child is interacting alone or in a dyad. The items are: 1.

⁴ Details on selection, feasibility and evaluation of different engagements scales in a science centre with younger children is detailed in Appendix D

Distraction/Attention, 2. Positive affect, 3. Touching of task materials, 4. Persistence, 5. Anger-Frustration, 6. Gross motor movements, 7. Fine motor movements, 8. Aggression, 9. Intrusiveness (only when another person is present), 10. Verbalizations during the task (to himself or others nearby), 11. Anxious, 12. Responsiveness to partner and 13. Independence/Autonomy. This scale also has four additional items that rate a children-dyad interaction which will not be used here because the study focuses on individual interaction. This coding system uses a 7-point Likert scale of how frequently each behaviour is happening, ranging from No instances to Constant instances. No information about validity/reliability was found for the full coding system (Deater-Deckard et al., 2014).

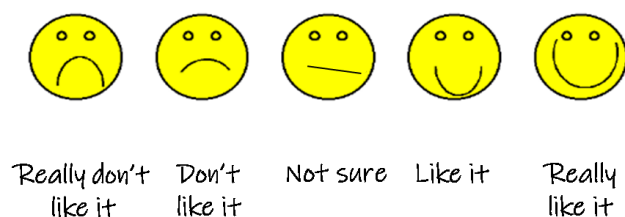
(2) *The Learning Engagement Coding Scale* (LECS – Halliday et al, 2018), since it was developed specifically for preschool children. This scale contains 7 indicators: 1. Attention to Instructions, 2. On-task behaviour, 3. Enthusiasm/Energy, 4. Persistence, 5. Monitoring Progress/Strategy Use, 6. Positive Affect, 7. Negative Affect. This coding relies on specific descriptors for rating each of their categories, which can range from no indicators to high indications of the behaviour. Predictive validity of this scale in an applied laboratory scenario is detailed in Halliday et al.(2018). Both scales were chosen because they can be aimed at children’s behaviour as well as tap into some of the components of engagement from an external perspective.

6) *Questionnaire with self-report and interview* - Self-report questionnaires are scarce when it comes to children’s engagement, the few found in the literature were either aimed at interacting with technology devices (Dietz et al., 2020), aimed at high-school students during class (Appleton et al., 2006), or for older children (Ben-Eliyahu et al., 2018). Given the lack of self-report questionnaires aimed at young children’s engagement in science centres, a bespoke questionnaire was developed. The questionnaire explored children’s previous experience with a science centre, their expectations of the science centre, and their attitude towards science and

towards the science exhibit. The questionnaire contained 25 items divided into three sections: I. Before playing with the exhibit (Pre-interaction) with 8 items, II. After playing with the exhibit (Post interaction) with 8 items, and III. After taking part in the exhibit with 9 items. The last section also collects feedback from children about the use of new tools, such as the head-mounted camera and E4 sensor, and how comfortable they perceived them. The questions used were informed by previous questionnaires used with older children to report engagement or user experience (Ben-Eliyahu, et al., 2017), the original version of this questionnaire required metacognitive skills that rely on an increased verbal component and a higher abstract and self-reflecting ability, which is not appropriate for all children within the early years age range. Appendix G shows the 25 question items used to build the questionnaire and the rationale behind each of these. To collect children's feelings and attitudes, the smiley-o-meter scale (Read & Macfarlane, 2006) was adapted and used throughout the questions, yellow faces ranging from sad to neutral to happy were used as a proxy to display the equivalent intensity of experience along with written words (Figure 3.13) making it easier for children to understand the construct and provide a more accurate account of their experience. The smiley-o-meter Likert scale was used as an evaluation parameter because it has been previously used with younger children to rate user-experience (Read & MacFarlane, 2006; Sridhar et al., 2018). The open-ended questions from the questionnaire aimed to detect particular features within the answers that could be categorised as engagement.

Figure 3.13

Example of Likert-Style Item with Smiley-O-Meter.



3.3.1.5 Procedure Firstly, a video-camera was set up before the interaction to a side of the exhibit, allowing both sides to be captured. Then, both parents and children provided verbal and written consent. The participant was interviewed to gather their interests and expectations of their participation in the study. After this, they were asked to wear the head-mounted camera and the EDA sensor. They were then positioned in front of the chosen science exhibit and were given five minutes of free interaction. Afterwards, participants were interviewed again to evaluate what they experienced after their interaction and gather their thoughts on the use of the tools. Finally, after their participation, they were rewarded with some stickers and a diploma for helping that says they are “a wee⁵ learning scientist”. Additionally, parents were present in the room but asked not to intervene while the child was interacting with the exhibit. Parents were provided with an iPad where they could see their child interact through a live-streamed video from the head-mounted camera.

3.3.1.6 Triangulation Strategy The strategy followed Azevedo’s (2015) proposed method for aligning in time multiple different sources of engagement. The measures that provide expectations and personal insight were intended to be analysed before and after the interaction, while the trace data from the EDA sensor, and the cameras and behavioural coding provided information about the process of engagement. The triangulation was carried out in the following steps:

1. Transforming raw data from the EDA sensor and exploring which type of results obtained would be better suited to triangulate information with the other tools. Ledalab provided two types of results: a list of points in time where the amplitude obtained was above an established threshold, highlighting Skin Conductivity Responses (SCRs) with amplitude and corresponding time-stamp. The other result that was provided was a plot with the decomposition of the phasic and tonic responses.

⁵ Scottish term for small or little

2., The video obtained from the head-mounted camera was edited by using a 3x3 grid to mark different locations within the frame. This was also used to highlight the focus of attention⁶. Coordinating starting points from both videos.

3. Behavioural categories derived from both engagement coding systems used were inserted into ELAN to be processed and then video-coded. The chosen behavioural categories along with each rationale are shown in Table 4.6. Each behavioural tier was then coded from the moment it was first observed until it was finished. For in-depth coding description, instructions and rationales go to Appendix H.

4. EDA measures obtained from Ledalab were added to ELAN software since they allowed time-series stamps to be added to the data, while also the obtained NS-SCRs were tracked within ELAN as an additional category.

Table 3.7

Behavioural Categories used from Engagement Scales

<i>Behavioural Code</i>	<i>Description</i>	<i>Rationale</i>
Time on task/ off- task	<i>eSOCS (Distraction /attention)</i> Total amount of time children do task-relevant activities.	A percentage of how much of the total time they did task-related activities as a proxy for “total engagement” (Serrell, 1997)
<i>cognitive engagement</i>	The extent to which the child maintains focus on-task and active productive engagement with the task throughout the session.	The time stamps can help triangulate quantitative information with EDA or qualitative description with head-camera

⁶ See Appendix H for more details about 3x3 grid and measures reflected from HMC

<p>Touching task-materials <i>behavioural engagement</i></p>	<p>eSOCS Amount of times child touches exhibit and/or task-relevant materials throughout total interaction time (5min).</p>	<p>Even if child is looking away, touching exhibit could provide an indication of child's passive engagement with exhibit as a spatial "holding place" for attention (Dancstep, 2018)</p>
<p>Fine motor manipulation <i>behavioural engagement</i></p>	<p>eSOCS Fine motor movements involving the device or task materials</p>	<p>This item could reflect active interaction with the exhibit (Heath & Vom Lehn, 2008).</p>
<p>Persistence <i>behavioural engagement</i></p>	<p>eSOCS Begins and completes task without breaking away, stopping or giving up.</p> <p>LECS This code comes into play when the child has some difficulty with a task.</p>	<p>An item that illustrates how many attempts a child is going to have before changing to a different activity of the exhibit/activity.</p>
<p>Enthusiasm or Energy <i>emotional engagement</i></p>	<p>LECS Quality of child's involvement with the task → appearing eager and enthusiastic to try each problem presented (regardless of success).</p> <p>eSOCS (<i>Positive Affect</i>) Smiling, laughing, excitement, enjoying the task.</p>	<p>Item that can highlight affective engagement and differentiate from "just doing" to "doing with motivation".</p>

Negative affect <i>emotional</i> <i>engagement</i>	<p style="text-align: center;"><i>LECS</i></p> <p>Shown through facial expressions, involving frowning, narrowing of eyes or a dejected look.</p>	<p>Useful to code other aspects of an emotion, such as facial expression or voice tone.</p>
	<p style="text-align: center;"><i>eSOCS</i></p> <p>May include frustration, anger, annoyance, sadness and boredom.</p>	

Monitoring progress-Strategy use <i>cognitive</i> <i>engagement</i>	<p style="text-align: center;"><i>LECS</i></p> <p>The extent to which the child is aware of their own progress solving the task and/or can recognise a problem preventing a correct solution and use problem-solving strategies.</p>	<p>Provides further insight into cognitive processing behind the observed interaction with exhibit.</p>
--	--	---

3.3.2 Results

Figure 3.14

Child Interaction with The Science Exhibit Using Equipment



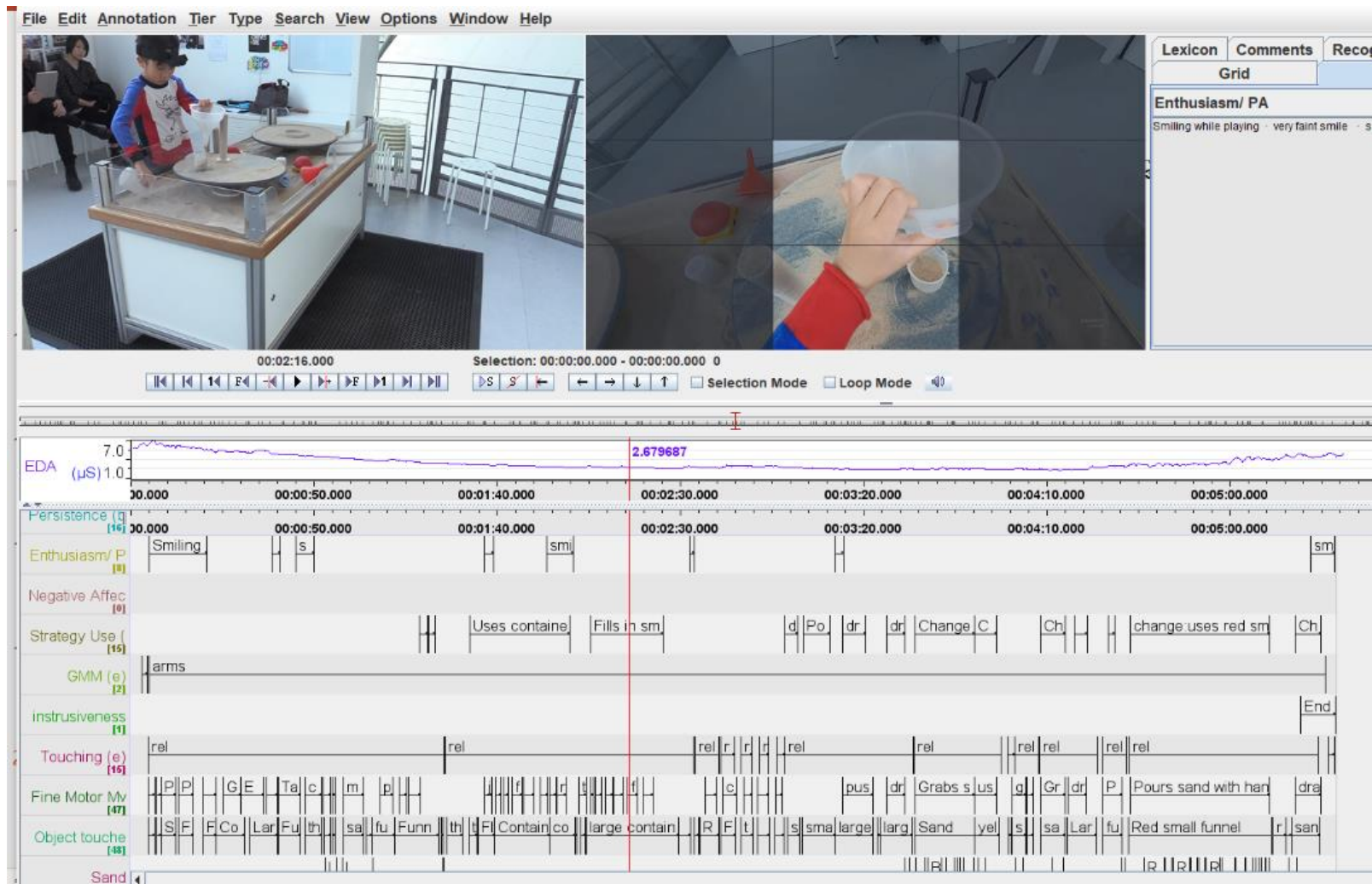
Note. Child is wearing head-mounted camera and E4 sensor on wrist.

3.3.2.1 Feasibility of Triangulation Tools. The results show that it is feasible for a young child to interact with a science exhibit while simultaneously wearing a head-mounted camera, an electrodermal sensor and a microphone (Figure 3.14).

When triangulating all the tools in the ELAN software, it was possible to triangulate both first-person and third-person view videos, whilst also using the head-mounted camera to track when the EDA sensor was pressed, marking the beginning of the interaction time. Figure 3.15 shows how the software looked with both video perspectives, which also facilitated coding the behaviours as they conveyed more information than either of them on their own. It can be noted that the grid used in the head-mounted camera might be over-obscuring the surrounding personal perspective of the child's view, for which it is recommended to either leave only the grid without shadow or reduce the opacity of the

Figure 3.15

Screenshot of ELAN Software Triangulating Trace Data



Note. ELAN coded interaction for 5 minutes. Screenshot shows both perspectives: third-person from fixed camera and first-person from head-mounted camera. First row shown is data from electrodermal activity sensor. Remaining rows show all recorded instances per code.

surrounding colour. Furthermore, ELAN allowed the download of the combined output, respecting time-stamp, category and description. This feature permitted the use of other software to analyse the time spent on each behaviour, the frequency of use and the frequency of the same events.

3.3.2.2 Feasibility of Behavioural Codes. While both scales used had similar items in common, such as *Positive/Negative Affect*, *Persistence* and *Attention to the Task*, they have different approaches when it came to the granularity of evaluating the behaviours. The first scale, eSOCS, was highly detailed in some sections, for example: about the type of movements, since it takes into account *Fine Motor Movements* and *Gross Motor Movements*, but it did not give room to differentiate between significant intentional movement related to the exhibit interaction and those that were just reassuring or unintentional gestures, particularly with the exhibit being a hands-on science exhibit. Also, the item *Touching*, which is useful to differentiate between intentional actions and just placing hands on the exhibit, when it comes to hands-on exhibit it does not give much more information on its own besides total touching time. Additionally, eSOCS differentiates between two types of Negative Affect, such as *Anger-Frustration* and *Nervous-Anxiety*, that in the current scenario might work better when combined as done in LECS. The second scale, LECS, allowed further exploration of the strategy behind children's actions and movements with its item *Strategy Use/Monitoring*, however, the scale to evaluate this item was oriented to a goal-driven kind of task with specific markers of task investigation and task achievement. Since Sand Patterns exhibit is open-ended, the descriptive scaling of this item would not be as helpful for the triangulation method, so it is recommended that this behavioural category is developed more.

Most of the codes used were easy and straightforward at the time of coding, except for the code "Persistence", which became ambiguous while coding since it was complicated to decide when the child had completed a task. The definition of 'task' also became

ambiguous, since it could mean both the whole interaction or a particular action the child was performing, however, the latter case could also have been part of another bigger “task”. It is then suggested that for future use of these coded, the coding definitions are revisited and operationally defined by what is meant by task, and whether to retroactively track the possible larger task so persistence can be coded with more precision. Developing a category that allows tracking of what the smaller goals children might be setting for themselves could become useful to follow whether these are new or repeated goals and how much they could be persisting on them.

When a category was tested to combine ELAN with EDA peaks, it was noted that NS-SCRs obtained from Ledalab had no specific duration. This hindered the triangulation of measures by not clarifying how many other behaviours happened around the time of an SCR. Since the natural response of EDA of 1-4 s (Boucsein, 2012), setting a specific duration that potentially matches this response might be ideal.

Finally, when the self-report questionnaire was used, the smiley-o-meter was shown to the child, and they were able to differentiate between the different states by choosing different intensity ones in different questions. However, some of the questions for the interview needed more than one re-explanation, which was taken as an indicator of an item that needed to be addressed for future versions. Since this child was on the higher end of the age range and was struggling with some of the items, it was assumed it might be a similar case with younger children. Thus re-wording of the selected items to a simpler version is recommended.

3.3.3 Discussion

This feasibility study aimed at evaluating whether it was possible for children to use all the tools simultaneously while interacting with a science exhibit. The study demonstrated that it is possible to use a head-mounted camera, an electrodermal sensor, a microphone and a recording of a young child interacting with a science exhibit. The participating child did

not show any signs of discomfort while taking part in the study. Therefore, tool-wise, it is possible to conduct a triangulation of different tools by having young children use tools that have been evaluated for comfort and practicality to capture measurable outcomes.

Moreover, this feasibility study also shed a positive light on triangulating the outcomes from the different tools and combining them for analysis using the software ELAN. The tools complemented each other, for example, the first-person view provided more detailed information about where the child was looking within their own scene and indicated the next objects they would interact or search for, which, when the 3-person view was used, was only noticeable when the movement was more pronounced. Furthermore, it was observed that there were not many instances where clear facial expressions of emotions were observed, however, the electrodermal activity sensor provided a higher number of NS-SCRs. In a full-scale study, considering the number of NS-SCRs peaks might be a better indication of what could be triggering these emotional responses.

These reported results should be taken with care since it was beyond the scope of this feasibility study to gather specific information that could lead to a first impression of the data, in this case of the time spent in each of the behaviours, as has been stated by Whitehead et al.(2014). Moreover, as part of the limitations of this study, the study was only done with a single participant who was in the upper end of the age range, however, it is assumed that since the feasibility of tools was done with younger children, the combined use of tools should not interfere with the data collection nor with any of the outcomes obtained from the tools. Furthermore, participation in this feasibility study was limited to one since the time for data processing and triangulation was unknown in advance and needed to be evaluated.

In conclusion, this study showed that it was feasible to use current and new tools to capture engagement with younger children in a science centre. This implies that it is possible to design and carry out a study that can help build up a scene for early years' engagement of how different measures of engagement were contributing individually to capturing and

understanding it as a minute-by-minute process. Particularly this feasibility study showed a potential to track some triggers within the environment that could be impacting engagement in a science centre, which could be used to facilitate and enhance early science learning. Since research has shown that increasing engagement in interactive exhibits can lead to an increase in learning (Dancstep née Dancu et al., 2015; Gutwill, 2008) the potential of this method becomes key for informal science learning.

Chapter 4

Understanding the Process of Individual Engagement in Children

4.1 Overview

The previous chapters have clarified that it is feasible to use emerging tools for informal science context research with younger children, such as an electrodermal sensor or a head-mounted camera, while also providing insight into what a triangulation method for multimodal tools would involve and what changes may be required when coding engagement through ELAN. After feasibility was demonstrated, the present chapter applies the triangulation method to capture engagement in a wider population of early years children interacting with the Sand Patterns exhibit at Glasgow Science Centre. This chapter seeks to understand the relationship between the different outputs from the tools, how these map to the components of engagement, and if there are any particular events detected that could be influencing engagement during the activity. Consequently, this chapter addresses RQ3 “What is the relationship between the different outputs of the tools that measure different components of engagement?” and RQ4 “Are there particular events that could be understood as drivers behind engagement during the activity?”. It then discusses the implications of using a combination of tools to detect cognitive and emotional drivers of engagement and how these could be acting reciprocally.

4.2 Capturing and Triangulating Multimodal

Approach

In the science learning literature, capturing different aspects of engagement has aimed to understand what drives engagement and how engagement drives science learning activities. While *Going Ape!* (Tisdal & Perry, 2004) was one of the first research projects in a museum that specifically considered engagement from four perspectives to understand what fosters active prolonged engagement in science exhibits, their measures were based only on observational data and in-depth interviews after an interaction had finished. Azevedo (2015) proposes the incorporation of tools, like psychophysiological sensors to current methods used, like video observation, to capture the specific nature of each of the three components of engagement in real-time, allowing the trace of engagement *during* an activity. However, this method has only been used to evaluate teenager and adult engagement when related to self-regulated learning during an educational science video game (Taub et al., 2017, 2018). Although a similar triangulation method has also been researched with kindergarten children, this was measuring mental effort during cognitive tasks and children were asked to always be seated in a quiet room.

The present study aims at understanding what happens during the process of engagement when a young child interacts with a science exhibit on their own. The multimodal triangulation method proposed by Azevedo (2015) was tested in the present study using tools that had been confirmed as feasible for the early years and for use in a controlled set-up within a science centre. The study also aimed to understand how the components of engagement might be interacting with one another during a science learning experience, while also evaluating whether the current tools used in informal scenarios can provide a depiction of engagement and under which circumstances would the triangulation method be a more appropriate choice than current methods used.

This research study seeks to answer the following questions:

RQ 3. What is the relationship between the different outputs of tools that measure the components of engagement?

- a. Is there a relationship between the tools capturing engagement?
- b. Can children's emotional response be predicted by the onset of cognitive-behavioural responses?

For this thesis to explore how engagement in a science centre can be captured, it was hypothesised that if the tools were accurately capturing different components of engagement, there would be an observable correlation between the tools measuring the same components of engagement. That is, for emotional engagement, there would be a correlation between the self-report items (i.e. items reporting their perception of enjoyment of the science exhibit), and the EDA outputs (i.e. amount of EDA NS-SCR or the amplitude of the NS-SCR peaks), as well as a correlation between either/both of these methods with behavioural observations of emotional engagement (i.e. facial expressions/vocalizations). Similarly, it was expected that the head-mounted camera provided more information and better accuracy when tracking behavioural and cognitive variables during the observational coding which could lead to an increase in the reliability of what was being observed. Furthermore, following the somatic marker hypothesis (Bechara & Damasio) and Azevedo's (2015) theory of synchronising patterns, it was of interest to explore whether there was a specific relationship to be found throughout the interaction between the psychophysiological markers used for emotional engagement that could have been driven by cognitive-behavioural markers of engagement. In other words, if during an engaging interaction, some behaviours would predict when a child experienced an emotional arousal response that perpetuated that engagement, which could be classified as emotional engagement. Lastly, it was expected to find some evidence of how the three components of engagement would be inter-relating during an exhibit interaction and whether there was a pattern behind some of

the outputs from the trace data.

4.3 Method

4.3.1 Participants

Twenty-eight children between 4-7 years old ($M = 5.39$, $SD = 1.06$, 14f) were invited from the visitors at the exhibit floor in the Glasgow Science Centre to take part in this study. Given that the Glasgow Science Centre is an open space for everyone, inclusion/ exclusion criteria for data collection, beyond the age restriction, were only limited by: a) the possibility that children could wear and use the equipment appropriately, b) being able to interact with the exhibit and c) by their willingness to participate.

Ethics - Written and verbal assent of each child, and accompanying adult was sought on every encounter, paying careful attention to any verbal or non-verbal signs of inhibition, reluctance, or discomfort. To ensure this, information was provided in child-friendly terms about the project when inviting them to them and their parents, showing them the tools and the exhibit they would be using. The study followed ethical guidelines from the British Psychology Society and was approved by the Research Ethics Committee of Moray House School of Education, University of Edinburgh.

4.3.2 Tools

The same tools as used in *Chapter 3, section 3.3. Feasibility of Method Study*:

Testing the Triangulation Multimodal Approach were employed:

- Self-Report Interview Questionnaire - In the Pre-interaction questionnaire, the item “*What do you think about science?*” was moved to be the second question to start contextualising children on the theme of the questionnaire. Whilst the item “*How do you feel about coming up to play in the science centre today*”

swapped third position with “*What do you want to see in the science centre*”. In the Post-interaction questionnaire, the item “*What did you have to do?*” was rephrased to “*What were you trying to do with the sand?*” to clarify what children had to answer. Then for the item “*What was your least favourite part?*”, *least favourite* was changed for *not so favourite* to decrease vocabulary complexity interfering in answering the question. Finally in the questionnaire post-interaction, the question rating the overall experience “*Did you enjoy taking part in this today?*” was moved after the feedback of the tools used to prevent biasing children before or after feedback from the tools used and to pair with the question regarding their attitudes and interests about science after taking part in the study⁷.

- *Fixed video-camera (Sony 4K-HD)*
- *Action camera (Yi4k+) on a head-mounted strap.*
- *Electrodermal sensor (Empatica E4)*

4.3.3 Science Exhibit – Sand Patterns

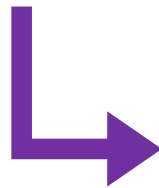
The same exhibit as *Chapter 3, section 3.3. Feasibility of Method Study: Testing the Triangulation Multimodal Approach* was used, however, the location was changed due to space availability in GSC. For this study, the exhibit was located on the 2nd floor area of the GSC, (Figure 4.1). The space was first tested in a pilot study using retractable queue dividers, however, other visitors crossed the lines and started interacting with the exhibit while the research was being carried out. Besides the aim of the project being focused on children’s individual engagement, having children approach the exhibit space while research was ongoing could also present ethical issues about confidentiality and sensitive data of children (and their parents) who have not given consent for being recorded. Wallboards

⁷ See full version of questionnaire in Appendix I “Self-Report Questionnaire”.

around two sides of the exhibit that faced the floor were placed, signs were set up on the walls to advise visitors about research being conducted inside.

Figure 4.1

Exhibit Space in GSC For Study



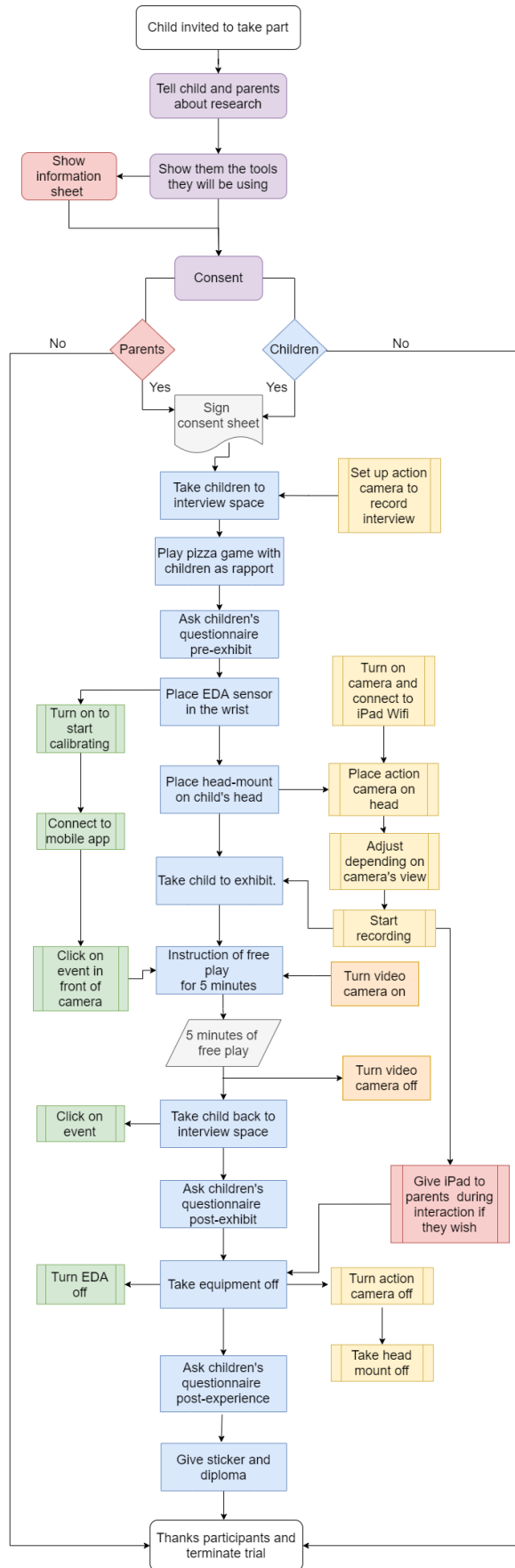
Note. Top image shows the space in the GSC where Sand Patterns was located. Bottom image shows how it looked for the study, enclosed by the wall boards and with all the tools set up.

4.3.4 Procedure

Children, along with their accompanying adults, were invited from the museum floor to take part in the research. When approached, they were told about the research, shown both the equipment and exhibit and if they agreed to take part, they were given full research details to allow for informed consent. In the case of the children, the researcher explained

Figure 4.2

Flowchart Showing Steps Followed by Participant from Recruitment Until End



the consent form and then if children agreed, the researcher would support children's written confirmation. For particular details of the procedure refer to Figure 4.2. Informed consent and placing of equipment were done at a table away from the science exhibit.

At this point, the video recording started to capture children's answers of the pre-test self-report questionnaire. Afterwards, they had the EDA sensor placed on their dominant wrist and waited until the E4 software was calibrated, for which they watched the iPad E4 interface that showed a "live-connection" to the sensor and indicated when calibration was done. Then, the head-camera was placed on their heads using the head-mount elastic band, and they were asked to look at their wrist while the researcher pressed the event button on the EDA E4 sensor to mark the beginning of the interaction time. Children were advised they could play however they wanted with the exhibit for 5 minutes, and the researcher would tell them when to stop. After 5 minutes passed, the researcher asked participants to stop and mark the E4 trigger event to signal the ending of the interaction. The head-camera was then removed from their heads. Finally, children were moved away from the exhibit and completed the post-interaction version of the self-report questionnaire. Children got a sticker and a "wee learning scientist certificate" as a reward for their contribution to this learning science study.

4.3.5 Data Processing

Before proceeding to the triangulation method, two different procedures were used to process parts of the data: an Engagement codebook used for video-coding behavioural and cognitive engagement, and the Convolution-Deconvolution Method for EDA data.

4.3.5.1 Engagement Codebook

I. Development of the Engagement Codebook

In Chapter 3, section 3.3 *Feasibility of Method Study Testing Triangulation of Multimodal Approach* two engagement scales – eSOCS and LECS - were chosen to track

and code children's cognitive and behavioural engagement in the multimodal method. However, as discussed earlier in that section, after coding the behavioural analysis it became evident that some items needed to be modified to provide a richer understanding of children's engagement during a continuous evaluation. This also meant adapting both grading approaches of each questionnaire for a more concise and descriptive one to capture the continuity of the behaviour.

In particular, more information was needed to track the item *Monitoring of Strategy Use* proposed by LECS. A restructure of this behavioural category was done where *Strategy Use* was no longer rated by specific task-related descriptors but was considered a sequence of observed goal-directed actions that were aimed to solve a problem (Börnert-Ringleb & Wilbert, 2018). Also, to improve the detection of strategies used, four different sub-classifications of strategies were added to this category: *New strategy*, *Adapted strategy*, *Going back or Previous strategy* (time-dependent) and *Repeated Strategy*. Along with this, a new behavioural category called Goal was created, which allowed the researcher to retroactively keep track of what the apparent goal the participant had behind a group of actions and strategies, this also had 3 subcategories: New Goal, Previous Goal, Same Goal. Since the science exhibit was hands-on it would be using fine motor movements constantly, thus the Fine Motor Movement category by eSOCS needed to be refined and constrained towards actions, where an action would be a fine-motor movement that revealed purpose/intention to an object in use. Finally, using the head-mounted camera to provide more information on visual attention, a category based on the cognitive markers obtained from early feasibility testing⁸ was created called Visuo-Tactile Attention Mismatch. This helped track whenever the participant looked away from the objects being manipulated while doing an action, which has been associated with anticipation of another action or expected reaction (Land et al., 1999). All the items were modified to be able to track behavioural, cognitive and emotional engagement accordingly. Appendix J describes how each of these

⁸ See Appendix E "Testing Feasibility of Head-Mounted Camera" for more details.

categories captures engagement, the assumptions of the underlying processes behind each of these categories and what was expected from each code. This was used as a guide to redefine the final labels described in the following section.

The development and adaptation of this codebook was an iterative process of evaluating the literature for the items identified as problematic in previous feasibility findings⁹. This process also involved deep discussions between three expert researchers, based on literature and group revision of in-depth coded examples of new proposed items.

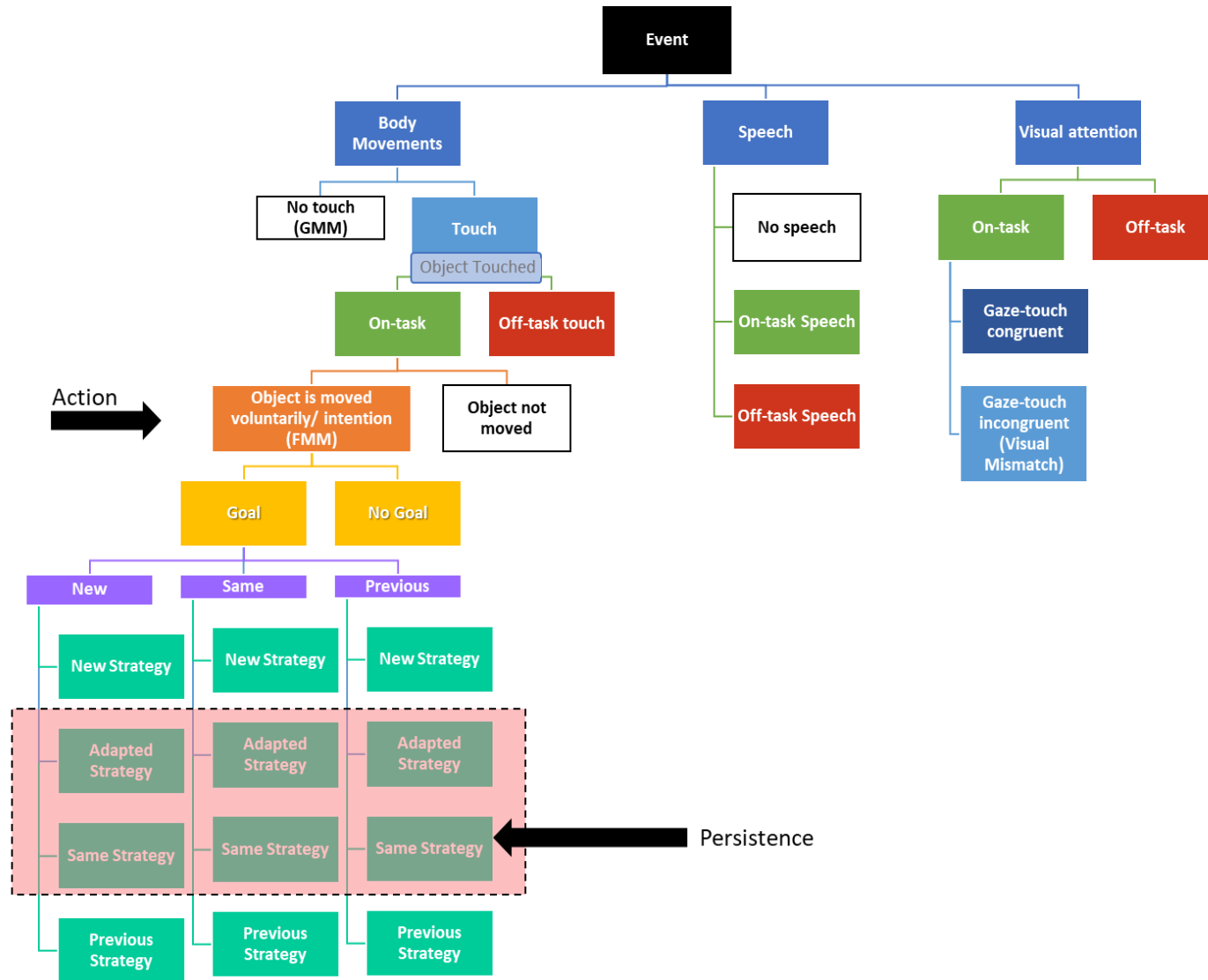
II. Engagement Codebook

The codebook developed and used to code the videos comprises 12 items: Touching, Objects touched, Actions-Fine Motor Manipulation, Strategy Use, Apparent Goal, Persistence, Enthusiasm/Positive Affect, Intrusiveness, Gross Motor Movement, Verbalizations, Visual Mismatch, Interesting Moments. Figure 4.3 shows a code-tree that depicts how each of these tiers is related to one another hierarchically and Table 4.1 shows an abbreviated version with the names, description, data entry and an example. Full details of the engagement scale can be found in Appendixes J and K.

⁹ See Appendix D “Testing feasibility of Engagement scales” and Chapter section 3.3 “Feasibility of Method Study: Testing the Triangulation Method”

Figure 4.3

Code Tree Showing Hierarchy of Codes from Engagement Codebook



Note. Code tree represents the different paths an event can be categorised once coded. Each colour represents the main event tiers.

Table 4.1

Codebook items (Abbreviated version)

ELAN tier 1: Touching (from eSOCS)	
Definition/ Description	Whenever children’s hands (either one or both) are positioned on any relevant part of the exhibit and/or touching task-relevant materials.
Data entry	There is no data entry here, just time selection.
Examples	<i>*When a child touches the red button in sand table. This counts whether they push/rub or just touch object.</i>
ELAN tier 2: Action -FMM (Fine Motor Manipulation) (modified from eSOCS)	
Definition/ Description	<p>Actions performed throughout the task that involve manipulation of task-materials.</p> <p>Action: A voluntary movement in the interaction using an object which "suggests" the intention of changing something in the environment, e.g. moving a hand in the sand</p>
Data entry	The description will always start with a verb representing the action, then how the object(s) were used.
Examples	<p>1. <i>Touches red button</i></p> <p>2. <i>Grabs Funnel</i></p> <p>3. <i>Pours sand through funnel</i></p> <p>4. <i>Puts funnel in flask and observes it.</i></p> <p>Exceptions:</p> <ul style="list-style-type: none"> • One hand is grabbing a container and is directly collecting sand from table (1) and using other hand pushes sand into the container too (2). <p style="text-align: right;">When NOT:</p> <p>* Changes item from one hand to other, fidgets with object in hand, when objects are ONLY touched but not manipulated.</p>

ELAN tier 3: Strategy Use
(modified from LECS)

Whenever the child is performing one or more actions that appears to be trying to solve a particular problem in the moment.

Or

That child appears to be reaching a particular goal/outcome

**Definition/
Description**

(e.g. Goal might be 'to see what happens when I do X after I have used Z' or it might be 'fill container'):

*Involves the extent to which the child seems to be aware of their own progress solving the task and/or can recognise a problem preventing a correct solution and use problem-solving strategies.

Data entry

- **New** - An action that has not been performed previously in the interaction
- **Adapted** - Whenever there is a change to the previously used - new (or adapted) - strategy that still appears to have the same goal.
- **Previous or Going back** - When returning to an action that has been used before at some point in the past of the interaction, but not immediately before.
- **Repeat** (previous action) - An action that has been used immediately before with no changes or actions in between at all.

Examples

- New* – First time a child has brushed sand off the table
- Adapted* - Adapt (object): goes from big funnel--> to smaller funnel.
- Previous or Going back* - Going back to using container to fill sand from table.
- Repeat* - Pours sand into container (graduated cylinder)

ELAN tier 4: Apparent Goal*
(New tier)

**Definition/
Description**

Higher-order category of Strategy - Describe whether the strategies used seem to be following a same goal/solution, a new goal/solution or are going back to an "unfinished" goal/solution.

Data entry

- **New Goal** - Whenever it seems to be a new goal/problem to solve
- **Same Goal** - Whenever the strategy seems to be following the exact same goal as the strategy used before (new or previous)

- **Previous Goal** - Whenever the strategy seems to be going back to a strategy used previously in the interaction to solve a previous goal or one that was left unfinished, but not immediately used before.

Examples

New goal: Fill in container

Same Goal: Fill in container

Previous: Brush off sand from spinning disc

**ELAN tier 5: Attention Off-task
(modified from eSOCS + LECS)**

**Definition/
Description**

The period when the child deviates attention (focus) from the task or task-relevant items. This could be in a form of visual attention, touch or verbalizations.

**Data
entry**

- Verbal (hands are on-task, but speech is not task relevant)
- Visual (hands are on-task, but visual attention is not task-relevant)
- Touch (vision is on task, but hands are touching task-irrelevant items).
- Mixed (when more than one modality is off-task).

Examples

1. *Someone approaches exhibit*
2. *They are talking with parents about unicorns.*

**ELAN tier 6: Enthusiasm/ Positive Affect
(modified from eSOCS + LECS)**

**Definition/
Description**

An expression of a positive emotion while interacting with the exhibit.

Data entry

Describe any expression of enthusiasm:

- Smiling, laughing, excitement, enjoying the task; vocalization and/or facial expression.
- Appearing eager and enthusiastic to try each problem presented

Examples

1. *Smile while playing*
2. *Excited vocalization after successful action*

**ELAN tier 7: Negative Affect
(modified from eSOCS + LECS)**

Definition/ Description	An expression of a negative emotion while interacting with the exhibit.
Data entry	Describe any expression of a negative emotion (frustration, sadness, annoyance, boredom, anger): <ul style="list-style-type: none"> • Yelling or tossing elements from the exhibits. • Loud vocalization with negative emotion • Expressing feeling a negative emotion • Tears
Examples	<ol style="list-style-type: none"> 1. <i>Frustration when disc won't stop turning</i> 2. <i>Forcefully tries to stop disc</i>

**ELAN tier 8: Intrusiveness
(from eSOCS)**

Definition/ Description	Physical and/or verbal events (whether voluntarily or involuntarily) that disrupt, manipulate, or control child. (Added here whenever there were external factors that disrupted child's interaction)
Data entry	Describe event (person/item/situation) that caused intrusion
Examples	<i>Uninvited child walks towards exhibit</i>

**ELAN tier 9: Body Movement – GMM (Gross Motor Movement)
(modified from eSOCS)**

Definition/ Description	This definition is grounded on developmental definition of gross motricity, but since Action- FMM always entails movement of arms, it will only track whenever child is moving around the exhibit or in their same place but changing position significantly (e.g. standing --> walking).
Data entry	<ul style="list-style-type: none"> • Walk • Whole body - Whenever a child is crouching/sitting/ standing up
Examples	<p style="text-align: center;"><i>Walking - When child circles around the exhibit</i></p> <p style="text-align: center;"><i>Whole body - when child sits down on stool</i></p>

**ELAN tier 10: Verbalizations
(modified from eSOCS)**

Definition/Description Task-related speech, either self-directed or directed at surrounding members.

Data entry Transcription of their speech.

Examples *Child mutters: "will this be full?"
*Child speaks to adult: "Whoops, it fell down!"

**ELAN tier 11: Visuo-Tactile Attentional Mismatch
(New Tier)**

Definition/Description The periods of time when the head-camera does not match the location of where the hands are being placed, yet it is still task related.

Data entry Description of what seems to be at the centre of the video.

Examples *Gaze at spinning disc while having hands on sand.
*Looking away because he is listening to container.

**ELAN tier 12: Interesting Moments
(New Tier)**

Definition/Description Any other behaviours or events that are not considered but that they seem interesting or relevant for the interaction.

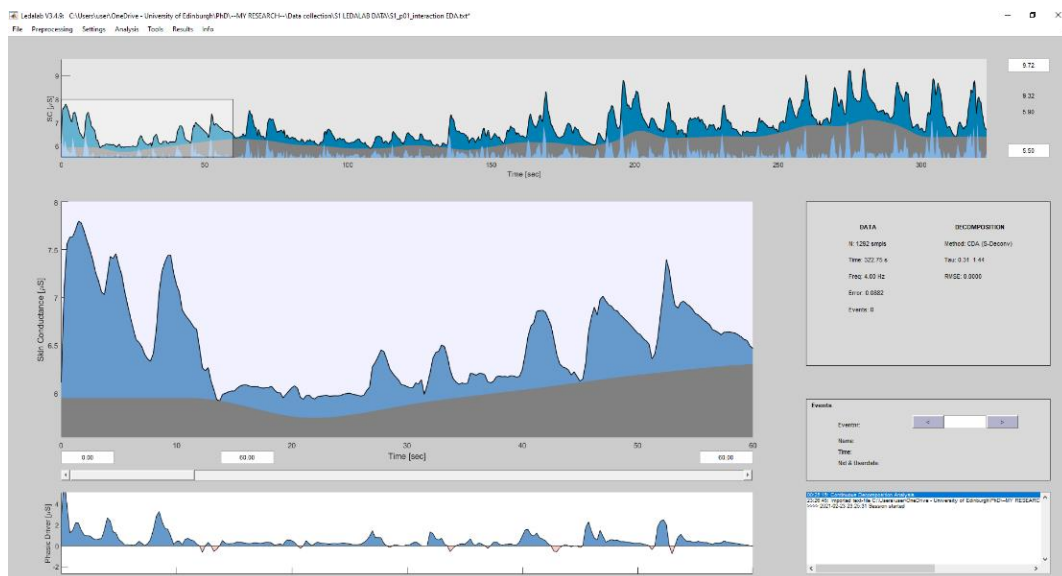
Data entry Description of moment

Examples *Tries to set up the cylinder and can't because of sand on spinning disc
---> Uses his hand to swipe sand off spinning disc
*Knows what part of container is easier to pour from

4.3.5.2 Convolution-Deconvolution Method This method by Benedek and Kaernbach (2010) proposes to use an impulse response function (IRF) that delayed the EDA wave pattern and uses this algorithm to map the raw EDA data and split the activity into its phasic and tonic components, making the SCRs more observable. To do this, their software Ledalab contains a function that transforms the data. Figure 4.4 shows an example, where the tonic response – in grey colour – can be observed slowly increasing, while the phasic responses are shown as blue peaks on top of the tonic response.

Figure 4.4

Example of EDA Response After Using Convolution Deconvolution Method



Note. Tonic activity can be seen in grey colour and phasic activity peaks are shown in blue.

4.3.5.3 Triangulation Method Once all the data had been captured, it was prepared and processed by aligning and triangulating it. The data had to be processed in 3 different ways: 1) coding behavioural engagement in videos, 2) processing EDA data to obtain both phasic and tonic activity, as well as the Skin Conductivity Responses (SCRs), 3) triangulation method. This meant that the different tools had to be prepared first: video data (both fixed and head-mounted camera), EDA data, and self-report questionnaire.

a. Video data. The fixed-camera videos were trimmed and split into the pre-post questionnaire and the interaction time, marked by the researcher instructing to start and stop playing with the exhibit, which also aligned with pressing the EDA button to record an event for Start/Stop. The head-mounted camera videos had a timestamp and were also edited and trimmed to start and end at the time the EDA button was pressed on camera. Being able to observe with one of the cameras the exact time in which the E4 is pressed, along with the time stamp from both the head-mounted camera and EDA sensor helped to align the videos with the EDA recording. The video-coding software ELAN (ELAN - Linguistic Annotator, n.d.) was used again to process and code the behaviours. To do this, both videos, fixed and head-mounted camera, were synchronised to the same point in time when the E4 sensor event button was pressed. The video coding was done in 3 different rounds where each round built on the previous one: 1) Coding and categorising all the touched objects from the exhibit, 2) Actions- FMM, Attention Off-task, GMM, Positive/Negative Affect, Verbalizations, Visuo-tactile Mismatch and Intrusions were then coded and described, 3) Goal behind the Actions and perceived Strategy, for this round it was useful to have the descriptions of the actions in advance, as it tracks retrospectively where one goal started and finished. The data was then extracted into a CSV file that contained the beginning, end and duration of each code in each of the event tiers. Between the different coding rounds, the videos were never coded in the same order, using a random order generator to avoid any possible fatigue bias. The videos were not coded by any other researcher so no inter-rater reliability was done.

b. EDA data. For this study's particular case, all activity peaks were considered non-specific skin conductivity peaks (NS-SCR) rather than SCRs peaks since there were no fixed events designed a priori to generate a response. The number of non-specific active peaks throughout the interaction can also reflect different levels of arousal depending on the amount found per minute, where having between 1-10 p/min reflects periods of quietness whilst higher arousal ranges from 20-25 p/min (Malmberg et al., 2019). EDA raw files were downloaded from Empatica Online Manager and processed through Excel to find the event

timestamps tagged during the interaction in order to align the data with the time-stamp in the videos. Data was then processed in MATLAB via the Ledalab interface (Benedek & Kaernbach, 2016) using Continuous Deconvolution Analysis (CDA). Participant data was optimised before doing the CDA, relying on a more restrictive threshold of the minimum amplitude of $\mu 0.05$ in NS-SCR peaks (Boucsein, 2012; Boucsein et al., 2012), which is higher than the 0.01 used when EDA is collected in a more controlled environment. However, since children moved and interacted with an exhibit, the chosen parameter reduced the possibility of confounding possible artefacts generated by hand movement. This resulted in a file of selected NS-SCRs that passed the amplitude threshold, with their respective amplitude and timestamp. The amplitude score was standardized and z-scores were obtained per participant to make a comparison between participants more accurate.

c. Self Report Data – Once the data from the questionnaires was transcribed, a change variable was computed for pre-interaction question 8: “*Do you like science*” and end-of-study question 24: “*Do you think it changed how much you like science now?*”. This change variable was obtained by subtracting the Likert-scale value of question 24 from the obtained result of question 8, if the outcome was 0, it meant there was no change, and any positive value meant there was an increase in how much children enjoyed science after the interaction, whilst negative values meant the enjoyed decreased after the interaction. A composite variable was created to capture self-report of enjoyment of the exhibit post-interaction, to do this question 9: “*How do you feel about the exhibit?*” and question 16: “*What do you think your friends would say about the exhibit?*” were combined. To calculate the reliability of this composite variable a Spearman-Brown correlation was used requiring a minimum value of 0.8, as recommended by Eisinga and colleagues (2013) for using only two variables for this.

4.4 Data Analysis

First, the descriptive statistics were calculated for the three main tools to explore what each of the tools was showing: a) Self-report questionnaire, averaged composite and change scores for self-report questionnaire; b) EDA mean amplitude per participant, and mean count of NS-SCRs; c) for each of the behavioural codes, mean and standard deviations were computed for frequency counts in total per participant and for each of the behavioural variables specified in Table 4.1, and the proportional time spent in each behaviour calculated per participant. The analysis was then divided into two parts to answer each of the specific research questions from this study.

The research question (RQ3) **“What is the relationship between the different outputs of the tools that measure different components of engagement?”** was answered through two different sub-questions with their respective quantitative analysis. The first sub-question looks at the results of engagement as an outcome because it uses measures aggregated by the end of the interaction, whereas the sub-question 2, focuses on the process of engagement and considers the markers of engagement as well as the time-point of each of these throughout the interaction.

a. Is there a relationship between the different tools that capture engagement?

It was first hypothesised that the outputs from the tools that captured similar components of engagement would show some relationship in their outputs. To explore this, correlations between the different outputs were carried out to evaluate if there was any relationship between them. The variables used for this analysis were: a) 8 Cognitive-behavioural markers obtained from the video-coding using the Engagement Codebook (“Off-task Attention”, “Enthusiasm”, “Action [FMM]”, “Body Movements [GMM]”,

“Verbalizations”, “Strategy/Goal”¹⁰, “Intrusiveness” and “Visuo-tactile Mismatch”)¹¹, b) Physiological markers obtained from the EDA sensor as the significant non-specific skin conductivity responses (NS-SCRs), c) Answers on self-reported engagement and enjoyment of exhibit from the self-report questionnaire. A series of Spearman correlation pairs were carried out after all the variables failed to pass a Shapiro-Wilk normality distribution test. The series of correlation pairs were: a) the average mean response from the Likert questions that evaluated children’s self-reported enjoyment of the exhibit: “What did you think about the exhibit?” was correlated with, b) the total count of EDA NS-SCRs per participant, c) the median EDA Amplitude (z-score), and d) the total percentage of time per participant spent in each of the 8 ELAN categories of interest. Additionally, an exploratory series of Spearman correlation pairs within the cognitive-behavioural markers was carried out to examine if there were any underlying relationships throughout the interaction with the exhibit between the different markers recorded from the combined fixed camera and head-mounted camera. That is, to explore if any of these behaviours would happen simultaneously (beyond those that were parent hierarchies – FMM → Strategy), the proportional time spent for each behaviour was used. The only correlation pairs that were excluded from the analysis were: Strategy-Use - FMM and Goal-FMM. The correlation pairs were explored with and without adjustments for multiple comparisons. For family-wise error, a Bonferroni-Holm correction was used to account for false positive results, whereas a false discovery rate adjustment was also used (i.e. Benjamini-Hochberg) to account for false negatives and because it has been suggested as a less conservative measure that is useful for screening analysis and exploratory hypothesis (Benjamini, 2010; Benjamini & Hochberg, 1995).

¹⁰ Since the behavioural codes of Strategy and Goal were always coded simultaneously (see section 4.3.5.1 or Appendix H), the proportional time was the same for both.

¹¹ Negative Affect was removed due to a low number of observations (see section 4.5.1.2)

b. Can children's emotional response be predicted by the onset of cognitive-behavioural responses?

To explore throughout the process of engagement whether specific cognitive-behavioural markers of engagement were generating an emotional arousal response that did not lead to dis-engagement (i.e. emotional engagement) a hierarchical mixed-effect model was used ($y \sim x + z \mid 1$). Nested or hierarchical models considered the effect from each unique observation of each type of behaviour, but also that each observation within each behaviour is nested within a unique individual participant (Gelman, 2006). For emotional arousal, it was used the number of NS-SCRS present within each observation of any of the cognitive-behavioural markers. In other words, the outcome variable was the total count of NS-SCRs that fell within the possible response range for an electrodermal response (1-4 s after an event starts, Boucsein, 2012) present in each of the behavioural observations per participant. Thus, any NS-SCR that appeared 1s after the behavioural observation had initiated and up to 4s after the behaviour had finished was considered. Since the outcome reflecting the emotional component was a count variable, the multilevel model structure was a Poisson generalised linear model (Rabe-Hesketh & Skrondal, 2008). The behavioural predictor variables to explore were Action – FMM, Positive Affect, Strategy/Goal, Verbalization, Attention Off-task, Visuo-tactile attention mismatch, General Body Movement – GMM, and Intrusiveness¹². The duration of each observation was also considered a fixed factor for this model since it was recognised that different behaviours had different lengths due to their nature¹³. A generalised linear mixed effect model (GLMM) with a random intercept was employed to predict whether the number of NS-SCRs was dependent on the type of behaviours the children were doing. For this model, the first level considered observations (e.g. starting

¹² Behavioural category of Touching was removed from main analysis since it was present throughout more than 90% of the interaction and it would confound analysis for other touch-dependent variables like FMM and Strategy. Negative Affect did not have enough observations for the model and Goal category had the exact same times as Strategy and would confound the model.

¹³ See section 4.5.2

time and duration) for each of the behavioural categories, which were nested within the second level, which were the 27 participants. The random intercept in the model considers the variance occurred by having repeated measures in the behavioural categories attributed to the different participants. To find the best model, models were incrementally built in R (R Core Team, 2021) using the package *lme4* (Bates et al., 2015). Starting from a null model with a random intercept at participant-level (Level 2) and incrementally adding the fixed factors: behavioural variables, duration and an interaction term between these two. Model estimation was done using Maximum Likelihood. Estimation parameters used for model comparison were Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) and log-likelihood. To evaluate random-effect, the intra-class correlation coefficient (ICC) was used. This statistic quantifies the proportion of variance explained by a grouping (random) factor in multilevel/hierarchical data (Nakagawa et al., 2017). The model was built using the Attention Off-Task variable as the referent level of the group because theoretically it is a behaviour where children are not completely immersed in what they are doing, and their attention has shifted (Akshoomoff, 2002). The exponentiated regression coefficients were presented as Incidence rate ratios (IRR), which also facilitate comparison between variables.

4.5 Results

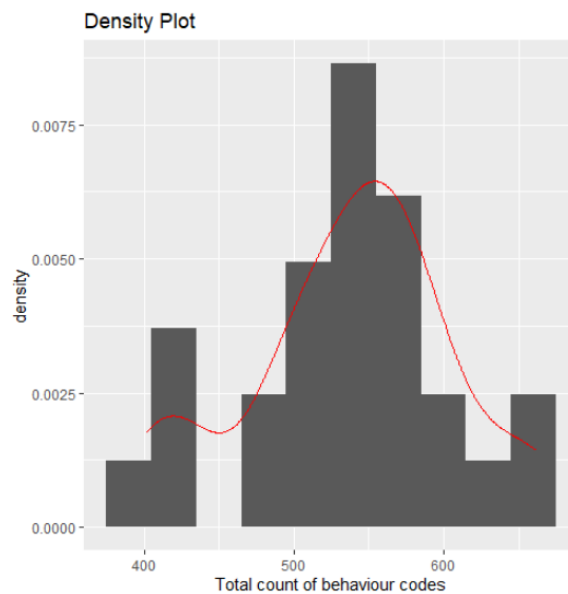
4.5.1 Descriptive Statistics

Behavioural Codes - The mean of the total count of behaviours per participant was $M = 537.41$, $SD=68.07$. The distribution of the total amount of behavioural codes per participant showed a normal distribution, passing a Shapiro test ($p = .453$), as seen in Figure 4.5 where the red density lines highlight the distribution despite not having a completely

homogeneous distribution in the histogram. The total count per behavioural tier is described in Table 4.2, whilst the behavioural code distribution between participants is shown in Figure 4.6 where each colour represents a different behaviour. It can be observed that almost all variables seem to follow a similar pattern in all the participants.

Figure 4.5

Distribution of Total Behavioural Codes and Total Count Per Behavioural Category



Note. Distribution is shown with a histogram of total count of behavioural codes. Density distribution is shown with a red line.

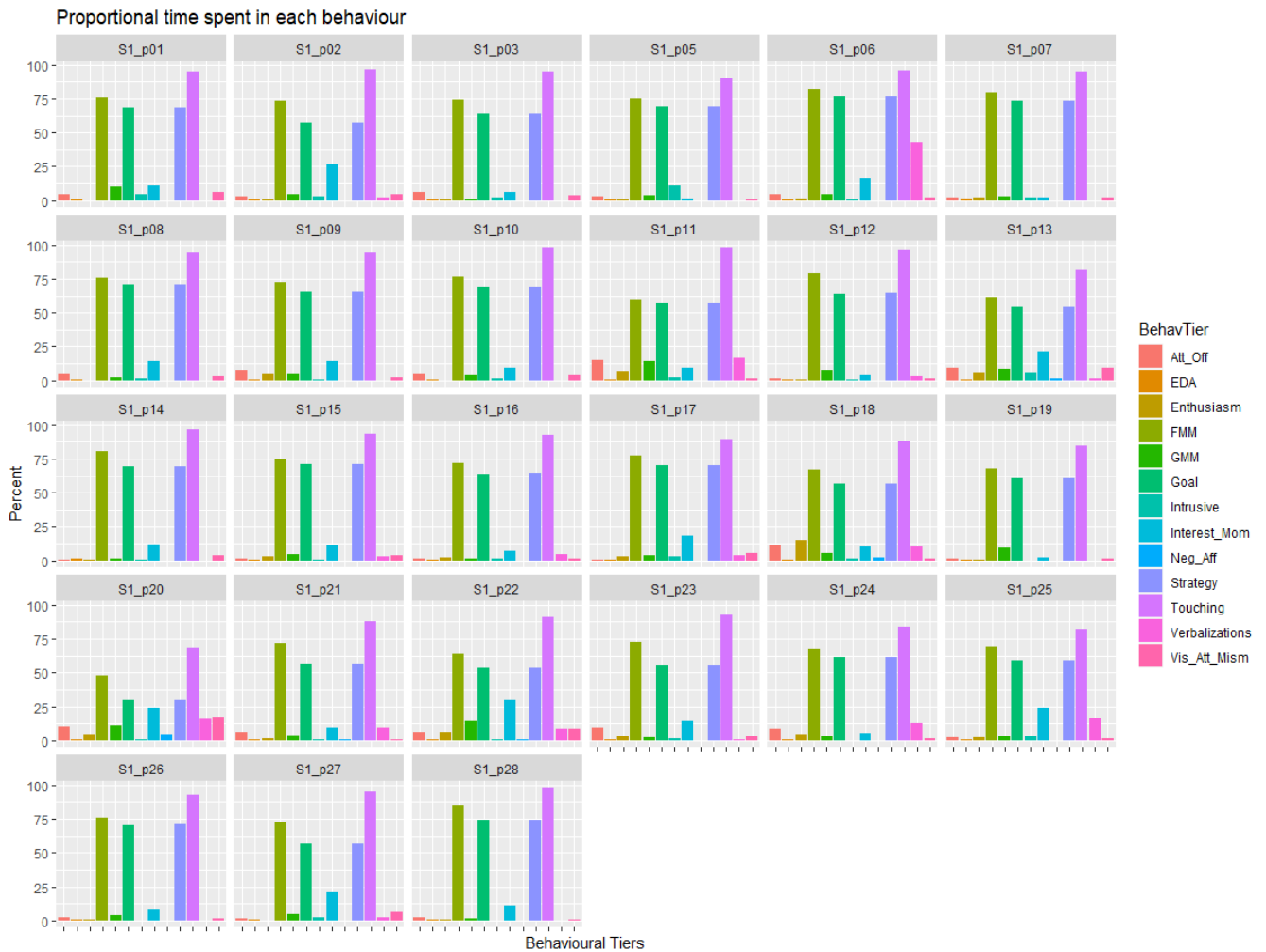
Touching was the behavioural category where the highest percentage of time was used (>90%), although it was expected considering it is a hands-on exhibit. Since this category was also present along with all the other variables most of the interaction time and did not provide any nuanced results regarding touch, this variable was not considered for future analysis because it could lead to confounding results. The second and third most recurrent behaviours were Action- Fine Motor Movement [FMM], and Strategy Use -Goal which were always coded together (as explained in Table 4.1).

Table 4.2*Descriptive Information of Behavioural Variables*

<i>Behavioural Tier</i>	<i>Total count</i>	<i>Codes per participant (M)</i>	<i>Time spent (%)</i>
EDA	6156	228	-
Action [FMM]	2846	105.407	72.4%
Goal	1902	70.444	63%
Strategy	1894	70.148	63%
Interesting Moments	398	14.741	12.7%
Touching	309	11.444	91.3%
Attention Off- task	236	8.741	4.8%
Verbalizations	226	8.37	9.50%
Visuo-Tactile Mismatch	194	7.185	3.58 %
Body Movement [GMM]	170	6.296	5.21%
Enthusiasm / Positive Affect	97	3.593	2.96%
Intrusiveness	40	1.481	2.19%
Negative Affect	16	0.593	0.32%

Figure 4.6

Distribution of Behavioural Counts Per Participant



Note. Distribution showing distribution of codes per behaviour in each participant. Behaviours have the same colours across participants shown in the legends at the right.

Electrodermal Activity - Participants had a standardised mean amplitude of $\mu = .0076$ (SD=.032) and a standardised median amplitude of $\mu = -.306$. There was a total of 6156 NS-SCRs that passed the amplitude threshold of $\mu = 0.05$ for all 27 participants, with an average of 228 per participant which, when divided by the average interaction time of 5.5 minutes, gives an average of around 41 NS-SCRs per minute. Table 4.3 shows the distribution of obtained NS-SCRs and their amplitude per participant. Appendix L also

contains an image of all the participants' skin conductivity responses and the results of signal decomposition into phasic and tonic.

Table 4.3

EDA Descriptive Statistics per Participant

<i>Participant</i>	<i>Mean (μs)</i>	<i>Md (μs)</i>	<i>count</i>	<i>Participant</i>	<i>Mean (μs)</i>	<i>Md (μs)</i>	<i>count</i>
S1_p01	-6.18 e-18	-0.446	207	S1_p16	3.51 e-07	-0.23	287
S1_p02	1.06 e-06	-0.385	191	S1_p17	-1.80 e-06	-0.402	169
S1_p03	-3.98 e-07	-0.206	253	S1_p18	2.33 e-06	-0.176	259
S1_p05	6.71 e-07	-0.368	151	S1_p19	-7.41 e-07	-0.169	272
S1_p06	1.18 e-06	-0.37	257	S1_p20	-3.16 e-07	-0.196	318
S1_p07	2.63 e-06	-0.399	192	S1_p21	-4.24 e-07	-0.353	238
S1_p08	2.47 e-06	-0.433	204	S1_p22	-5.78 e-07	-0.508	175
S1_p09	-4.81 e-07	-0.254	210	S1_p23	6.55 e-06	-0.299	231
S1_p10	4.067 e-18	-0.423	172	S1_p24	1.67 e-01	0.003	250
S1_p11	-2.13 e-06	-0.308	237	S1_p25	-1.03 e-18	-0.299	241
S1_p12	1.37 e-18	-0.321	212	S1_p26	2.22 e-02	-0.317	219
S1_p13	3.15 e-06	-0.248	256	S1_p27	-2.71 e-06	-0.365	223
S1_p14	-2.74 e-06	-0.302	221	S1_p28	-3.64 e-18	-0.138	270
S1_p15	-1.67 e-06	-0.359	241				

Self-Report Questionnaire - A change variable was obtained for the item “Do you like science?” considering scores before (M = 4.59, SD = 0.56) and after interaction (M = 4.71, SD = 0.59), however, there was a ceiling effect in the results with no changes observed in the results due to a similarity before and after the interaction (M = 0.19, SD = 0.57). The composite variable considering whether children enjoyed the exhibit (and what their friends had also thought), was tested for inner reliability using a Spearman-Brown approach, which did not show a level above 0.8 needed for reliability ($\Lambda=0.64$), thus it was decided to use responses to the individual question “How do you feel about the exhibit?” for the triangulation analysis, where the average answer was also close to ceiling (M = 4.75, SD=0.51).

4.5.2 RQ 3.a - Is There a Relationship Between the Tools That Capture Engagement?

A Pearson correlation between the chosen EDA variables (NS-SCRs total frequency and median Amplitude) showed these were related to the same construct ($R = .70$, $p < 0.001$).

A series of Spearman correlations were used to look for relationships between the EDA Amplitude, the count of NS-SCRs, the self-report item and the proportional percent of time used for each of the 9 main coded behaviours of interest: Action-FMM, Verbalization, Attention Off-Task, Strategy, Goal, Body Movements [GMM], Visuo-Tactile Attentional Mismatch, Enthusiasm and Intrusiveness. Negative Affect had very few observations (16 instances out of 14,484), where not all the participants had at least one event, and for which the average of time spent in this behaviour was close to zero (0.32%). The low number of participants restraints the use of this variable for any correlation analysis, and the low number of events also impact any possible multilevel modelling planned since the total number does not reach a minimum of 20 observations (Bolker et al., 2009). Thus this variable was not considered for this or any other analysis from this point forward. An in-depth analysis of the correlation pairs representing different tools was prioritised and it can be seen in Table 4.4, with its un-adjusted value, and the adjustments for family wise error and false discovery rate.

This analysis showed at first a significant negative correlation between both measures of EDA and the Intrusiveness category. That is, the total count of NS-SCRs and the proportional percentage time used by Intrusiveness events was $R = -.43$, $p = .026$, indicating that a higher percentage of time used by Intrusive episodes was associated with a lower count of NS-SCRs. A significant negative correlation was also found between the EDA amplitude and the proportional time used in Intrusive events $R = -.40$, $p = .037$, which shows that children had a smaller EDA amplitude as total time used with intrusion increased. However, when accounting for multiple comparisons with adjusting for family-wise-error

(f.w.e.) and for false-discovery-rate (f.d.r.), the adjusted *p* was not significant for either of these categories.

Table 4.4

Correlation Coefficients Behind Tool Triangulation

<i>Behavioural variable</i>	<i>NS-SCR</i>				<i>EDA</i>			
	<i>mean frequency</i>				<i>Median Amplitude (μ)</i>			
	<i>R</i>	<i>p</i>	<i>f.w.e.</i>	<i>f.d.r.</i>	<i>R</i>	<i>p</i>	<i>f.w.e.</i>	<i>f.d.r.</i>
		<i>p adjust</i>	<i>p adjust</i>			<i>p adjust</i>	<i>p adjust</i>	
Attention Off-Task	.19	.35	1	.47	.20	.3	1	.45
Enthusiasm	.26	.18	1	.36	.34	.081	1	.25
Action [FMM]	-.37	.055	.94	.22	-.39	.047	.84	.22
Strategy/ Goal	-.22/ -.23	.27 / .25	1	.45	-.27 / -.28	.17 / .16	1	.36
Movement [GMM]	-.02	.9	1	.90	-.16	.42	1	.52
Visuo-Tactile Mismatch	-.20	.32	1	.45	-.34	.086	1	.25
Verbalization	.37	.056	.94	.22	.13	.53	1	.59
Intrusiveness	-.43	.026	.52	.22	-.4	.037	.71	.22
Self -Report Post-exhibit	-.10	.63	1	.67	.14	.5	1	.59

Note: f.w.e was done with a Bonferroni-holm correction, while f.d.r was done using Benjamini- Hochberg correction.

To understand how some of the different cognitive-behavioural markers appeared throughout the interaction, a Spearman correlation analysis was also done with the proportional time spent in each of the behaviours (Table 4.5). Some variable pairs were excluded since the correlation was expected due to category hierarchy. The excluded pairs were actions [FMM] and Strategy/Goal, as well as Attention Off-task and Visual Attention Mismatch¹⁴. These analyses were also adjusted for multiple comparisons using only false discovery rate, due to the exploratory nature of it.

¹⁴ See section 4.3.5 “Data Processing” for more information

Table 4.5

Correlation Coefficients of Coded Behaviours after FDR adjustment

p (adjust) R^2	Att	Str	Go	Ent	FMM	Ver	Int	Vis	GMM
Attention Off-Task (%time)	1.00	.01	.01	.01	.01	.19	.84	.84	.19
Strategy (%time)	-.56**	1.00	–	.08	–	.08	.83	.37	.11
Goal(%time)	-.56**	–	1.00	.08	–	.08	.83	.37	.12
Enthusiasm(%time)	.56**	-.41	-.42	–	.01	.01	.99	.84	.25
Action [FMM] (%time)	-.58**	–	–	-.57**	1.00	.04	.88	.84	.07
Verbalizations (%time)	.32	-.41	-.41	.59**	-.48*	1.00	.84	.83	~(.02)
Intrusive (%time)	.06	-.09	-.09	.00	-.04	-.07	1.00	.37	.84
Visuo-Tactile Mismatch (%time)	.05	-.23	-.23	-.06	-.06	-.10	.23	1.00	.25
GMM (%time)	.32	-.38	-.37	.28	-.44	.29	.06	.29	1.00

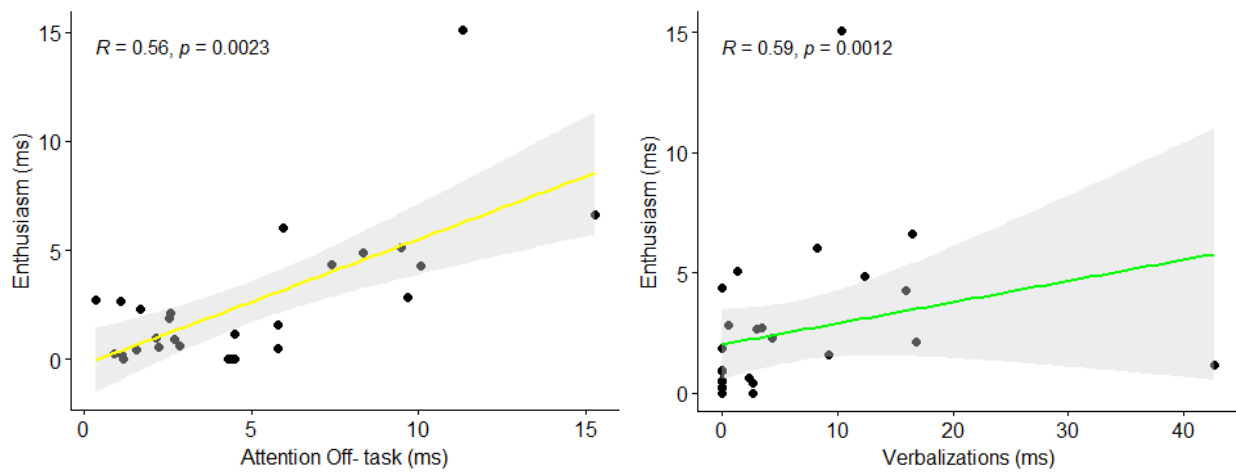
Note: The table shows in blue negative correlations, in red positive correlations and in black the remaining correlation and p coefficients.

Results showed significant positive correlations between time spent in Attention Off-Task with time spent in Enthusiastic behaviours ($R = .56$, $p = .01$) and Verbalizations with Enthusiasm ($R = .59$, $p = .01$) which can be seen in Figure 4.7.

Significant negative correlations were found between time spent with Attention Off-task and time spent doing actions [FMM] ($R = -.58$, $p = .01$), and with Strategy-Goal driven episodes ($R = -.56$, $p = .01$), showing that the more time children spent with their attention off-task, the less time they did actions. Significant negative correlations were also found between time used in Enthusiastic behaviours and time doing actions [FMM] ($R = -.57$, $p = .01$), that is, the more time children performed any actions, the less time they showed Enthusiasm. Similarly, significant negative correlations were observed between Verbalising time, time doing actions [FMM] ($R = -.48$, $p = .04$). This indicates that the more time children spent in different actions, the less time they used talking. These are all shown in Figure 4.8.

Figure 4.7

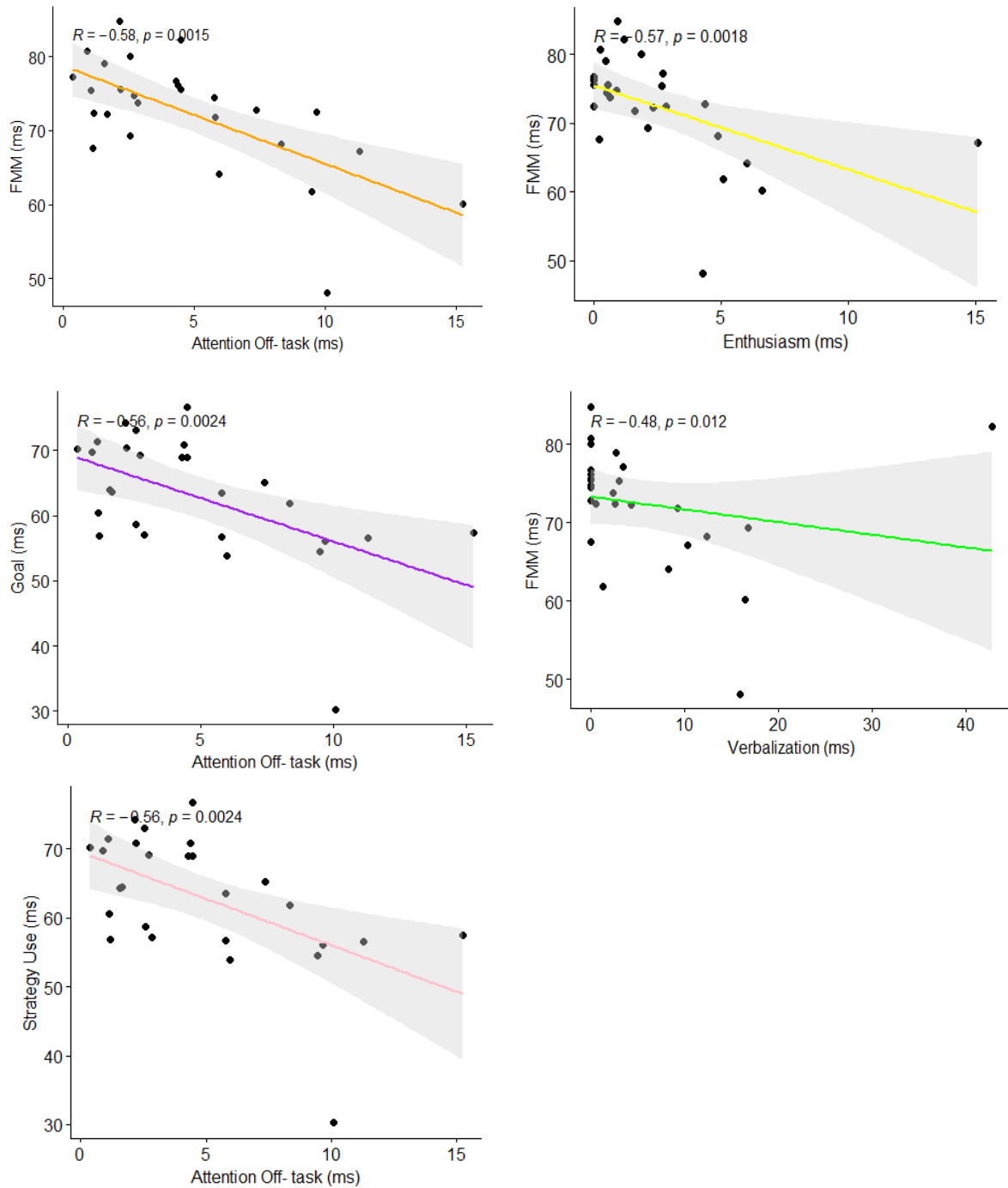
Positive Correlation plots between cognitive-behavioural markers



Note. *Left:* Positive correlation between proportional percentage time of Enthusiasm and Attention Off-task behaviours (yellow). *Right:* Positive correlation between proportional percentage time of Enthusiasm and Verbalizations (green).

Figure 4.8

Negative Correlation plots between cognitive-behavioural markers



Note. *Left:* Negative correlation between proportional percentage time of Attention-Off task and actions [FMM] (top, orange), Strategy Use (middle, pink) and Goal (bottom, purple) Enthusiasm and Attention Off-task behaviours (yellow).

Right: Negative correlation between proportional percentage time spent doing actions [FMM] and Enthusiasm (top, yellow), and with Verbalizations (middle, green).

4.5.3 RQ 3.b - Can Children's Emotional Responses be Predicted by the Onset of Cognitive-Behavioural Responses?

To answer this question, a more detailed analysis was used for simultaneous trace data. The number of total NS-SCRs associated with each behavioural category is described in Table 4.6 and Figure 4.9, which shows an expected Poisson distribution for count data: skewed to the left and with a long tail to the right. Appendix L also shows how the NS-SCRs are distributed in each participant per variable.

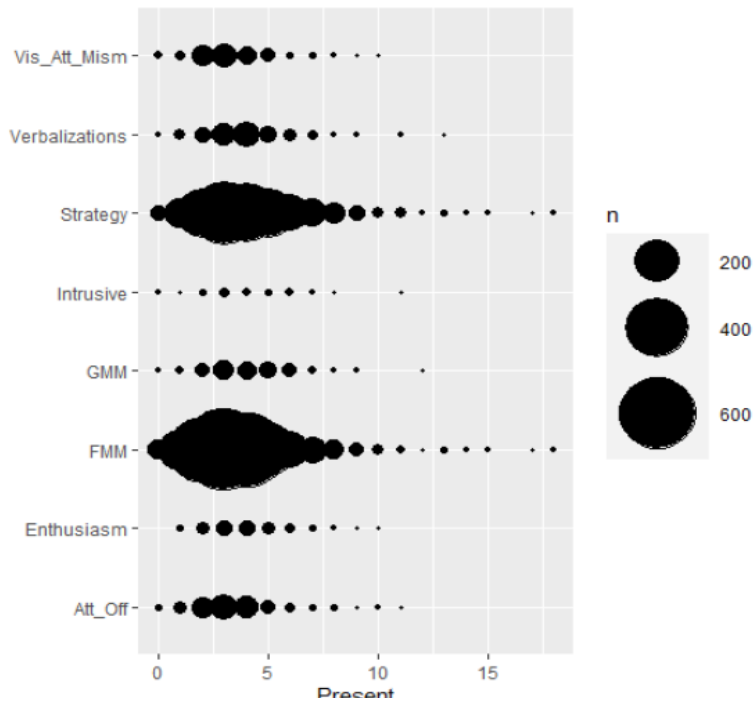
Table 4.6

NS-SCRs Descriptive Statistics Per Behaviour

<i>Behaviour</i>	<i>Total count NS-SCR</i>	<i>NS-SCR (freq. mean)</i>	<i>NS-SCR (min)</i>	<i>NS-SCR (max)</i>	<i>Duration (s)</i>
FMM	10209	3.59	0	18	2.23
Strategy	7640	4.04	0	18	2.92
Verbal	889	3.93	0	13	2.18
Attention Off-task	805	3.41	0	11	1.8
GMM	684	4.02	0	12	2.7
Visual Attention Mismatch	638	3.29	0	10	1.64
Enthusiasm	386	3.98	1	10	2.31
Intrusive	160	4.10	0	11	4.23

Figure 4.9

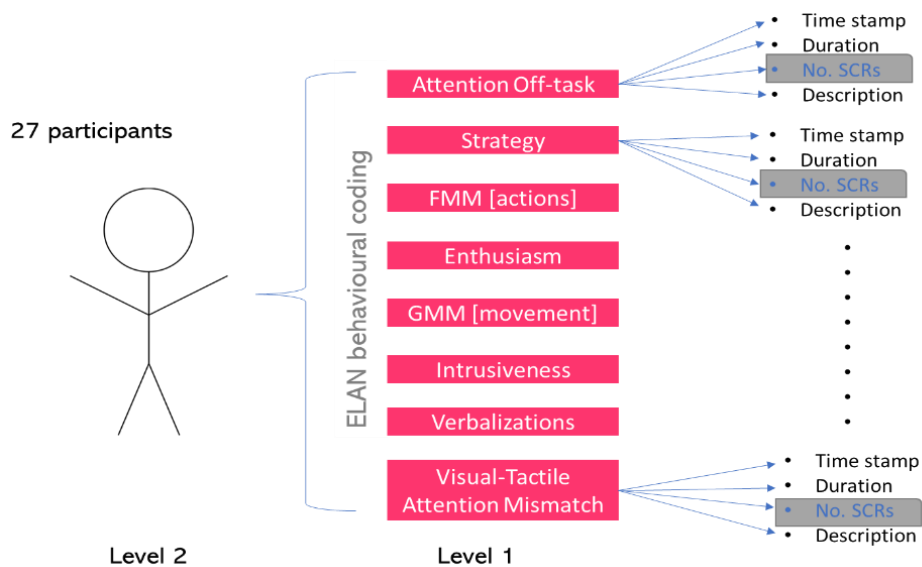
Distribution of NS-SCRs Count per Behavioural Variable



Note. Distribution of NS-SCRs shows a skew to the left and a long right tail, indicating a Poisson distribution. The amount is represented by the size of the circle. Legends to the right shows scale of size.

Figure 4.10

Nested Levels Behind Multilevel Structure



A Poisson generalised linear mixed effect model [GLMM] was used to evaluate the observations nested within each of the 8 chosen behavioural codes, which were also nested within the 27 participants. Figure 4.10 shows the multilevel structure of the data. The model aimed to determine whether the behaviour children were doing predicted the number of EDA responses.

Comparison between the four models showed the best-fitting model was the one that included all fixed factors and random intercept. Table 4.7 shows the estimation parameters and model criteria for comparison (AIC, BIC, log-likelihood). It can be noted that as each term is added, the model fit improves significantly. The best model to answer the research question was *No. of NS-SCR Present ~ Behavioural Variable * Duration + (1 | PartCode) (X(7) = 21.495, p = .003)*.

The model had a marginal R^2 which explained 16.8% of the variance was due to the fixed-factors of behaviour type, duration and their interaction ($R^2_{GLMM(m)} = 0.168$), whereas the conditional R^2 showed a variance of 26.5% due to the combination of the fixed factors with the random intercept per participant ($R^2_{GLMM(c)} = 0.265$) (Nakagawa & Schielzeth, 2013).

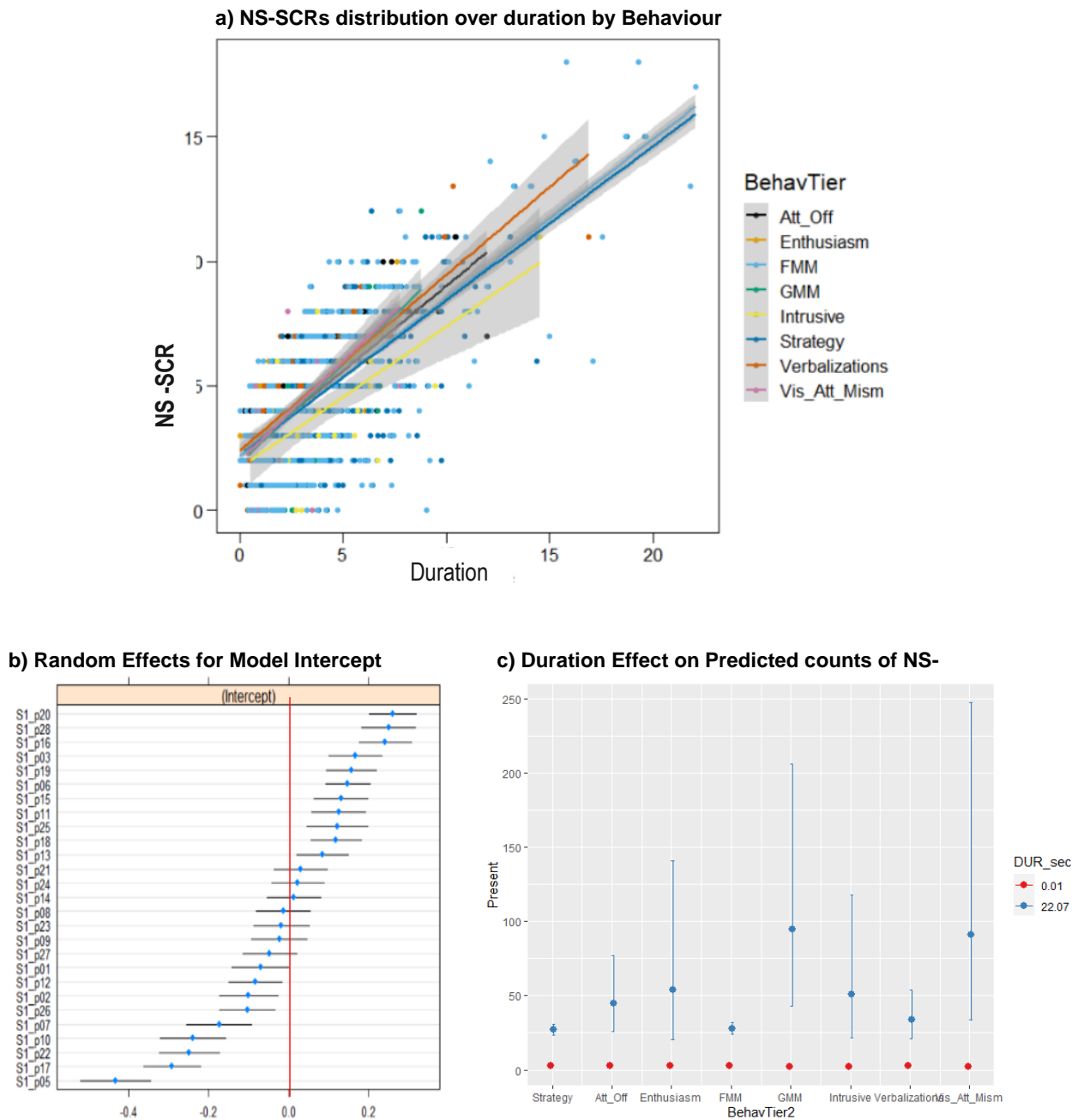
Table 4.7

Model Comparison

	<i>npar</i>	<i>AIC</i>	<i>BIC</i>	<i>logLike</i>	<i>Deviance</i>	<i>Chi²</i>	<i>Df</i>	<i>p</i>
1. Null model	2	22317	22330	-11156	22313			
2. Main predictor: <i>Behavioural variables</i>	9	22246	22306	-11114	22228	84.78	7	< .001
3. Adding predictor: <i>Duration</i>	10	20478	20545	-10229	20458	1769.92	1	< .001
4. Interaction of both predictors	17	20471	20584	-10218	20437	21.49	7	.003

Figure 4.11

Visualisation of Model's Results



Note. Visualisation of model's result showing: a) NS-SCR distribution depending on events duration and coloured depending on Behavioural Variable; b) distribution of participants intercept variance (random intercept) and c) the interaction term between the minimum (red) and maximum (blue) duration depending on the behaviour.

The model shows that the number of NS-SCRS per behavioural variable varied significantly ($\chi^2 = 84.78$), where considering Duration of behavioural variables contributed significantly to an increase in NS-SCRS ($\chi^2 = 1769.92$) (Figure 4.11a). Additionally, an interaction term between the duration and each behavioural variable was significant ($\chi^2 = 21.49$). Figure 4.11c shows the duration range for each variable and how many NS-SCRS are elicited within this range, thus confirming that duration has a different impact on the number of NS-SCRS depending on the behavioural variable. Differences in participants' random intercept showed a variance of 3% ($\sigma^2 = 0.031$) and the proportional variance explained by grouping the participants was 12% ($ICC_{GLMM} = 0.12$) with Figure 4.11b showing the distribution of the random effects in each participant.

The model built using the Attention Off-Task variable as the referent level (Table 4.8) showed that the ratio for the number of NS-SCRS when Attention is off-task increased significantly the incidence rate by 14% (CI =11-17%) whenever the duration increased by 1 s ($Z = 9.32$, $p < 0.001$). That is, if 100 NS-SCRS were observed as events of Attention Off-Task with an average duration of 1.8s, and if the duration of these events increased by 1s, the expected number of NS-SCRS would be 114. Attention Off-Task only showed a significant difference in IRR with Strategy. The ratio of NS-SCRS when Strategy had a similar starting point as Attention Off task showed a significant advantage of 14% (CI: 3-26%) in the incidence ratio ($Z = 2.45$, $p = .014$), which indicates that the starting number of NS-SCRS was higher than for Attention Off-task. However, when the duration of the Strategy behaviour increased by 1s, the difference between them reduced significantly by 3% (CI:0-6%) each second ($Z = -2.01$, $p = .443$). These results indicated that the rate of NS-SCRS for the rest of the behavioural variables did not change significantly compared to Attention Off-task, and the rate associated with an increase in their duration was also not significantly different from the one reported for Attention Off-task.

Strategy showed to have a longer duration, more observations and a pronounced slope on the plot (Figure 4.11a), it was the only variable to have shown any effect compared

to Attention Off-task. Since Strategy is a high-order cognitive process used for exploratory behaviour (Börnert-Ringleb & Wilbert, 2018) and this is a skill needed for inquiry and scientific reasoning (Van Schijndel et al., 2010) it became of interest how it compared to the other behavioural variables. Strategy was set as a referent level to compare to other behaviours. The main model estimates and random effects were not affected when the referent level was modified (as shown in the table results 4.7/4.8 for both referent variables).

Table 4.8

Model Results for Attention Off-task

Attention Off Task as Referent	NS-SCRs			
	<i>Predictors</i>	<i>Incidence Rate Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)		2.52	2.24 – 2.84	<.001
BehavTier [Enthusiasm]		1.07	0.88 – 1.31	.485
BehavTier [FMM]		1.09	0.99 – 1.20	.095
BehavTier [GMM]		0.97	0.81 – 1.16	.736
BehavTier [Intrusive]		0.90	0.66 – 1.22	.503
BehavTier [Strategy]		1.14	1.03 – 1.26	.014
BehavTier [Verbalizations]		1.07	0.93 – 1.23	.325
BehavTier [Vis_Att_Mism]		0.95	0.81 – 1.12	.559
DUR_sec		1.14	1.11 – 1.17	<.001
BehavTier2 [Enthusiasm] *Duration		1.01	0.95 – 1.06	.859
BehavTier2 [FMM] * Duration		0.97	0.95 – 1.00	.076
BehavTier2 [GMM] *Duration		1.04	0.99 – 1.09	.160
BehavTier2 [Intrusive] *Duration		1.01	0.96 – 1.07	.711
BehavTier [Strategy] * Duration		0.97	0.94 – 1.00	.044
BehavTier2 [Verbalizations] * Duration		0.98	0.95 – 1.02	.388
BehavTier2 [Vis_Att_Mism] * Duration		1.04	0.98 – 1.10	.232

Random Effects (Attention Off-task referent)	
σ^2	0.24
τ_{00} PartCode	0.031
ICC	0.12
N PartCode	27
Observations	5701
Marginal R ² / Conditional R ²	.168 / .265

Table 4.9

Model Results for Strategy

Strategy as Referent	NS-SCRs			
	<i>Predictors</i>	<i>Incidence Rate Ratios</i>	<i>CI</i>	<i>p</i>
(Intercept)		2.87	2.66 – 3.09	<.001
BehavTier2 [Att_Off]		0.88	0.79 – 0.97	.014
BehavTier2 [Enthusiasm]		0.94	0.79 – 1.13	.526
BehavTier2 [FMM]		0.96	0.92 – 1.00	.048
BehavTier2 [GMM]		0.85	0.73 – 1.00	.044
BehavTier2 [Intrusive]		0.79	0.59 – 1.06	.121
BehavTier2 [Verbalizations]		0.94	0.85 – 1.04	.249
BehavTier2 [Vis_Att_Mism]		0.84	0.74 – 0.96	.009
Duration		1.11	1.10 – 1.11	<.001
BehavTier2 [Att_Off] * Duration		1.03	1.00 – 1.06	.044
BehavTier2 [Enthusiasm] *Duration		1.03	0.98 – 1.09	.185
BehavTier2 [FMM] * Duration		1.00	0.99 – 1.01	.445
BehavTier2 [GMM] *Duration		1.07	1.02 – 1.11	.002
BehavTier2 [Intrusive] *Duration		1.04	0.99 – 1.09	.115
BehavTier2 [Verbalizations] * Duration		1.01	0.99 – 1.04	.305
BehavTier2 [Vis_Att_Mism] * Duration		1.07	1.01 – 1.12	.013

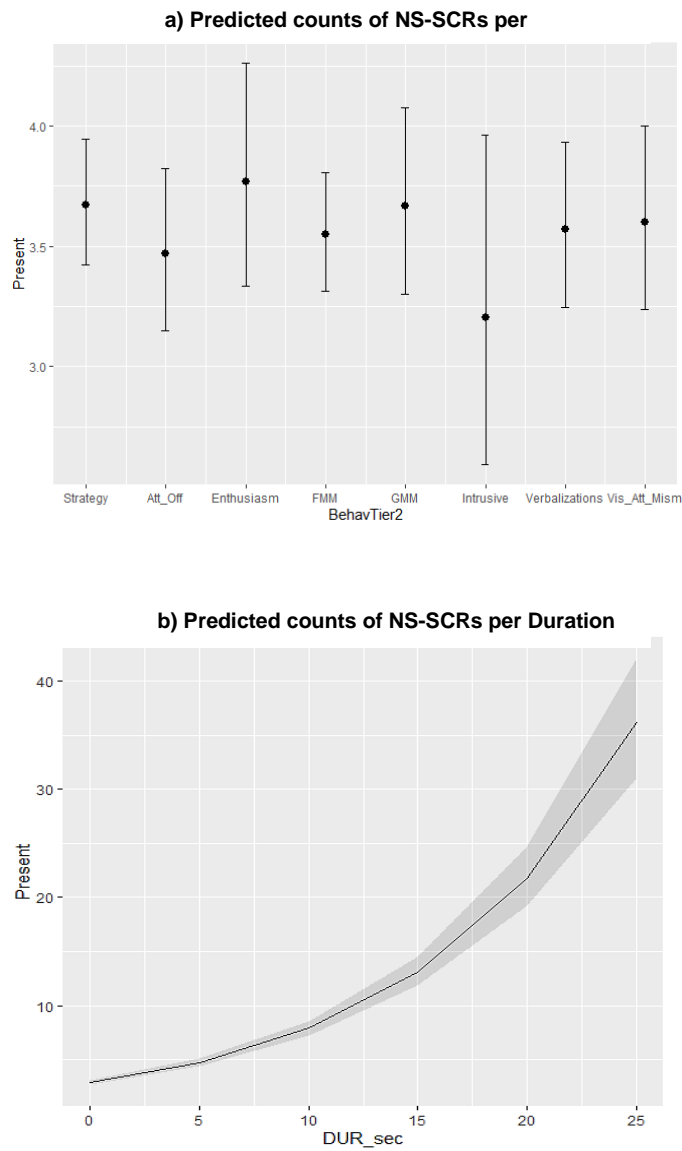
<i>Random Effects (Strategy Referent)</i>	
σ^2	0.24
τ_{00} PartCode	0.03
ICC	0.12
N PartCode	27
Observations	5701
Marginal R2 / Conditional R2	.168 / .265

The results reported in Table 4.9 show that whenever Duration is increased for Strategy, the ratio for the number of NS-SCRs significantly increases its incidence rate by 11% (CI: 10-11%) each second ($Z = 31.29$, $p < .001$). When the incidence rate for Strategy is compared to Attention Off-task in a similar duration, the ratio of NS-SCRs significantly decreases by 12% (CI: 3-21%, $Z = -2.45$, $p = .014$), but whenever duration increases 1 s for both these behaviours, the difference decreases by 3% per second. When compared to the other behaviour variables under the same duration, the ratio of NS-SCRs decreased significantly its incidence rate when it was a gross motor movement [GMM] by 15% (CI: 0-27%, $Z = -2.01$, $p = .044$), when it was an action [FMM] it decreased by 3% (CI: 0-8%, $Z = -1.97$, $p = .48$), and when the visual attention did not match the action [Vis_Attn_Mism], the incidence rate of NS-SCRs decreased its ratio by 16% (CI: 4-26%, $Z = -2.63$, $p = .008$). Whenever duration increased 1 s for each variable compared to an increase in 1 s in Strategy, the incidence rate increased significantly by 7% for GMM (CI: 2-11%, $Z = 3.02$, $p = .002$), and by 7% (CI: 1-12%, $Z = 2.45$, $p = .012$) for any mismatch between visual attention and action, but none for FMM.

These results combined help predict that on average, despite the referent level, the behaviours would have an NS-SCR count of around 3 (Figure 4.12a), and the duration of behaviours, in general, would increase the incidence rate of NS-SCRs (Figure 4.12b). It also shows that each behavioural variable has a different interaction with the NS-SCRs incidence rate that is modulated by the duration of the behaviours (Figure 4.13).

Figure 4.12

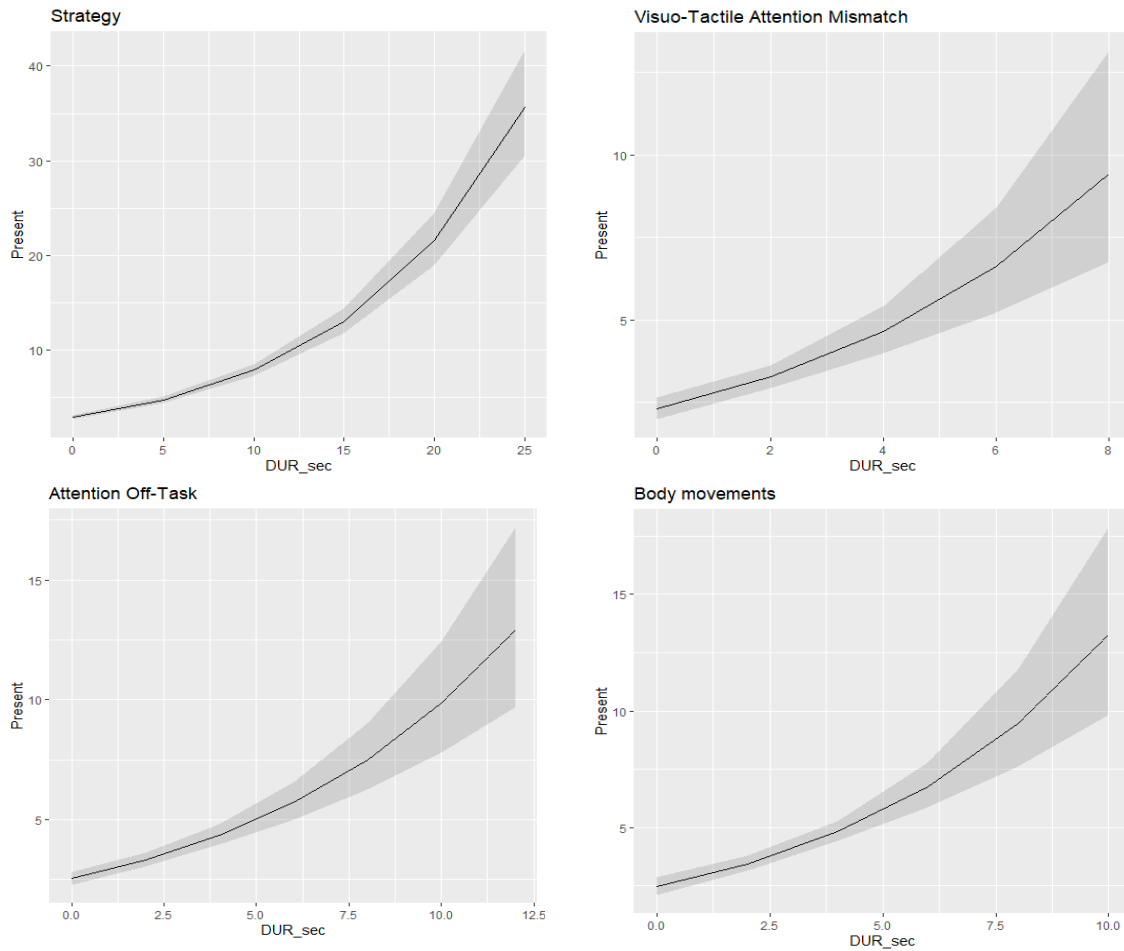
Plots of Model Predictions



Note. Model predictions estimating future distribution of (a) count of NS-SCR for each behavioural variable and (b) increase in count of NS-SCRs depending on the duration of any given variable.

Figure 4.13

Plots Showing Influence of Duration in Significant Behaviours



Note. Plot comparison of the Duration slope for each of the 4 main variables where effects were found: Strategy (top-left), Attention-Off task (bottom left), Body movements [GMM] (top-right), and visuo-tactile attention mismatch (bottom-right).

4.5.3.1 – Illustrative Description of Use of Strategies and Goals

Since Strategy/Goal showed a significant increase in NS-SCRs, the particular use of these cognitive-behavioural markers was examined to illustrate how children were using them during their exploration of an open-science exhibit. These categories comprised 63% of the total interaction time. Both of these categories require deeper cognitive engagement where children explored new ideas, adapted some of the ideas already tested, or returned to

previous ideas and repeated them¹⁵. Table 4.10 shows the total count of events within the Goal and Strategy sub-categories, the total interaction time spent in each (percentage), and the proportional time spent within each category.

Table 4.10

Description of Interaction Times and Frequency Counts Within Behavioural Categories

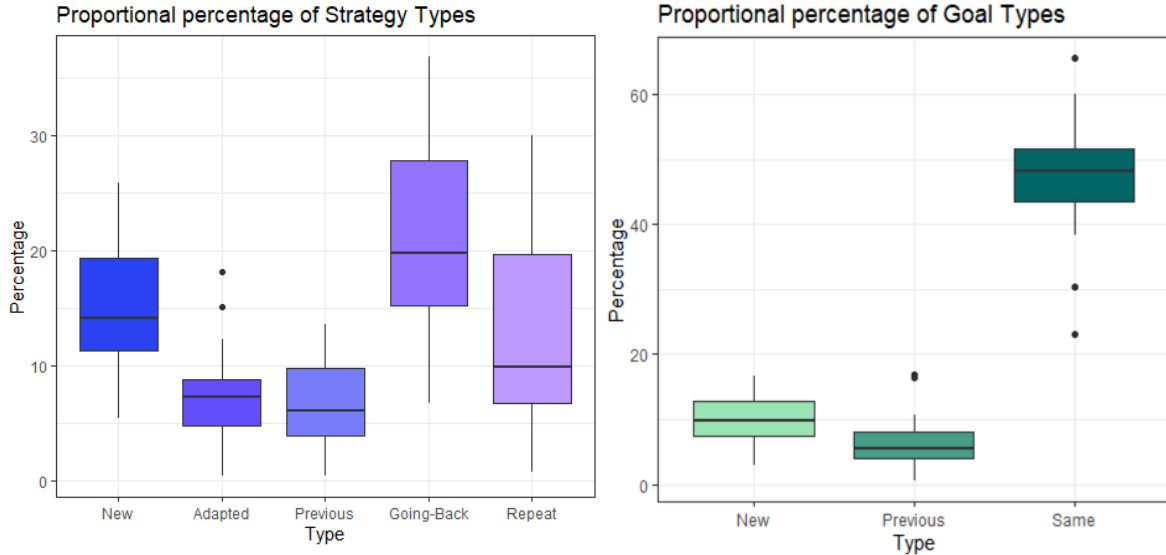
<i>Behavioural Tier</i>	<i>Type</i>	<i>count</i>	<i>% Total interaction time</i>	<i>% Proportional time within category</i>
Goal	Previous	197	6.42%	10.19%
Goal	New	309	9.84%	15.62%
Goal	Same	1396	46.7%	74.13%
Strategy	Previous	195	6.58%	10.44%
Strategy	Adapted	226	7.44%	11.81%
Strategy	Repeat	319	13.2%	21.02%
Strategy	New	505	15%	23.86%
Strategy	Going back	646	20.7%	32.87%

Children used the different types of strategies differently throughout the interaction (Figure 4.14). *Going back* to a strategy previously used (i.e. more than 30 seconds from last use, having other strategies in between) was the most used type of strategy (20.7% of total interaction time), particularly when compared to the use of *Adapted* strategies (i.e. something changed from the strategy used immediately before, 7.44% of total interaction time), or *Previously used* strategies (i.e. less than 30 seconds between the last use of the strategy - 6.58% total interaction time). *New* Strategies (i.e. a strategy not used so far in this interaction) were the second most used strategy (15% of total interaction time) with *Repeated* Strategies (i.e. the exact same strategy as before) following closely; both of these strategies were used twice as much as *Adapted* strategies or *Previously used* ones.

¹⁵ For more details refer to Table 4.1 in Methods section (4.3) of this chapter or Appendix K.

Figure 4.14

Boxplots showing Proportional Percentage Time Used in Different Strategies and Goals Sub-categories



Note. Percentages of time are considered within the total time used in the general category of Strategy (left) and Goal (right).

Regarding how children seemed to have set their own Goals throughout the interaction also varied, showing that they showed persistence in a single goal for long periods of time, with the *Same Goal* category being the most used one (46.7% of total interaction time) when compared to the others. Also, children came up with *New Goals* (9.84% total interaction time) slightly more than they returned to a *Previous goal* once this was finished or interrupted.

4.6 Discussion

This study explored the relationship between the different outputs from the tools capturing engagement, how these measures related to the components of engagement in a science centre context, and if particular events could be associated directly to a change in

engagement markers during the activity. The different outputs matching the components of engagement consisted of: a) Cognitive-behavioural markers obtained from the different behaviours tracked using video code derived from the *engagement scales* and the video recording (*fixed camera and head-mounted camera*), b) Emotional markers obtained as the significant non-specific skin conductivity responses (NS-SCRs) from an *electrodermal activity sensor* as well as from answers on engagement from a *self-report questionnaire*.

The findings from this study suggest the possibility of an underlying relationship between the cognitive, behavioural, and emotional engagement markers captured by different tools. This was found when engagement was measured as an outcome, that is, using the total percentage of time a cognitive-behaviour was used throughout the interaction, where there was a negative relationship between tools where intrusive events were related with a lower number of total non-specific skin conductivity responses (i.e. emotional responses), also showing a standardised lower amplitude; while these relationships were not significant once multiple comparisons were adjusted for, the high correlation coefficient ($R > .40$) could be indicating towards an underlying relationship that was not powered enough due to the smaller population used or smaller number of intrusive events recorded. Positive relationships were also observed between the different cognitive-behavioural markers captured using the engagement codebook, where children showed Enthusiasm more times whenever they spent time with their Attention off-task, and children Verbalized more when Enthusiasm was shown. On the other hand, more time used in actions [FMM], Strategies and Goals was related to less time spent with Attention off task, showing Enthusiasm, or Verbalizing. It was also observed that children preferred re-using strategies they had tested before the most, trying new ones next, then repeating strategies and adapting them; whereas for self-setting Goals, children maintained the same goal for more than 40% of the interaction, and they enjoyed more changing to new goals once the goals were finished than returning to incomplete goals.

When engagement was captured as a process, considering the number of EDA responses throughout the interaction time and triangulating them with coded cognitive-

behavioural markers at similar times, it was observed that the number of NS-SCRs could be significantly predicted by some of these cognitive-behavioural markers, particularly more by the use of strategy and goals, and less so by behaviours showing attention off-task, moving around the exhibit [GMM] or by visually anticipating events. It was also observed that the duration of any behaviour increased the likelihood of triggering a psychophysiological response (i.e. NS-SCRs).

This study answered its main research question (RQ3) by examining two different sub-questions, each of them capturing engagement differently (outcome vs process), seeking to understand to what extent the tools capture engagement by the triangulation of different tools measuring the different components of engagement from multimodal data.

4.6.1 Relationship Between the Different Outcomes of Engagement After the Exhibit Interaction.

This section considered engagement as the total time percentage used in the different behaviours, as well as the total frequency of electrodermal events and the averaged amplitude of it. The relationship between the outputs of these tools was explored through a correlation analysis to explore whether the outputs showed any pattern of occurrence.

Firstly, there was no significant correlation between the two different outputs used as emotional markers: the EDA sensor (measuring the emotional state) and the self-report measure (measuring the emotional experience). One reason could be attributed to the prevalence of the ceiling effect in the responses from the self-report questionnaire increasing the difficulty of observing any variations. However, it could have also been because the specific markers used for EDA (total frequency count of NS-SCRs or median amplitude of SCRs) and the items from the self-report measured different parts of an emotional experience (emotional state vs feelings), thus they were not measuring emotion a similar dimension. This would make sense since the self-report measure from the child requires them to average out their overall experience by using the Smiley-o-meter to select a specific

facial expression with a certain intensity, whilst the EDA sensor was a resulting measure from constantly tracking the physiological arousal of the child, where more nuances and variance was encountered. Furthermore, Fowles and colleagues (2000) mention in their research that it is possible to have events that could be psychologically relevant which appear less effective in eliciting a physiological response. When both EDA measures were correlated with the cognitive-behavioural marker used for positive emotion¹⁶ (i.e. *Enthusiasm*), a positive relationship was hinted at by the medium effect size ($R=0.34$) with EDA amplitude, and low-medium effect size with the number of NS-SCR peaks ($R=0.26$), whilst these were not statistically significant, they could still point towards a relationship between the observed emotional expressions and experienced emotional markers that the data collected was not powered enough to show.

Relationships between the measures obtained from the EDA sensor and the cognitive-behavioural measures (camera and head-mounted camera) showed 3 significant relationships that were no longer significant once multiple comparisons were accounted for, however, the effect sizes observed could be pointing at an underpowered relationship that would need more data: Two of these were negative correlations between Intrusive events and both the electrodermal measures that could be indicating that having intrusive episodes was not driving the increase in the NS-SCRs frequency ($R= -.43$), nor with their amplitude ($R = -.40$). The third correlation showed a medium-effect decrease in the median EDA amplitude whenever more time was used performing different actions [FMM] ($R = -.39$). This could be indicating that performing more actions could be related with smaller amplitudes or with a decrease in amplitude from the electrodermal response. This would be in line with results found by Fowles and colleagues (2000) where they found smaller amplitude response for positive stimuli, and less SCR responses for stimuli with which participants were more familiar. Translating this into the results found for this research, the sand exhibit

¹⁶ There were not enough cognitive-behavioural markers observed for Negative Expressions of Emotion as discussed in section 4.5.1.

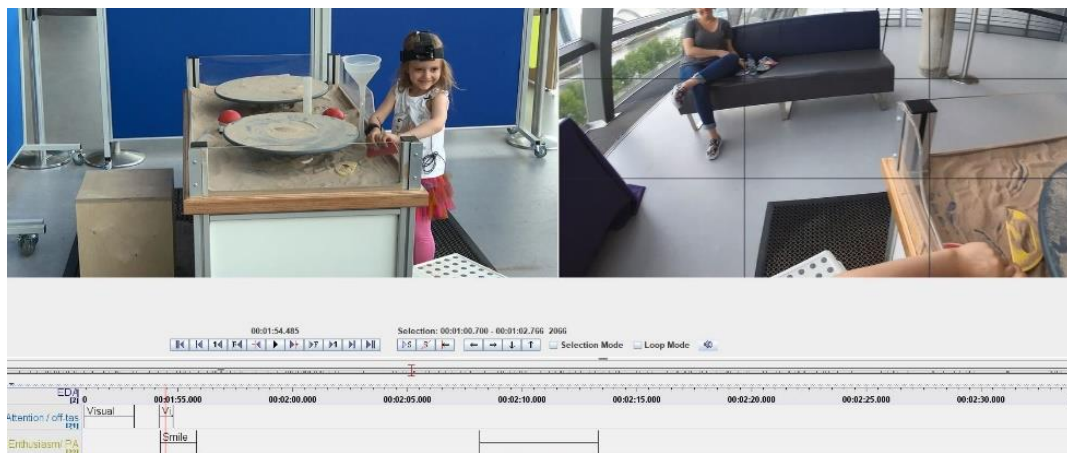
allows a limited number of possible actions that can be done throughout the interaction (i.e. affordances). Thus if more actions were performed it could lead children into either familiarising themselves faster with these actions, or might lead to reaching the affordance limit sooner than children who did not perform as many actions. If this were the case, fewer novelty items and moments would potentially mean fewer emotional reactions from the physiological system (Davidson & Smith, 1991; Fallahpour et al., 2010; van den Bosch et al., 2013). Similarly, other behaviours did not show statistical significance when related to EDA measures, but when examining their effect size in Table 5.4, these could be suggesting a series of underlying relationships that could be further explored, for example relationships with medium effect sizes ($R > .3$) were found between EDA NS-SCR frequency and the time spent Verbalizing ($R = .37$), which points to a possible relationship that NS-SCR events might be more likely to increase when children are Verbalizing more. Also, another underlying negative relationship of medium size ($R = -.34$) was observed whenever visual attention did not match between what is touched and what is observed with the EDA median amplitude, with a similar yet lower effect size ($R = -.20$) when correlated with NS-SCRs events. With more data to power this analysis, this last relationship could be suggesting that whenever children's attention is divided, the median amplitude of EDA might be lower, along with the number of response peaks. Nonetheless, having 27 participants single aggregated score for each behaviour was not enough to clarify how the outcome of engagement could show a relationship between the different tools, but it points to future relationships that could be investigated further.

On the other hand, exploring within the cognitive-behavioural markers to explore how these could be relating differently to one another throughout the interaction, showed more fruitful results. It was intriguing to find a significant positive relationship between the amount of time children spent with their attention off-task, their time spent showing enthusiasm and the time spent verbalising something. A possible explanation arose when the data was coded and explored, since it was common to notice through the head-mounted camera

(child's perspective) that children looked away from the exhibits to look at parents (or accompanying adults) as well as towards the researcher, where sometimes they would make remarks about the exhibit towards any of the adults (shown in Figure 4.15). These remarks often contained excitement towards something that had just happened at the exhibit.

Figure 4.15

Example Image of Child with Attention Off-Task Looking at Parent



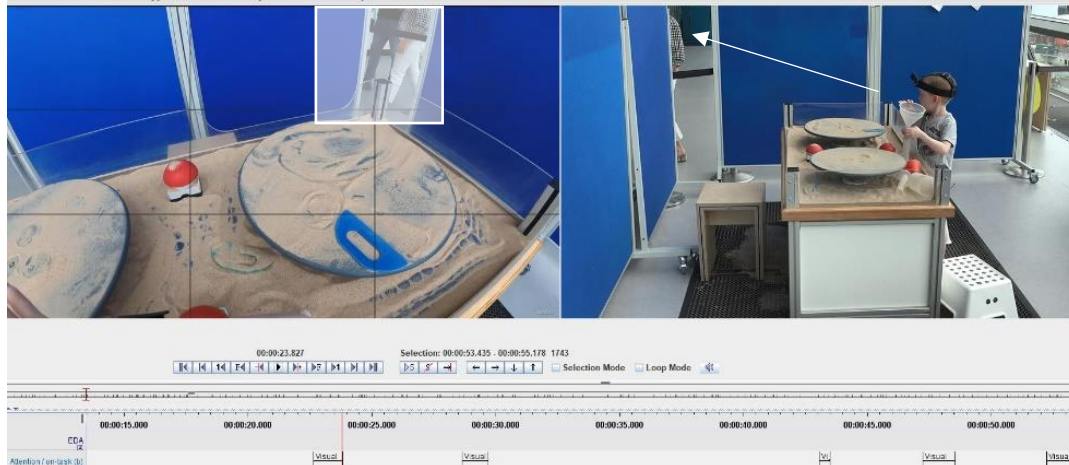
Note. Example shows child smiling and looking away from the sand exhibit while still grabbing a funnel. The panel on the right shows the child is looking at parent while smiling. In the ELAN code underneath it can be observed the synchronous code of Enthusiasm with Attention Off-task

It was interesting to observe this naturally emerging behaviour where children visually and sometimes verbally looked for reassurance to their parents first. These results align with research findings that younger children tend to gaze more at parents, and where gender might also be playing a role with girls under 5 year olds relying on and gazing more at parents compared to boys (Schofield et al., 2008; Vlietstra & Manske, 1981). Moreover, this could be an already established behaviour driven by the release of oxytocin and other peptides that help with relationship building when children and parent cross their gaze

(Heinrichs & Domes, 2008; Wittfoth-Schardt et al., 2012). These synchronicities of events could also be observed in the effect sizes of the correlations between these behaviours.

Figure 4.16

Example Image of Child with Attention Off-Task Looking Away from Exhibit



Note. Example shows child looking away from the sand table into the gap between panels. As shown in the associated ELAN code, it can be observed this behaviour does not last very long.

Two main factors could be responsible for driving children's gaze off task: an unrelated event that shortly distracted them from the task to evaluate what the external stimuli were (Figure 4.16), or alternatively, children averted their gaze to process cognitively the events they had observed in a "visually neutral" scene. Similar gaze aversion has been observed with eye-tracking where participants move their gaze away when processing cognitive tasks, or when answering difficult questions (Laeng et al., 2014; A. McCarthy et al., 2008). If the latter was the case, there is a possibility that children might have expressed a positive emotion during or after processing what they were doing in the exhibit, which could explain the relationship found between these Attention, Enthusiasm and Verbalizations. It should be noted that the verbalizations also included exclamations of emotions or emotional vocalizations, singing and humming to allow for a broader range of emotional expressions.

These results would also serve as an explanation for the underlying positive relationships found exactly between EDA measures, Attention Off Task, Enthusiasm and Verbalizations.

Not surprisingly, negative correlations were found between the behaviour triad of Attention Off task, Enthusiasm and Verbalizations and the action triad of Strategy/Goal/FMM, where the more actions and strategies were performed throughout the task, the less time was invested in looking away, expressing positive emotions or making any remarks about it. One explanation for this could be associated with children's age¹⁷, children performed more actions and strategies throughout the interaction but of shorter duration as age increased. Having a simultaneous increase in the number of actions performed whilst having a decrease in the duration of each of these actions as age progress could be showing part of children's development where they turn towards an inner speech and self-validation, becoming less reliant in seeking external support, validation and scaffolding (Alderson-Day & Fernyhough, 2015; Brinthaup et al., 2009). There could also be an increase in children's exploratory behaviour due to their age, in comparison to a familiar and controlled approach of younger children who performed a lesser number of actions and strategies yet with longer duration. Research has shown age-related differences in the use of strategies (Börnert-Ringleb & Wilbert, 2018; Siegler & Ellis, 1996). This could be associated with a decreased use of strategies, whether it is due to a difficulty in knowing which strategy is more appropriate, or struggle to spontaneously use a strategy, which has been related to poorer working memory (Dupont-Boime & Thevenot, 2018) and low metacognitive skills, which are both skills that develop with age (Bryce & Whitebread, 2012; Moriguchi et al., 2016).

The correlations found within the cognitive-behavioural markers could be used as evidence that a triangulation of tools can convey more information underlying children's engagement and interaction with a science exhibit. These results could also be pointing to

¹⁷ An exploratory age analysis was carried out to examine whether age was driving any of the results, since this was out of the scope of the project it is only reported in Appendix M

similar evidence about engagement as outcome of different components as found by Ben-Eliyahu et al.(2018) when comparing structural equation models regarding the multidimensionality of engagement. These results also build on evidence provided by the Going Ape! project (Tisdal & Perry, 2004), whilst their research aim was to compare exhibits that elicited active prolonged engagement against those that did not, one of their main predictors was time spent in the exhibit and the emotional engagement observed through what they call “driving questions” driving their behaviour and the emotional response observed (in their case it relied on observed facial expressions, self-report and voice expression). This study contributes to their results by focusing on a single exhibit, and not only labelling the particular behaviours visitors do, but also recording the amount of time spent on each of these behaviours. It also contributes by showing some connections between emotional markers of engagement and the cognitive-behavioural ones, which go beyond those that are observable. That is, markers that are captured thanks to the addition of a physiological tool that allow the exploration of possible underlying co-occurring patterns of emotion-related processes that could potentially be linked to engagement, which involved more information than a single observational variable would have provided (e.g. facial expressions).

By capturing the finished product of what happened throughout the interaction, the results show engagement both as a process, by being coded during the interaction, and as a resulting outcome, by summarising what engagement looks like after the interaction was finished. Recalling the high possibility of children being engaged given the high number of NS-SCRs per minute, would lead to believe that the codes reflected measures that captured engagement. The findings of this section show that a relationship between some of the tools was observed once the total engagement was considered, but it was not enough in some other cases to yield results that were powered enough to make substantial claims. It is also necessary to explore how these events could be interacting with one another, if at all, *during* the process of engagement, that is, when the relationships are explored throughout the

interaction.

4.6.2 Emotional Engagement Markers Predicted by Cognitive and Behavioural Responses throughout Exhibit interaction.

Following the direction from the previous research question, a different way of exploring a data relationship amongst the varying tools was to look at whether events co-occurred throughout the interaction during the process of engagement, and if so, how. More specifically, looking at whether events from the different cognitive-behavioural markers of engagement had any effect on a psychophysiological level (i.e. NS-SCRs) that could be partially reflecting children's emotional arousal. This was done by taking non-specific skin conductivity responses (NS-SCRs), where the amplitude value was above the minimum threshold of 0.5 μ s, and using their time-stamps to look for any behavioural and cognitive co-occurrences within the EDA response windows of 1-4 s (Boucsein, 2012).

The action [FMM] and Strategy categories had the highest counts of total behavioural events, which also showed to have the highest counts of total NS-SCRs that fell within the timing of the behaviours. These outcomes are congruent for a hands-on exhibit since these are expected to elicit visitors to act with them, however, they also present an interesting finding since they provide physiological evidence that this particular hands-on, but also open-ended exhibit, fostered great emotional arousal. This would align with observations from Gutwill (2008), where he found open-ended exhibits attracted visitors for longer, eliciting a more active engagement through a playful exploration with more opportunities for creation when compared to planned-discovery counterintuitive exhibits. Also, Tisdal & Perry (2004) found that open-ended exhibits enhanced what they called a "complete engagement" where their interaction had: 1. Beginning – where visitors discover a driving question they would like to answer with the exhibit; 2. Middle – where strategies and efforts are used to answer the question; and 3. End – once the question has been answered, and positive

emotions and information has been obtained as an emotional and intellectual reward. It could be said that because of the extended interaction time in this research study, the cycle of engagement could easily have been fulfilled multiple times with different driving questions, simultaneously recording physiological evidence that could be reflecting the emotional arousal observed after doing exploratory strategies.

The data from the chosen hierarchical model showed that the number of emotional arousal markers (i.e. NS- SCRs) per cognitive-behavioural event was influenced by the type of cognitive-behaviour being done, the duration of each behavioural event, and that this latter interacted with the type of cognitive-behaviour differently. More specifically, it showed that the number of psycho-physiological activity peaks elicited (i.e. NS-SCRs) varied depending on the type of cognitive-behavioural event, and that the number of NS-SCR peaks would also vary differently depending on the duration of each type of event. This echoes previous research that has triangulated cognitive markers with some emotional arousal markers during mental effort tasks using EDA sensors in young children by Sridhar et al. (2018). This could also be observed in Figure 4.11a, where the slopes and starting points were different for the different behaviours, showing also variability in the duration axis dependent on the number of the NS-SCRs of each behaviour. These results could be reflecting children's individual differences in psychophysiological reactivity to the different cognitive-behavioural markers they did, which could also be associated with differences in emotional arousal elicited due to personal preferences of the chosen actions and behaviours they did throughout their interaction with the exhibit.

From all the different behaviours, Strategy Use and Goals elicited the most psychophysiological responses per event, while also being the ones carried out for a longer duration compared to the other cognitive-behavioural markers. This is not a surprising result for an open-ended hands-on exhibit that facilitates "engagement" like *Sand Patterns* has shown (see section 3.2). These types of exhibits encourage visitors to invest in their own self-driven inquiry of the space, questioning and hypothesising to connect to previous

knowledge but also to prepare for future experimentation (Allen & Gutwill, 2004; Gutwill, 2008; Tisdal & Perry, 2004). It has also been shown that another driver behind engagement with an exhibit, is the visitor choosing to engage instead of walking away (Tisdal & Perry, 2004).

However, it was interesting to find that even though the action category [FMM] had the largest count of associated NS-SCRs, they did not appear to stand out as significantly different from other behaviours, except for Strategy, and even then, the effect size was marginal. It could well be that other behaviours did not have a higher NS-SCRs count when compared to FMM. On the other hand, it could also be that Strategy/Goals events implied a higher count of NS-SCRs because they involve higher cognitive processing where planning, hypothesising and managing self-expectations played a critical role (Gutwill, 2008; Rennie & McClafferty, 1996; E. Wood & Wolf, 2008). Thus, if the assumption is that electrodermal activity is the sympathetic response associated with arousal driven by cognitive effort (Critchley, 2002), then Strategy, being an effortful higher-order cognitive process, aligns well as the main behaviour eliciting the highest number of arousal peaks. Similarly, following the somatic marker hypothesis (Bechara et al., 2005), it could be that these NS-SCR peaks being observed could be a product of children's decision-making process of choosing some Strategies and Goals over others, and then these markers would accompany any future-decisions to re-use them, choose something different or use when the Strategies and Goals were similar. These findings could also be demonstrating the model proposing the cycle of engagement (O'Brien & Toms, 2008), where it is suggested that once a point of engagement has been established, there will be different attributes, in this case engaging behaviours, with variation in their intensity levels which are shaping the process of engagement. In this particular application of the model, the relationship between those NS-SCRs and the cognitive-behavioural markers could be representing those variation points, while the intensity could be reflected in the amplitude of each of the NS-SCRs. Moreover, this would also be supported by the influence of the amygdala's response to EDA when stimuli (i.e.

actions and strategies) obtain emotional meaning through learning, learning that in this scenario would be done through self-inquiry (Antoine Bechara et al., 2000; Critchley, 2002).

It has also been discussed in research that high-explorative level skills, like whenever children are able to manipulate and add variation to their manipulation, can be compared to “scientific reasoning in action” (Van Schijndel et al., 2010). Thus it could be argued that the use of strategy and goal setting in children was evidencing their scientific reasoning by setting and exploring their own hypothesis. Whenever children engage in these scientific skills, their skin conductivity potentially reflected an emotional arousal increase the more they use these skills.

The triangulation model also allowed to look more in-depth at how children were using the different sub-types of Strategy and Goals throughout the interaction, which provided a more comprehensive view of the underlying cognitive processing guiding these inquiry behaviours. The differences found in the style of use of these skills reflected not only children’s individual differences and ability to plan ahead of the current action (Börnert-Ringleb & Wilbert, 2018), but also their degree of cognitive, behavioural and emotional engagement towards preferred actions (and not others) allowed throughout the exhibit interaction. The most commonly used sub-category of Strategy was children Going back to previously used strategies, which shows that children preferred returning to successful and familiar strategies used in the past, whilst also reflecting that certain strategies were more useful for children to return to them. This specific finding aligns with arguments presented by Van Schijndel and colleagues (2010), where they argue that particularly repeating behaviours with variation is part of children’s higher-level exploration process. These authors also support that children between 4 and 5 years discover how an item works: they rotate, inspect and transform the item. To further exemplify this argument, evidence from a participant from this thesis progressing through a series of actions that responded to different strategies can be seen in Figure 4.17 (For more examples of Strategies and Goals using different items in the exhibit refer to Appendix N).

Figure 4.17

Example Image of Child Progressing Through Series of Strategies



Note. Example shows the progression of most common combination of actions: placing containers which received sand (top), then grabbing sand from a surface (middle) and finally depositing sand into container (lower). The codes used are highlighted to show the progression of the action and the associated use of Strategy and Goal.

In hand with this, children also performed more New strategies, whether that was new uses for the same objects or using new objects for the same actions, however, this exploration tended to be more spaced in time. This could be explained by research suggesting that children can discover new strategies by just performing a task, even though these might not be as effective (Börnert-Ringleb & Wilbert, 2018; Pressley & Hilden, 2006). This would also explain the high use of strategies that had been previously used, if children were trying new strategies without testing their effectiveness in advance, it is more likely for them to return to identified successful/useful strategies. Very similar results were observed for overarching perceived Goals, where children leaned significantly more to remain with the Same Goal for longer durations of time whilst exploring different actions and strategies. Children also set new Goals more than they returned to previous/ incomplete ones. Since the exhibit is an open-ended exhibit, which means there is no particular aim or goal to fulfil, it provides room for children to explore, test hypotheses about different objects and consequences and, if successful, repeat them to confirm the results (Gutwill, 2008).

Results from the model showing an increase in NS-SCRs whenever Strategies/Goals were used, could be linked to an increase in children's emotional arousal related to the novelty and familiarity of each strategy. This could be supported by research where arousal was measured by an event-related P300 wave, after hearing familiar and unfamiliar percussive events (Ueda et al., 2021; Yanagisawa et al., 2019). In both these studies, results showed that the uncertainty that comes with novelty encountered, enhances and raises arousal whilst repetitive tasks tend to generate habituation and reduce that uncertainty, thus decreasing the emotional arousal caused by any novel stimuli. These authors have also argued that initial uncertainty depends on the individual's prior experience, when this prior knowledge is increased, by continuous exposition, the arousal accompanying uncertainty generated by new stimuli, decreases. Similar results concerning arousal with novelty and habituation have been observed with EDA, in a study evaluating the influence of caffeine when encountering novel and repeated tasks, where it was found that caffeine

contributed to maintain arousal in non-novel tasks, but most importantly, that electrodermal activity was higher when related to novelty (Davidson & Smith, 1991). Therefore, there seems to be evidence that could indicate that the use of novel or familiar strategies could modulate some of the emotional arousal associated with each of these. In future research it would be interesting to understand why these events, and not the others, showed a higher number of psycho-physiological markers of emotional engagement, or which unique actions also had higher emotional arousal and what other lessons could be abstracted from discerning between the more emotionally arousing actions and the ones that did not elicit anything. The preference and use of Strategies and Goals also reflect part of the scientific thinking children develop from an early age (Gopnik, 2012), mirrored in the observed preferred choice of strategy and goal shown by children: Returning to previously used, New strategies and Repeated strategies. These results also align with the suggestions Van Schijndel and colleagues (2010) gave to parents as a tutorial to increase exploratory behaviour while visiting a science exhibit with younger children: 1) getting to know the material, 2) investigating, 3) drawing conclusions. In their study, it was shown that children with parents who were given this tutorial showed higher exploratory levels than those whose parents did not have the video tutorial. What is interesting in the findings of this study, is that children seemed to have followed the recommended steps on their own as they engaged with the exhibit. Although this may not always happen when they are exploring a science centre or a science exhibit on their own free time, whether that is because of all the other competing stimuli and exhibits or due to time restraints on their visit from their parents (Tisdal, 2004; Tisdal & Perry, 2004) .

Another significant result from the model was that children having their attention off-task also elicited a high count of NS-SCRs, whereas the probability of having more psychophysiological responses increased when the duration of these off-task events was longer in comparison to the increase when the duration of strategies or actions was longer. One explanation for this steeper slope with Attention Off-task could be related to the

correlation results found previously, where an association was found between attention off-task, enthusiasm, and verbalizations. Thus, in this triad, it could be possible that children looking away and placing their attention off-task seemed to be more rewarding when this duration is longer. Interestingly, when children were looking and talking to their parents, this would entail a longer duration of an off-task period and it would also elicit a higher emotional response in comparison to other events where they placed their attention elsewhere. It could be likely that children were looking over at parents for support as shown in Figure 4.15, since it has been reported that when children interact in museums with their parents, these tend to guide exploration and point to relevant features to children while providing possible explanations to observed phenomena (Crowley et al., 2001).

However, an alternative explanation is that greater frequency of NS-SCRs could be attributed to short bursts of mind-wandering, whether task-related or unrelated, particularly since these had shorter durations and the effects were still observed up to four seconds afterwards. Research combining both eye-tracking and EDA measures found that mind-wandering episodes detected by following participant's visual attention, triggered delayed emotional arousal (Brishtel et al., 2020). Although for this study mind-wandering was prompted by music being played while reading a text, finding that having music when reading prompted more episodes of mind-wandering than not having music. Their results still evidence that the emotional response to mind-wandering can appear after the episode happened. It could also be that these results with a higher EDA could be reflecting children's self-detection of an error after which they increase their effort in the task to compensate for it. Supporting this explanation, a study conducted by Smallwood and colleagues (2004) showed that when adults reported being engaged with a sustained attention to response task, their errors were higher, but so was their EDA response, whilst they did not observe this effect in EDA when participants reported mind-wandering thoughts. The authors argue that having short episodes of reported self-awareness of errors support an "Oops phenomenon", where attentional efforts are redirected to the task after the detection of an

error. This theory again would link well with the somatic marker hypothesis (Bechara et al., 2005), where the psychophysiological response is a consequence of the decision made and the result of said decision, which in this case could be the awareness of having made an error.

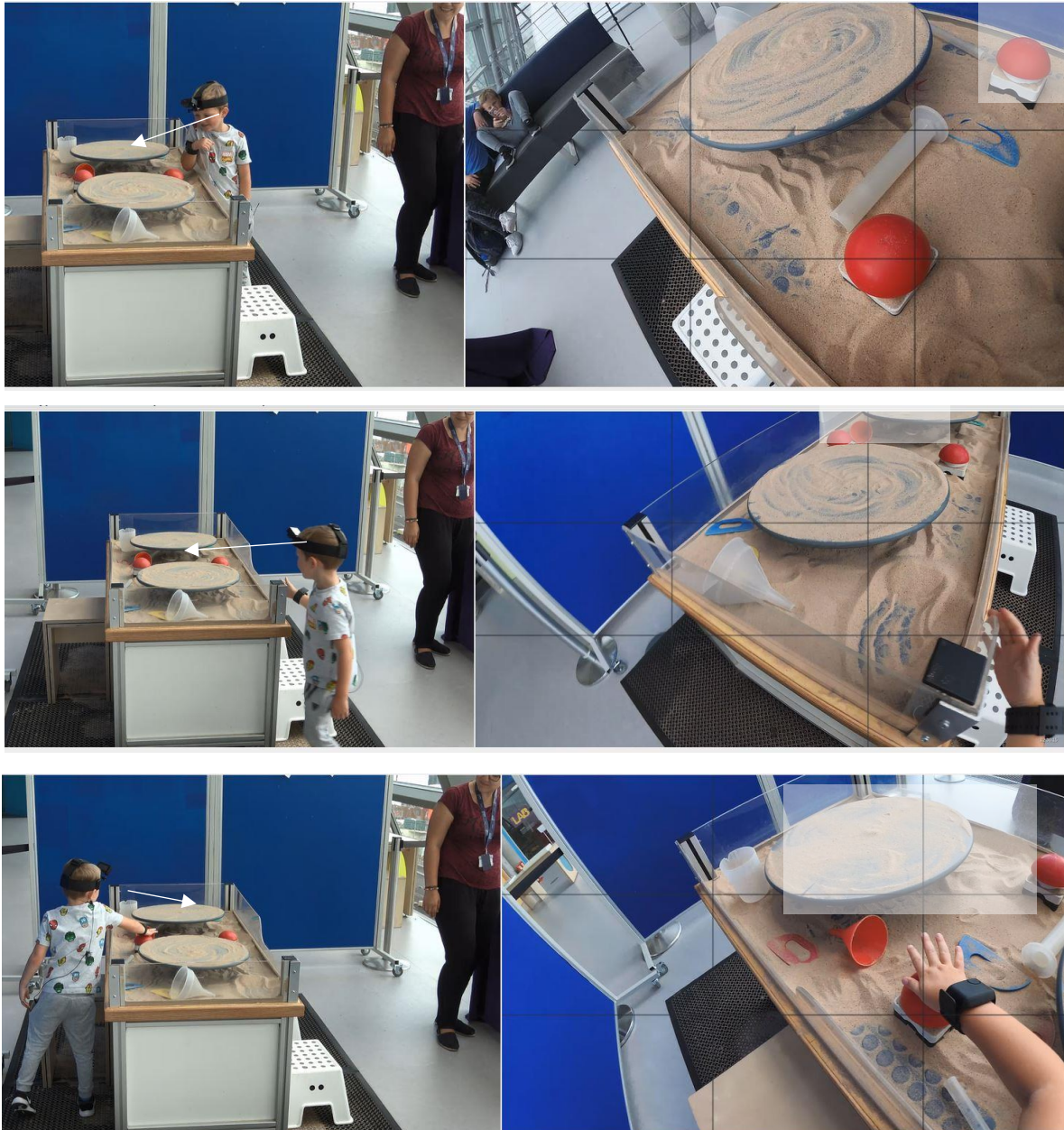
Related to changes in children's attention, another variable that interestingly showed an increase in NS-SCRs when its duration increased, was Visual Attention Mismatch, which was coded whenever children's attention was split between what their hands were holding and where their eyes were fixated. Research has shown that individuals tend to gaze constantly at the objects being manipulated for monitoring purposes, however, they also move in anticipation to the objects or locations that will be needed next in order to plan and guide the body towards it (Land et al., 1999). Figure 4.18 shows an example of a child looking away from where his hands are placed, and it can be appreciated that the mismatch between his gaze and his actions is related to exploration of space and planning for next action. In this particular example shown in Figure 4.18, the child was visually exploring the exhibit and after pressing the red button closer to him he directs his attention to the button farther from him and then explored the other big red button. It can be observed how they direct their attention to the other disc while pressing the button, already expecting an effect in the disc that starts spinning.

One unanticipated result was that the observed psychophysiological responses to the Enthusiasm cognitive-behavioural events were not significantly different from the other observed behaviours. This could be attributed to the smaller number of observed cognitive-behavioural markers related to an emotional expression, whether this was verbal or non-verbal, in comparison with the higher numbers of observations behind actions and strategies. The few amount of observed events of Enthusiasm (and even fewer of Negative expression of emotion) was possibly due to not being able to detect many obvious behavioural changes in the participants whilst they experienced any sort of additional

emotional arousal, since it is possible that subtle emotional changes are easier to pass undetected by the human eye while more pronounced emotional behaviours are observed

Figure 4.18

Example Image of Visuo-Tactile Attention Mismatch



Note. Example shows how the child has his hands in the edge of the table, but his gaze seems to be directed at the red button in the other side of the sand exhibit (top). He moves towards that red button without moving his gaze (middle) and when he presses the button, his gaze is no longer directed at the button but at the expected movement of the spinning disc beyond the red button (bottom). During this example, the child had previously tested the other button so he was aware of the effect it had on the disc that was closer to him and therefore expected certain movement from the opposite disc.

less frequently (Martinez & Du, 2012; Matsumoto & Hwang, 2011). This could be used as evidence that not all emotional responses can be easily detected by observational coding of behaviour (i.e. facial expressions, voice tone), since there was a high number of NS-SCRs recorded on average per participant (approx. <200) which did not map onto observed Enthusiasm events. Furthermore, it could also be likely that the NS-SCRs were not attributed to a positive emotional arousal even if no negative expressions of emotions were observed. Evidence of this has been shown in research by Hedman (2008) where he used EDA sensors to capture a child's emotional arousal when a child played with Lego on their own or with a help of a parent, where their results pairing EDA with observational analysis reflected that children's frustration or stress for building Lego with uncertainty showed certain SCR peaks if the child was not reassured after looking at the parent. Research, however, has also found that facial expressions of emotion do not always co-occur with the expected facial signal, particularly for positive affective expressions (Duran & Fernandez-Dols, 2021). These results highlight the limitations of using only observable expressions of engagement to try to determine a child's emotional state and emotional experience from their interaction with the exhibit. It leads to questioning how much we should rely on using observable facial/bodily expressions of emotions as the sole indicators of whether children are enjoying an experience or not. This would push arguments in favour of automatization of tools that could capture, recognise and identify the micromovements behind facial expressions of emotion (Malmberg et al., 2019; Wilhelm & Grossman, 2010).

It is important to emphasize that the NS-SCR count for the associated behaviours all needed to have passed a specific threshold, reducing the possibilities of noise or artefacts found, in other words increasing the likelihood that the count of NS-SCR could have been a consequence of each of the specific cognitive-behavioural markers considering the response window. However, one alternative result that could be driving the results observed between the number of NS-SCRs could be due to the constant movements during interaction with the hands-on exhibit. Movement has been shown to correlate with bursts or peaks of NS-SCRs

in EDA sensors, which if it were the case, when pairing hands-on exhibits that heavily rely on movement could lead to an artificial increase in NS-SCRs (Garbarino et al., 2014; Taylor et al., 2015). While this has hopefully been controlled for by having a higher threshold for considering significant NS-SCRs, the possibility remains and using data from an additional accelerometer and heart-rate monitors could provide additional information to clarify whether the EDA markers observed were due to movement or emotional arousal. Alternatively, EDA sensors can also capture temperature regulation in the body coming from the autonomic nervous system, however, this might be measured more through slow-response tonic activity, which was not used during this research project.

Overall it could be said that the answers to this question demonstrate that a triangulation method can capture some of the relationship between cognitive, behavioural and emotional markers behind early years children's science engagement. Particularly it has revealed that the emotional arousal measured by the EDA sensor (i.e. defined in this research as part of the emotional component of engagement) seems to be more likely influenced by cognitive-behavioural markers related to exploratory inquiry, such as using strategies and self-determined goals to achieve. While these findings are limited to EDA outputs, this could be highlighting some of the underlying reasons behind the prominent role emotion plays with engagement has when it comes to science learning, which mainly encourages exploration and curiosity.

4.6.3 Reflection on the Tools Capturing Engagement

Since this was an exploratory study, it was important to ensure there was reliability in the different behaviours and interaction patterns across all children since there was no full inter-rater reliability done. According to Rothman, (2007) reliability ideally should reveal similar scores using the same procedure, this is particularly important for coded data that is acquired after a level of interpretation, since the difference in the results are expected to be due to random error (e.g. individual differences). Reliable data should approach a normal

distribution, if the distribution is too flat it would be pointing to unreliable results that have an equal probability as chance. Overall, a total count of behavioural codes per participant (regardless of the type of behaviour) showed a normal distribution of the number of behaviours done per child, reflecting reliability in the engagement codebook potentially showing individual differences in behaviour patterns of how children interacted with the exhibit. Once the type of behaviour was considered, the different behaviours showed similar distribution patterns, a possible indicator of the nature of how each of the behaviours coded was working for this particular hands-on open-ended sand exhibit. Understandably, touching behaviours and those that involved object manipulation, like actions [FMM], strategies, and goals, were the most frequently found. On the other hand, behaviours like attention off-task, verbalizations and enthusiasm expressions showed more variability across the participants, which could be an indicator of the participant's unique style of exploring the exhibit.

EDA markers per participant¹⁸ showed variability in their times of arousal, some had constant NS-SCR peaks throughout the interaction, others showed increasing patterns of both phasic and tonic activity, and others showed a decreasing pattern of activity. The mean number of NS-SCR peaks per participant was 41 NS-SCR peaks per minute, confirming that in average children were experiencing high arousal during the interaction, which according to Malmberg et al. (2019) can be marked from 20-25 response per minute. Little research exists on young children's EDA parameters in the last decades (Cole et al., 1996; Fowles et al., 2000; Scarpa et al., 1997), whilst the oldest research with children and skin conductivity dates between 70-90 years back (Jones, 1930, 1935, 1950, 1960), which hinders the interpretation of the variability in EDA responses from being attributed to one single reason. However, there are some possible reasons that could explain the variability. For example, variability could be reflecting young children's personality traits or temperament (Fowles et al., 2000; Jones, 1950, 1960), as research has found that when children are more anxious or inhibited their electrodermal reactivity tends to be higher. It could also be attributed as a

¹⁸ Individual images per participant found in the Appendix I

response to the different stimuli being experienced, both the ones used in the exhibit as well as some of the science centre's intrinsic ones (e.g. loud constant noises), since children could have responded with large peaks for louder sounds and smaller yet significant ones for positive visual stimuli (e.g. spinning disc or watching sand fall). This is supported by research that has found that children between 2-4 years old, have responded more strongly to loud sounds (e.g. doorbell or car's horn) and less strongly to turning lights on/off or to hearing minimal verbal stimuli (e.g. "look there") (Wechsler et al., 1930), which was also confirmed in other research showing that 4year old children show large SCRs with loud and moderately loud noises and smaller yet significant SCRs found with positive visual stimuli (e.g. glow in the dark stickers) (Fowles et al., 2000). Some of this variability could have also been caused by different levels of hydration in their skin at the beginning, which probably increased throughout the interaction. Fowles, and colleagues (2000) hypothesise that higher resistance in the skin when it is not hydrated could be related to a notable increase in the course of the research session as the skin gets more hydration. Lastly, movement has also been found to influence artifacts in SCR (Garbarino et al., 2014; Taylor et al., 2015), and as discussed in section 4.6.2, this could have been heavily influencing the NS-SCRs given that the exhibit used was hands-on. However, the E4 sensor used, was designed to be used as a mobile research tool, which would have been designed to account for mobile artifacts, whereas research pointing out at movement artifacts has used wired EDA sensors that require the participant to stay as still as possible. Furthermore, since the scope of the study was not to compare EDA activity before or after the interaction but to triangulate EDA information during the interaction with the rest of the tools, a baseline was not considered and instead measures were used using individual standardized scores. All the different applications presented here where mobile EDA has been used speaks to the potential this tool has for providing a glimpse into the unobservable emotional processes an individual goes through and might be using as a 'gut-feeling' reference for deciding how to engage while interacting with a learning activity. Thus, while EDA seems to be a valuable psychophysiological source that can track some markers of children's emotional

engagement throughout the interaction, it is still not enough on its own to explain what happens throughout the interaction. Therefore to understand what could be driving these arousal peaks it is necessary to complement with observed behavioural data and additional psychophysiological tools.

Regarding the self-report questionnaire, all answers scored with the visual analogue scales of enjoyment (i.e. Smiley-o-meter) reached ceiling effects. In general, self-report questionnaires in children are not very common, the one found for children in science centres (Ben-Eliyahu et al., 2018) seemed to be inadequate at capturing younger children's engagement, particularly since it was designed for older children, and does not consider as much younger children's tendency to agree with the researcher (Hernandez et al., 2014). This led to include in the questionnaire the Smiley-o-meter evaluation, which like the Fun Toolkit, was designed to compensate for the verbal component weight by helping children identify feelings or opinions using a visual depiction of inner states (Read et al., 2001; Read, 2008). However, the elevated results for the smiley-o-meter items could be suggesting that this style of self-report measures tends to generate the opposite effect by overcompensation, supported by previous research where younger children tend to choose the highest score on the scale when asked about single opinions about a topic (Read et al., 2002). The smiley-o-meter has also been criticized for facilitating the agreeability bias children are already prone to, which has also been discussed in other studies that use similar tools (Delprino et al., 2018; Sridhar et al., 2018). The latter could have been enhanced when the researcher helped read and explain the items out loud to the participants before the interviews, which could have led to children choosing happier faces to please the researcher afterwards. Other researchers have also argued that this tool does not help or allow for an accurate reflection of children's emotions or attitudes (Huan, 2020). The lack of nuance found in the self-report scale could also be related to providing children with too many "emotive" options to choose from, when a different narrower approach (e.g. "Yes/No/ Maybe") could have captured that nuance better by simplification (Read, 2008). Nonetheless, the current research project was

aimed at using and/or adapting tools and measures that have been previously used in research, allowing a critical reflection of the tools once the study was done. The questionnaire used was accompanied by open-ended questions that allowed children to explain their intentions and experience with the exhibit which informed the retroactive coding of some of the self-set goals. A difference as well in this questionnaire when compared to the fun Toolkit (Read, 20002), is that it captures single open-ended experiences, contrasting the Fun Toolkit which measures and compares children's fun with different objects/activities, the latter having a specific goal to be achieved and more than one experience. Also, to circumnavigate children's agreeability bias, an item was added where children had to self-report what a friend who was not in the room would think/feel. Nonetheless, this self-report questionnaire could be improved by limiting the emotional options available in the smiley-faces and adding specific depiction for more accurate image of "like/dislike". While the visual analogue scale fails to provide nuance, it still captures children's expectations and feelings of this particular experience.

Answers to this study's research question also encompass how the same tools can capture the temporality of the continuum of engagement proposed in section 2.2.5, where engagement is captured as both a resulting outcome and the process of engagement. The outcome was captured in the interaction by considering the total time spent used for certain behaviours and the total frequency/amplitude of electrodermal responses, whereas the process of engagement considered the same outputs from the tools but in the shape of trace-data throughout the interaction with the exhibit. Particularly since the same behaviours showed a similar relationship when the data was averaged and when this was analysed as continuous data. Furthermore, following the co-occurrence of single events throughout the interaction rather than the total sum of these events provides more detailed information that facilitates the identification of specific behaviours that could be key drivers of engagement, like the use of strategies and goal-setting and its likelihood in eliciting an emotional arousal via SCR.

4.6.4 Limitations and Future Directions

This research study also has its limitations to be considered. The absence of additional relationships between the different tools' outputs could be attributed firstly to a lack of influence of the cognitive-behavioural markers on emotional arousal. Secondly, limitations in the coding process of the video data, in how the data was processed for the analysis, such as proportional percentage of time which sums the events driving the process of engagement, or to a relatively small sample underpowered to compare between total averaged outcomes ($n=27$). On the other hand, the absence of significative relationships between the tool's outputs could be pointing towards a limitation of the different measures being used, such as using the proportional percentage of time spent in the exhibit, total frequency count of NS-SCRS or median amplitude as reliable measures for tracking overall cognitive/emotional engagement on their own. It could also have been because the measures used for EDA could potentially be driven by all the different behaviours simultaneously and not by a specific one, thus leading to confounding results.

Additionally, since the E4 does not provide its own analysis software while, it could be possible that some of the parameters used to extract the raw data with Ledalab were not ideal for the type of data obtained with children in a hands-on scenario that has no specific SCR events, which could be increasing noise or failing to detect artefacts related to movement as it has been suggested before (Garbarino et al., 2014; Taylor et al., 2015). If these were wrong, they would have impacted the number and amplitude of NS-SCRs per participant. Moreover, since there are no existing parameters of children's skin conductivity, the amplitude was z-scored to facilitate comparison between participants. However, because standardisation centres data around 0 as the mean, a lack of variance in the standardised amplitude complicates statistical analysis when it comes to working with number values closer to zero and requires more complex solutions outwith the time-frame of this project to be considered. There may also be a limitation in using only the frequency of NS-SCR peaks

and not the amplitude of each of these peaks, since these could have provided more information about the intensity of the events.

Along these lines, future work could use not only the EDA phasic peak activities but complement it with the skin conductance level (tonic activity) which might be less prone to movement artefacts. Tracking tonic electrodermal activity during an interaction has been shown to have the potential to deliver an automated longer-term perspective on children's emotional arousal as shown in Hernandez et al.(2014). However, Hernandez and colleagues (2014), investigated the automatic recognition of EDA parameters using machine-learning algorithms to detect whether children can be engaged during social situations, which was not done for this project since the aim was to explore an already engaging exhibit and triangulate results between tools. While there is no normative data on children's SCL, using tonic data could show the gradual change of SCL throughout the interaction time, which can then be standardised to account for children's individual variability¹⁹ allowing a more appropriate comparison of SCL change over time between the different children. Nonetheless, the challenges associated with using a mobile EDA sensor and the multiple co-occurring factors that come with examining a context "in-the-wild" should still be considered when interpreting the data.

The E4 sensor used also captures other psychophysiological markers like temperature, heart-rate and accelerometer data, while adding these additional measures was beyond the scope of this project, future research could consider how these other markers could track behaviour and emotional arousal and how they relate to the data obtained from EDA sensor. Particularly since research has shown, for example, that although electrodermal activity increases with task engagement, heart-rate increases with mind-wandering or dis-engagement episodes (Smallwood et al., 2004). An accelerometer could also be useful in hands-on scenarios where the data obtained from it could help account for sudden abrupt movements and decrease even more the likelihood of noise or

¹⁹ as shown in Appendix L "Electrodermal Decomposition per Participant"

artefact interference (Blanchard et al., 2014). Thus, it could be possible that a constant coupling of simultaneous behaviours that probably influence one another cascading into co-occurring simultaneous psychophysiological changes reflects more accurately what is an average experience in a science centre. For example research by Wass et al. (2016) found that some psychophysiological signals can predict another physiological response that takes longer to act. In their findings, they show heart-rate predicting an onset of EDA responses in infants. Then, caution should also be placed on the possible causality driving the NS-SCRs since one of the limitations of working in a science centre context is having uncertainty over the causality of chained events, particularly if overlapping stimulating events could all be enhancing the captured arousal as discussed earlier in section 4.6.3.

The triangulation method presented here, using trace data of the interaction can be an efficient method to keep track of co-occurring cognitive-behavioural markers happening during the process of engagement, but also to triangulate how these cognitive-behavioural markers could be reciprocally influencing a child's emotional arousal. Nonetheless, the triangulation presented here could be further enhanced if, like in cognitive-behavioural markers, more co-occurring psychophysiological markers were included to clarify the role of emotional arousal in emotional engagement.

Regarding the statistical analysis, different analyses could have helped find additional relations between the cognitive-behavioural markers and the psychophysiological time-series from the different tools, such as cross-correlations between the different outputs, or creating time epochs or trend-analysis throughout the whole interaction as observed in Wass et al. (2015, 2016). However, these types of analysis have been used with more fixed scenarios that involve a controlled change of activities and events happening in a laboratory that could not be established due to the exploratory nature of the science exhibit. Whereas the generalised hierarchical model takes into account the type of data (i.e. count data with Poisson distribution) and facilitates the explanation of variables that are not equally distributed (Boucsein, 2012; Rabe-Hesketh & Skrondal, 2008), also accounting for the

different hierarchies in the behavioural categories and presenting a novel method to analyse triangulated data. Moreover, the hierarchical model used as an analysis method for the trace data presented an advantage in untangling some of the confounding factors, like considering different behaviours might have similar patterns, but also accounting for those different children that would present similar patterns within their own behaviours.

Lastly, behavioural in-depth coding can be time-consuming and prone to human error, especially when using a newly developed tool, like the engagement codebook presented here. However, both the in-depth version revision done while developing the code, and the results obtained can confirm that the method and codes used seem to point to findings that have also been supported by research regarding engagement. Whilst having a time-consuming method could also be a limitation moving forward with other science exhibits, a method like the triangulation approach discussed here provides not only the successful feasibility to capture a multimodal approach to engagement in younger children but would be useful as an in-depth tool used for evaluation. For future work, partnering with human-computer interaction researchers and computing engineers who are more experienced in automation and detection of specific events, such as facial and voice tone expression, could reduce the coding workload and enhance coding accuracy.

4.7 Conclusion

The findings from this study highlight the importance of considering both the multicomponent of engagement and the temporality of engagement (i.e. engagement as process/outcome) within a task when designing research to understand what drives it. This study has shown that, as proposed by Azevedo (2015), a triangulation approach with a multimodal perspective on engagement is a method that can help navigate a complex scenario, such as a science centre, by unpicking and separating some of confounding variables that normally impact engagement in science learning. In particular, this study found

that cognitive-behavioural markers of engagement, such as goal-setting, use of strategies or planning of next actions, are more likely to elicit an emotional arousal response, such as an increase of NS-SCRs, during interaction with a science exhibit. This likelihood of eliciting electrodermal NS-SCRs increases the longer children do a behaviour. Children also showed higher use of New strategies when compared to strategies previously used or adapted. Children showed persistence in keeping a goal until this was appeared to be fulfilled. There might also be an indication of an underlying relationship worth exploring between children's engagement and parents' facilitation.

The overall results of this study showed that having both a triangulation of different measures but also a conceptualisation of engagement as a process, can provide richer information about the specific processes that are relevant to engagement. It can provide enough fine-grain detail to understand in-depth particular activities of interest to science learning, like exploratory behaviour, that could help identify the most common interaction patterns of said activity as shown in the results. Furthermore, this study also found that the method evaluated here is also feasible to use with children as young as 3 years old, adding to the literature of a group that has been understudied when it comes to science learning, and even more when it comes to engagement (Andre et al., 2017). While there is still work to be done to improve the accuracy of the tools used within the triangulation model, the overall proof-of-concept shows that the combination of different tools broadens the perspective behind how engagement might look like.

These types of results can contribute to museum and science centre practice by providing a new conceptualisation and method to capture engagement which can help guide design and evaluation as suggested by Allen (2004), and E. Wood and Wolf (2008), whether it is of science exhibits, science visits or learning activities that can identify beneficial or detrimental features that could be misleading children's interactions.

Taken together, the implications of conceptualising engagement as both an outcome, the summarised result of all the time and behaviours elicited by the end of the interaction,

and as a process, the time-series of event co-occurrence, lead to pondering whether this concept can lead to differences in capturing engagement through simplified methods. That is, would there be differences in results when engagement in an activity is specifically asked to be evaluated as a discrete summative score, or whether there would be differences when asked to be conceptualised as a dynamic process. This guides the aim of the next chapter evaluating expert practitioner's dynamic perceptions of engagement with a science exhibit.

Chapter 5

Science Practitioner's Perceptions of Engagement

5.1 Overview

By applying the triangulation multimodal approach, the previous chapter contributed to our understanding of the conceptualisation and evaluation of engagement as both a process and a short-term outcome. This previous work highlighted the importance of distinguishing between evaluations focused on summarising engagement, often unidimensionally, at the end of interaction and evaluations that attempt to capture the dynamic and multifaceted dimensions of engagement throughout an interaction. It particularly emphasized the need to consider dynamic changes when evaluating engagement as this could help identify how the physical and social context could be shaping children's experiences during their interaction with a science learning activity. This also highlights the limitations of summarising engagement post-interaction, since it prevents the documentation of *how* these factors might be influencing engagement and the learning experience. Museum and science practitioners rely heavily on capturing engagement during summative evaluations of exhibits or activities, which makes them experts in acknowledging and identifying how engagement with the environment could be shaping children's experiences. Therefore it would be valuable to explore differences when these experts are asked to judge engagement dynamically throughout the interaction or when asked to rate it at the end of it.

This chapter addresses this exploration with a study relying on expert science practitioners who are experienced in detecting and judging daily engagement with visitors. They watched videos with different levels of engagement and were asked to rate engagement with a tool specially developed for this study to capture the dynamic scoring of engagement as well as a discrete score at the end. Also, to capture their understanding of the term engagement, they were asked to define engagement along with markers they used to track it. Finally, the chapter discusses the benefits and limitations of these differences in capturing engagement, whether practitioners agree on their ratings of engagement between them and what new directions this capturing conceptualisation of engagement could entail. This study adds an additional view to the triangulation method by aligning findings from the previous chapter to practitioner's perceptions by evaluating whether there was agreement between their perceptions and some of the results obtained from the triangulation method, which could help consolidate whether the triangulation method is capturing and measuring similar processes to those of the experienced practitioners and how useful it can be in the future.

5.2 Study of Practitioners' Perceptions of Children's Engagement

Informal Science Education institutions are required by their funders and stakeholders to provide summative evaluations of their spaces, exhibits, visitors and activities (Callanan, 2012). According to a review of designs and methods used in Informal Science by Fu et. al (2016), summative evaluations are concerned with producing practical knowledge that informs decision-making and improvements. These evaluations contribute on three different fronts to the community: 1) they show the impact these activities or exhibits are having on visitors and help justify the value of informal education to society, 2) they

contribute knowledge to the informal education field, and 3) when having high-quality standards, these evaluations can build theory and inform practice (Fu et al., 2016). One of the evaluation purposes is to describe specific events, activities, exhibits or programs, contributing also to identify the characterization of specific programs; for example the *Going Ape!* project was about finding active prolonged engagement in certain exhibits (Tisdal & Perry, 2004). Another purpose is to find relationships across variables involved in these activities, preferably relationships that can lead to discovering causal mechanisms that can also link back to the theory (Fu et al., 2016). Informal science institutions also carry out internal evaluations and use them to improve their practice and design of activities and exhibits. Observation is one of the main methods used to carry out these evaluations (Fu et al. 2016). Observation used in museums can be unstructured, where notes are taken in-depth of what practitioners are observing without any particular set of guidelines (Grack Nelson & Cohn, 2015), for example, a researcher that sits close to an exhibit during a school visit and registers all the behaviours observed (Shaby et al., 2019). Observation can also be structured, identifying specific behaviours to observe and using a protocol or checklist for this, for example using scales or questionnaires that can help guide an observation regardless of this observation tracking the amount of time spent in an exhibit, the population that approaches the exhibit, or the behaviours displayed when being at the exhibit (Allen, 2004; Falk & Dierking, 2000; vom Lehn et al., 2002). Observation can be used in a live scenario either by following visitors through a museum or by unobtrusively observing from a distance and keeping a record of the main variables of interest. It can also be used through video-analysis, although this requires more time-allocation when researching visitor's behaviours in a museum/ science centre setting (Allen & Gutwill, 2004; Barriault & Pearson, 2010; Schneider & Cheslock, 2003). For example, observation could be used to evaluate how visitors engaged with videos in a museum (Serrell, 2002), or to detect whether visitors were engaged with a science exhibit and how (Barriault & Pearson, 2010). However, E. Wood and Wolf (2008) have also acknowledged that practitioners' interpretation of visitors' observed and measured behaviour (including engagement) is guided by their own

preconceptions and values of engagement, along with institutional goals and by the visitor's own attitudes.

Furthermore, the result of these guided observations is a discrete score (or a sum of scores) that summarises the perceived intensity of the selected cognitive process during the interaction. Following the previous example, the Visitor Engagement Framework (Barriault & Pearson, 2010) is a scale with a set of behaviours that capture whether a child goes through 3 different stages of engagement, from initiation into a breakthrough. However this scale is scored in the number of behaviours present of each of these three categories by the end of the interaction, with one of the specific items called engagement where the observer decides whether children were engaged or not. This type of summary blurs the evaluation of time-dependent variables, such as learning or engagement, since it reduces them to a final discrete score that relies heavily on an average of the child's interaction throughout the activity, which would encumber the exploration and contribute little to an understanding of their underlying processes. There is a need from the museum and science centre practice to be able to quantify visitors' behaviours in order to compare experiences in different exhibits, or with the same exhibit over time (Serrell, 2020). But it is proposed in this section that the comparison of these different experiences should account for the temporal dynamics and the variability of the process of engagement that is associated with each type of exhibit. This is important as Wolf (2002 in E. Wood & Wolf, 2008) states that more precise measures of engagement can clarify differences between features that attract the most visitors versus the ones that can promote or elicit higher measures of engagement.

As argued previously in this thesis, engagement being both an outcome and a process²⁰, it would be important to have tools that can be used by practitioners in messy contexts to capture more information about the process of engagement. Mainly because these summative and practical evaluations have the potential to examine plausible causal-

²⁰ See section 2.2.5 "Continuum of Engagement"

effects mechanisms behind engagement related to science learning. This can inform practice and help build theory for science engagement in a science centre, as proposed by Fu et. al (2016). However, as recognised in the previous chapter, in-depth methods of evaluation, like the triangulation method, can also be a time-consuming process that requires resources museums might not always be able to spare (Callanan et al., 2017; Fu et al., 2016; Grack Nelson & Cohn, 2015). Despite these limitations, the findings obtained from this method regarding the importance of capturing the dynamic variability of changes behind engagement remain of value.

A solution for this would be developing a tool that demands less time, or less training but still allows practitioners or researchers to monitor the process of engagement, tracking the continuous subtle changes in the intensity of engagement. Whereas tools that capture an intensity rating from an observer have been used in the field of human-computer interaction (Laurans et al., 2009), these have not been found in the literature for museum or learning studies. One of the tools found in the literature is the emotion slider, which allows participants to rate with a slider how intense they felt an elicited emotion (positive/negative) after looking at a photograph. Tools like this could be a foundation for developing other tools that work with interactive systems in dynamic intensity ratings. Commercially, marketing tools have been developed, such as the Dial Perception Analyzer (Schlesinger Group, n.d.) to help companies measure the “enjoyment” of a live audience after watching a specific advertisement or show, however, costs for marketing research tend to be higher and less affordable for smaller research-practices. Using a slider as a tool has the potential to enhance live observation, providing a more detailed assessment of what is being observed. This could go from accessing more information about an event or capturing dynamic changes throughout the interaction. Even more so, if this data capture method were combined with video-observation it could generate triangulation markers based on science practitioner’s expertise and detect if there are possible points during an interaction where engagement varies for all visitors. It could also provide a more quantifiable method to reflect

individual differences in interaction patterns as expressed by Gutwill (2008) to be relevant when it comes to visitor's engagement.

Considering the expertise science facilitators have when it comes to monitoring children's engagement and interest at a science exhibit it would be valuable to compare their appraisal of children's engagement when asked to evaluate engagement as a continuous process versus being asked to consider a summative score by the end of the interaction. Their judgements have the potential to illustrate a common understanding in practice of engagement (E. Wood & Wolf, 2008) and the different markers used in practice to track children's engagement. This could help consolidate whether the triangulation method evaluated in Chapter 4 is capturing and measuring similar processes to those of the experienced practitioners. For this, a real-time dynamic rating tool that measures continuous engagement was developed in order to address the gap lacking practical tools that can capture the dynamic intensity of a variable (i.e. engagement) in a science centre. Using this tool with practitioners adds further value to this study by capturing a more nuanced perception of expert practitioners and contrasting these with their own summative perceived outputs.

This chapter examines science practitioners' judgements when engagement is rated dynamically versus as a discrete score, and whether practitioners showed any agreement in their perceived engagement using videos selected from the triangulation method. This was done through the developed slider tool that captures continuous and summative scores of engagement was used. Data was also captured around practitioners' views where they were asked to define engagement and the behaviours they used to detect and track engagement in the videos. This chapter also aimed to evaluate whether there was an agreement between practitioners' appraisal of children's engagement from what was observed in Chapter 4 and whether this could be influenced by using a tool that allows a continuous evaluation of children's engagement that is time-sensitive.

This chapter seeks to answer RQ4: ***Do science practitioners' perceptions of engagement align with results from the triangulation method?*** by exploring the following sub-questions:

a) Are there differences in practitioners' perceived ratings of engagement as a dynamic process vs engagement as an outcome (mean dynamic and overall discrete score)?

b) How much agreement is there between practitioners' perceptions of engagement?

5.3 Method

5.3.1 Participants

Six science practitioners (3 male;3 female) from 3 different institutions related to informal science learning took part in this study (National Museum of Scotland, Glasgow Science Centre and Edinburgh Science Festival). The main inclusion criterion was to have working experience as a science facilitator in a UK science centre.

Ethics - Participants were all asked for signed consent for participation. The study followed ethical guidelines from the British Psychology Society and was approved by the Research Ethics Committee of Moray House School of Education.

5.3.2 Tool- Slider for Tracking Engagement

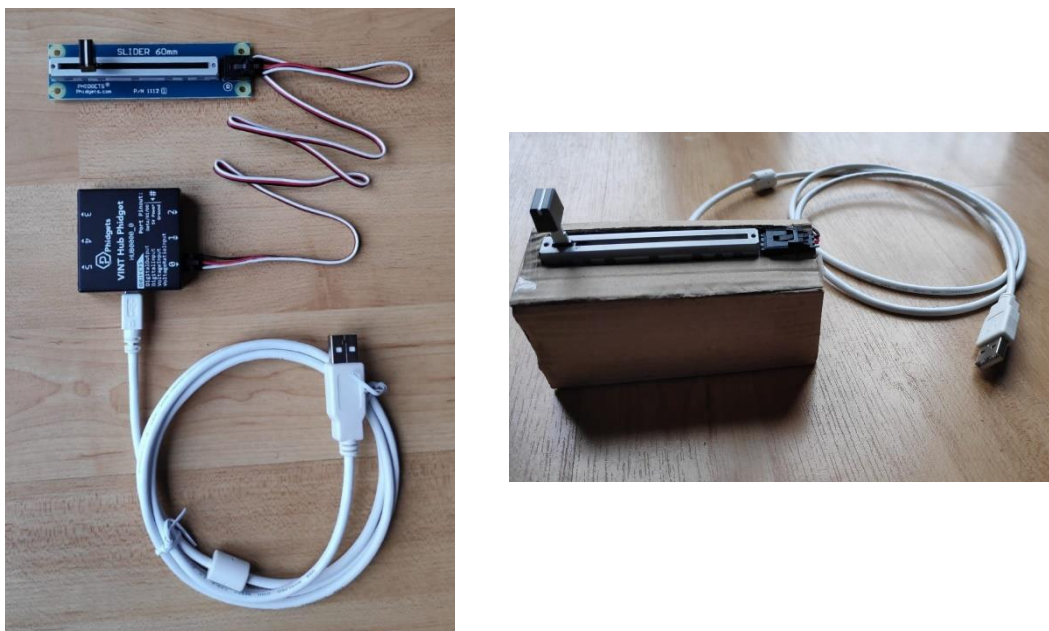
Development - Since the aim was to work with practitioners in different locations, the tool needed to be non-intrusive and as mobile as possible. Using a hand-held slider presented a good opportunity to become an embodied tool for which users could off-load some of their cognitive thinking into a natural increase-decrease hand movement. This would avoid constant monitoring while using it or adjusting a particular value, and potentially

missing key information. The developed slider is intuitive to use, sliding only in two possible directions, from left to right (or up-down depending on how the observer positions it), representing an increase of intensity like that displayed in other common devices (e.g. volume symbol). The slider tool works along a computer-screen interface where different videos can be uploaded and rated, but only allows playing the videos once and without pausing to simulate some of the restraints with live-observation. Lastly, the slider tool is meant to be used with a screen to allow synchronisation of the continuous score with the video and to compensate for and reduce distracting factors that could be impacting if done in-the-wild.

Description - The slider is a screen-based but hand-held tool that was developed specifically for this research to continuously measure the variability in the intensity of engagement while observing a recorded video (Figure 5.1). It is a tool that combines a computer interface with a Phidget® slider, a linear sensor (or potentiometer) that measures

Figure 5.1

Composition of Slider Tool



Note. Left: Phidget slider and Phidget VINT hub. Right: Box used to hold and use slider.

the movement of an object within two points and transforms it into an electric signal that ranges from 0 to 1. The slider is connected to a Phidget® VINT Hub that allows the analogue information from the slider to be stored and relayed digitally onto a computer as an output. The open-source interface was programmed using Python (Revueltas Roux & McClure, 2020). The interface allows a user to choose a video, calibrate the sensor to test if is working, and then record the user's continuous perceptions of engagement for the duration of the chosen video. Once the video finishes, it requests the user for a discrete score of the engagement observed and any additional relevant feedback that should be considered.

5.3.3 Engagement Videos

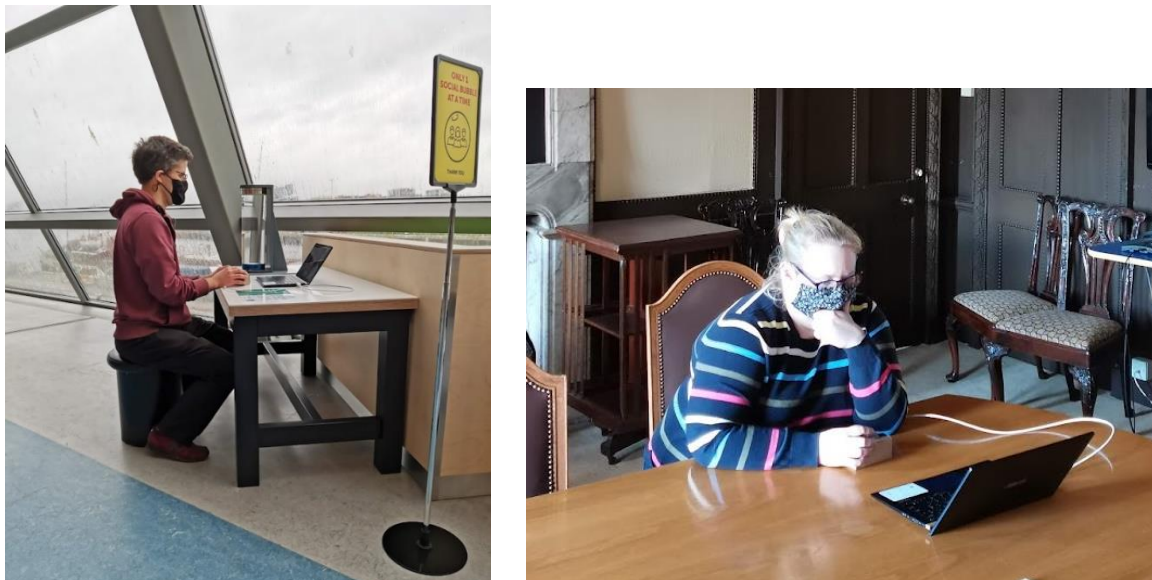
Three videos were selected using the results from chapter 4 for use in the present study. The videos showed children interacting with the Sand Patterns science exhibit and three were chosen to show variability in engagement depiction because they represented a low (V1), moderate (V2) and high (V3) engagement according to the previous chapter, using the proportional time spent in each of the relevant categories: Strategy, Actions and Attention Off-Task. To do this the 28 videos were rated from highest to lowest proportional times. The highest engagement level had a 76.08% of time spent using *Strategy*, 81.65% using *Actions* and only 4.49% time spent *off-task*, whereas the lowest level had 29.72% of time spent using *Strategy*, 47.22% using *Actions* 9.88% time spent *off-task*. To decide on the moderate level of engagement, the mean and the median for each category were obtained: *Strategy* M = 60.46%, Md = 62.76%, *Actions* M = 69.46%, Md = 72.83%; *Attention Off-Task* M = 10.58%, Md = 4.27%, and the participant's video who was closer in all three behaviours was selected. Thus, the moderate engagement video had 62.19% time spent using *Strategy*, 72.97% using *Actions* and only 5.66% time spent *off-task*.

5.3.4 Procedure

The participants were asked to rate the three engagement videos. These were counterbalanced for each participant to minimise bias. In order to familiarise participants with the slider, they were initially shown a 1-min video of rain falling with different intensities which they had to score using the slider. This demonstrated participant testing and understanding of the tool before evaluating the videos. Participants were then asked to “use the slider tool to continuously evaluate how engaged you consider the child is throughout the interaction video”. They were not provided with a working definition of engagement intentionally to minimise priming, since part of the aim was to understand what different science practitioners understood as engagement. Participants were then seated in front of a laptop ASUS ZenBook where the slider was connected, and the program started running (Figure 5.2).

Figure 5.2

Examples of Practitioners Using the Slider Tool While Watching Interaction Videos



The interface registered the participant and their general information, then it allowed videos to be selected. Before starting the recording, the program had a "Calibrate" button,

which checked if there was an input from the slider and showed how the intensity in the sensor shifted, depending on the direction the user will move it ranging from (0-100% or 100-0%). Once the device had been calibrated and the video was chosen, the interface synchronised the data recording from the slider with the beginning of the video. The video played without being able to pause in order to emulate attentional demands during real-time observation. Once the video finished, a screen asked the participant to enter an overall score for the intensity perceived throughout the video. When all the videos were shown, a final screen asked participants to describe what they had understood as engagement, the cues they relied on to grade children's engagement, their definition of engagement, and any additional feedback they might have had. A csv file was saved as an output for each video observed which contained the video time stamp and the corresponding measure from the slider every 4Hz to match the EDA sensor sampling frequency; it also showed the assigned discrete score and the answers provided at the end. Table 6.1 shows the engagement measures the slider captured.

5.3.5 Data Processing

The slider tool provided 3 different outputs: dynamic rating per video, discrete score per video and written feedback from the overall experience. A more detailed description of the obtained measures is shown in Table 5.1. The measures were all extracted using a .csv file and then exported into RStudio for cleaning, processing and analysis using R version 4.1.1 "Kick Things" (R Core Team, 2021)

Table 5.1*Measures Obtained from Experts' Observation*

Measure Obtained	Description	Output	Rationale
<i>Dynamic Rating</i>	A continuous measure of perceived engagement from the expert, obtained through the slider device for the duration of child's interaction with exhibit.	-One continuous rating from 0-100% of the perceived engagement for every 0.25s throughout the duration of the video. -one averaged score of the continuous dynamic rating	Capture a continuous measure throughout the length of the process of engagement
<i>Discrete rating</i>	At the end of the video a single discrete score is asked from practitioners to summarise the perceived engagement throughout the interaction.	One score per participant between 0-100%	A perceived average of the child's engagement experience, which when correlated with the dynamic report allows evaluation of differences in how engagement is perceived depending on the moment of the interaction.
<i>Experts' Definition of Engagement</i>	Experts will be asked to describe the markers they look for when identifying engagement during an interaction	Small definition of engagement	Understanding what practitioners define as engagement can provide grounding for future understanding of conceptualization of engagement from their perspective

5.4 Data Analysis

To answer the RQ4: “**Do science practitioners’ perceptions of engagement align with results from the multi-modal approach?**” It was important to first explore the specific sub-questions.

(4.a) Are there differences in the practitioner’s perceived rating of engagement as a dynamic process vs engagement as an outcome (mean dynamic and overall discrete score)? Given the small sample of participants, a non-parametric repeated measures test (i.e. Wilcoxon signed rank exact test) was used to compare the practitioners’ averaged dynamic score with their discrete score. The dynamic rating was averaged here to examine whether the averaged score of the continuous rating was similar or different from their self-perception of an overall discrete score for engagement.

(4.b) How much agreement is there between practitioners’ perceptions of engagement? To understand if practitioners’ perception of engagement aligned when asked to judge a child interacting with an exhibit, it was important to account for each of the types of data obtained from the slider and whether practitioners were both consistent in their rating and whether they agreed in what they were all observing. It was also of interest to explore whether practitioners’ perceptions of different types of engagement aligned with the classification of the three videos using the triangulation model.

1) Inter-rater agreement was analysed between participants to evaluate whether there was absolute agreement and consistency in how they had rated the three videos – this was done using their discrete scores, their averaged dynamic scores, and the continuous dynamic scoring data for each video. The latter was done using every 0.25 data captured as multiple examples cases where practitioners had to agree on what they were observing, as an equivalent of observing many different subjects. As a measure of inter-rater agreement between the scores from the 6 practitioners an intra-class correlation (ICC) analysis was

used (this was chosen because the data used is metric rather than nominal or ordinal). The ICC compared their scores for the three videos in the aggregated dynamic score and the discrete score. Then the agreement was analysed separately for each video using all the trace dynamic data, looking for agreement within the process of engagement. Absolute agreement and consistency between the practitioners' scores are reported. Following the guidelines to select and report ICC for reliability by Koo & Li (2016), each of the ICC analyses was done using a two-way random model, because it was the same set of raters who had been selected randomly from a specific population. The type of inter-rater agreement used averaged data since the comparison between raters' scores was sought rather than having a single rater as a base-measure. Absolute agreement was used to evaluate agreement among raters, while consistency was used to evaluate scoring consistency within the raters. Scores range from 0-1, 0 being no agreement and 1 being absolute agreement. ICC was interpreted according to Koo & Li (2016), who suggest that scores below 0.50 are poor, between 0.50 and 0.75 are moderate, between 0.75 and 0.90 are good and anything above 0.90 is excellent. ICC estimates and their 95% confidence intervals were calculated using the '*icc*' function from the '*irr*' package (Gamer et al., 2019) was used in R.

2) Analysis of inter-rater agreement between participant's assignment of the highest, moderate and lowest engagement score to each of the videos, considering single dynamic and discrete scores and whether this aligned with the selection from the triangulation model. For this analysis, a contingency table was created where it was recorded whether each practitioner's score aligned with the classification of the engagement videos using only "Yes/No" categories. To analyse this nominal data, a Fleiss Kappa²¹ was used for inter-rater agreement on whether they had agreed or not with the triangulation model. To interpret the results of this analysis, Fleiss et al.(2003) were used as guiding values.

²¹ This was chosen rather than Cohen's Kappa because agreement was being sought among more than two raters (i.e. 6 practitioners)

5.5 Results

Descriptive results were obtained for each engagement video (low, moderate, high) within each type of scoring (dynamic vs discrete) and are shown in Table 5.2.

Table 5.2

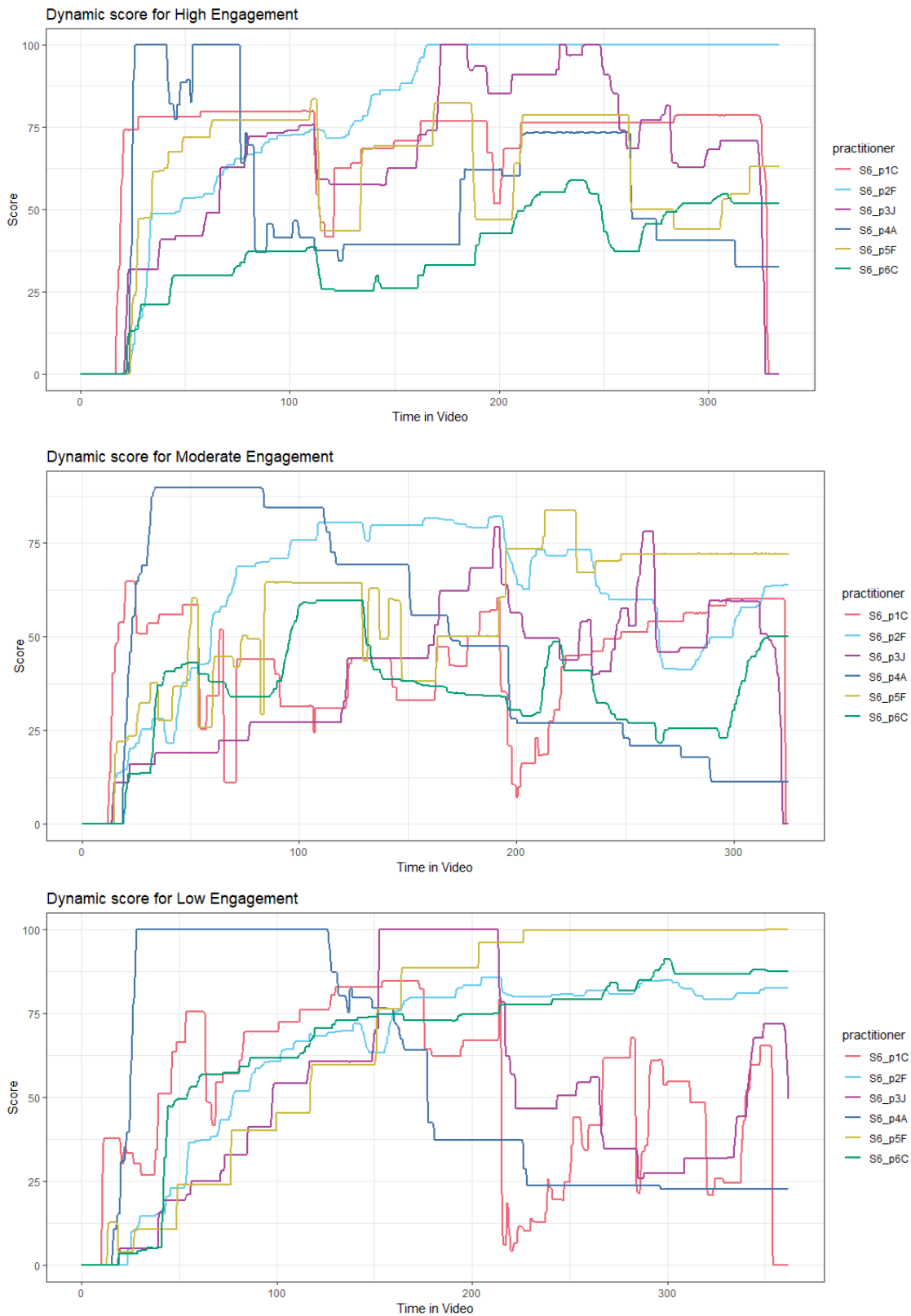
Descriptive Results for Three Levels of Video-Coded Engagement

Engagement level	Discrete (outcome)			Dynamic (process)		
	Mean	SD	Median	Mean	SD	Median
Low	64.83	(26.47)	69.5	58.76	(8.5)	58.38
Moderate	53.66	(23.60)	62.5	46.63	(9.5)	44.84
High	71.5	(18.53)	74.5	60.26	(14.55)	62.21
Total	63.33	(22.98)	68.5	55.21	(12.23)	54.60

The process of engagement implies a continuous set of scores throughout the interaction time rather than as a single score, a visualization of the practitioners' complete dynamic score was plotted across time for each of the engagement videos to describe the practitioner's perceived judgement for engagement and to examine whether there were differences in practitioner's pattern for perceiving engagement throughout time (Figure 5.3). This visualization also compares whether these patterns of perceived engagement change depending on the different engagement levels for each practitioner shown in the same colour. There were no specific patterns that stood out from the visualization, however, practitioners did not seem to rate the high-engagement video below 25% intensity, whereas the other two levels present more variability in the ranges considered. Additionally, in low-engagement, most participants seem to show similarity in a perceived gradual increase of engagement in the first half of the interaction.

Figure 5.3

Visualization of Dynamic Scores in Each Engagement Level



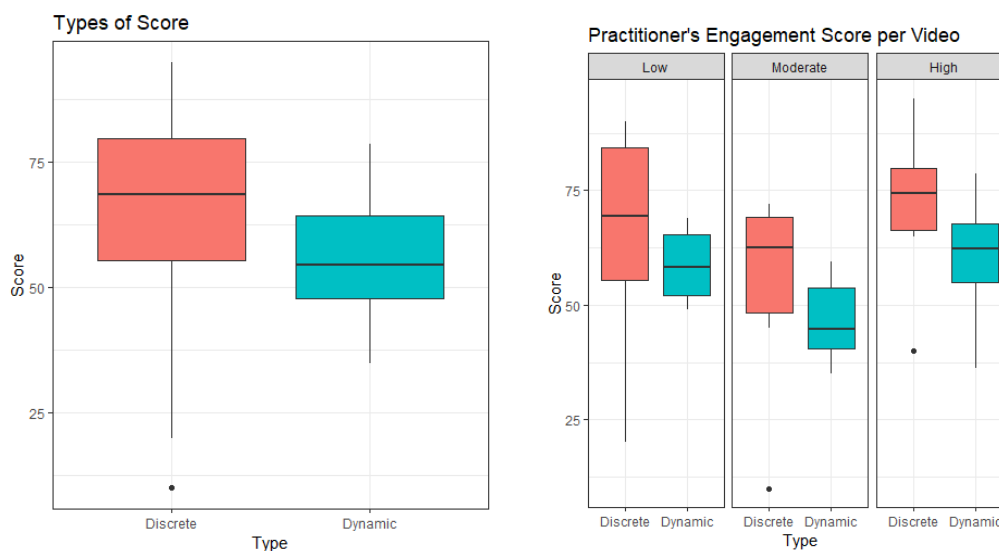
Note: All practitioners have the same colour throughout the 3 plots.

5.5.1 Differences in Dynamic vs Discrete Scoring

It was of interest to compare whether practitioners scored engagement similarly when they were able to capture a fluctuation of engagement intensity throughout the interaction (i.e. dynamic scoring) in comparison to when they were required to give a discrete overall score by the end of the task. Discrete engagement was rated on average at about 63.33% of intensity (SD = 22.978) with a median of 68.5%, whereas Dynamic engagement was rated on average at 55.219% (SD = 12.233) with a median of 54.606. When both types of overall scores were compared using a Wilcoxon paired test, no differences were found between the scoring ($V = 129$, $p = .059$), although this did approach conventional levels of statistical significance. Figure 5.4 shows a comparison between the range of overall discrete/dynamic scores, and how this looks for each of the videos used.

Figure 5.4

Plots Showing Engagement Scores Depending on Type of Scoring



Note. Plot showing the difference between total discrete and dynamic single score, both for all videos (left) and for each of the videos (right).

5.5.2 Participant's Agreement Scoring Engagement

Agreement and consistency between the 6 practitioners were sought between their three types of measures from the engagement videos: averaged dynamic scores, between their discrete scores, and between all their scores. Table 5.3 shows the ICC coefficient for absolute agreement and consistency, their lower and upper bounds, as well as their statistical significance.

Table 5.3
Interrater Agreement and Consistency for Single Scores

	Intraclass Correlation	Raters	Items Videos	95% Confidence Interval		F Test with True Value 0				
				Lower Bound	Upper Bound	Value	df1	df2	Sig	
<i>Agreement</i>										
Discrete	.404	6	3	-.096	.973	3.9	2	7.61	.068	
Dynamic	.66	6	3	-.244	.991	3.6	2	13.8	.055	
Overall	.442	6	6	-.196	.89	2.4	5	17.9	.078	
<i>Consistency</i>										
Discrete	.744	6	3	-.399	.993	3.9	2	10	.056	
Dynamic	.722	6	3	-.517	.993	3.6	2	10	.067	
Overall	.584	6	6	-.302	.934	2.4	5	25	.065	

First, to examine whether the 6 practitioners had shown absolute agreement in their discrete scoring of the three videos, the computed ICC of a 2-way random effects model with mean-rating (k=6) showed poor agreement between the raters (ICC = .40), however, this was not statistically significant (p = .068). Consistency was computed using a 2-way fixed effects model with mean-rating between the practitioners' discrete scores, which showed moderate reliability (ICC = 0.74), but again this was not statistically significant (p = 0.056).

For the dynamic averaged single score, the mean-rating 2-way random effects model showed moderate agreement between the raters (ICC = .66), however, this was not

statistically significant ($p = .055$). Consistency between the averaged dynamic scores of the 6 practitioners using a 2-way fixed effects model with mean rating also reflected moderate consistency ($ICC = 0.72$), however this was also not statistically significant ($p = .067$).

When all the practitioner's scores were considered, regardless of these being discrete or dynamic, their ICC showed poor reliability ($ICC = 0.442$) which was not statistically significant ($p = 0.078$). Consistency within the 6 ratings of the 6 practitioners showed moderate reliability ($ICC = 0.58$), but this was also not significant ($p = .065$). Figure 5.5 shows the discrete and dynamic for the three videos per participant.

Figure 5.5

Plots Showing Practitioner's Dynamic and Discrete Scores



Note. Plot showing the individual discrete (red) and dynamic (blue) ratings for each video per participant

Since the dynamic scoring represents practitioners' continuous perception of engagement, the absolute agreement between the 6 practitioners for the complete trace data of each video was computed, but only within each of the videos. Table 5.4 shows the

ICC coefficients, the lower and upper bounds and the statistical significance for each of the videos. For all the videos, the same 2-way random effects model with mean-average was used when looking for absolute agreement. Consistency was not computed here because the nature of the data was continuous and variable. Practitioners showed significant moderate reliability for video 1 and video 2 with ICC scores ranging between .50 and .75, whereas video 3 showed significant good reliability with their score $>.75$ but lower than .90.

Table 5.4

Interrater Agreement of Dynamic Scores Between Practitioners for Each Video

	Intraclass Correlation	Raters	Items <i>0.25 ms</i>	95% Confidence Interval		F Test with True Value 0			
				Lower Bound	Upper Bound	Value	<i>df1</i>	<i>df2</i>	Sig
Video 1 (low)	.67	6	1449	.62	.712	3.26	1448	510	<.0000
Video 2 (moderate)	.571	6	1301	.456	.655	2.68	1300	135	<.0000
Video 3 (high)	.781	6	1337	.624	.858	7.34	1336	27.3	<.0000

Lastly, Fleiss kappa was computed to assess the agreement between practitioners and researchers showed was analysed by recording whether practitioners' score (both dynamic and discrete) matched the classification of the intensity of the engagement videos. That is whether practitioners' lowest score matched with the triangulation's classification of the lowest engagement video, the middle score with the moderate video and the highest score with the high engagement video (Table 5.5). Practitioners' discrete scores showed poor agreement beyond chance on the intensity of the videos (kappa = 0.0182, $p = .903$), but this was not statistically significant. When using their discrete scores, only 17% of practitioners agreed that the lowest video reflected the lowest engagement, 33% that the moderate video was not the highest nor the lowest score, while the majority agreed (67%) on the highest rating given to the video classified as "high engagement". Interestingly, this

changed when the averaged dynamic score was considered, where practitioners showed a fair to good agreement ($\kappa = .486$, $p = 0.0012$), but there was no agreement at all on the video classified as the lowest, nor on the video classified as moderate, but the majority (67%) still agreed on the highest video being ranked the highest. Thus, when a dynamic score was used, practitioners unanimously agreed that the video classified as “low” or “moderate” engagement did not represent the lowest and middle scores respectively. This can also be observed in Figure 5.5

Table 5.5
Practitioner’s Score Versus Score Assigned from Triangulation Model

<i>Practitioners Choice</i>	<i>Discrete Scores Match</i>			<i>Dynamic Scores Match</i>		
	<i>Engagement Videos</i>			<i>Engagement Videos</i>		
	<i>V1 Low</i>	<i>V2 Moderate</i>	<i>V3 High</i>	<i>V1 Low</i>	<i>V2 Moderate</i>	<i>V3 High</i>
p1	Low	Moderate	High	Moderate	Low	High
p2	Moderate	Low	High	Moderate	Low	High
p3	Moderate	Low	High	Moderate	Low	High
p4	Moderate	Low	High	Moderate	Low	High
p5	High	Moderate	Low	High	Low	Moderate
p6	High	Low	Moderate	High	Low	Moderate
Total Matches	17%	33%	67%	0%	0%	67%

5.5.1 Practitioners’ Written Answers

Two main questions were asked to capture the differences in practitioners’ understanding of engagement across the informal science learning field. Answers to the first question explored how each practitioner understood engagement and are shown verbatim in Table 5.6. Whereas the second question examined what practitioners considered as key

markers of engagement which they relied on for detecting engagement throughout the three exhibit interactions. Answers to these questions are also shown verbatim in Table 5.7. It can be observed that practitioners vary in their definitions of engagement, where some have more abstract conceptualizations of it (e.g. p3J), and others use more practical observable behaviours (e.g. p4A). Moreover, the markers provided by practitioners revealed a diverse variety of indicators that sometimes rely on observable behaviours (e.g. “trying different ways of using something” / “looking at their gaze”), whilst some other markers relied more on subjective appraisals from the observer (e.g. “concentration” / “how engrossed they are”).

Table 5.6

Practitioner’s Conceptualization of Engagement

<i>Participant</i>	<i>Q1. What does engagement look like/mean to you?</i>
<i>S6_p1C</i>	Active interest in the task or activity at hand
<i>S6_p2F</i>	Using your imagination, senses, language and losing yourself in a task/activity.
<i>S6_p3J</i>	Interacting with an exhibit in a focused way so that most outside distractions are ignored. Engagement can be short but often goes on for longer and involves enjoyment, either of using / playing with an exhibit or of discovering (learning) something new through using it.
<i>S6_p4A</i>	Participants asking questions. Moving and exploring an exhibit.
<i>S6_p5F</i>	Continuous 'play', i.e. not distracted by external noises or humming.

S6_p6C Engagement is an active process, where you feel motivated to discover/explore/learn something new about the world.

Table 5.7

Practitioner's Markers of Engagement

Participant Q2. What particular signs did you use to determine engagement?

S6_p1C Concentration, focusing on manually dextrous tasks, creative thinking - making the pizza, problem solving around why the button did or didn't work

S6_p2F Experimenting, asking questions, looking for permission, singing and imaginative play, dialogue.

S6_p3J Focused on the exhibit, ignoring outside distractions, trying different ways of using something, asking specific questions about how something works or how to make it do a certain thing.

S6_p4A How engrossed they are. If they are fully immersed in the experience and not noticing anything other than the activity.

S6_p5F Eye / gaze if the child looked up.
If the child makes hypothesis "what will happen if" - these can be verbal or not (e.g. one of the children was looking underneath the table.

S6_p6C - Asking questions aloud e.g. how does this not stop? Why is it moving?
- Moving beyond simple observations towards concrete explanations ("It must have batteries")
- Depth of exploration (playing with all the items, pressing all the buttons etc)

- Evidence of trying to get others (Mum) involved as well to share in the experience
 - Lack of/prevalence of distraction e.g. looking away from the activity
 - Levels of energy exhibited by children engaging with the exhibition (child #3 had way more energy than children #1 and #2)
 - Explicit use of imagination e.g. Making Pizza.
-

5.6 Discussion

This study aimed to explore whether the expert practitioners would judge a child's engagement differently if they were asked to give a discrete overall score (i.e. outcome), compared to having a tool that allows them to rate engagement continuously (i.e. process).

When comparing score differences depending on whether practitioners had judged engagement as an outcome (discrete score), or as a process (dynamic score), no significant differences were observed. The visualization of the data seems to indicate that discrete scores show more variance in the range of answers, which could explain the lack of differences found. It could have also been that there were not enough sample videos for practitioners to rate, which could be contributing to the analysis being underpowered to detect differences. However, it can be appreciated that discrete scores were perceived as higher than the averaged dynamic score. This score difference can also be observed in Figure 5.4 showing each participant's score for the three videos, showing a variability in practitioner's rating for the different videos, while also portraying dynamic scores lower in average than discrete scores. However, dynamic scores showed less variance in the range of answers provided, both overall and for each of the videos used. In the agreement analysis, it can also be observed that practitioners' averaged dynamic scores showed higher interrater agreement ($ICC = .66$) with a lower probability coefficient than the discrete scores ($ICC = .404$). Contrarily, the consistency of practitioners' ratings was similar for discrete and dynamic scores, thus it is unlikely that practitioners are being inconsistent in their rating,

rather it could be that is the nature of the conceptualization of engagement driving their perception variance.

Another explanation for the lack of differences is that given that the nature of the continuous score is inherently different from the summative outcome, it erases the nuances observed throughout the interaction. This was echoed when examining practitioners' agreement, since the interrater agreement scores obtained for the continuous dynamic data for each of the videos showed higher reliability and statistical significance than the single score. This could possibly be explaining that having more items (i.e. 0.25 ms events) that the practitioners can rate, increases the reliability of the data.

The second aim of this study was to evaluate whether the results observed here would align with the classification of engagement obtained from the triangulation method. Firstly, looking at their agreement on how they scored the continuous videos, they showed a higher agreement ($ICC > .75$) for the video classified as "High" engagement ($ICC = .78$), meaning that practitioners agreed on the ratings they had been assigning for this video. While the other two videos showed the same moderate reliability ($.50 < ICC < .75$) in practitioners' agreement, the amount of agreement was different for both videos. The second video where practitioners showed more reliability on their scores was the video classified as "low" ($ICC = .67$), while video 2 classified as "moderate" showed the lowest reliability of the three ($ICC = .57$). When comparing this agreement to the classification made from the triangulation method to capture and measure engagement, a high number of participants (67%) agreed on giving their highest score to the highest engagement-level video. However, practitioners' scoring of the low and moderate engagement intensity was not similar to the classification obtained from the triangulation method. There was little to no agreement on the low and moderate videos being respectively the lowest and middle scores. While practitioners showed slightly more agreement with the model for the lowest (17%) and middle (33%) scores when they provided a discrete score, this changed to no agreement (0%) when the averaged dynamic score was considered. Interestingly, the classification for these two

videos was often swapped, giving the lowest score to the video classified as moderate engagement by the triangulation method and vice versa. When dynamic scores were considered, there was 100% agreement on giving the lowest score to the video classified as moderate from the triangulation method. However, in some cases the difference between the lowest and moderate scores was minimal (refer to Figure 5.5). Interestingly, two participants gave the highest score to the video classified as low from the triangulation method. Thus, the agreement shown in the ICC for the continuous judgement of engagement reflects practitioners' agreement on the intensity of engagement being perceived, which does not necessarily match with the classification obtained from the triangulation method, particularly for those classified as low and moderate engagement.

This result potentially highlights the subjectivity used by the different participants to score more nuanced levels of engagement. This echoes the suggestion by E. Wood and Wolf (2008) who emphasize practitioners' observations will be guided by their own assumptions and values, and that there is no guarantee of a shared understanding of features of engagement just because everyone uses the same term. The dynamic nature of the process of engagement and how this can vary throughout a single interaction is well represented in Figure 5.3, which again highlights the variability in practitioners' scoring throughout an interaction even when the same level of engagement is being observed. It also highlights the need for measures that can help detect whether these differences can still align in particular points of the experience, or provide a base for future automated analysis methods using machine-learning algorithms like the ones observed for automated detection of EDA in (Hernandez et al., 2014). In Hernández's research, young children interacted with the researcher during a social task, where their engagement was recorded and rated by a human video-coder, post-interaction, characterising how easy to engage children were. Then a machine-learning algorithm was trained using children's EDA outputs and the classification from observers, which lead the algorithm to search and automatically detect patterns that could be associated with each of the classification levels provided. When applied to the

outcomes of this tool, this could mean that enough sessions where engagement is evaluated by practitioners for different levels of engagement previously classified could be to test and train a machine-learning algorithm that potentially detects if there are any underlying patterns in the data captured of engagement.

The results from Table 5.7 show variability in the markers that practitioners rely on when looking for engagement in children. These findings could be reflecting that the detection of different levels of engagement via some specific markers, particularly when the intensity is lower, would be more challenging than when the intensity is high. Similar results have been observed when facial expressions are rated by external observers, where high-intensity expression of emotion is easier to recognise whereas a change in the intensity makes the markers more subtle and requires more training to improve their detection (Duran & Fernandez-Dols, 2021; Matsumoto & Hwang, 2011). Practitioners' answers in Tables 5.6 and 5.7 show that some of the behaviours mentioned by practitioners as relevant markers to track engagement were similar to the behaviours detected as relevant in the triangulation model. For example, Attention to the task, where participants mentioned: "active interest in the task", "losing yourself in the task", "Most outside distractions are ignored", "not distracted by external noises", and "lack of distraction" whereas Attention off-task was considered in the triangulation model as the referent point for children being dis-engaged following research by (Akshoomoff, 2002). Regarding the use of high-end exploratory behaviours, such as Strategy or Goal Use based on (Börnert-Ringleb & Wilbert, 2018), participants mention that engagement "makes you feel motivated to discover and explore", "moving and exploring the exhibit" and "using your imagination", where more specific markers related to this behaviour were: "problem-solving situations", "experimenting", "trying different ways of using something", "the child hypothesising *what will happen if..(verbally or physically)*", and "depth of exploration". At least two practitioners referred to observing imaginative processes as a marker of engagement while interacting, where practitioners mentioned: "explicit use of imagination" or "using your imagination" which could be related to an equivalent of "self-

setting” goals as proposed by Renninger and Bachrach (2015) when studying triggers behind interest and engagement. These findings seem to reflect that the definitions used in chapter 5 for the engagement codebook encompass a broader categorisation of relevant engagement markers.

One of the key strengths of this study was the development of the slider tool which is the first study to capture the nuanced variation of practitioners’ perceptions of different levels of engagement, which could become a solid starting point for future tools aimed at capturing time-related process in education, particularly in informal learning contexts. The findings here emphasized the need for practical accessible methods that capture the continuity behind engagement. The slider tool has the potential to be used in practice as a screening device that allows practitioners to identify key features in a series of recorded exhibit interactions and increase understanding of visitors’ engagement. This tool could also be used during the design and evaluation of science exhibits or experiences as a first evaluation that could identify specific moments that appear beneficial or problematic, which then could be analysed in depth by the use of the engagement codebook developed for the previous study.

Also, the additional benefit of being an open-sourced interface (Revueltas Roux & McClure, 2020) that can be built with components available online is that it allows input and room for improvement from a wider community. This has the potential of being adapted into other useful tools, for example integrating live-video from a remote connection to a head-mounted camera; or to be used in different scenarios, for example following children’s day visit in an informal learning scenario. Thus, the slider tool could benefit practitioners’ training, highlighting the hues underlying the process of engagement, and refining practitioners’ perceptions of engagement or any other continuous unidimensional variable (e.g. learning).

This study has limitations, the key one being the number of participants and the number of videos used, which added more variability to the data, potentially obscuring patterns that could be more pronounced with a higher pool of participants and with a larger

sample of videos. However, a smaller number of participants and videos permitted evaluating the feasibility of the bespoke slider tool while also exploring more in-depth the variability in the results obtained from it, similar to a focus group. Moreover, analysis of the dynamic time-series could have revealed unobserved statistical differences (or lack of) in the patterns of interaction that could not have been fully appreciated with just the visualization, Future research should consider comparing the dynamic results from the practitioners throughout the interaction without averaging it. It is also recommended to recruit a higher number of participants and ideally an equal number from different institutions, to evaluate whether their background institution influences their understanding and scoring of engagement.

The observed differences between the practitioners' definitions and their scores of engagement, both in dynamic and discrete scores, lead to the conclusion that the use of a single summative score is not ideal for capturing nuances around engagement. There is a need to develop and refine methods that can be used by practitioners to address time-cost limitations found in current methods like observation-in-the-wild or a unique reliance on summative scores. Therefore tools that can accessibly capture and measure intensity variation as part of the process of engagement could improve current museum practice when it comes to internal research, for enhancing training and facilitation, as well as enriching science exhibit design.

Chapter 6

General Discussion

6.1 Overview

Engagement has been shown to influence and enhance science learning for children (Dancstep née Dancu et al., 2015; Halliday et al., 2018; E. Wood & Wolf, 2008), and is particularly valuable for children in the early years to establish solid foundations for their future cognitive skills (Cabe Trundle, 2010; Falk, 2005; Shwe Hadani et al., 2018). During the early years, children have minimal if any contact with formal school learning, which makes informal learning contexts ideal places for them to connect with science. However, researching engagement with science learning has not been widely studied with young children in science centres (Fu et al., 2016). One of the reasons behind this scarcity could be because the conceptualization of engagement itself presents a lack of agreement upon its definition, even when narrowed to the learning sciences field (Boekaerts, 2016; Christenson & Reschly, 2012; E. Wood & Wolf, 2008). This lack of agreement also hinders how engagement is measured, which consequently limits both our understanding of the concept and our ability to evaluate how different pedagogical and design interventions can enhance children's engagement (Boekaerts, 2016). This thesis addressed this challenge by investigating engagement as a multi-component construct where its cognitive, emotional and behavioural components are at a constant interplay, driving what is overtly perceived as engagement as proposed by Fredricks et al. (2004). This thesis also expanded the definition of engagement to consider the situated context and the intended learning outcome of the activity to be examined. In the particular context of interacting in a science centre, this research proposes that engagement is defined as a process that results from an active interplay of interest, curiosity, sustained attention, situated thinking and emotional attachment that allows an individual to ground themselves physically, mentally and

emotionally during the time and space of an activity. Moreover, after examining some of the differences in the literature behind the term engagement this thesis also proposed a dual continuum underlying the functionality and conceptualization of engagement: an intensity continuum to capture subtle hues regarding how a learner moves between engagement and disengagement, and a temporality continuum, where engagement can be examined as the process elicited throughout the duration of an activity or as a resulting outcome from an activity. Either the process or the outcome of engagement could also be examined with a short-term lens for engagement, for example, whether a particular object in a science exhibit enhanced engagement during an interaction, or a long-term focus where engagement is measured as a result of a longitudinal activity, such as engagement of visitors over a day visit in a science centre.

This thesis presents a body of work that thoroughly explored the process of engagement in early years children while interacting at a science centre. Four studies were carried out to increase understanding of methods that could improve the capture and measurement of engagement in early years children whilst they interact with science exhibits. Each of these studies helped to address the main research question of this research project: How could a triangulation multimodal approach capture engagement for early science learning at a science centre exhibit? Study 1 observed children's natural behaviour when interacting with science exhibits and showed that younger children gravitated more and for longer periods to an open-ended exhibit when compared to a planned-discovery exhibit. Study 2 demonstrated the feasibility of using current and new tools to capture young children's engagement during interaction with a science exhibit. The study showed it was feasible for early years children to comfortably wear a head-mounted camera and an electrodermal sensor as they freely interacted with a science exhibit, neither of which obstructed their play. This study also showed it was feasible to combine the sensors with video to capture multiple sources of data on children's interaction which were then synchronised to portray engagement from different angles. The key contribution of this study was demonstrating it was feasible to triangulate data from all the tools capturing

engagement in early years' children while they interacted in a science centre context. Findings from Study 3 showed there is a relationship between the outputs from tools measuring the different components of engagement, particularly between tools that measured trace data throughout the interaction. Findings also revealed that cognitive-behavioural markers, such as using strategies and self-setting goals, are more likely to elicit an emotional arousal response (i.e. skin conductivity) during interaction, particularly the more that children use these cognitive-behavioural markers. Study 3 also demonstrated that capturing engagement as both a process (using data reflecting the temporality of events) and an outcome (using scores summarising the events of an interaction) provided a more complete overview of what factors could be influencing engagement and how these interrelate, particularly when compared to representations obtained from an individual use of each of the tools. For example, the findings showed that overall, emotional facial and vocal expressions tend to appear more frequently and intensely when children looked away from the exhibit and looked at one of the adults present. On the other hand, this study also showed that the use of different strategies and goals throughout the interaction (i.e. engagement as a process) influenced children's emotional response (i.e. EDA). Finally, Study 4 showed variability in how expert practitioners perceived different levels of engagement classified by the number of strategies/goals, particularly when they judged engagement dynamically using the slider tool. Although this study did not reveal any differences between practitioners' judgements when scoring engagement dynamically or discretely, it still showed differences in the variability and nuance of judging engagement when practitioners rated dynamically as compared to discretely. Also, practitioners' scores showed more agreement in their continuous dynamic score than in their discrete assigned score. Practitioners' perceptions aligned when engagement was classified by the triangulation model as high, this did not stand for videos classified as moderate and low, demonstrating that scoring of engagement cannot be identified solely on subjective judgement, even if these come from experts. These results highlighted the need for tools

that can capture less evident levels of engagement in order to support engagement in those learners that might not be as evident.

6.2 Capturing Engagement

Understanding engagement in depth, whether it is conceptualised as a natural single occurring phenomenon, as a composition of independent phenomena that are often observed together, or somewhere in the middle, requires addressing first what is known of engagement and the event, and what is observed or captured from an event. A triangulation method of multimodal tools for capturing and measuring engagement was theoretically proposed by Azevedo (2015) which presented a promising solution to examine what is known of engagement more thoroughly. Aligning different data sources that are commonly used to capture separate cognitive, behavioural and emotional processes can potentially capture some of the causal mechanisms behind engagement by conceptualising engagement as a process, while it can also provide alternative explanations of what is being observed and captured.

This thesis captured engagement using this triangulation method for the first time in a science centre context with early years children. It also employed two key features : it considered the multi-components of engagement and it captured engagement *during* an interaction. Addressing the multi-component of engagement provided a good rationale for attempting to use tools that can capture each of these components separately. This meant reaching out for new tools that are not as commonly used in the science learning field, but that could be more sensitive to illustrate the intrinsic nature of each component and capture data that cannot be typically detected through observation. Exploring a possible underlying continuum of engagement, justified the value of using a method that captured the temporality of the process of engagement with trace data when using a science exhibit. Significantly, this work also emphasized the evaluation of this approach with early years children, which is an under-explored population in science learning research in this context due to methodological

challenges (Yarosh et al., 2011), despite evidence of the importance of improving science engagement and learning at this crucial developmental period (McClure et al., 2017; Tisdal & Perry, 2004). The findings from this thesis have also demonstrated the value of implementing a triangulation of multimodal tools capturing the different components of engagement as a method that more comprehensively represents the interplay between different factors influencing engagement, as theoretically proposed by Azevedo (2015). Although similar attempts have been tested before focusing on self-regulated learning when playing a science videogame on a screen (Taub et al., 2017, 2018), these relied on adult participation and more costly equipment like eye-trackers. This thesis adapted the proposed method using tools that are more accessible and affordable for researchers and museum practitioners to use for evaluation such as a head-mounted camera or an EDA sensor. This method also proved to be feasible for use in the early years since, a key developmental age when it comes to science learning (McClure et al., 2017; Shwe Hadani et al., 2018). Since this method relies more on children's behaviour rather than their verbal communication it reveals a potential for a more in-depth understanding of how children choose to engage with particular science learning activities or which factors are relevant for a specific learning goal (scientific skills or activity features).

The research presented here evaluated a novel method for capturing engagement in an understudied group (i.e. early years children) while they interacted in a science centre context. It particularly highlights the benefits of considering engagement as a temporal continuum to define the research design, the tools and the analysis of the outputs obtained. In this thesis, this continuum of engagement was observed throughout Chapters 4 and 5 using two different techniques: the triangulation method of different tools and the slider tool. Both these studies highlighted the existing differences that arise when the conceptualisation of engagement changes from a static single measure to a dynamic and continuous evaluation of the process of engagement. Focusing on the process of engagement revealed more information about how the different cognitive, behavioural and emotional markers of engagement could be interaction, driving children's different steps through their interaction

with an exhibit, but also showed practitioners' judgement of the perceived engagement of an exhibit interaction changed depending on whether they evaluated engagement as an outcome or as a process. Thus, it encapsulates the advantages of considering both the process and the outcome of engagement for a more rounded understanding of the observed engagement.

One of the main aims of this thesis was to evaluate existing tools that were used in common science centre/museum practice – such as time tracking, observation, behavioural coding, self-report measures and the use of engagement scales; as well as tools that already exist in other fields, but which are new to the science centre research context – electrodermal activity sensor, and head-mounted camera. Evaluation of these tools meant examining their strengths, their limitations, the unique contribution each tool could potentially make and whether they would capture nuances in engagement, capture different information between them or whether they would provide sufficient information about children's engagement when working with a) early year's children, b) uncontrolled interaction of a child with a science exhibit. Table 6.1 shows all the tools used with each of their strengths, limitations and unique contribution each of these can provide when capturing engagement with children. While there is still work to be done with each of the independent tools when using them for research and adapting them for a triangulation model, the findings from this research project are a good starting point for broadening the use of combined tools for capturing a construct from different perspectives in-the-wild.

Table 6.1

Evaluation of Tools to Capture Engagement in Children

Tool	Strength	Limitation	Unique Insight into Engagement
<i>Self-report Measures</i>	-Easy and fast to administer.	- Young children might not be able to fully articulate or be self-	-Window into children's self-perception of their

	<ul style="list-style-type: none"> -Children can highlight what they considered important in their experience. -Visual Analogues for Emotional Report reduce children's cognitive load and agreeability bias - It can be modified according to the research aims. -Multiple ways to administer, child can do it on their own, with a parent or the researcher next to them. -Does not require additional training to administer 	<ul style="list-style-type: none"> aware of their own engagement. -Heavily relies on children's verbal skills - Smiley-o-meter might not adequately capture children's appraisal of specific statements. -Relies on children's memory of their experience and is susceptible to recency effect -Prone to agreeability bias with researcher 	<ul style="list-style-type: none"> engagement, learning and enjoyment -Open questions give room for children to explain some of their thoughts, feelings and actions during their interaction.
<i>Time-tracking</i>	<ul style="list-style-type: none"> -Simple and easy to use -Used along affordable tools (i.e. chronometer) -It is malleable to suit different research designs (e.g. follow complete visit of 1 participant, or track multiple visitors in a single exhibit) -Non-obtrusive -Good for screening engagement 	<ul style="list-style-type: none"> -Limited input in underlying behaviours of time spent in an exhibit -Easy to jump to assumptions behind the use of time spent in an exhibit -Requires supporting details to understand what was happening throughout the learning experience. -Hard to "re-do" if there is any mistake throughout 	<ul style="list-style-type: none"> -Shows screening results about attraction (i.e. number of times someone visited something) and holding power (i.e. length of time spent) of exhibits
<i>Engagement Codebook</i>	<ul style="list-style-type: none"> -Multiple cognitive skills and behaviours related to engagement that can be tracked 	<ul style="list-style-type: none"> -Time consuming -Requires video recording to analyse more in-depth data. 	<ul style="list-style-type: none"> -Comprehensive set of cognitive and behavioural markers of engagement that offer rich details to

	<ul style="list-style-type: none"> -Specific definitions and coding instructions of the different behaviours. -Allows for simultaneous tracking of behaviours. -Observed Markers can be chosen according to research design -When used with video-coding software can provide in-depth rich data from milliseconds to minutes -Can be used independently 	<ul style="list-style-type: none"> -Video-coding software enhances accuracy 	<ul style="list-style-type: none"> ground other trace data.
<i>Fixed Video recording</i>	<ul style="list-style-type: none"> -Provides detailed information about what the child is doing throughout the interaction from a third-person perspective -Shows what else is happening around the child and their interaction -If a microphone is incorporated can record detailed audio from the child during the interaction -Output can be seen many times 	<ul style="list-style-type: none"> - Prone to interference if other people are around -View is limited and can be obstructed when participant moves or changes position. - Audio is not always reliable and might be dependent on the distance from the camera 	<ul style="list-style-type: none"> - Provides a visual overview of both the participant as well as their environment.
<i>Head-Mounted Camera</i>	<ul style="list-style-type: none"> -Provides first-person perspective of how an activity is experienced by the participant -Shows participant's field of vision and when angled also tracks the participant's hands. -Insight into participant's visual attention patterns 	<ul style="list-style-type: none"> - Not suitable for detailed information of visual cognition on specific scenes -Cannot track eye-movements -If head-mount is not secure or hair is long, it can slip off the head or visual angle can be modified or interfered. 	<ul style="list-style-type: none"> -Provides first-person perspective of how the child is experiencing an activity

	<ul style="list-style-type: none"> -Takes advantage of head-eye reflex movement for attention in large spaces -Children have not fully developed control of operating eye-movements separate from head-movement. -Can be used in any context and on its own -Freedom of movement and action while using it -Affordable 	<ul style="list-style-type: none"> -Needs researcher to define how this camera will be used and whether additional measures will be incorporated to the video (e.g. field of vision grid). -Easy for children to remove 	
<p><i>Mobile Electrodermal Activity Sensor</i></p>	<ul style="list-style-type: none"> -Insight into unobserved details of psycho-physiological arousal -Captures emotional-arousal associated with decision-making -Slow-response waves pair better with uncontrolled behaviours that are longer lasting. -Captures response from the sympathetic nervous system -Freedom of movement and action -Not fixed to particular context 	<ul style="list-style-type: none"> -Requires training to be analysed and interpreted -Requires special software to process the data -Cannot be used on its own -Does not distinguish between positive/negative stimuli -Prone to artefacts due to movement or other psychophysiological regulation processes -No standardised data or reference values for children -Expensive -Children can easily move it and uncalibrated it - Results are limited to what the sensor may (or may not) capture 	<ul style="list-style-type: none"> -E4 specifically has 4 sensors that measure sympathetic nervous activity. This can provide greater insight into the different psycho-physiological responses from participant to evaluate emotional arousal.

<p><i>Slider Tool</i></p>	<ul style="list-style-type: none"> -Captures the dynamic process of engagement -Can capture an overall description of engagement -Uses common and affordable programming devices such as Arduino and Phidget -Open-access -Reproducible -Easily transportable -Can be used with any desired video -Clear data outputs 	<ul style="list-style-type: none"> -Requires Windows as OS. -Displaying of the slider needs improving -Requires python programming skills to make any changes to the interface design 	<p>A tool that can capture a single continuous measure of someone's perception of process of engagement throughout an interaction.</p>
---------------------------	---	--	--

In sum, the findings from this thesis cannot provide a verdict on whether engagement is a “real” construct, a meta-construct, or an umbrella term for different processes, however, the findings did show that having tools that can capture trace-data allows concepts to be explored more as processes which can point to a more comprehensive view of the observed phenomena. Whether what was observed was a “real” engagement, or a different latent process that arises during an interactive learning experience, there seems to be agreement between researchers and practitioners that there is an additional factor influencing learning that drives children to initiate, remain and finish any learning activity.

Furthermore, the results from this thesis also confirm it was feasible to use more complex methods for data capture within a challenging context (i.e. a science centre) by closely collaborating with practitioners within the context that can help plan and co-design the research process from the beginning. Whilst the triangulation method would not be a method recommended for screening diagnostic research in-the wild, rather it would be suggested as useful for in-depth summative evaluations and exhibit design because of its

potential to identify key features and behaviours related to engagement. Contrarily, for short screening or training on evaluation of engagement, the slider tool shows potential for development as a tool that captures nuances of engagement with a more accessible equipment and less time-consuming requirements. Improving science engagement in science facilitation, activities and exhibit design could enhance and tailor the overall science learning of children, which has been shown as relevant for impacting long-term on their interests, attitudes and aspirations towards science paths, (DeWitt et al., 2013).

Overall, this body of research provides strong evidence of the potential of triangulating different sources of data for understanding engagement in the early years while interacting at a science centre. It particularly highlights how multiple sources of data provide a more comprehensive portrayal of engagement compared to subjective judgements. Doing this has the potential to improve practice, both in facilitation and exhibit design, by understanding engagement processes and tailoring them specifically to different age groups. It also focuses on a key developmental age for science learning that is relevant for future contact with science content, for developing a positive attitude towards science and contributing towards aspirations of children with science careers.

6.3 Contributions

6.3.1 Methodological

This thesis contributes methodologically by demonstrating the feasibility of using a triangulation method that allows the thorough examination of a multicomponent construct, such as engagement, as proposed by Azevedo (2015). More importantly, this thesis contributes with the design, implementation, and interpretation of a multimodal triangulation approach that captures detailed markers of engagement in early years science learning and discusses the advantages and disadvantages of using this method. This meant the research

had to address three significant methodological challenges that when combined, increased the complexity for design, implementation and interpretation of results. First, the challenge of adapting, testing and evaluating a new methodology, as the triangulation of multimodal tools proposed, which involved using tools such as a head-mounted camera or an electrodermal sensor not commonly used in a science centre context nor with early years children. This challenge also involved adapting to work with a concept like engagement that encapsulates the interplay of its different components and with variables that could influence it.

Second, testing a new method while working with an early years population involves constraints surrounding the ethics and participation invitation, since it needed to account for children's verbal comprehension (Forrester, 2017). This also helps ensure their consent is real and not biased by noticing non-verbal cues and being aware of the dynamics that arise between children and the researcher (Einarsdóttir, 2007). Moreover, the interaction required careful planning to account for their shorter attention spans (Akshoomoff, 2002), as well as considering methods of facilitating gathering feedback from them while avoiding biasing children in a specific direction pre-set by the researcher (Yarosh et al., 2011). One of the advantages of the triangulation method to address this challenge is its limited reliance on children's verbal skills and more on children's non-verbal behaviour.

Thirdly, the challenge of researching in-the-wild scenarios (i.e. the science centre), in contrast to a controlled context like a designated space in a laboratory, since it increases the complexity by acknowledging there will be many uncontrolled confounding variables which may influence engagement (Chamberlain et al., 2012; Chamberlain & Crabtree, 2020). For example, in a science centre some of these confounding variables could be multiple visitors interfering with each other in the exhibit space; visitors who have paid to visit and expect to use all the exhibits in the space; noise levels higher than in a laboratory which can both influence participation and prevent some data to be captured; or controlling when video recording that other visitors who are not participating in the study do not get recorded, particularly if they are children.

Previous literature had only addressed either one of these challenges, for example: Lee-Cultura et al. (2021) evaluated the use of sensors with children; Ben-Eliyahu et al. (2018) evaluated science learning in-the-wild in science centres and schools; and Halliday et al. (2018) evaluating engagement in preschool age children, or two of these challenges, such as the research done by Sridhar et al. (2018) where they also triangulated behavioural coding and psychophysiological tools in kindergarten children - this, however, was done researching mental effort in a quiet space. Therefore, this thesis is the first research to embrace the complexity of these three methodological challenges while researching engagement.

Furthermore, this thesis also presents feasible tools that can be used with a very young population while interacting in an informal science learning scenario. It also contributed by presenting a method to extract the co-occurrences of events that represent the temporality of the process of engagement. Along with this, it contributes methodologically by using non-specific skin conductance responses (NS-SCR) as markers that can provide a proxy measure for emotional engagement, tracking how emotional arousal occurs in contexts where there are no pre-design events throughout an activity, facilitating the identification of these. Finally, this thesis demonstrates two unique tools that can help guide future research: a codebook for tracking engaging behaviours and an open-source slider tool that captured the dynamic rating of recorded interactions. The engagement codebook can be used by researchers and practitioners who need an in-depth exploration of children's nuanced behavioural patterns associated with engagement during an interaction with a science learning activity. This codebook builds on previous engagement scales such as eSOCS (Deater-Deckard et al., 2011) and LECS (Halliday et al., 2018) that consider the multicomponent of engagement and behaviours needed for the early years, but it adds categories to track higher-order cognitive skills such as goal-setting and strategy use that are not context or activity-dependent. An in-depth exploration of these engaging behaviours offers value as a guide to improve engagement while designing science exhibits and activities. On the other hand, the slider tool is based on the combination of previous tools

that measure the self-reported intensity of a process in a screen (e.g. an emotion) or the self-reported intensity of an experience during a live event (e.g. enjoyment of a movie). The novelty of the slider tool, however, is that it combines the capacity to rate intensity of a process with the capability of an interface that allows the synchronisation of the score to a desired video. This makes it a tool that can capture a continuous measure of the intensity of a unidimensional variable, in this case engagement. This tool is unique because it is a Python-based open-source interface that uses affordable components that only need to be assembled and that can be adapted by other researchers to fit in their design with other dynamic unidimensional variables. Using this slider tool in a science centre can help identify key moments of engagement during a recorded interaction that may need detailed exploration and can be complemented with the engagement codebook or tested in depth with the triangulation method.

6.3.2 Theoretical

This thesis contributes to the development of a theory regarding learning engagement as called upon by Boekaerts (2016) by proposing a conceptualization of engagement as a continuum. This continuum should consider the intensity and the temporality of an event, seeing engagement as both a process and an outcome, that can also be determined by the short or long-term focus of the research. Sinatra and colleagues (2015) had already proposed engagement as a continuum of engagement, but their model depended on the grain size of unit of analysis, ranging from person-oriented to context-oriented. This means engagement would be different when focusing on what influences a person when interacting with an activity, than understanding engagement when considering a context, such as a science centre. The continuum presented here complements Sinatra's model by adding a temporality and an intensity axis to each of the units of analysis, which can help researchers track whether engagement is being measured as a process/outcome in a short/long-term view.

Findings also support the conceptualization of engagement as a process since most of the literature has focused on engagement as an outcome (as discussed in section 2.2.5), which could be reflecting the tools available to capture engagement as such. Thus, an increased availability of tools that can capture the dynamic process of engagement, as shown in this thesis, may help to reframe and reconceptualise engagement as a process as well as an outcome. Furthermore, evidence from the countless cognitive-emotional markers (i.e. NS-SCRs) found in Study 4 and from the dynamic variation of engagement perceived from practitioners in Study 5, supports the model observed by O'Brien and Toms (2008). This model proposed that during a cycle of engagement there would be multiple points of engagement, all of which would have a similar starting point, and at some point, might finish naturally or abruptly. In the findings from this research this interruption was observed by the attention-off-task events, and it was also observed that engagement with the interaction was also recaptured on multiple occasions. Therefore, this study strengthened the theory of engagement as a dynamic process that can be found and captured outside of user experience with digital platforms. Also providing evidence that this cycle of engagement can also be found in children's engagement experiences with science learning.

The findings of this research also provide evidence to the framework established by Fredricks et al. (2004) and Tisdal and Perry (2004) that considers engagement as a multi-component construct where the interplay of its emotional, behavioural and cognitive components need to be examined to increase understanding of the overall result. Ben-Eliyahu et al. (2018) also evidenced that according to their structural equation model, engagement in science learning is a construct that is more than just the sum of the outputs from each component.

This thesis also contributes to the field of informal science learning by increasing understanding of engagement during science learning in a science centre for the early years, which is a construct that has not been as studied as much as other age groups in childhood (Fu et al., 2016) or as in other learning contexts, such as formal student engagement

(Mahatmya, Lohman, Matjasko, & Feldman Farb, 2012). Specifically, this thesis showed that early years children were more inclined towards and interacted for longer with an open-ended exhibit compared to a planned-discovery one which echoes results observed by Gutwill (2008) when examining counter-intuitive and open-ended exhibits at the Exploratorium in San Francisco. It also supports views about open-ended exhibits fostering more engagement as found by Tisdal and Perry (2004) who found that exhibits that promote what they called active prolonged engagement (APE) were more open-ended and allowed visitors to explore more freely. It also showed a connection between children's age where younger children used less variety in the number of different higher-order cognitive functions such as actions, strategies and goals, compared to older children, however, these behaviours in older children had a shorter duration than those done by younger children. This mirrors arguments from Pressley and Hilden (2006) where they mention that children could discover new strategies performing a task although they may be deemed ineffective because they have not been practised and the outcome is unknown, which potentially also explains the lower duration of these. Thus, as children get older, they could test more strategies. Finally, the findings also evidence a support role parents could play with children in the early years even when children are interacting individually. Since children looked for parents during interaction, facilitation from the adult to the child can support their understanding according to Van Schijndel et al.(2010). They argue that at this early age, facilitation could be shaped by prompting the child by asking them questions about the activities, pointing them to key actions or events, encouraging children to ask questions and providing them with guiding answers. This would drive children into making their own predictions of what could happen before/after some key event, modelling their actions in advance or helping them in see some causal relationship between certain. Research has shown that when teaching parents how to guide their children's interaction significantly benefitted children's high-level exploratory behaviour, related to scientific inquiry (Van Schijndel et al., 2010).

6.4 Implications and Future Directions

6.4.1 For Research

The methodology tested here responds to a call from a review of museums' last decade on research by Andre et al. (2017) on how to support learning, particularly for preschool-aged children in museums, a population which has not been as studied. The evidence for triangulating information from tools that capture the different components of engagement presented here has significant implications for studying other complex constructs that require multiple points of data and less reliance on verbal communication. One concept strongly related to engagement that can benefit from this approach would be learning, which has been recognised to be both an interplay of different factors and to have a duality of conceptualization as process/outcome (Falk, 2005; Falk & Dierking, 2000). That is, this method could better capture the social, affective and cognitive dimensions associated with learning while also detecting additional variables of interest such as curiosity. There is evidence of attempting the use of this triangulation method with self-regulated learning of adults as used by Taub et al.(2017), as well as triangulating psychophysiological markers for learning using EEG and eye-tracking during a learning game on a screen (Cowley & Ravaja, 2014). Therefore, an implication of this research is the possibility of incorporating additional mobile psychophysiological tools to the triangulation method, such as accelerometers, heart-rate monitors or EEG, to the study of both engagement and learning in-the-wild. This would also provide additional psychophysiological markers which would be of assistance in evaluating the relationship with video-coded cognitive-behavioural markers as was done with EDA and other emotional expressions or cognitive-behavioural markers. Using a triangulation of an EDA sensor with cognitive-behavioural markers of engagement provided additional evidence with respect to the use of emotional arousal observed by skin conductivity (i.e. phasic and tonic activity) as one of the possible markers for emotional engagement in early years children interacting in-the-wild. While EDA sensors had been

used before to measure either emotional arousal in adults interacting with museums (Mitas et al., 2020); emotional arousal of children in different controlled scenarios (Hedman, 2008), cognitive effort in children (Nourbakhsh et al., 2012; Sridhar et al., 2018), or the more recent use for capturing engagement in older children interacting with an augmented reality learning activity (Lee-Cultura et al., 2021); none of these had used EDA sensors to explore emotional arousal as a dimension of emotional engagement with young children in a science centre context. Findings showed that using an EDA sensor provides an informative account of emotional arousal peaks that could be the results of specific explorative behaviours done, however, NS-SCRs are also prone to artefacts, and EDA can be influenced by additional factors beyond decision-making (such as temperature regulation). Thus, using an EDA sensor is not enough, but it can be complemented by additional layers of psychophysiological markers that are also captured by the E4 sensors. Most importantly, the findings from this body of work help establish a methodological foundation for the integration of tools that can capture complementary processes from internal and external sources in a real-world context that could account for some of the varied factors normally encountered in this type of context (Chamberlain & Crabtree, 2020).

Future work is needed to fully understand the implications of conceptualising engagement as a process and as an outcome when designing research. There are three main suggestions for further research to improve the triangulation multimodal methodology around engagement: Firstly, more exploration of engagement using this method with early years children using: a) different science exhibits, since research has shown that interaction can be different depending on different types of exhibits(Allen, 2004; Dancstep née Dancu et al., 2015; Gutwill, 2008); b) interacting in different informal science learning scenarios such as zoos, aquariums or art museums, since behavioural engagement might be shaped differently due to the intrinsic nature of each of these (National Research Council, 2009; Schwan et al., 2014); c) adding a social component to their exhibit interaction, whether this is interacting with parents, a science facilitator, or peers, since research has also pointed to the

relevance that social interaction adds to a museum visit (Allen & Gutwill, 2004; Falk et al., 2004; Shaby et al., 2017). Secondly, the use of this methodology with older children and teenagers while exploring a science centre scenario would be particularly valuable to identify and extend our knowledge on any developmental patterns of engagement with science exhibits. Thirdly, adapting the triangulation method and/or the slider tool to use in complete museum visits and not restricted only to a single science exhibit. For example, incorporating the use of the head-mounted camera and an E4 sensor to record the process of engagement with the exhibits in both short and long-term throughout the visit. This video could then be evaluated using the engagement codebook with limited behaviours or the slider tool could help screen for key moments of engagement throughout the visit. These suggested studies would shed more light on the different aspects of the continuum of engagement that could be influencing a child's visit to the science centre.

6.4.2 For Theory

This thesis builds up the theoretical definition of engagement by illustrating the conceptualization of engagement as a multicomponent process during learning experiences as much as an outcome of them. It expands the conceptualization of engagement to consider how its temporality can accommodate the different, and sometimes conflicting, definitions of engagement (Azevedo, 2015; Boekaerts, 2016), to be understood with a more nuanced framework in different learning contexts. The continuum proposed here, paired with the continuum that considers grain-size of unit of analysis by Sinatra et. al (2015), raise important theoretical issues that bear on the understanding of the construct of engagement from multiple perspectives. A theoretical implication if both of these continuums were to be considered, would open new paths for researchers to explore and explain underlying processes of engagement considering both the grain-size of a learning activity and the temporality in which this is happening. Further research is needed to better understand the extent to which these continuums of engagement can accommodate the complex interaction between engagement and the different contexts where it occurs. This means a better

mapping of the temporal, intensity and person-context positioning of different states of engagement when it comes to science learning experiences. When working with a complex construct like engagement or learning, having more information regarding the captured data can facilitate the delimitation of some depictions of engagement. By doing this, there could be an implication for reducing the notion of engagement as an umbrella term, since clearer definitions bring with them better accuracy in measuring and representing of a complex idea.

Moreover, understanding the temporality of engagement combined with its multicomponents could prove to be a key factor to understand how engagement influences learning experiences and prompt future research into exploring how to influence and modulate engagement in science learning. This can be particularly advantageous for underserved populations with limited access to some informal science learning experiences, because it can shape their future engagement with science as shown by research from Essex and Haxton (2018) who researched different levels of engagement taking into account participants with different levels of science capital. This potential improvement in science learning experiences can also lead to increasing their individual and community science capital (DeWitt et al., 2013, 2016).

One unexpected additional result from this thesis hinted at the prominent role children give to parents, even if just for emotional support throughout an individual interaction and the positive reaction that doing this elicited in children. This finding builds on evidence from Van Schijndel et al.(2010) regarding the prominent role adults could have in supporting children's exploratory behaviours. However, what creates interesting questions for future work is that children searched for parents' support even during an engaging interaction. This could guide future studies that examine whether children can identify how to support their exploration with science learning by including their parents, or if this characteristic would be attributable to a particular parenting style and typical interaction during learning experiences.

Future studies of the triangulation method could integrate the role of social interaction to increase understanding of how and to what extent it impacts the interplay of the individual components of engagement. This would provide a comprehensive assessment of engagement in a science centre that facilitates comparison of how different social agents can influence and shape engagement throughout an interaction with a science experience.

6.4.3 For Practice

The evidence from this thesis can support how practitioners capture and measure engagement. Since practitioners need to evaluate engagement to improve their experiences and exhibit designs, developing methods that improve a nuanced representation of engagement could contribute to enhancing these experiences. Firstly, the feasibility of use of a head-mounted camera and a wearable sensor that can capture up to four different measures (EDA, HR, accelerometer and temperature) demonstrated here for early years children during a hands-on interaction with a science exhibit. The use of wearable devices for practice that are becoming easier to use and analyse opens the possibility of increasing richer multifaceted explorations of complex constructs, like engagement, in real-world contexts. The triangulation method can be used in summative evaluations, contributing to detailed descriptions of a specific exhibit or activity, as it has been needed for research like the *APE! project* looking to characterise features that promote engagement (Tisdal, 2004; Tisdal & Perry, 2004). This method could also unveil potential causal-mechanisms for engagement and science learning, which could lead to evidence-based changes, whether in facilitation or exhibition design. This would address one of the main needs raised in the review of designs and methods in *Informal Science Education Evaluation* by Fu et al. (2016) to “develop better measures, particularly those that are direct measures of outcomes, embedded in the informal experience” (p.33), where they also ask researchers to go beyond descriptive studies and address causal questions. Therefore, the findings from this research address the need by capturing direct outcomes within the science centre context from the

participants and using markers of emotional arousal (i.e. proxy for emotional engagement) as the leading measure to guide cognitive-behavioural markers.

However, this in-depth method would not be suggested for smaller evaluations where the focus is to provide a quick descriptive overview of how a visitor interacts with a specific activity or throughout a visit. Particularly since findings from Chapter 6 showed great variability in the subjective scoring of children's lower levels of engagement by expert practitioners. Instead, it is suggested that the bespoke slider-tool is used, which is a new hand-held device that can capture the dynamic temporality and variability of intensity of engagement with less time-constraints. The slider tool can be very useful in practice since it can serve as a screening instrument that helps identify how and why engagement is varying during a recorded interaction or that can be used in practice for training other staff to detect dynamic engagement in visitors. Further research should focus on the improvement of the hand-held slider tool and assess whether it is possible to combine with recording from a head-mounted camera an electrodermal sensor to measure engagement during a live-interaction. This could provide practitioners with a device that can monitor the first-perspective of the child's, or the visitor's, point of view and triangulate with emotional arousal along with the practitioner's judgment of engagement peaks. Doing this can help tackle the additional constraint of doing controlled or recorded research in science centres, resulting in more ecologically valid results and better understanding of the real-world context.

Although in order to achieve this, communication between researchers and practitioners is key to understanding how best these findings can be put into practice (Ghazai-Mohammed & Meikleham, 2020). A better communication about this body of work would also entail greater knowledge regarding engagement, the importance of considering it a process and also gaining a perspective from practitioners around the practicalities of the methods proposed. The potential benefits of this research could lead to improving and tailoring informal learning experiences in younger children, which can lead to enriching their overall future engagement with science ideas (DeWitt et al., 2013; Essex & Haxton, 2018).

6.5 Limitations

While this thesis provided greater insight into engagement by examining it thoroughly from multiple perspectives, demonstrating the value of the triangulation method approach used in science learning scenarios, some limitations need to be acknowledged. First, although this research method can provide rich detailed information about some of the markers behind engagement, the process needed to obtain that information is costly in time. This type of limitation echoes criticism made towards research that involves video-observation (Grack Nelson & Cohn, 2015; Rennie & McClafferty, 1996).

A second limitation regarding the triangulation method is limiting the psychophysiological marker to EDA only and not the additional 3 sensors, temperature, accelerometer and blood volume). Moreover, the ease of use of the E4 sensor, which although easy to use for recording data, entails a higher-level of expertise in the processing and analysis than the marketing proposes, along with a lack of straightforward directions from the sensor's providers on what type of software or parameters might work best for the analysis. This could limit the future use of the device in less controlled scenarios, or with researchers with less expertise in psychophysiological devices.

Thirdly, since this thesis project grounded all the results upon a single exhibit, this might limit the generalizability of the obtained results. Although the selected hands-on open-ended exhibit showed similar results regarding longer interactions when compared to a planned-discovery exhibit as argued previously by Gutwill (2008), as well as finding the chosen exhibit promoted self-guided exploration and inquiry as found in previous research for this type of exhibit (Dancstep née Dancu et al., 2015; Tisdal & Perry, 2004).

A fourth limitation questions the validity of the set-up used within the science centre as being truly "in-context", it is acknowledged that the context it lacked the availability of other exhibits competing for attention that guides science centre visitors' free-choice learning, although the feel of the set-up was controlled by it being on the museum floor with

semi-closed walls as some exhibitions have. However, since this triangulation method had not been done before with early years population and within a science centre, it was necessary to first explore and set a precedent for what interaction in young children would look like in a science centre when there are no competing stimuli and children were able to interact individually and without disruptions. A future study could use these findings as a base and follow a child on a visit to a science centre using the head-mounted camera or while interacting freely with an exhibit but with fewer restrictions in time and allowing other visitors to also interact with the exhibit.

This leads to the final limitation of not examining any social interaction during the encounter with the science exhibit, which has been argued to be a key factor influencing children's engagement in a science exhibit (Allen & Gutwill, 2004; Falk & Dierking, 2000; Gutwill, 2008; Schwan et al., 2014; Shaby et al., 2017). Since the main aim was to first understand children's individual engagement, this would be out of the scope of the main research, however, the most important next step would be to research how the influence of peers or parents interacting along with the child could shape engagement and whether the cognitive-emotional markers found in this thesis would show a similar relationship when the interaction is not individual. Therefore, future directions of this research would lead to an in-depth inspection of social influence over engagement using the triangulation multimodal method.

6.6 Conclusions

The aim of this doctoral thesis was to examine how and to what extent the process of engagement can be captured by a triangulation multimodal method for children in the early years interacting at a science exhibit. This thesis confirms the possibility proposed, but not tested, by Azevedo (2015) of capturing engagement by using a triangulation approach and builds on this work by extending it into a real-world context, a science centre, with a different population, the early years. The findings from this body of work indicate that engagement

can be captured both as a process and as an outcome through the triangulation of different tools measuring markers from the cognitive, emotional and behavioural components of engagement. Evidence from this research shows that this method can capture the process of engagement by considering the temporal co-occurrence of cognitive-behavioural events with a lagged response from the emotional arousal component. The obtained results from the triangulation method showed that engagement in children in the early years is more likely to be driven by the emotional arousal caused by inquiry exploratory processes. The triangulation method allowed a capture of not only the interplay between the three components of engagement but also indicated some underlying relationships surrounding children's engagement and their interaction with their parents. The results from this thesis also point to the importance of asking observers to conceptualise and capture engagement as a dynamic process, since it provided a more nuanced view of how engagement varies throughout an interaction. Although the triangulation method proposed is a time-consuming method to use in everyday research, it is a method that can yield strong results when in-depth detail is needed of how engagement is working in a particular context or activity. On the other hand, the slider-tool developed for this research shows potential as an immediate application to screen and identify triggers of engagement during an interaction.

Finally, given the importance of incorporating science learning during the early years to positively impact children's future outcomes, enhancing engagement in science learning experiences is critical. The results from this thesis broaden the understanding of engagement and provide an approach to capture engagement more efficiently in a science centre context that can be used in both research and practice. This body of work lays the groundwork for future research that can ultimately enrich science learning experiences through engagement and ultimately improve science learning during key influential stages, such as the early years.

References

- Adams, M., Luke, J. & Moussouri, T. (2004). Interactivity: Moving beyond terminology. *Curator: The Museum Journal*, 47(2), 155–170.
<https://doi.org/10.1111/j.2151-6952.2004.tb00115.x>
- Adolphs, R., Tranel, D. & Damasio, A. R. (2003). Dissociable neural systems for recognizing emotions. *Brain and Cognition*, 52(1), 61–69.
[https://doi.org/10.1016/S0278-2626\(03\)00009-5](https://doi.org/10.1016/S0278-2626(03)00009-5)
- Aguiar, C. & McWilliam, R. A. (2013). Consistency of toddler engagement across two settings. *Early Childhood Research Quarterly*, 28(1), 102–110.
<https://doi.org/10.1016/j.ecresq.2012.04.003>
- Akshoomoff, N. (2002). Selective Attention and Active Engagement in Young Children. *Developmental Neuropsychology*, 22(3), 625–642.
<https://doi.org/10.1207>
- Alderson-Day, B. & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, 141(5), 931–965. <https://doi.org/10.1037/bul0000021>
- Allen, S. (2004). Designs for learning: Studying science museum exhibits that do more than entertain. *Science Education*, 88(SUPPL. 1), S17–S33.
<https://doi.org/10.1002/sce.20016>
- Allen, S. & Gutwill, J. (2004). Designing with multiple interactives: five common pitfalls. *Curator: The Museum Journal*, 47(2), 199–212.
<https://doi.org/10.1111/j.2151-6952.2004.tb00117.x>
- Anderson, M. L. (2003). Embodied Cognition: A field guide. In *Artificial Intelligence* (Vol. 149, Issue 1, pp. 91–130). [https://doi.org/10.1016/S0004-3702\(03\)00054-7](https://doi.org/10.1016/S0004-3702(03)00054-7)
- Andre, L., Durksen, T. & Volman, M. L. (2017). Museums as avenues of learning for children: a decade of research. *Learning Environments Research*, 20(1), 47–76.
<https://doi.org/10.1007/s10984-016-9222-9>
- Ansari, D., De Smedt, B. & Grabner, R. H. (2012). Neuroeducation – A Critical Overview of An Emerging Field. *Neuroethics*, 5(2), 105–117.

<https://doi.org/10.1007/s12152-011-9119-3>

- Appleton, J. J., Christenson, S. L. & Furlong, M. J. (2008). Student Engagement with School: Critical Conceptual and Methodological Issues of the Construct. *Psychology in the Schools, 45*(5), 369–386. <https://doi.org/10.1002/pits.20303>
- Appleton, J. J., Christenson, S. L., Kim, D. & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology, 44*(5), 427–445. <https://doi.org/10.1016/J.JSP.2006.04.002>
- Archer, L., Dawson, E., DeWitt, J., Seakins, A. & Wong, B. (2015). “Science capital”: A conceptual, methodological, and empirical argument for extending bourdieusian notions of capital beyond the arts. *Journal of Research in Science Teaching, 52*(7), 922–948. <https://doi.org/10.1002/TEA.21227>
- Archer, M., Decoteau, C., Gorski, P., Little, D., Porpora, D., Rutzuo, T., Smith, C., Steinmetz, G. & Vandenberghe, F. (2016). *What is Critical Realism? - Theory Section*. Perspectives: A Newsletter of the ASA Theory Section. <http://www.asatheory.org/current-newsletter-online/what-is-critical-realism>
- ASDC. (n.d.). *The Association for Science and Discovery Centres*. Retrieved November 15, 2021, from <https://www.sciencecentres.org.uk/about/>
- Aslin, R. N. (2009). How Infants View Natural Scenes Gathered From a Head-Mounted Camera. *Optometry and Vision Science, 86*(6), 561–565. <https://doi.org/10.1097/OPX.0b013e3181a76e96>
- Azevedo, R. (2015). Defining and Measuring Engagement and Learning in Science: Conceptual, Theoretical, Methodological, and Analytical Issues. *Educational Psychologist, 50*(1), 84–94. <https://doi.org/10.1080/00461520.2015.1004069>
- Azevedo, R. (2018). Using hypermedia as a metacognitive tool for enhancing student learning? the role of self-regulated learning. In S. I. Educational Psychologist (Ed.), *Computers as Metacognitive Tools for Enhancing Learning: A Special Issue of Educational Psychologist* (Vol. 40, pp. 199–209). Taylor and Francis. <https://doi.org/10.4324/9781315866239-2/USING-HYPERMEDIA-METACOGNITIVE-TOOL-ENHANCING-STUDENT-LEARNING-ROLE-SELF->

REGULATED-LEARNING-ROGER-AZEVEDO

- Azevedo, R. & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia? *Journal of Educational Psychology, 96*(3), 523–535. <https://doi.org/10.1037/0022-0663.96.3.523>
- Bailey, C., Madden, A., Alfes, K. & Fletcher, L. (2017). The Meaning, Antecedents and Outcomes of Employee Engagement: A Narrative Synthesis. *International Journal of Management Reviews, 19*(1), 31–53. <https://doi.org/10.1111/ijmr.12077>
- Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology, 4*(3), 359–373.
- Baniyamin, N. & Rashid, M. M. (2016). Understanding Science Centre Engagement in Nurturing Visitor Interest and Curiosity. *Procedia - Social and Behavioral Sciences, 222*(222), 235–243. <https://doi.org/10.1016/j.sbspro.2016.05.152>
- Barrett, L. F. (2011). Constructing emotion. *Psychological Topics, 20*(3), 359–380.
- Barriault, C. & Pearson, D. (2010). Assessing exhibits for learning in science centers: A practical tool. *Visitor Studies, 13*(1), 90–106. <https://doi.org/10.1080/10645571003618824>
- Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bathgate, M. E., Schunn, C. D. & Correnti, R. (2014). Children's motivation toward science across contexts, manner of interaction, and topic. *Science Education, 98*(2), 189–215. <https://doi.org/10.1002/sce.21095>
- Bechara, A., Damasio, H., Tranel, D. & Damasio, A. R. (2005). The Iowa Gambling Task and the somatic marker hypothesis: Some questions and answers. *Trends in Cognitive Sciences, 9*(4), 159–162. <https://doi.org/10.1016/j.tics.2005.02.002>
- Bechara, Antoine & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior, 52*(2), 336–372. <https://doi.org/10.1016/j.geb.2004.06.010>

- Bechara, Antoine, Damasio, H. & Damasio, A. R. (2000). Emotion, Decision Making and the Orbitofrontal Cortex. *Cerebral Cortex*, 10(3), 295–307.
<https://doi.org/10.1093/cercor/10.3.295>
- Ben-Eliyahu, A., Moore, D., Dorph, R. & Schunn, C. D. (2018). Investigating the Multidimensionality of Engagement: Affective, Behavioral, and Cognitive Engagement Across Science Activities and Contexts. *Contemporary Educational Psychology*, 53(January), 87–105.
<https://doi.org/10.1016/j.cedpsych.2018.01.002>
- Benedek, M. & Kaernbach, C. (2010a). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods*, 190(1), 80–91.
<https://doi.org/10.1016/j.jneumeth.2010.04.028>
- Benedek, M. & Kaernbach, C. (2010b). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47(4), 647–658.
<https://doi.org/10.1111/J.1469-8986.2009.00972.X>
- Benedek, M. & Kaernbach, C. (2016). *Ledalab* (3.4.9). Free Software Foundation.
<http://www.ledalab.de/documentation.htm>
- Benjamin, N., Haden, C. A. & Wilkerson, E. (2010). Enhancing Building, Conversation, and Learning Through Caregiver-Child Interactions in a Children’s Museum. *Developmental Psychology*, 46(2), 502–515.
<https://doi.org/10.1037/a0017822>
- Benjamini, Y. (2010). Discovering the false discovery rate. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4), 405–416.
<https://doi.org/10.1111/J.1467-9868.2010.00746.X>
- Benjamini, Y. & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300.
<https://doi.org/10.1111/J.2517-6161.1995.TB02031.X>
- Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D. & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks.

Aviation, Space, and Environmental Medicine, 78(5), 231–244.

Bhaskar, R. & Hartwig, M. (2016). *Enlightened common sense: The philosophy of critical realism*. Routledge.

Blanchard, N., Bixler, R., Joyce, T. & D'Mello, S. (2014). Automated physiological-based detection of mind wandering during learning. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8474 LNCS, 55–60. https://doi.org/10.1007/978-3-319-07221-0_7

Boekaerts, M. (2016). Engagement as an inherent aspect of the learning process. *Learning and Instruction*, 43, 76–83.
<https://doi.org/10.1016/j.learninstruc.2016.02.001>

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H. & White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3), 127–135.
<https://doi.org/https://doi.org/10.1016/j.tree.2008.10.008>

Borghini, A. M. & Cimatti, F. (2010). Embodied cognition and beyond: Acting and sensing the body. *Neuropsychologia*, 48(3), 763–773.
<https://doi.org/10.1016/j.neuropsychologia.2009.10.029>

Börnert-Ringleb, M. & Wilbert, J. (2018). The Association of Strategy Use and Concrete-Operational Thinking in Primary School. *Frontiers in Education*, 3, 38.
<https://doi.org/10.3389/feduc.2018.00038>

Boucsein, W. (2012). *Electrodermal Activity* (2nd ed.). Springer US.
<https://doi.org/10.1007/978-1-4614-1126-0>

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E. & Fillion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, 49(8), 1017–1034.
<https://doi.org/10.1111/j.1469-8986.2012.01384.x>

Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., Weiner, D., Bakken, S., Kaplan, C. P., Squiers, L., Fabrizio, C. & Fernandez, M. (2009). How We Design Feasibility Studies. In *American Journal of Preventive Medicine*

(Vol. 36, Issue 5, pp. 452–457). Elsevier.
<https://doi.org/10.1016/j.amepre.2009.02.002>

Bowers, B., Brightful, D., Heflin, C., Hindley, A., Kiehl, K. L., Pruckno, E., Raso, C. & Wolfe, J. (2015). Museums Providing Opportunities for Promoting a Positive Sense of Self in the Early Years. *Museum and Society*, 13(2), 142–157.
<https://doi.org/10.29311/mas.v13i2.323>

Brieber, D., Nadal, M. & Leder, H. (2015). In the white cube: Museum context enhances the valuation and memory of art. *Acta Psychologica*, 154, 36–42.
<https://doi.org/10.1016/j.actpsy.2014.11.004>

Brinthaup, T. M., Hein, M. B. & Kramer, T. E. (2009). The self-talk scale: Development, factor analysis, and validation. *Journal of Personality Assessment*, 91(1), 82–92. <https://doi.org/10.1080/00223890802484498>

Brishtel, I., Khan, A. A., Schmidt, T., Dingler, T., Ishimaru, S. & Dengel, A. (2020). Mind Wandering In A Multimodal Reading Setting: Behavior Analysis & Automatic Detection Using Eye-Tracking And An Eda Sensor. *Sensors (Switzerland)*, 20(9), 2546. <https://doi.org/10.3390/s20092546>

Broaders, S. C., Cook, S. W., Mitchell, Z. & Goldin-Meadow, S. (2007). Making Children Gesture Brings Out Implicit Knowledge and Leads to Learning. *Journal of Experimental Psychology: General*, 136(4), 539–550.
<https://doi.org/10.1037/0096-3445.136.4.539>

Bruner, J. S. (1966). *Toward a theory of instruction* (Vol. 59). Harvard University Press.

Bryce, D. & Whitebread, D. (2012). The development of metacognitive skills: Evidence from observational analysis of young children's behavior during problem-solving. *Metacognition and Learning*, 7(3), 197–217.
<https://doi.org/10.1007/s11409-012-9091-2>

Buchanan, D. W. & Sobel, D. M. (2011). Mechanism-Based Causal Reasoning in Young Children. *Child Development*, 82(6), 2053–2066.
<https://doi.org/10.1111/j.1467-8624.2011.01646.x>

Buelow, M. T. & Suhr, J. A. (2009). Construct validity of the Iowa gambling task. In

- Neuropsychology Review* (Vol. 19, Issue 1, pp. 102–114). Springer.
<https://doi.org/10.1007/s11065-009-9083-4>
- Burris, A. (2017). A Child's-Eye View: An Examination of Point-of-View Camera Use in Four Informal Education Settings. *Visitor Studies*, 20(2), 218–237.
<https://doi.org/10.1080/10645578.2017.1404352>
- Cabe Trundle, K. (2010). Teaching science during the early childhood years. *Best Practices and Research Base*.
http://www.ngspscience.com/profdev/monographs/SCL22-0429A_SCI_AM_Trundle_lores.pdf
- Cain, R. & Lee, V. R. (2016). Measuring electrodermal activity to capture engagement in an afterschool maker program. *ACM International Conference Proceeding Series*, 78–81. <https://doi.org/10.1145/3003397.3003409>
- Callanan, M. A. (2012). Conducting Cognitive Developmental Research in Museums: Theoretical Issues and Practical Considerations. *Journal of Cognition and Development*, 13(2), 137–151. <https://doi.org/10.1080/15248372.2012.666730>
- Callanan, M. A., Castañeda, C. L., Luce, M. R. & Martin, J. L. (2017). Family Science Talk in Museums: Predicting Children's Engagement From Variations in Talk and Activity. *Child Development*, 88(5), 1492–1504.
<https://doi.org/10.1111/cdev.12886>
- Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E. & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, 15(8), 1243–1249.
<https://doi.org/10.1093/CERCOR/BHI007>
- Cassotti, M., Ania, A., Osmont, A., Houdé, O. & Borst, G. (2014). What have we learned about the processes involved in the Iowa Gambling Task from developmental studies? *Frontiers in Psychology*, 5(AUG), 915.
<https://doi.org/10.3389/fpsyg.2014.00915>
- Chakraverty, D., Newcomer, S. N., Puzio, K. & Tai, R. H. (2020). It Runs in the Family: The Role of Family and Extended Social Networks in Developing Early Science Interest. *Bulletin of Science, Technology and Society*, 38(3–4), 27–38.

<https://doi.org/10.1177/0270467620911589>

- Chamberlain, A. & Crabtree, A. (2020). Research 'In the Wild'. In A. Chamberlain & A. Crabtree (Eds.), *Into the Wild: Beyond the Design Research Lab* (pp. 1–6). Springer International Publishing. https://doi.org/10.1007/978-3-030-18020-1_1
- Chamberlain, A., Crabtree, A., Rodden, T., Jones, M. & Rogers, Y. (2012). Research in the wild: Understanding “in the wild” approaches to design and development. *Proceedings of the Designing Interactive Systems Conference, DIS '12*, 795–796. <https://doi.org/10.1145/2317956.2318078>
- Chapman, P., Selvarajah, S. & Webster, J. (1999). Engagement in Multimedia Training Systems. *Hawaii International Conference on System Sciences, 00(c)*, 1–9. <https://doi.org/10.1109/HICSS.1999.772808>
- Chita-Tegmark, M. (2016). Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Research in Developmental Disabilities*. <https://doi.org/10.1016/j.ridd.2015.10.011>
- Choi, B. C. K. & Pak, A. W. P. (2005). A Catalog of Biases in Questionnaires. *Preventing Chronic Disease*, 2(1), 1–13. https://www.cdc.gov/pcd/issues/2005/jan/04_0050.htm
- Choi, Y. K. (1999). The morphology of exploration and encounter in museum layouts. *Environment and Planning B: Planning and Design*, 26(2), 241–250. <https://doi.org/10.1068/b4525>
- Christenson, S. L. & Reschly, A. L. (2012). Jingle, Jangle, and Conceptual Haziness : Evolution and Future Directions of the Engagement Construct. In *Handbook of Research on Student Engagement* (pp. 3–19). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_1
- Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences*, 3(9), 345–351. [https://doi.org/10.1016/S1364-6613\(99\)01361-3](https://doi.org/10.1016/S1364-6613(99)01361-3)
- Cole, P. M., Fox, N. A., Zahn-Waxler, C., Usher, B. A. & Welsh, J. D. (1996). Individual differences in emotion regulation and behavior problems in preschool children. *Journal of Abnormal Psychology*, 105(4), 518–529. <https://doi.org/10.1037/0021-843X.105.4.518>

- Cowley, B. & Ravaja, N. (2014). Learning in balance: Using oscillatory EEG biomarkers of attention, motivation and vigilance to interpret game-based learning. *Cogent Education*, 1(1), 1–23.
<https://doi.org/10.1080/2331186X.2014.962236>
- Cresswell, J. (2009). *engage*. The Oxford Dictionary of Word Origins (2 ed.).
<https://doi.org/10.1093/acref/9780199547920.013.1812>
- Critchley, H. D. (2002). Electrodermal Responses: What Happens in the Brain. *The Neuroscientist*, 8(2), 132–142. <https://doi.org/10.1177/107385840200800209>
- Crowley, K., Callanan, M. A., Jipson, J. L., Galco, J., Topping, K. & Shrager, J. (2001). Shared scientific thinking in everyday parent - Child activity. *Science Education*, 85(6), 712–732. <https://doi.org/10.1002/sce.1035>
- Cruz-Garza, J. G., Brantley, J. A., Nakagome, S., Kontson, K., Megjhani, M., Robleto, D. & Contreras-Vidal, J. L. (2017). Deployment of Mobile EEG Technology in an Art Museum Setting: Evaluation of Signal Quality and Usability. *Frontiers in Human Neuroscience*, 11, 527.
<https://doi.org/10.3389/fnhum.2017.00527>
- Csikszentmihalyi, M. & Hermanson, K. (1995). Intrinsic Motivation in Museums: Why Does One Wants to Learn? *Public Institutions for Personal Learning, Establishing a Research Agenda*, 67–77. <https://doi.org/10.12691/education-2-3-10>
- D’Mello, S., Dieterle, E. & Duckworth, A. (2017). Advanced, Analytic, Automated (AAA) Measurement of Engagement During Learning. *Educational Psychologist*, 52(2), 104–123. <https://doi.org/10.1080/00461520.2017.1281747>
- D’Mello, S. & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145–157.
<https://doi.org/10.1016/j.learninstruc.2011.10.001>
- Dabney, K. P., Chakraverty, D. & Tai, R. H. (2013). The Association of Family Influence and Initial Interest in Science. *Science Education*, 97(3), 395–409.
<https://doi.org/10.1002/sce.21060>
- Damasio, A. R. (2001). Emotion and the human brain. *Annals of the New York*

Academy of Sciences, 935(1), 101–106. <https://doi.org/10.1111/j.1749-6632.2001.tb03475.x>

Dancstep née Dancu, T., Gutwill, J. P. & Sindorf, L. (2015). Comparing the Visitor Experience at Immersive and Tabletop Exhibits. *Curator*, 58(4), 401–422. <https://doi.org/10.1111/cura.12137>

Dancstep née Dancu, T. & Sindorf, L. (2018). Exhibit Designs for Girls' Engagement (EDGE). *Curator: The Museum Journal*, 61(3), 485–506. <https://doi.org/10.1111/cura.12267>

Davidson, R. A. & Smith, B. D. (1991). Caffeine and novelty: Effects on electrodermal activity and performance. *Physiology & Behavior*, 49(6), 1169–1175. [https://doi.org/10.1016/0031-9384\(91\)90346-P](https://doi.org/10.1016/0031-9384(91)90346-P)

de Houwer, J., Barnes-Holmes, D. & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin and Review*, 20(4), 631–642. <https://doi.org/10.3758/s13423-013-0386-3>

Deater-Deckard, K., Chang, M. & Evans, M. E. (2013). Engagement states and learning from educational games. *New Directions for Child and Adolescent Development*, 2013(139), 21–30. <https://doi.org/10.1002/cad.20028>

Deater-deckard, K., Chen, N., McClelland, E., Nguyen, C. & Yamine, J. (2011). *Engagement States Observational Coding System (eSOCS)*. 1–7.

Deater-Deckard, K., El Mallah, S., Chang, M., Evans, M. A. & Norton, A. (2014). Student behavioral engagement during mathematics educational video game instruction with 11-14 year olds. *International Journal of Child-Computer Interaction*, 2(3), 101–108. <https://doi.org/10.1016/j.ijcci.2014.08.001>

Deci, E. L. & Ryan, R. M. (1985). Cognitive evaluation theory. In *Intrinsic motivation and self-determination in human behavior* (pp. 43–85). Springer.

Dejonckheere, P. J. N., De Wit, N., Van de Keere, K. & Vervaet, S. (2016). Exploring the classroom: Teaching science in early childhood. *International Electronic Journal of Elementary Education*, 8(4), 537–558. <https://doi.org/10.12973/eu-jer.5.3.149>

Delprino, F., Piva, C., Tommasi, G., Gelsomini, M., Izzo, N. & Matera, M. (2018, May

- 29). ABBOT: A smart toy motivating children to become outdoor explorers. *Proceedings of the Workshop on Advanced Visual Interfaces AVI*.
<https://doi.org/10.1145/3206505.3206512>
- Denham, S. A., Bassett, H. H., Way, E., Mincic, M., Zinsser, K. & Graling, K. (2012). Preschoolers' emotion knowledge: Self-regulatory foundations, and predictions of early school success. *Cognition & Emotion*, 26(June 2013), 667–679.
<https://doi.org/10.1080/02699931.2011.602049>
- Dewey, J. (1986). Experience and Education. *The Educational Forum*, 50(3), 241–252. <https://doi.org/10.1080/00131728609335764>
- DeWitt, J., Archer, L. & Mau, A. (2016). Dimensions of science capital: exploring its potential for understanding students' science participation. *International Journal of Science Education*, 38(16), 2431–2449.
<https://doi.org/10.1080/09500693.2016.1248520>
- DeWitt, J., Osborne, J., Archer, L., Dillon, J., Willis, B. & Wong, B. (2013). Young Children's Aspirations in Science: The unequivocal, the uncertain and the unthinkable. *International Journal of Science Education*, 35(6), 1037–1063.
<https://doi.org/10.1080/09500693.2011.608197>
- Dietz, G., Pease, Z., McNally, B. & Foss, E. (2020). Giggle Gauge: A Self-Report Instrument for Evaluating Children's Engagement with Technology. *Proceedings of the Interaction Design and Children Conference*.
<https://doi.org/10.1145/3392063>
- Ding, L., Erkan, E. & Orey, M. (2018). An exploratory study of student engagement in gamified online discussions. *Computers & Education*, 120(January), 213–226.
<https://doi.org/10.1016/j.compedu.2018.02.007>
- Dupont-Boime, J. & Thevenot, C. (2018). High working memory capacity favours the use of finger counting in six-year-old children. *Journal of Cognitive Psychology*, 30(1), 35–42. <https://doi.org/10.1080/20445911.2017.1396990>
- Duran, J. I. & Fernandez-Dols, J. M. (2021). Do Emotions Result in Their Predicted Facial Expressions? A Meta-Analysis of Studies on the Co-Occurrence of Expression and Emotion. *Emotion, undefined(undefined), undefined*.

<https://doi.org/10.1037/emo0001015>

Early Childhood STEM Working Group. (2017). *Early STEM Matters. Providing High-Quality STEM Experiences for All Young Learners* (Issue January).

<http://ecstem.uchicago.edu.acknowledgements>

Eccles, J. S. (2016). Engagement: Where to next? *Learning and Instruction*, 43, 71–75. <https://doi.org/10.1016/j.learninstruc.2016.02.003>

Eccles, J. S. & Wang, M. (2012). Part I Commentary: So What Is Student Engagement Anyway? In *Handbook of Research on Student Engagement* (pp. 133–145). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_6

Eccles, J. S. & Wigfield, A. (2002). Motivational Beliefs, Values, and Goals. *Annual Reviews in Psychology*, 53, 109–132.

Eghbal-Azar, K. & Widlok, T. (2013). Potentials and Limitations of Mobile Eye Tracking in Visitor Studies: Evidence From Field Research at Two Museum Exhibitions in Germany. *Social Science Computer Review*, 31(1), 103–118. <https://doi.org/10.1177/0894439312453565>

Einarsdóttir, J. (2007). Research with children: Methodological and ethical challenges. *European Early Childhood Education Research Journal*, 15(2), 197–211. <https://doi.org/10.1080/13502930701321477>

Eisinga, R., Grotenhuis, M. Te & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, 58(4), 637–642. <https://doi.org/10.1007/s00038-012-0416-3>

ELAN - Linguistic Annotator (5.7 FX). (n.d.). The Language Archive, Max Planck Institute. <https://archive.mpi.nl/tla/elan>

Emerson, A., Cloude, E. B., Azevedo, R. & Lester, J. (2020). Multimodal learning analytics for game-based learning. *British Journal of Educational Technology*, 51(5), 1505–1526. <https://doi.org/10.1111/bjet.12992>

Erickson, L. C., Thiessen, E. D., Godwin, K. E., Dickerson, J. P. & Fisher, A. V. (2015). Endogenously and exogenously driven selective sustained attention: Contributions to learning in kindergarten children. *Journal of Experimental Child Psychology*, 138, 126–134. <https://doi.org/10.1016/j.jecp.2015.04.011>

- Eshach, H. & Fried, M. N. (2005). Should science be taught in early childhood? *Journal of Science Education and Technology*, 14(3), 315–336.
<https://doi.org/10.1007/s10956-005-7198-9>
- Essex, J. & Haxton, K. (2018). Characterising patterns of engagement of different participants in a public STEM-based analysis project. *International Journal of Science Education, Part B: Communication and Public Engagement*, 8(2), 178–191. <https://doi.org/10.1080/21548455.2017.1423128>
- Eysenck, M. W. & Brysbaert, M. (2018). Cognition and Emotion. In *Fundamentals of Cognition* (Vol. 11, Issue 1, pp. 1–12).
- Falk, J. H. (2005). Free-choice environmental learning: framing the discussion. *Environmental Education Research*, 11(3), 265–280.
<https://doi.org/10.1080/13504620500081129>
- Falk, J. H. & Dierking, L. D. (1998). Free-Choice Learning: An Alternative Term to Informal Learning? *Informal Learning Environments Research Newsletter*, 2.
http://ls-tlss.ucl.ac.uk/course-materials/ARCLG034_62453
- Falk, J. H. & Dierking, L. D. (2000). Learning from museums: Visitor experiences and the making of meaning. In *American Association for State and Local History Book Series*. AltaMira Press.
- Falk, J. H., Dierking, L. D. & Foutz, S. (2007). In principle, in practice: museums as learning institutions. *Choice Reviews Online*, 45(04), 45-1829-45–1829.
<https://doi.org/10.5860/choice.45-1829>
- Falk, J. H., Scott, C., Dierking, L., Rennie, L. & Jones, M. C. (2004). Interactives and Visitor Learning. *Curator: The Museum Journal*, 47(2), 171–198.
<https://doi.org/10.1111/j.2151-6952.2004.tb00116.x>
- Falk, J. H., Storksdieck, M. & Dierking, L. D. (2007). Investigating public science interest and understanding: Evidence for the importance of free-choice learning. *Public Understanding of Science*, 16(4), 455–469.
<https://doi.org/10.1177/0963662506064240>
- Fallahpour, K., Clarke, S. D., Goldberg, E., Hermens, D. F., Falconer, E. M. & Gordon, E. (2010). Alterations in theta activity associated with novelty and

- routinization processing in ADHD. *Clinical Neurophysiology*, 121(8), 1336–1342. <https://doi.org/10.1016/j.clinph.2010.02.152>
- Faria, C., Guilherme, E., Gaspar, R. & Boaventura, D. (2015). History of Science and Science Museums: An Enriching Partnership for Elementary School Science. *Science and Education*, 24(7–8), 983–1000. <https://doi.org/10.1007/s11191-015-9773-7>
- Farrell, A., Kagan, S. L. & Tisdall, K. (2015a). Early Childhood Research: An Expanding Field. In *The SAGE Handbook of Early Childhood Research* (pp. 1–11). SAGE Publications Ltd. <https://doi.org/10.4135/9781473920859.n1>
- Farrell, A., Kagan, S. & Tisdall, E. (2015b). The SAGE Handbook of Early Childhood Research. In *The SAGE Handbook of Early Childhood Research*. SAGE Publications Ltd. <https://doi.org/10.4135/9781473920859>
- Finn, J. D. & Zimmer, K. S. (2012). Student engagement: What is it? Why does it matter? In *Handbook of Research on Student Engagement* (pp. 97–131). Springer US. <https://doi.org/10.1007/978-1-4614-2018-7>
- Fisch, S. M., Lesh, R., Motoki, E., Crespo, S. & Melfi, V. (2014). Cross-platform learning: How do children learn from multiple media. In *Learning by playing: Video gaming in education* (pp. 207–219). Oxford University Press New York, NY.
- Fleiss, J. L., Levin, B. & Paik, M. C. (2003). The Measurement of Interrater Agreement. In *Statistical Methods for Rates and Proportions* (pp. 598–626). Wiley. <https://doi.org/10.1002/0471445428.ch18>
- Fletcher, A. J. (2017). Applying critical realism in qualitative research: methodology meets method. *International Journal of Social Research Methodology*, 20(2), 181–194. <https://doi.org/10.1080/13645579.2016.1144401>
- Flowerday, T. & Shell, D. F. (2015). Disentangling the effects of interest and choice on learning, engagement, and attitude. *Learning and Individual Differences*, 40, 134–140. <https://doi.org/10.1016/j.lindif.2015.05.003>
- Forrester, M. (2017). Participation and engagement. *Research on Children and Social Interaction*, 1(1), 55–76. <https://doi.org/10.1558/rcsi.30588>

- Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J. & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition*, 117(3), 319–331. <https://doi.org/10.1016/j.cognition.2010.09.003>
- Fowles, D. C., Kochanska, G., Murray, & Kathleen. (2000). Electrodermal activity and temperament in preschool children. *Psychophysiology*, 37(6), S0048577200981836. <https://doi.org/10.1017/S0048577200981836>
- Fredricks, J. A., Blumenfeld, P. C. & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Fredricks, J. A., Filsecker, M. & Lawson, M. A. (2016). Student engagement, Context, And adjustment: Addressing definitional, Measurement, And methodological issues. In *Learning and Instruction* (Vol. 43). <https://doi.org/10.1016/j.learninstruc.2016.02.002>
- Fredricks, J. A. & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In *Handbook of Research on Student Engagement* (pp. 763–782). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_37
- Fredricks, J. A., McColskey, W., Meli, J., Mordica, J., Montrosse, B. & Mooney, K. (2011). Measuring student engagement in upper elementary through high school: A description of 21 instruments. *Issues and Answers Report*, 098, 26–27. https://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL_2011098.pdf
- Fu, A. C., Kannan, A., Shavelson, R. J., Peterson, L. & Kurpius, A. (2016). Room for Rigor: Designs and Methods in Informal Science Education Evaluation. *Visitor Studies*, 19(1), 12–38. <https://doi.org/10.1080/10645578.2016.1144025>
- Gamer, M., Lemon, J. & Puspendra Singh, I. F. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement* (R package version 0.84.1). <https://cran.r-project.org/package=irr>
- Garbarino, M., Lai, M., Tognetti, S., Picard, R. & Bender, D. (2014). *Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition*. <https://doi.org/10.4108/icst.mobihealth.2014.257418>

- Gashi, S., Di Lascio, E. & Santini, S. (2019). Using Unobtrusive Wearable Sensors to Measure the Physiological Synchrony Between Presenters and Audience Members. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(1), 1–19. <https://doi.org/10.1145/3314400>
- Gelman, A. (2006). *Multilevel (Hierarchical) Modeling: What It Can and Cannot Do*. <https://doi.org/10.1198/004017005000000661>
- Ghazai-Mohammed, Z. & Meikleham, S. (2020). Cautious collaboration to perfect partnership'. *Digital Education Seminar*.
- Ghazali-Mohammed, Z., Meikleham, S., Macnab, S. & Manches, A. (2020). Video Analysis to Real-Time Observation of Adult-Child Interaction: The Challenge of Translation [Poster Session]. *AERA Annual Meeting [Conference Cancelled]*. https://convention2.allacademic.com/one/aera/aera20/index.php?program_focus=view_paper&selected_paper_id=1577598&cmd=online_program_direct_link&sub_action=online_program
- Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C. & Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience*, 2(10), 861–863. <https://doi.org/10.1038/13158>
- Godec, S., Archer, L. & Dawson, E. (2021). Interested but not being served: mapping young people's participation in informal STEM education through an equity lens. *Research Papers in Education*. <https://doi.org/10.1080/02671522.2020.1849365>
- Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Nugent, T. F., Herman, D. H., Clasen, L. S., Toga, A. W., Rapoport, J. L. & Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences of the United States of America*, 101(21), 8174–8179. <https://doi.org/10.1073/pnas.0402680101>
- Goldin-Meadow, S. (2018). Taking a Hands-on Approach to Learning: <https://doi.org/10.1177/2372732218785393>, 5(2), 163–170. <https://doi.org/10.1177/2372732218785393>

- Gopnik, A. (2012). Scientific Thinking in Young Children: Theoretical Advances, Empirical Research, and Policy Implications. *Science*, 337(6102), 1623–1627. <https://doi.org/10.1126/science.1223416>
- Grack Nelson, A. & Cohn, S. (2015). Data collection methods for evaluating museum programs and exhibitions. *Journal of Museum Education*, 40(1), 27–36. <https://doi.org/10.1080/10598650.2015.11510830>
- Gutwill, J. P. (2008). Challenging a Common Assumption of Hands-on Exhibits How Counterintuitive Phenomena Can Undermine Inquiry. *Journal of Museum Education*, 33(2), 187–198. <https://doi.org/10.2307/40479643>
- Gutwill, J. P. & Dancstep (née Dancu), T. (2017). Boosting Metacognition in Science Museums: Simple Exhibit Label Designs to Enhance Learning. *Visitor Studies*, 20(1), 72–88. <https://doi.org/10.1080/10645578.2017.1297132>
- Halderman, L. K., Finn, B., Lockwood, J. R., Long, N. M. & Kahana, M. J. (2021). EEG Correlates of Engagement During Assessment. *ETS Research Report Series*. <https://doi.org/10.1002/ets2.12312>
- Halliday, S. E., Calkins, S. D. & Leerkes, E. M. (2018). Measuring preschool learning engagement in the laboratory. *Journal of Experimental Child Psychology*, 167, 93–116. <https://doi.org/10.1016/j.jecp.2017.10.006>
- Harel, G. & Koichu, B. (2010). An operational definition of learning. *Journal of Mathematical Behavior*, 29(3), 115–124. <https://doi.org/10.1016/j.jmathb.2010.06.002>
- Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R. & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. *Computers in Human Behavior*, 48, 615–625. <https://doi.org/10.1016/j.chb.2015.02.013>
- Huan, N. P. & Kolstør, S. D. (2014). Exhibitions as learning environments: a review of empirical research on students' science learning at Natural History Museums, Science Museums and Science Centres. *Nordic Studies in Science Education*, 10(1), 90–104. <https://doi.org/10.5617/nordina.652>
- Heath, C. & Vom Lehn, D. (2008). Configuring “interactivity”: Enhancing engagement

- in science centres and museums. *Social Studies of Science*, 38(1), 65–93.
<https://doi.org/10.1177/0306312707084152>
- Hedman, E. B. (2008). Thick Psychophysiology for Empathic Design. *Massachusetts Institute of Technology*, 2008. <http://dspace.mit.edu/handle/1721.1/91436>
- Heinrichs, M. & Domes, G. (2008). Neuropeptides and social behaviour: effects of oxytocin and vasopressin in humans. In *Progress in Brain Research* (Vol. 170, pp. 337–350). [https://doi.org/10.1016/S0079-6123\(08\)00428-7](https://doi.org/10.1016/S0079-6123(08)00428-7)
- Hernandez, J., Riobo, I., Rozga, A., Abowd, G. D. & Picard, R. W. (2014). Using electrodermal activity to recognize ease of engagement in children during social interactions. *UbiComp 2014 - Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 307–317.
<https://doi.org/10.1145/2632048.2636065>
- Hidi, S. & Renninger, K. A. (2006). The Four-Phase Model of Interest Development. *Educational Psychologist*, 41(2), 87–98.
<https://doi.org/10.1207/s15326985ep4102>
- Hoad, T. F. (1996). *engage*. The Concise Oxford Dictionary of English Etymology.
<https://doi.org/10.1093/acref/9780192830982.013.5042>
- Huan, R. (2020). *Evaluating Child Engagement in Digital Story Stems using Facial Data* [University of Glasgow]. <https://doi.org/10.5525/gla.thesis.82040>
- Humphrey, T. & Gutwill, J. P. (2005). *Fostering Active Prolonged Engagement* (T. Humphrey & J. P. Gutwill (Eds.); 1st ed.). Routledge.
<https://doi.org/10.4324/9781315428291>
- Jarodzka, H., Scheiter, K., Gerjets, P., Gog, T. Van & van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2), 146–154.
<https://doi.org/10.1016/j.learninstruc.2009.02.019>
- Johnson, M. H. & de Haan, M. (2011). Developmental Cognitive Neuroscience. In *Wiley-Blackwell*. <https://doi.org/10.1007/s13398-014-0173-7.2>
- Jones, H. E. (1930). The Galvanic Skin Reflex in Infancy. *Child Development*, 1(2), 106. <https://doi.org/10.2307/1125585>

- Jones, H. E. (1935). The Galvanic Skin Reflex as Related to Overt Emotional Expression. *The American Journal of Psychology*, 47(2), 241–251.
- Jones, H. E. (1950). The study of patterns of emotional expression. In M. L. Reymert (Ed.), *Feelings and Emotions: The Amsterdam Symposium* (pp. 161–168). McGraw-Hill.
- Jones, H. E. (1960). The longitudinal method in the study of personality. In *Personality development in children* (pp. 1–27). University of Texas Press.
- Jorgensen, D. (2011). The Methodology of Participant Observation. In *Participant Observation* (pp. 12–25). <https://doi.org/10.4135/9781412985376.n1>
- Jung, Y. J., Zimmerman, H. T. & Pérez-Edgar, K. (2018). A Methodological Case Study with Mobile Eye-Tracking of Child Interaction in a Science Museum. *TechTrends*, 62(5), 509–517. <https://doi.org/10.1007/S11528-018-0310-9/FIGURES/8>
- Kamzanova, A. T., Matthews, G., Kustubayeva, A. M. & Jakupov, S. M. (2011). EEG Indices to Time-On-Task Effects and to a Workload Manipulation (Cueing). *International Journal of Psychological and Behavioral Sciences*, 5(8), 19–22. <https://pdfs.semanticscholar.org/693e/f16179e7859d3d2631c015096238bba563bd.pdf>
- Kano, A., Horton, M. & Read, J. C. (2010). Thumbs-Up Scale and Frequency of Use Scale for use in self reporting of children’s Computer Experience. *NordiCHI 2010: Extending Boundaries - Proceedings of the 6th Nordic Conference on Human-Computer Interaction*, 699–702. <https://doi.org/10.1145/1868914.1869008>
- Khedher, A. Ben, Jraidi, I., Frasson, C., Khedher, A. Ben, Jraidi, I. & Frasson, C. (2019). Tracking Students’ Mental Engagement Using EEG Signals during an Interaction with a Virtual Learning Environment. *Journal of Intelligent Learning Systems and Applications*, 11(1), 1–14. <https://doi.org/10.4236/JILSA.2019.111001>
- Knickmeyer, R. C., Gouttard, S., Kang, C., Evans, D., Wilber, K., Smith, J. K., Hamer, R. M., Lin, W., Gerig, G. & Gilmore, J. H. (2008). A Structural MRI Study

- of Human Brain Development from Birth to 2 Years. *Journal of Neuroscience*, 28(47), 12176–12182. <https://doi.org/10.1523/JNEUROSCI.3479-08.2008>
- Koo, T. K. & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kress, E. E. (2016). *Cognitive Curating : Using Brain and Behavioral Science to Design Engaging Art Exhibitions for the Everyday Visitor*. Sotheby's Institute of Art.
- Krigolson, O. E., Williams, C. C., Norton, A., Hassall, C. D. & Colino, F. L. (2017). Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. *Frontiers in Neuroscience*, 11(MAR), 109. <https://doi.org/10.3389/fnins.2017.00109>
- Krogh-Jespersen, S., Quinn, K. A., Krenzer, W. L. D., Nguyen, C., Greenslit, J. & Price, C. A. (2020). Exploring the awe-some: Mobile eye-tracking insights into awe in a science museum. *PLoS ONE*, 15(9 September), e0239204. <https://doi.org/10.1371/journal.pone.0239204>
- Kuhn, D. (2000). Developmental origins of metastrategic awareness and control. *Current Directions in Psychological Science*, 9(5), 178–181. <http://ezproxy-prd.bodleian.ox.ac.uk:7216/doi/pdf/10.1111/1467-8721.00088>
- Kumpulainen, K., Karttunen, M., Juurola, L. & Mikkola, A. (2014). Towards children's creative museum engagement and collaborative sense-making. *Digital Creativity*, 25(3), 233–246. <https://doi.org/10.1080/14626268.2014.904370>
- Lachman, S. J. (1997). Learning is a Process: Toward an Improved Definition of Learning. *The Journal of Psychology*, 131(5), 477–480. <https://doi.org/10.1080/00223989709603535>
- Laeng, B., Bloem, I. M., D'Ascenzo, S. & Tommasi, L. (2014). Scrutinizing visual images: The role of gaze in mental imagery and memory. *Cognition*, 131(2), 263–283. <https://doi.org/10.1016/j.cognition.2014.01.003>
- Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., Lee, M. H., Chiou, G. L., Liang, J. C. & Tsai, C. C. (2013). A review of using eye-tracking

- technology in exploring learning from 2000 to 2012. *Educational Research Review*, 10, 90–115. <https://doi.org/10.1016/j.edurev.2013.10.001>
- Lancaster, G. A. (2015). Pilot and feasibility studies come of age! In *Pilot and Feasibility Studies* (Vol. 1, Issue 1, pp. 1–4). BioMed Central Ltd. <https://doi.org/10.1186/2055-5784-1-1>
- Land, M., Mennie, N. & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11), 1311–1328. <https://doi.org/10.1068/p2935>
- Laurans, G., Desmet, P. M. A. & Hekkert, P. (2009). The emotion slider: A self-report device for the continuous measurement of emotion. *Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009*. <https://doi.org/10.1109/ACII.2009.5349539>
- LeDoux, J. E. & Phelps, E. (2008). Emotional Networks in the Brain. In M. Lewis, J. M. Haviland-Jones & L. Feldman Barret (Eds.), *Handbook of Emotion* (pp. 159–179). The Guilford Press.
- Lee-Cultura, S., Sharma, K., Cosentino, G., Papavlasopoulou, S. & Giannakos, M. (2021). Children’s Play and Problem Solving in Motion-Based Educational Games: Synergies between Human Annotations and Multi-Modal Data. *Interaction Design and Children (IDC '21)*, 408–420. <https://doi.org/10.1145/3459990.3460702>
- Leister, W., Tjøstheim, I., Joryd, G., Andersson, J. A. & Heggelund, H. (2019). Strengthening Engagement in Science Understanding with Learning Trails. *Multimodal Technologies and Interaction*, 3(3), 48. <https://doi.org/10.3390/mti3030048>
- Leister, W., Tjøstheim, I., Schulz, T., Joryd, G., Larssen, A. & De Brisis, M. (2016). Towards Assessing visitor engagement in science centres and museums. *International Journal on Advances in Life Sciences*, 8(1–2), 50–64.
- Lerner, R. M., Easterbrooks, M. A. & Mistry, J. (2012). *Developmental Science Across the Life Span. An Introduction* (I. B. Weiner, R. M. Lerner, M. A. Easterbrooks, J. Mistry, & I. Weiner (Eds.)).

- Li, X., Lu, Z. L., D'Argembeau, A., Ng, M. & Bechara, A. (2010). The Iowa Gambling Task in fMRI images. *Human Brain Mapping*, 31(3), 410–423.
<https://doi.org/10.1002/hbm.20875>
- Lindah, B. (2007, April). A longitudinal study of students' attitudes towards science and choice of career. *NARST Annual Conference*.
<http://urn.kb.se/resolve?urn=urn:nbn:se:hkr:diva-6093>
- López, F., Menez, M. & Hernández-Guzmán, L. (2005). Sustained attention during learning activities: An observational study with preschool children. *Early Child Development and Care*, 175(2), 131–138.
<https://doi.org/10.1080/0300443042000230384>
- M2L Team. (n.d.-a). *Move2Learn: Engaging Preschool Scientists through Embodiment and Technology*. Retrieved February 26, 2018, from
<http://www.move2learn.ed.ac.uk/wp-content/uploads/2016/04/Move2Learn-Overview-Brochure-March-2016.pdf>
- M2L Team. (n.d.-b). *Move2Learn*. Embodied Science Learning for Children in Museums. www.move2learn.net
- Mahatmya, D., Lohman, B. J., Matjasko, J. L. & Farb, A. F. (2012). Engagement across developmental periods. In *Handbook of Research on Student Engagement* (pp. 45–63). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_3
- Mahatmya, D., Lohman, B. J., Matjasko, J. L. & Feldman Farb, A. (2012). Engagement Across Developmental Periods. In *Handbook of Research on Student Engagement* (pp. 1–840). <https://doi.org/10.1007/978-1-4614-2018-7>
- Mahoney, G. & Wheeden, C. A. (1999). The effect of teacher style on interactive engagement of preschool-aged children with special learning needs. *Early Childhood Research Quarterly*, 14(1), 51–68. [https://doi.org/10.1016/S0885-2006\(99\)80004-0](https://doi.org/10.1016/S0885-2006(99)80004-0)
- Malmberg, J., Järvelä, S., Holappa, J., Haataja, E., Huang, X. & Siipo, A. (2019). Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Computers in Human*

Behavior, 96, 235–245. <https://doi.org/10.1016/j.chb.2018.06.030>

- Mantzicopoulos, P., Patrick, H. & Samarapungavan, A. (2008). Young children's motivational beliefs about learning science. *Early Childhood Research Quarterly*, 23(3), 378–394. <https://doi.org/10.1016/j.ecresq.2008.04.001>
- Marín-Morales, J., Higuera-Trujillo, J. L., Greco, A., Guixeres, J., Llinares, C., Gentili, C., Scilingo, E. P., Alcañiz, M. & Valenza, G. (2019). Real vs. immersive-virtual emotional experience: Analysis of psycho-physiological patterns in a free exploration of an art museum. *PLoS ONE*, 14(10), 1–24. <https://doi.org/10.1371/journal.pone.0223881>
- Martin, A. J. (2009). Motivation and engagement across the academic life span: A developmental construct validity study of elementary school, high school, and university/college students. *Educational and Psychological Measurement*, 69(5), 794–824. <https://doi.org/10.1177/0013164409332214>
- Martinez, A. & Du, S. (2012). A Model of the Perception of Facial Expressions of Emotion by Humans: Research Overview and Perspectives. *Journal of Machine Learning Research*, 13, 1589–1608.
- Massachusetts, A., Museum, P. E. & Asher, T. (2019). *Smithsonian . com The Neuroscientist in the Art Museum At Massachusetts ' s Peabody Essex Museum , Tedi Asher is using neuroscience research to create.* 1–8.
- Matsumoto, D. & Hwang, H. S. (2011). Reading facial expressions of emotion. *Psychological Science Agenda*, 2011(May), 1–6. <http://www.apa.org/science/about/psa/2011/05/facial-expressions.aspx>
- McCarthy, A., Lee, K., Itakura, S. & Muir, D. W. (2008). Gaze display when thinking depends on culture and context. *Journal of Cross-Cultural Psychology*, 39(6), 716–729. <https://doi.org/10.1177/0022022108323807>
- McCarthy, C., Pradhan, N., Redpath, C. & Adler, A. (2016). Validation of the Empatica E4 wristband. *2016 IEEE EMBS International Student Conference: Expanding the Boundaries of Biomedical Engineering and Healthcare, ISC 2016 - Proceedings*, 1–4. <https://doi.org/10.1109/EMBSISC.2016.7508621>
- McClure, E. R., Guernsey, L., Clements, D. H., Bales, S. N., Nichols, J., Kendall-

- Taylor, N. & Levine, M. H. (2017). STEM Starts Early: Grounding Science, Technology, Engineering, and Math Education in Early Childhood. In *Joan Ganz Cooney Center at Sesame Workshop*. www.joanganzcooneycenter.org
- McCormack, T., Bramley, N., Frosch, C., Patrick, F. & Lagnado, D. (2016). Children's use of interventions to learn causal structure. *Journal of Experimental Child Psychology*, 141, 1–22. <https://doi.org/10.1016/j.jecp.2015.06.017>
- McGuigan, F. & Schoonover, R. (1973). *The Psychophysiology of Thinking: Studies of Covert Processes* (F. J. McGuigan & R. A. Schoonover (Eds.)). Academic Press.
- McManus, P. M. (1992). Topics in museums and science education. *Studies in Science Education*, 20(1), 157–182. <https://doi.org/10.1080/03057269208560007>
- McWilliam, R. A. & De Kruif, R. E. L. (1998). E-Qual III: Children's engagement codes. Chapel Hill, NC: Frank Porter Graham Child Development Center, University of North Carolina at Chapel Hill.
- Mercier, J. & Charland, P. (2013). An agenda for neuroeducation: relating psychophysiological and behavioral data across time scales of learning. *Neuroéducation*, 2(1), 71–86. <https://static1.squarespace.com/static/520e383ee4b021a19fa28bf7/t/52991ab7e4b0d8d4512b4ed0/1385765559571/Mercier2013.pdf>
- Meyer, D. K. & Turner, J. C. (2006). Re-conceptualizing emotion and motivation to learn in classroom contexts. *Educational Psychology Review*, 18(4), 377–390.
- Michie, S., West, R. & Hastings, J. (2019). Creating ontological definitions for use in science. *Qeios*. <https://doi.org/10.32388/ygif9b>
- Mitas, O., Cuenen, R., Bastiaansen, M., Chick, G. & van den Dungen, E. (2020). The War from both Sides: how Dutch and German Visitors Experience an Exhibit of Second World War Stories. *International Journal of the Sociology of Leisure*, 3(3), 277–303. <https://doi.org/10.1007/s41978-020-00062-3>
- Montgomery, H. (2014). Participant Observation. In *Understanding Research with Children and Young People* (pp. 122–135). SAGE Publications, Inc.

<https://doi.org/10.4135/9781526435637.n8>

- Moriguchi, Y., Zelazo, P. D., Chevalier, N. & Zelazo, P. D. (2016). Development of Executive Function During Childhood. In *Frontiers in Psychology* (Vol. 7, Issue JAN). <https://doi.org/10.3389/fpsyg.2016.00006>
- Murphy, B. L., Arnsten, A. F., Goldman-Rakic, P. S. & Roth, R. H. (1996). Increased dopamine turnover in the prefrontal cortex impairs spatial working memory performance in rats and monkeys. *Proceedings of the National Academy of Sciences of the United States of America*, 93(3), 1325–1329. <https://doi.org/10.1073/pnas.93.3.1325>
- Nakagawa, S., Johnson, P. C. D. & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of The Royal Society Interface*, 14(134). <https://doi.org/10.1098/RSIF.2017.0213>
- Nakagawa, S. & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. <https://doi.org/10.1111/J.2041-210X.2012.00261.X>
- National Research Council. (2009). Learning Science in Informal Environments. In P. Bell, B. Lewenstein, A. W. Shouse & M. A. Feder (Eds.), *Learning Science in Informal Environments*. The National Academies Press. <https://doi.org/10.17226/12190>
- Nourbakhsh, N., Wang, Y., Chen, F. & Calvo, R. A. (2012). Using Galvanic Skin Response for Cognitive Load Measurement in Arithmetic and Reading Tasks. *Proceedings of the 24th Australian Computer-Human Interaction Conference*, 1–4. <https://doi.org/10.1145/2414536.2414602>
- Novack, M. A. & Goldin-Meadow, S. (2015). Learning from gesture: How our hands change our minds. In *Educational Psychology Review* (Vol. 27, Issue 3, pp. 405–412). Springer New York LLC. <https://doi.org/10.1007/s10648-015-9325-3>
- Nubani, L., Puryear, A. & Kellom, K. (2018). Measuring the effect of visual exposure and saliency of museum exhibits on visitors' level of contact and engagement. *Behavioral Sciences*, 8(11), 100. <https://doi.org/10.3390/bs8110100>

- O'Brien, H. L., Cairns, P. & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human Computer Studies*, 112(January), 28–39. <https://doi.org/10.1016/j.ijhcs.2018.01.004>
- O'Brien, H. L. & Toms, E. G. (2008). What is user engagement? A Conceptual Framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 59(6), 938–955. <https://doi.org/https://doi.org/10.1002/asi.208011>
- Ohlendorf, S., Sprenger, A., Speck, O., Glauche, V., Haller, S. & Kimming, H. (2010). Visual motion, eye motion, and relative motion: A parametric fMRI study of functional specializations of smooth pursuit eye movement network areas. *Journal of Vision*, 10(2010), 1–15. <https://doi.org/10.1167/10.14.21>.
- Osborne, J., Simon, S. & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, 25(9), 1049–1079. <https://doi.org/10.1080/0950069032000032199>
- Oudeyer, P. Y., Gottlieb, J. & Lopes, M. (2016). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. In *Progress in Brain Research* (Vol. 229, pp. 257–284). Elsevier B.V. <https://doi.org/10.1016/bs.pbr.2016.05.005>
- Paris, S. G. S. G. (1997). Situated Motivation and Informal Learning. *The Journal of Museum Education*, 22(2/3), 22–27. <https://doi.org/10.2307/40479100>
- Pekrun, R. & Linnenbrink-garcia, L. (2012). Academic Emotions and Student Engagement. In *Handbook of Research on Student Engagement* (pp. 259–284). <https://doi.org/10.1007/978-1-4614-2018-7>
- Perski, O., Blandford, A., West, R. & Michie, S. (2017). Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Translational Behavioral Medicine*, 7(2), 254–267. <https://doi.org/10.1007/s13142-016-0453-1>
- Pham, M. T. & Avnet, T. (2009). Rethinking Regulatory Engagement Theory. *Journal of Consumer Psychology*, 19(2), 115–123.

<https://doi.org/10.1016/j.jcps.2009.02.003>

- Piaget, J. (1970). *Science of education and the psychology of the child*. New York Orion Press.
- Pinto, A. I., Aguiar, C., Barros, S. & Cruz, O. (2004). O Sistema de Avaliação da Qualidade do Envolvimento III: Um procedimento de avaliação do envolvimento da criança em contexto de creche. *Avaliação Psicológica: Formas e Contextos*.
- Poppe, R., van Delden, R., Moreno, A. & Reidsma, D. (2014). Interactive Playgrounds for Children. In *Playful user interfaces* (pp. 99–118). Springer, Singapore. https://doi.org/10.1007/978-981-4560-96-2_5
- Porpodas, C. D. (1987). The One-question Conservation Experiment Reconsidered. *Journal of Child Psychology and Psychiatry*, 28(2), 343–349. <https://doi.org/10.1111/j.1469-7610.1987.tb00216.x>
- Pressley, M. & Hilden, K. (2006). Cognitive Strategies. In *Handbook of child psychology: Cognition, perception, and language, Vol. 2, 6th ed.* (pp. 511–556). John Wiley & Sons Inc.
- Price, L. & Martin, L. (2018). Introduction to the special issue: applied critical realism in the social sciences. In *Journal of Critical Realism* (Vol. 17, Issue 2, pp. 89–96). Routledge. <https://doi.org/10.1080/14767430.2018.1468148>
- Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, 17(9), 458–470. <https://doi.org/10.1016/j.tics.2013.06.004>
- R Core Team. (2021). *R: A Language and Environment for Statistical Computing* (R version 4.0.5 “Shake and Throw”). R Foundation for Statistical Computing. <https://www.r-project.org/>
- Rabe-Hesketh, S. & Skrondal, A. (2008). *Multilevel and longitudinal modeling using Stata*. STATA press.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. In *Quarterly journal of experimental psychology (2006)* (Vol. 62, Issue 8). <https://doi.org/10.1080/17470210902816461>

- Rayner, K., White, S. J., Johnson, R. L. & Liversedge, S. P. (2006). Reading words with jumbled letters: There is a cost. *Psychological Science*, 17(3), 192–193. <https://doi.org/10.1111/j.1467-9280.2006.01684.x>
- Read, J. C. (2008). Validating the Fun Toolkit: An instrument for measuring children's opinions of technology. *Cognition, Technology and Work*, 10(2), 119–128. <https://doi.org/10.1007/s10111-007-0069-9>
- Read, J. C. & MacFarlane, S. (2006). Using the fun toolkit and other survey methods to gather opinions in child computer interaction. *Proceeding of the 2006 Conference on Interaction Design and Children - IDC '06*, 81. <https://doi.org/10.1145/1139073.1139096>
- Read, J. C., Macfarlane, S. & Casey, C. (2002). Endurability, Engagement and Expectations: Measuring Children's Fun. *Interaction Design and Children*, 189–198. [file:///C:/Users/alexi/OneDrive - University of Edinburgh/PhD/--MY RESEARCH-- /Reading/Endurability_Engagement_and_Expectations_Measuring \(smileyometer\).pdf](file:///C:/Users/alexi/OneDrive%20-%20University%20of%20Edinburgh/PhD/--MY%20RESEARCH--/Reading/Endurability_Engagement_and_Expectations_Measuring(smileyometer).pdf)
- Rennie, L. J., Feher, E., Dierking, L. D. & Falk, J. H. (2003). Toward an agenda for advancing research on science learning in out-of-school settings. *Journal of Research in Science Teaching*, 40(2), 112–120. <https://doi.org/10.1002/tea.10067>
- Rennie, L. J., Mc Clafferty, T. & Johnson, D. (1993). Interactive Science and Technology Centres: Helping Teachers Make Best Use of Them. *Annual Conference of the Australian Association for Research in Education*.
- Rennie, L. J. & McClafferty, T. P. (1993). Learning in science centres and science museums: A review of recent studies. *Research in Science Education*, 23(1), 351.
- Rennie, L. J. & McClafferty, T. P. (1996). Science centres and science learning. *Studies in Science Education*, 27(1), 53–98. <https://doi.org/10.1080/03057269608560078>
- Renninger, K. A. & Bachrach, J. E. (2015). Studying Triggers for Interest and

- Engagement Using Observational Methods. *Educational Psychologist*, 50(1), 58–69. <https://doi.org/10.1080/00461520.2014.999920>
- Revueltas Roux, A. & McClure, D. (2020, March). *Slider-evaluation-tool*. *GitHub Repository*. <https://github.com/alerevroux/slider-evaluation-tool>
- Roberts, J. A., Wallis, G. & Breakspear, M. (2013). Fixational eye movements during viewing of dynamic natural scenes. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2013.00797>
- Rolls, E. T. (2019). The cingulate cortex and limbic systems for emotion, action, and memory. *Brain Structure and Function*, 224(9), 3001–3018. <https://doi.org/10.1007/S00429-019-01945-2/FIGURES/4>
- Roth, W. M. (2002). From action to discourse: The bridging function of gestures. *Cognitive Systems Research*, 3(3), 535–554. [https://doi.org/10.1016/S1389-0417\(02\)00056-6](https://doi.org/10.1016/S1389-0417(02)00056-6)
- Rothman, S. B. (2007). Understanding Data Quality through Reliability: A Comparison of Data Reliability Assessment in Three International Relations Datasets. *International Studies Review*, 9(3), 437–456. <https://doi.org/10.1111/j.1468-2486.2007.00698.x>
- Saçkes, M., Trundle, K. C., Bell, R. L. & O’Connell, A. A. (2011). The influence of early science experience in kindergarten on children’s immediate and later science achievement: Evidence from the Early Childhood Longitudinal Study. *Journal of Research in Science Teaching*, 48(2), 217–235. <https://doi.org/10.1002/tea.20395>
- Samarapungavan, A., Mantzicopoulos, P. & Patrick, H. (2008). Learning science through inquiry in kindergarten. *Science Education*, 92(5), 868–908. <https://doi.org/10.1002/sce.20275>
- Sammons, P. (2010). Does pre-school make a difference? Identifying the impact of preschool on children’s cognitive and social behavioural development at different ages. In K. Sylva, E. Melhuish, P. Sammons, I. Siraj-Blatchford & B. Taggar (Eds.), *Early Childhood matters: Evidence from the effective preschool and primary education project* (1st ed., pp. 92–113). Taylor & Francis Group.

- Samuel, J. & Bryant, P. (1984). Asking Only One Question in the Conservation Experiment. *Journal of Child Psychology and Psychiatry*, 25(2), 315–318. <https://doi.org/10.1111/j.1469-7610.1984.tb00152.x>
- Sandifer, C. (2003). Technological novelty and open-endedness: Two characteristics of interactive exhibits that contribute to the holding of visitor attention in a science museum. *Journal of Research in Science Teaching*, 40(2), 121–137. <https://doi.org/10.1002/TEA.10068>
- Sarter, M., Givens, B. & Bruno, J. P. (2001). The cognitive neuroscience of sustained attention: Where top-down meets bottom-up. In *Brain Research Reviews* (Vol. 35, Issue 2, pp. 146–160). Elsevier. [https://doi.org/10.1016/S0165-0173\(01\)00044-3](https://doi.org/10.1016/S0165-0173(01)00044-3)
- Sawyer, R. K. (2005). *The Cambridge handbook of the learning sciences*. Cambridge University Press.
- Scarpa, A., Venables, P. H., Raine, A. & Mednick, S. A. (1997). Heart rate and skin conductance in behaviorally inhibited Mauritian children. *Journal of Abnormal Psychology*, 106(2), 182–190. <https://doi.org/10.1037/0021-843X.106.2.182>
- Scherer, K. R. (2001). Emotions, Psychological Structure of. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 4472–4477). Pergamon. <https://doi.org/10.1016/b0-08-043076-7/017111-3>
- Schlesinger Group. (n.d.). *Dials Testing | Perception Analyzer*. Retrieved March 24, 2020, from <https://www.schlesingergroup.com/en/solutions/qualitative/research-labs/dials-perception-analyzer/>
- Schmitow, C., Stenberg, G., Billard, A. & von Hofsten, C. (2013). Using a head-mounted camera to infer attention direction. *International Journal of Behavioral Development*, 37(5), 468–474. <https://doi.org/10.1177/0165025413495749>
- Schneider, B. & Cheslock, N. (2003). *Measuring Results - A report of the Coevolution Institute "Understanding Metrics Project."* http://www.pollinator.org/PDFs/websitemaster_pdf_final.pdf
- Schofield, T. J., Parke, R. D., Castañeda, E. K. & Coltrane, S. (2008). Patterns of gaze between parents and children in European American and Mexican

- American families. *Journal of Nonverbal Behavior*, 32(3), 171–186.
<https://doi.org/10.1007/s10919-008-0049-7>
- Schotter, E. R., Tran, R. & Rayner, K. (2014). *Don ' t Believe What You Read (Only Once) : Comprehension Is Supported by Regressions During Reading*.
<https://doi.org/10.1177/0956797614531148>
- Schwan, S., Grajal, A. & Lewalter, D. (2014). Understanding and Engagement in Places of Science Experience: Science Museums, Science Centers, Zoos, and Aquariums. *Educational Psychologist*, 49(2), 70–85.
<https://doi.org/10.1080/00461520.2014.917588>
- Science Museum. (2017). *A Brief History of the Science Museum*.
<https://www.sciencemuseum.org.uk/sites/default/files/2017-10/science-museum-history.pdf>
- Scottish Government. (2009). *The Early Years Framework* .
<https://www.gov.scot/publications/early-years-framework/pages/4/>
- Seifert, A. M. & Metz, A. E. (2017). The Effects of Inflated Seating Cushions on Engagement in Preschool Circle Time. *Early Childhood Education Journal*, 45(3), 411–418. <https://doi.org/10.1007/s10643-016-0797-7>
- Serrell, B. (1997a). Time Is Indeed of the Essence. *Curator: The Museum Journal*, 40(4), 253–255. <https://doi.org/10.1111/j.2151-6952.1997.tb01310.x>
- Serrell, B. (1997b). Paying Attention: The Duration and Allocation of Visitors' Time in Museum Exhibitions. *Curator: The Museum Journal*, 40(2), 108–125.
<https://doi.org/10.1111/j.2151-6952.1997.tb01292.x>
- Serrell, B. (2002). Are They Watching? Visitors and Videos in Exhibitions. *Curator: The Museum Journal*, 45(1), 50–64. <https://doi.org/10.1111/j.2151-6952.2002.tb00049.x>
- Serrell, B. (2020). The Aggregation of Tracking-and-Timing Visitor-Use Data of Museum Exhibitions for Benchmarks of “Thorough Use.” *Visitor Studies*, 23(1), 1–17. <https://doi.org/10.1080/10645578.2020.1750830>
- Shaby, N., Assaraf, O. B. Z. & Tal, T. (2017). The Particular Aspects of Science Museum Exhibits That Encourage Students' Engagement. *Journal of Science*

Education and Technology, 26(3), 253–268. <https://doi.org/10.1007/s10956-016-9676-7>

Shaby, N., Ben-Zvi Assaraf, O. & Tal, T. (2019). A Student's-Eye View: What 4th Grade Students Describing Their Visit to a Science Museum Recall as Significant. *Research in Science Education*, 49(6), 1625–1645. <https://doi.org/10.1007/s11165-017-9669-4>

Shapiro, L. & Stolz, S. A. (2019). Embodied cognition and its significance for education. *Theory and Research in Education*, 17(1), 19–39. <https://doi.org/10.1177/1477878518822149>

Sheppard, S. L. (2011). School Engagement: A “Danse Macabre”? *Journal of Philosophy of Education*, 45(1), 111–123. <https://doi.org/10.1111/j.1467-9752.2010.00782.x>

Shettel, H. H. (1997). Time-Is It Really of the Essence? *Curator: The Museum Journal*, 40(4), 246–249. <https://doi.org/10.1111/j.2151-6952.1997.tb01308.x>

Shwe Hadani, H., Rood, E., Eisenmann, A., Foushee, R., Jaeger, G., Jaeger, G., Kauffmann, J., Kennedy, K. & Regalla, L. (2018). *The Roots of STEM Success: Changing Early Learning Experiences to Build Lifelong Thinking Skills*. http://centerforchildhoodcreativity.org/wp-content/uploads/sites/2/2018/02/CCC_The_Roots_of_STEM_Early_Learning.pdf

Siegler, R. S. & Ellis, S. (1996). Piaget on Childhood. *Psychological Science*, 7(4), 211–215. <https://doi.org/10.1111/j.1467-9280.1996.tb00361.x>

Silpasuwanchai, C., Ma, X., Shigemasu, H. & Ren, X. (2016). Developing a Comprehensive Engagement Framework of Gamification for Reflective Learning. *Proceedings of the 2016 ACM Conference on Designing Interactive Systems - DIS '16*, 459–472. <https://doi.org/10.1145/2901790.2901836>

Sinatra, G. M., Heddy, B. C. & Lombardi, D. (2015, January 2). The Challenges of Defining and Measuring Student Engagement in Science. *Educational Psychologist*, 50(1), 1–13. <https://doi.org/10.1080/00461520.2014.1002924>

Skinner, E. A., Kindermann, T. A. & Furrer, C. J. (2009). A motivational perspective

on engagement and disaffection: Conceptualization and assessment of children's behavioral and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement*, 69(3), 493–525. <https://doi.org/10.1177/0013164408323233>

Skinner, E., Furrer, C., Marchand, G. & Kindermann, T. (2008). Engagement and Disaffection in the Classroom: Part of a Larger Motivational Dynamic? *Journal of Educational Psychology*, 100(4), 765–781. <https://doi.org/10.1037/a0012840>

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R. & Obonsawin, M. (2004). Subjective experience and the attentional lapse: Task engagement and disengagement during sustained attention. *Consciousness and Cognition*, 13(4), 657–690. <https://doi.org/10.1016/j.concog.2004.06.003>

Smith, L. B., Yu, C., Yoshida, H. & Fausey, C. M. (2015). Contributions of Head-Mounted Cameras to Studying the Visual Environments of Infants and Young Children. *Journal of Cognition and Development*, 16(3), 407–419. <https://doi.org/10.1080/15248372.2014.933430>

Smith, T. J. & Mital, P. K. (2013). Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, 13(8), 1–24. <https://doi.org/10.1167/13.8.16>

Sparman, A. (2005). Video recording as interaction: Participant observation of children's everyday life. *Qualitative Research in Psychology*, 2(3), 241–255. <https://doi.org/10.1191/1478088705qp041oa>

Sridhar, P. K., Chan, S. W. T. & Nanayakkara, S. (2018). Going beyond performance scores: Understanding cognitive-affective states in kindergarteners. *IDC 2018 - Proceedings of the 2018 ACM Conference on Interaction Design and Children*, 253–265. <https://doi.org/10.1145/3202185.3202739>

Stebbins, R. (2001). What is Exploration? In *Exploratory Research in the Social Sciences*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412984249>

Swedberg, R. (2020). On the use of definitions in sociology. *European Journal of Social Theory*, 23(3), 431–445. <https://doi.org/10.1177/1368431019831855>

Tao, Y., Oliver, M. & Venville, G. (2012). Long-term outcomes of early childhood

- science education: Insights from a cross-national comparative case study on conceptual understanding of science. *International Journal of Science and Mathematics Education*, 10(6), 1269–1302. <https://doi.org/10.1007/s10763-012-9335-2>
- Taub, M., Azevedo, R., Bradbury, A. E., Millar, G. C. & Lester, J. (2018). Using sequence mining to reveal the efficiency in scientific reasoning during STEM learning with a game-based learning environment. *Learning and Instruction*, 54, 93–103. <https://doi.org/10.1016/j.learninstruc.2017.08.005>
- Taub, M., Mudrick, N. V., Azevedo, R., Millar, G. C., Rowe, J. & Lester, J. (2017). Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with C rystal I sland. *Computers in Human Behavior*, 76, 641–655. <https://doi.org/10.1016/j.chb.2017.01.038>
- Taylor, S., Jaques, N., Chen, W., Fedor, S., Sano, A. & Picard, R. (2015). Automatic identification of artifacts in electrodermal activity data. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2015-Novem*, 1934–1937. <https://doi.org/10.1109/EMBC.2015.7318762>
- Thomas Jha, R., Price, S. & Motion, A. (2020). Designing ‘Embodied’ Science Learning Experiences for Young Children. *International Conference on Human-Computer Interaction*, 12205 LNCS, 207–225. https://doi.org/10.1007/978-3-030-50513-4_16
- Tickle-Degnen, L. (2013). Nuts and bolts of conducting feasibility studies. *American Journal of Occupational Therapy*, 67(2), 171–176. <https://doi.org/10.5014/ajot.2013.006270>
- Tisdal, C. (2004). *Phase 2 summative evaluation of “Active Prolonged Engagement” at the Exploratorium* (Issue September).
- Tisdal, C. & Perry, D. L. (2004). Going APE! at the Exploratorium. In (*unpublished manuscript*) *Interim summative report* (Issue April). http://exploratorium.edu/vre/pdf/APE_Summative_Phase_1.pdf
- Tsuchiya, N. & Adolphs, R. (2007). Emotion and consciousness. *Trends in Cognitive*

- Sciences*, 11(4), 158–167. <https://doi.org/10.1016/j.tics.2007.01.005>
- Tuckey, C. (1992). Children's informal learning at an interactive science centre. *International Journal of Science Education*, 14(3), 273–278. <https://doi.org/10.1080/0950069920140304>
- Turner, P. (2010). The Anatomy of Engagement. *Proceedings of ECCE 2010*, 44(August), 59–66. <https://doi.org/10.1145/1962300.1962315>
- Ueda, K., Sekoguchi, T. & Yanagisawa, H. (2021). How predictability affects habituation to novelty. *PLoS ONE*, 16(6 June), e0237278–e0237278. <https://doi.org/10.1371/journal.pone.0237278>
- UNESCO. (n.d.). *Early childhood care and education*. Retrieved September 25, 2020, from <https://en.unesco.org/themes/early-childhood-care-and-education>
- van den Bosch, I., Salimpoor, V. N. & Zatorre, R. J. (2013). Familiarity mediates the relationship between emotional arousal and pleasure during music listening. *Frontiers in Human Neuroscience*, 07(September), 1–10. <https://doi.org/10.3389/fnhum.2013.00534>
- Van Schijndel, T. J. P., Franse, R. K. & Raijmakers, M. E. J. (2010). The Exploratory Behavior Scale: Assessing young visitors' hands-on behavior in science museums. *Science Education*, 94(5), 794–809. <https://doi.org/10.1002/sce.20394>
- Vartiainen, J. & Kumpulainen, K. (2020). Playing with science: manifestation of scientific play in early science inquiry. *European Early Childhood Education Research Journal*, 28(4), 490–503. <https://doi.org/10.1080/1350293X.2020.1783924>
- Vile Junod, R. E., DuPaul, G. J., Jitendra, A. K., Volpe, R. J. & Cleary, K. S. (2006). Classroom observations of students with and without ADHD: Differences across types of engagement. *Journal of School Psychology*, 44(2), 87–104. <https://doi.org/10.1016/J.JSP.2005.12.004>
- Vlietstra, A. G. & Manske, S. H. (1981). *Looks to Adults, Preferences for Adult Males and Females, and Interpretations of an Adult's Gaze by Preschool Children*. 27(1), 31–41.

- vom Lehn, D., Hindmarsh, J. & Heath, C. (2002). Video based field studies in museums and galleries. *Visitor Studies Today!*, V(Fall/Winter), 15–17; 20–23. <https://www.researchgate.net/publication/228603483>
- Wakefield, E., Congdon, E. L., Novack, M. A., Goldin-Meadow, S. & James, K. H. (2019). Learning math by hand: The neural effects of gesture-based instruction in 8-year-old children. *Attention, Perception, and Psychophysics*, 81(7), 2343–2353. <https://doi.org/10.3758/s13414-019-01755-y>
- Walker, F., Bucker, B., Anderson, N. C., Schreij, D. & Theeuwes, J. (2017). Looking at paintings in the Vincent Van Gogh Museum: Eye movement patterns of children and adults. *PLoS ONE*, 12(6), e0178912. <https://doi.org/10.1371/journal.pone.0178912>
- Wang, M. Te, Fredricks, J. A., Ye, F., Hofkens, T. L. & Linn, J. S. (2016). The Math and Science Engagement Scales: Scale development, Validation, And psychometric properties. *Learning and Instruction*, 43, 16–26. <https://doi.org/10.1016/j.learninstruc.2016.01.008>
- Ward, L. (2008). Attention. *Scholarpedia*, 3(10), 1538. <https://doi.org/10.4249/scholarpedia.1538>
- Wass, S. V., Clackson, K. & de Barbaro, K. (2016). Temporal dynamics of arousal and attention in 12-month-old infants. *Developmental Psychobiology*, 58(5), 623–639. <https://doi.org/10.1002/dev.21406>
- Wass, S. V., de Barbaro, K. & Clackson, K. (2015). Tonic and phasic co-variation of peripheral arousal indices in infants. *Biological Psychology*, 111, 26–39. <https://doi.org/10.1016/j.biopsycho.2015.08.006>
- Webb, C. A., DelDonno, S. & Killgore, W. D. S. (2014). The role of cognitive versus emotional intelligence in Iowa Gambling Task performance: What's emotion got to do with it? *Intelligence*, 44(1), 112–119. <https://doi.org/10.1016/j.intell.2014.03.008>
- Wechsler, D., Crabbs, L. M. & Freeman, R. G. (1930). Galvanic Responses of Preschool Children. *Pedagogical Seminary and Journal of Genetic Psychology*, 38(1–4), 203–222. <https://doi.org/10.1080/08856559.1930.10532260>

- Weisleder, A. & Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*, 24(11), 2143–2152. <https://doi.org/10.1177/0956797613488145>
- Weston, A., Morrison, L., Yardley, L., Kleek, M. Van & Weal, M. (2015). Measurements of engagement in mobile behavioural interventions? *Digital Health*, May, 1–8. https://hip.cat/emax/papers/Anna_Weston_ACMDH.pdf
- Whitehead, A. L., Sully, B. G. O. & Campbell, M. J. (2014). Pilot and feasibility studies: Is there a difference from each other and from a randomised controlled trial? *Contemporary Clinical Trials*, 38(1), 130–133. <https://doi.org/10.1016/j.cct.2014.04.001>
- Whitton, N. (2011). Game engagement theory and adult learning. *Simulation and Gaming*, 42(5), 596–609. <https://doi.org/10.1177/1046878110378587>
- Whitton, N. & Moseley, A. (2014a). Deconstructing Engagement. *Simulation & Gaming*, 45(4–5), 433–449. <https://doi.org/10.1177/1046878114554755>
- Whitton, N. & Moseley, A. (2014b). Deconstructing Engagement. *Simulation & Gaming*, 45(4–5), 433–449. <https://doi.org/10.1177/1046878114554755>
- Wiebe, E. N., Lamb, A., Hardy, M. & Sharek, D. (2014). Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Computers in Human Behavior*, 32, 123–132. <https://doi.org/10.1016/j.chb.2013.12.001>
- Wilhelm, F. H. & Grossman, P. (2010). Emotions beyond the laboratory: Theoretical fundamentals, study design, and analytic strategies for advanced ambulatory assessment. *Biological Psychology*, 84(3), 552–569. <https://doi.org/10.1016/j.biopsycho.2010.01.017>
- Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, 9(4), 625–636. <https://doi.org/10.3758/BF03196322>
- Wineman, J. D. & Peponis, J. (2010). Constructing spatial meaning: Spatial affordances in museum design. *Environment and Behavior*, 42(1), 86–109. <https://doi.org/10.1177/0013916509335534>
- Wittfoth-Schardt, D., Gründing, J., Wittfoth, M., Lanfermann, H., Heinrichs, M.,

- Domes, G., Buchheim, A., Gündel, H. & Waller, C. (2012). Oxytocin modulates neural reactivity to children's faces as a function of social salience. *Neuropsychopharmacology*, 37(8), 1799–1807.
<https://doi.org/10.1038/npp.2012.47>
- Wood, D. (2018). Commentary: Contribution of Scaffolding to Learning and Teaching: Interdisciplinary Perspectives. *International Journal of Educational Research*, 90, 248–251. <https://doi.org/10.1016/j.ijer.2018.03.005>
- Wood, E. & Wolf, B. (2008). Between the Lines of Engagement in Museums : Indiana University and the Children ' s Museum of Indianapolis. *The Journal of Museum Education*, 33(2), 121–130. <https://doi.org/10.2307/40479637>
- World Health Organization. (2020). *Improving early childhood development: WHO guideline*. <https://www.urban.org/research/publication/improving-early-childhood-development-policies-and-practices>
- World Health Organization, United Nations Children's Fund & World Bank Group. (2018). *Nurturing care for early childhood development: Linking survive and thrive to transform health and human potential*.
- Worsley, M. (2018). (Dis)engagement matters. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge - LAK '18*, 5, 365–369.
<https://doi.org/10.1145/3170358.3170420>
- Yalowitz, S. S. & Bronnenkant, K. (2009). *Visitor Studies Timing and Tracking: Unlocking Visitor Behavior*. <https://doi.org/10.1080/10645570902769134>
- Yanagisawa, H., Kawamata, O. & Ueda, K. (2019). Modeling emotions associated with novelty at variable uncertainty levels: A bayesian approach. *Frontiers in Computational Neuroscience*, 13, 2. <https://doi.org/10.3389/fncom.2019.00002>
- Yarosh, S., Radu, I., Hunter, S. & Rosenbaum, E. (2011). Examining values: An analysis of nine years of IDC research. *Proceedings of IDC 2011 - 10th International Conference on Interaction Design and Children*, 136–144.
<https://doi.org/10.1145/1999030.1999046>
- Yoshida, H. & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, 13(3), 229–248.

<https://doi.org/10.1080/15250000802004437>

Zaman, B., Vanden Abeele, V. & De Grooff, D. (2013). Measuring product liking in preschool children: An evaluation of the Smileyometer and This or That methods. *International Journal of Child-Computer Interaction*, 1(2), 61–70.
<https://doi.org/10.1016/j.ijcci.2012.12.001>

Zimmerman, C. (2000). The Development of Scientific Reasoning Skills. *Developmental Review*, 20(1), 99–149. <https://doi.org/10.1006/drev.1999.0497>

Appendix

A - Engagement Observation Tools in Children

Name of scale:	Engagement states Coding system eSOCS	
Authors:	<i>Deater-Deckard, Chang, Evans</i>	
Where is used?		
How is it used?	In game learning	
Population aimed at:	Children	
1. DIST/ATT: Distraction/Attention ("off task" or "on task")	7. A/N: Anxious/Nervous (vocalizations and/or facial expressions)	
1) Constant distraction (DIST) throughout task	1) no instances of A/N	
2) Substantial DIST, with only one or two instances of ATT	2) one or two instances of A/N	
3) DIST for more than half of task, with several instances of ATT	3) several instances of A/N, for less than half of task	
4) DIST for about half of task, ATT for about half of task	4) A/N for about half of the task	
5) ATT for more than half of task, with several instances of DIST	5) A/N for more than half of task, with several instances of non-A/N	
6) Substantial ATT, with only one or two instances of DIST	6) Substantial A/N, only one or two instances of non-A/N	
7) Constant ATT throughout task	7) Constant A/N throughout task	
2. PA: Positive affect (smiling, laughing, excitement, enjoying the task; vocalization and/or facial expression)	8. FMM: Fine motor movement (fine motor movements involving the device or task materials)	
1) no instances of PA	1) no instances of FMM	
2) one or two instances of PA	2) one or two instances of FMM	
3) several instances of PA	3) several instances of FMM, for less than half the task	
4) PA for about half of the task	4) FMM for about half of the task	
5) PA for more than half of the task, several instances of non-PA	5) FMM for more than half of the task, with several instances of inactivity	
6) Substantial PA, only one or two instances of non-PA	6) Substantial FMM, only one or two instances of inactivity	
7) Constant PA throughout the entire task	7) Constant FMM throughout task	
3. TCH: Touching of device and/or task materials	9. AGG: Aggression (physical and/or verbal) In INDIVIDUAL CONTEXT: AGG directed at device In DYADIC CONTEXT: AGG directed at device or partner	
1) No instances of TCH	1) no instances of AGG	
2) One or two instances of TCH	2) one or two instances of AGG	
3) Several instances of TCH, for less than half of task	3) several instances of AGG, for less than half the task	
4) TCH for about half of task	4) AGG for about half of the task	
5) TCH for more than half of task, with several exceptions	5) AGG for more than half of the task, with several instances of non-AGG	
6) Substantial TCH, only one or two exceptions	6) Substantial AGG, only one or two instances of non-AGG	
7) Constant TCH throughout task 3	7) Constant AGG throughout task 5	
4. PERS: Persistence (begins and completes task without breaking away, stopping, or giving up)	10. VERB: Verbalizations during task (task-related speech) In INDIVIDUAL CONTEXT: self-directed speech In DYADIC CONTEXT: self-directed speech and/or partner-directed speech	
1) no PERS; does not begin task	1) no instances of VERB	
2) Substantial non-PERS, begins task but stops soon after initiating	2) one or two instances of VERB	
3) PERS for less than half of task, with several instances of PERS	3) several instances of VERB, for less than half of task	
4) PERS for about half of the task	4) VERB for about half of task	
5) PERS for over half of the task with several instances of stopping	5) VERB for more than half of task, with several instances of non-VERB	
6) Substantial PERS, only one or two instances of stopping	6) Substantial VERB, only one or two instances of non-VERB	
7) Constant PERS, completes task with no instances of stopping	7) Constant VERB throughout task	
5. A/F: Anger/Frustration (vocalizations and/or facial expressions)	11. INT: Intrusiveness (physical and/or verbal attempts to disrupt, manipulate, or control partner)	
1) no instances of A/F		
2) one or two instances of A/F		
3) several instances of A/F		
4) A/F for about half of the task		
5) A/F for more than half of the task, with several instances of non-A/F		
6) Substantial A/F, only one or two instances of non-A/F		
7) Constant A/F throughout the entire task		

6. GMM: Gross motor movement (minor body movements [moving arms, pointing to stimuli] and major body movements [jumping, getting up and sitting down])

- 1) no instances of GMM
- 2) one or two instances of GMM
- 3) several instances of GMM, for less than half the task
- 4) GMM for about half of the task
- 5) GMM for more than half of the task, with several instances of inactivity
- 6) Substantial GMM, only one or two instances of inactivity
- 7) Constant GMM throughout task 4

12. RESP: Responsiveness to partner (contingent responding to partner's verbalizations and actions)

CODE ONLY in DYAD CONTEXT n/a: not coded (individual context)

- 1) no instances of RESP
- 2) one or two instances of RESP
- 3) several instances of RESP
- 4) RESP to about half of partner's questions/comments
- 5) RESP to more than half of partner's questions/comments, with several delays
- 6) Substantial RESP, only one or two instances of delay
- 7) Constant RESP throughout task 6

CODE ONLY in DYAD CONTEXT n/a: not coded (individual context)

- 1) no instances of INT
- 2) one or two instances of INT
- 3) several instances of INT
- 4) INT for about half of the task
- 5) INT for more than half of the task, with several instances of non-INT
- 6) Substantial INT, only one or two instances of non-INT
- 7) Constant INT throughout task

13. IND: Independence/Autonomy (leads and controls task; does not include off-task behaviours)

CODE ONLY in DYAD CONTEXT n/a: not coded (individual context)

- 1) no instances of IND; relies on partner throughout task
- 2) one or two instances of IND
- 3) several instances of IND, for less than half of task
- 4) IND for about half of task
- 5) IND for more than half of task, with several instances of following partner's lead
- 6) Substantial IND, only one or two instances of following partner's lead
- 7) Constant IND throughout task, controls task from beginning to end

Name of scale:	Observable indicators of learning engagement
Authors:	(Halliday et al., 2018)
Main criteria it uses:	<p>The tool for learning engagement observation has seven indicators of engagement:</p> <ol style="list-style-type: none"> 1. attention to instructions – how attentive children are during the initial task description and other interactions with experimenter 2. on-task behaviour – children’s maintained focus on task materials, task-relevant actions and the amount of time children remained task oriented. Success of task is not required to be considered “on task”. <ul style="list-style-type: none"> -Focus: visual attention to materials and to the examiner if he/she is talking or the child is requesting help. - Engagement with task: manipulation of materials in a way that would result in completing the task. 3. persistence – Whether children maintained engagement even when the task became demonstrably difficult 4. monitoring progress and strategy use – How flexible children were in their strategy use and how aware they appeared to be of specific problems or progress. <i>Strategic behaviour</i> - focusing on specific problems rather than aimlessly moving task materials around the table, checking work before claiming completion, asking for help on specific problem, using feedback in constructive ways. 5. enthusiasm and energy – How interested or eager children appeared to be in the task/ how energetic vs passive while engaging. (sitting up straight or expressing eagerness) 6. positive affect – Assessed the amount and intensity of physical and verbal cues of pleasure and enjoyment 7. negative affect. (sadness, anxiety and frustration – arms crossed or complains)
Population aimed at:	Children
Where is used?	In a lab-context
How is it used?	It is used in a laboratory environment to test engagement in younger children. They are presented with a Tangram task
<p>They have descriptive rationale of why each of these labels are important. The rating was a 5-Likert Scale. 1(no indication of behaviour) to 5 (high indication of behaviour)</p> <p>They describe affective-behavioural component: (enthusiasm)</p>	

Name of scale:	Visitor Engagement Framework (+ 2 additional learning behaviours)
Authors:	Shaby, Assaraf & Tal, 2017
Where is used?	Science centre
How is used?	Observed by the researchers to explore learning in science museums through interaction. What makes people engage?
Aims:	To assess the behaviour of school students, determining what sorts of behaviours they engage in and whether different exhibits elicit different behaviours
<p>Initiation behaviours</p> <p>1. <i>Doing the activity - BEHAVIOURAL</i></p> <ul style="list-style-type: none"> -In passing, not done completely -Doing the activity somewhat completely -Doing the activity without further exploration or testing of variables -Looking at the exhibit working, or someone doing the activity <p>2. <i>Spending time watching others engaging in activity or observing the exhibit – COGNITIVE?</i></p> <ul style="list-style-type: none"> -Watching the exhibit or person using the exhibit with expressed interest in the activity (facial expression or verbal) -Interested in learning outcome or in learning the activity; visitor does the activity after observing <p>3. <i>Expressing negative emotional response in reaction to engaging in activity – EMOTIONAL?</i></p> <ul style="list-style-type: none"> - Displeased with the exhibit, making negative remarks -Leaving the exhibit after a short experience or after watching others engage, showing marks of displeasure <p>Transition behaviours</p> <p>4. <i>Repeating the activity</i></p> <ul style="list-style-type: none"> -Doing the activity two to three times to attain desired outcome, to master the exhibit's function. -Enjoyment of outcome -Changing the variables once looking for a difference in outcome; becoming involved/engaged <p>5. <i>Expressing positive emotional response in reaction to engaging in activity</i></p> <ul style="list-style-type: none"> -Smiling, pleased with exhibit -Stronger signs of enjoyment such as laughter; verbal references to enjoyment -Obvious signs of eagerness to participate, excited disposition <p>6. <i>Asking others, consulting</i></p> <ul style="list-style-type: none"> Asking questions regarding the operation of the exhibit or the outcome Making general comments about the operation Not necessarily waiting for an answer <p>Breakthrough behaviours</p> <p>7. <i>Referring to past experiences while engaging in the activity</i></p> <ul style="list-style-type: none"> -Reference to past experience with exhibit or science centre -Simple reference to comparable experience in visitor's life -Reference to comparable experience in their life as well as making comparisons and deductions based on observations of similarities and differences <p>8. <i>Seeking and sharing information</i></p> <ul style="list-style-type: none"> -Calling someone over to look at exhibit, or to ask them to explain an exhibit, asking a question of staff or family member without lengthy discussion or exploration of topic -Reading signage; having conversations about exhibit and related science with staff or family member -Sharing experience and information with others by explaining the exhibit to them, giving them details about gained information and observations; discussions and questions about exhibit with staff or family member/friend <p>9. <i>Engaged and involved: testing variables, making comparisons, using information gained from activity</i></p> <ul style="list-style-type: none"> -Engaging in inquisitive behaviour, exploratory actions such as repeating the activity several times, reading signage, and asking questions; remaining on task for 2–3min 	


-Concentration and motivation are obvious; doing the activity as a means to an end, or meeting a challenge; length of interaction significant, 3 to 5 min; outcome or result of activity important

-Experimenting, testing different variables, looking for different outcomes; engages in discussion with others (visitors or staff) about the various outcomes; experience— 'flow'; involved in activity for long period of time, i.e., more than 5 min

Name of scale:	Engagement Quality Observation system (E-Qual III)
Authors:	Aguiar & MacWilliam, 2013 (Aguiar & McWilliam, 2013)
Where is used?	Kindergarten
Population aimed at:	Toddlers (1-3 yrs.)
How is it used?	
<p>Momentary time sampling procedure in which child engagement is coded every 15s during 15 min observation sessions. (the aim is to have 60 intervals at each session)</p> <p>In each observation, two codes are entered:</p> <ol style="list-style-type: none"> 1) a code for the engagement level (the quality of the child's behaviour) 2) a code for the engagement type (the target of focus of the child's behaviour) <p>At the end of each observation session, frequencies of each code are summed, and the result is divided by the total number of observations,</p> <p>Nine mutually exclusive and exhaustive engagement levels (From very engaged (1) to non-engaged (9):</p> <ol style="list-style-type: none"> 1. <i>persistent behaviour</i>: i.e. problem solving 2. <i>symbolic behaviour</i>: <i>decontextualized behaviour, pretend play and referent-free language</i> 3. <i>encoded behaviour</i>: <i>use of understanding language</i> 4. <i>constructive play</i>: <i>manipulation of objects in an intentional manner to create or build</i> 5. differentiated behaviour: active participation that does not meet criteria for higher or lower sophistication levels 6. focused attention: sustained attention at one feature of the environment for at least 3s 7. undifferentiated behaviour: repetitive behaviour 8. casual attention: attention to a sequence or range of events, object, people 9. non-engagement: passive or active undesirable behaviour such as lack of occupation, crying or aggressive or destructive acts. <p>Although it was based on the 9 scales, this study only used Sophisticated Behaviours (sum of 1,2,3,4) and non-engagement</p>	
<p>Rationale for not using it: This scale although it is aimed at children it is being discarded because it is mainly aimed at toddlers rather than younger children, and it more focused on play/free play, which could not hinge so well with science centre activities.</p>	



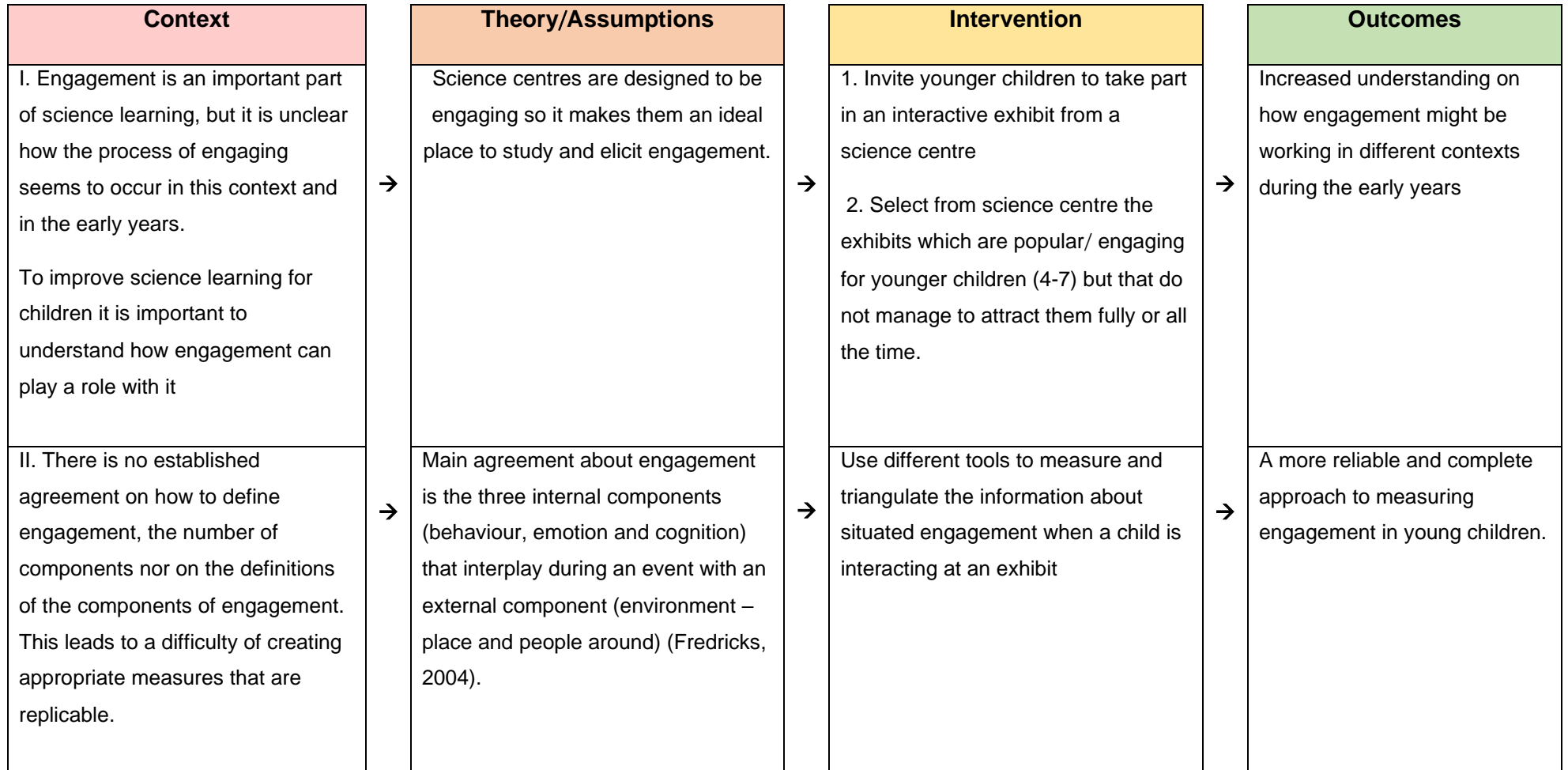
Name of scale:	Child Behaviour Rating Scale
----------------	-------------------------------------

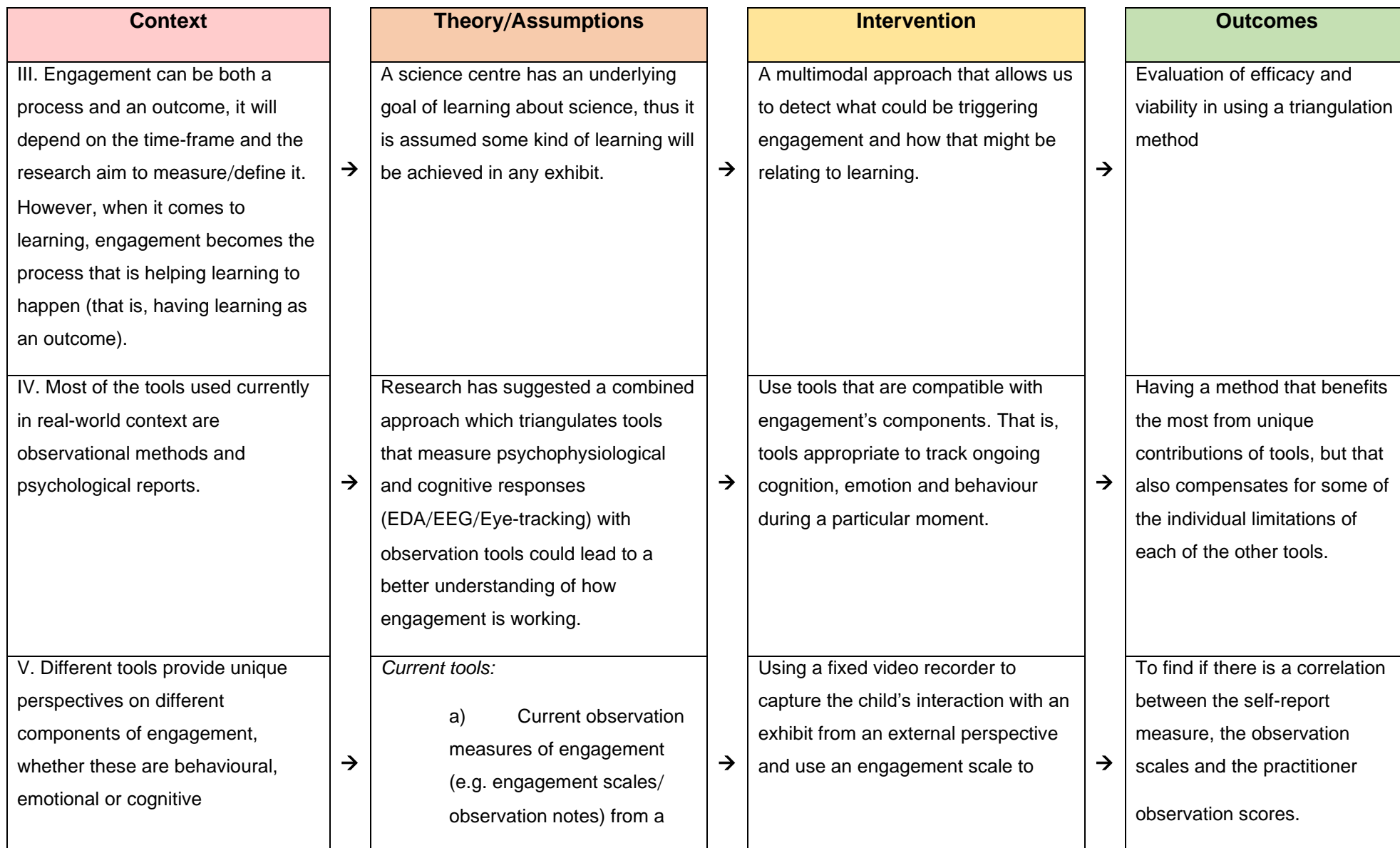
Authors:	Mahoney & Wheeden, 1999
Where is used?	School with special learning needs
Population aimed at:	Children
How is it used?	Examine effects of different styles of teacher interaction on engagement of preschool children through videotapes.
<p>It is a combination of diverse scales: scales reported by Meisels, Plunkett, Roloff, Pasick & Stiefel, 1986 and Egeland & Sroufe, 1981</p> <ol style="list-style-type: none"> 1. <i>Persistence</i>. The degree to which the child attempts or repeats actions and vocalizations. 2. <i>Attention to Activity</i>. The extent to which the child attends to the activity, independent of the quality of the child's participation or satisfaction. 3. <i>Involvement</i>. The degree to which the child participates in play activities 4. <i>Initiation: Activity</i>. The extent to which the child initiates different activities during the course of the observation 5. <i>Compliance/Cooperation</i>. The degree to which the child attempts to comply with the requests or suggestions of the teacher. 6. <i>Initiation: Teacher</i>. The extent to which the child initiates interaction with the teacher. 7. <i>Affect</i>. The child's general emotional state during the interaction. 	
<p>Rationale for not choosing: This scale is being discarded because it is mostly aimed at preschool children behaviour in a classroom, where they have more than one activity, they can take part on and where there is more interaction with other children and the teacher.</p>	

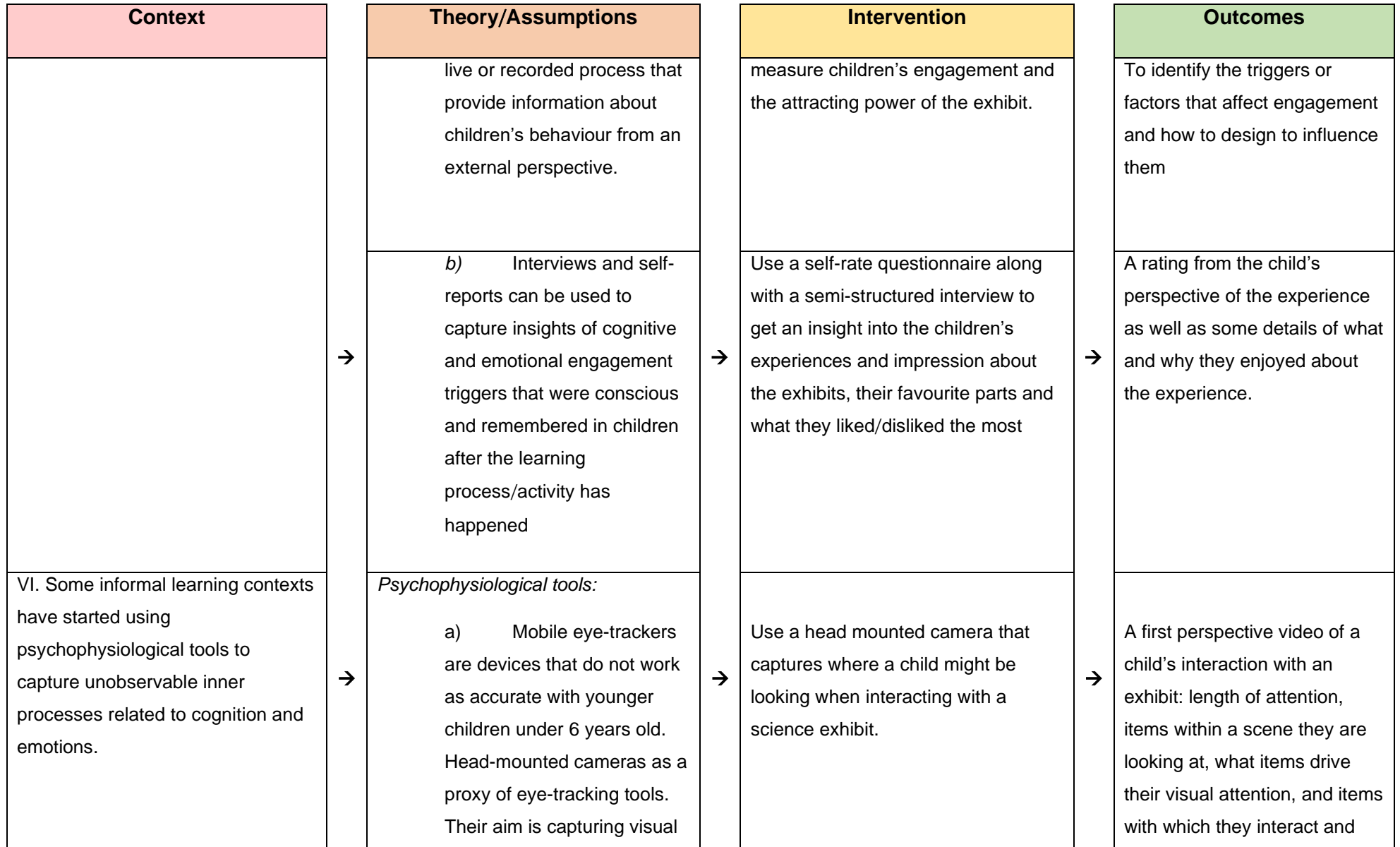
Name of scale:	Activity Engagement Survey
Authors:	(Ben-Eliyahu et al., 2018)
Main criterions it uses:	
Population aimed at:	Children (older 5 th /6 th grade)
Where is used?	Science centre and classes re science
How is it used?	17 items that encompass “affective-behaviour and cognitive engagement”
<p>During today’s activity</p> <ol style="list-style-type: none"> 1. I felt happy or excited 2. I felt relaxed or calm 3. I felt frustrated or annoyed 4. I felt tired or sad 5. I felt bored 6. I was thinking during the activity 7. I explained things to others 8. I tried out my ideas to see what would happen 9. I thought about how ideas in the activity related to other things 10. I was paying attention during the activity 11. I was doing what I was supposed to be doing 12. I did more than was required of me 13. I worked hard on the activity 14. I asked questions or talked with an adult 15. I asked questions or talked with another student <p>Did you do any of these things during today’s activity?</p> <ol style="list-style-type: none"> 16. I figured out something about science 17. I checked to make sure I understood what we were doing 	
<p>Rationale: This tool could be useful as an interview once children have finished their interaction with the exhibit. → I still think this has to be adapted to a younger children’s language/behaviour</p>	

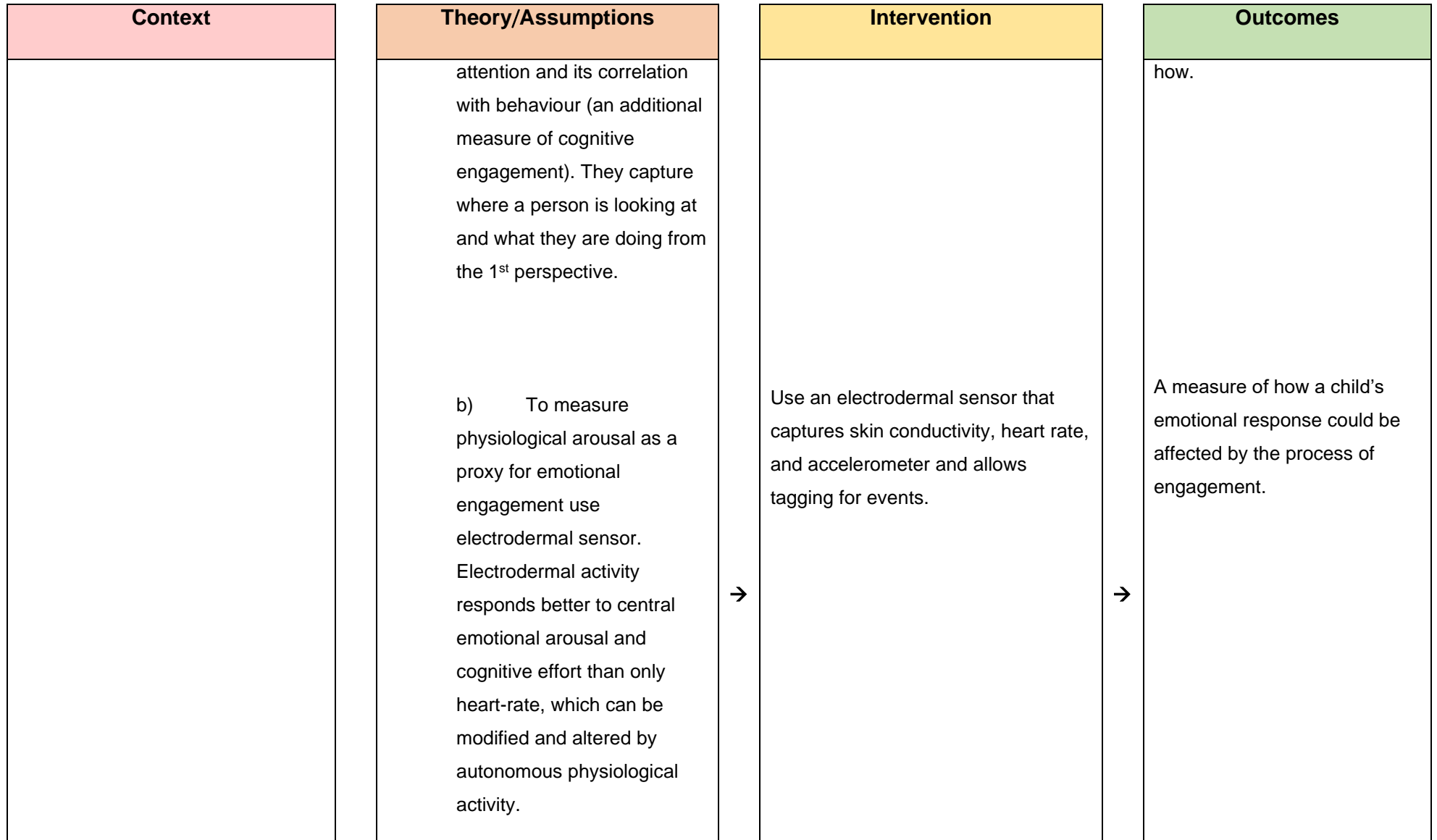


B - Rationale: Logic Model









C - Exhibit Selection From Engaging Exhibit Criteria

Criteria	Perception /Illusion	Ball through air hoop	Vital organ matching	Water table	Thermal Imaging Camera	Dino-Sand	Building genetic chains	Sand table	Cycloid Collision Physics
<i>Exhibit Aim</i>	Experience perception alterations/ wonder why it works like that	See how a constant air current can maintain something a float?	Understand where the organs are positioned	Understand how water moves in different ways and the effects this movement can have	Experience and understand that some items are warm or cold even if they don't look like it	Experience what it feels like to be a palaeontologist and discover fossils	Discovering what genes are, how they are arranged together and maybe what they can do. genes are built.	Experience how fluids work. Learning about particles. Cause-effect	Experience and understand what happens when two objects moving at the same speed collide at the same time
<i>Engagement Levels (high/low)</i>	✓	✓	✓	✓ ✓	✓ ✓	✓ ✓	✓ ✓	✓ ✓	✓
<i>Location</i>	✓	✓	✓ ✓	X	✓	✓ ✓	X	✓ ✓	✓
<i>Type of exhibit</i>	X	X	✓	✓ ✓	✓ ✓	✓ ✓	X	✓ ✓	X
<i>Theme</i>	✓	✓	✓ ✓	✓ ✓	✓	X	✓	✓	✓
<i>Interaction</i>	✓	✓ ✓	✓ ✓	✓ ✓	✓ ✓	✓ ✓	✓ ✓	✓ ✓	✓ ✓
<i>Labelling / Instructions</i>	✓ ✓	✓	X	X	X	X	✓	X	✓
<i>Inquiry allowed</i>	✓	✓	X	✓	✓	✓	X	✓ ✓	✓
<i>Attraction Power (Look and Feel)</i>	✓ ✓	✓	✓ ✓	✓ ✓	✓	✓ ✓	✓	✓ ✓	✓ ✓
<i>Exclusion criteria.</i>	✓ ✓	✓ ✓	✓ ✓	X	✓	✓ ✓	✓ ✓	✓ ✓	✓ ✓
<i>Capacity</i>	✓	✓	✓	✓ ✓	✓ ✓	✓ ✓	✓ ✓	✓ ✓	✓ ✓

D - Testing Feasibility of Engagement Scales and Fixed Video Camera

Engagement scales paired with observation have been developed and used in museums and science centres to assess visitors' engagement whilst they are at an exhibit (Barriault & Pearson, 2010; Dancstep née Dancu & Sindorf, 2018). Different engagement scales have been created lately to capture engagement when learning about science (see Chapter 2.3) where their advantages and disadvantages have been discussed (also Appendix A shows comparison table with all the engagement scales found in the literature so far), however only a few have been designed for a science centre and with young children in mind. For the purpose of this study only two will be considered: VEF (Barriault & Pearson, 2010), and eSOCS (Deater-deckard et al., 2011). Each of these engagement scales capture the components of engagement during an interaction and have been used either in a science centre/museum context or with young children, which makes them ideal tools to use with the triangulation method proposed by Azevedo (2015). Furthermore, the use of these scales could address capturing engagement directly while children interact with the exhibit and without needing any verbal communication, which is an advantage when working with early years children (Forrester, 2017). However, since none of the engagement scales above has been used both in a science centre and with a population of young children, their feasibility needed to be evaluated.

The current study aimed to evaluate the feasibility of using engagement scales with younger children during live observation and during video-observation of them interacting with a science exhibit. The data was collected to evaluate and compare how informative and useful two different engagement scales are when used in museum contexts with children in the early years interacting with different exhibits.

D.1 Method

Participants - Twelve children between 4 and 5 years old were recruited from a nursery in Glasgow. They visited the Glasgow Science Centre (GSC), accompanied by their teachers. Children were divided into four groups of three children and each team was given a specific colour for identification purposes.

Ethics - The study followed ethical guidelines from the British Psychology Society and was approved by the Research Ethics Committee of Moray House School of Education and Sport, University of Edinburgh. This study was carried out with the children who were taking part in a distinct study for the Move2Learn research project. To minimise demands on museum and children, this study used the same space and exhibits. Written and verbal assent of each child, and the accompanying adult was sought on every encounter, paying careful attention to any verbal or non-verbal signs of inhibition, reluctance, or discomfort. Parents were also asked to provide written consent prior to the visit.

Location - The study was located at GSC's 'Space X' – a separate space within the museum floor where exhibits can be moved into. This separate space was chosen to enable video recording without capturing other children. Whilst this space is separated and slightly quieter than the shop floor of the museum it is similar and arguably authentic of the museum context. The exhibits used were: (1) a *Balance Board* (BB) which requires collaboration to balance the circular board using 5 blocks. The balance board includes a spirit level with a spirit in the centre of the balance board. (2) a *Thermal Imaging Camera* (TIC) which shows the temperature of anything that is being captured by the camera, this could be either objects or people. Hand-warmers and frozen bottles of water were provided for children to explore the extreme temperature properties. This exhibit allows individual or cooperative exploration.

Tools - Two different engagement scales were chosen out of ten other scales that have been used or related to different research studies related to situated engagement (see Appendix 1 a table described the rest of the scales considered that capture engagement and the rationale for which they were not chosen). These were considered the best option to use in a science centre live-observation because they can be used in a science centre context.

- Visitor Engagement Framework - VEF (Barriault & Pearson, 2010). The VEF counts seven learning behaviours, divided into three categories that are thought of being progressive from one to another: (a) *Initiation Behaviours* - 1. Doing the activity, 2. Spending time watching others engaging in the activity. In this category participants gain some information through interacting with the exhibit potentially leading to further learning. (b) *Transition behaviours* – 1. Repeating the activity, 2. Expressing positive emotional responses in reaction to engaging in the activity. At this stage, participants are enjoying the exhibit and are showing signs of willingness to continue interaction. (c) *Breakthrough Behaviours* – 1. Referring to past experiences while engaging in the activity, 2. Seeking and sharing information with others, 3. Engaged and involved. For this final category, visitors have solidified their new obtained knowledge by making links with previous experiences and creating a meaningful experience. The rating of this framework of observed behaviours is done by an external observer through coding a paper sheet which tallies whether a behaviour was present or not. No information around validity/reliability was found for this framework.
- Engagement Scale Observation Coding System – eSOCS (Deater-deckard et al., 2011; Deater-Deckard et al., 2013). eSOCS contains 13 items used individually, whether a child is interacting alone or in a dyad. The items are: 1. Distraction/Attention, 2. Positive affect, 3. Touching of task materials, 4. Persistence, 5. Anger-Frustration, 6. Gross motor movements, 7. Fine motor movements, 8.

Aggression, 9. Intrusiveness (only when another person is present), 10. Verbalizations during the task (to himself or others nearby), 11. Anxious, 12. Responsiveness to partner and 13. Independence/Autonomy. This scale also has four additional items that rate a children dyad interaction which will not be used here because the study focuses on individual interaction. This coding system uses a 7-point Likert scale of how frequently each behaviour is happening, ranging from *No instances* to *Constant instances*. No information around validity/reliability was found for the full coding system (Deater-Deckard et al., 2014).

Procedure- Children were divided into four teams, each team was given a colour as a name. The groups visited three different exhibits in total and had 5 minutes of interaction with each exhibit after which they were interviewed by researchers from the M2L research project. All groups of children were led by a museum facilitator who would stay with them while they interacted with the exhibit and trying to facilitate as little as possible. Each group of children was only observed once. Children were first observed during their live interactions using the two engagement scales until their turn had finished and before the next team was observed. After the visit, the video recording of children's observation was coded using the same engagement scales. Due to a technical difficulty, the video-observation was only done with one of the exhibits.

D.2 Results

Feasibility during Live-Observation

During observation of the first group with the Thermal Imaging Camera it became evident that attempting to code the eSOCS and the VEF during the interaction risked information being missed since parts of the interaction were not observed while looking for items in scale and deciding what scoring criteria aligned better. Instead, a decision was made to collect data during the interaction about the relevant behaviours observed from each of the three groups. Relevant behaviours were considered as all those actions that

involved either exploring the cause-effect of their actions in the TIC, any verbalizations of their reactions, questions/comments they made to their facilitator/ peers about the TIC, and markers of any expressing emotion. Data collection of relevant behaviours was used throughout the whole study with the other children's groups. By the end of each observation the relevant notes were matched with the different criteria from the engagement scales. During the observation of children with the Balance Board exhibit, children relied more on the facilitator to guide them on their interaction, whilst they also began addressing the researcher to help them with questions or additional guidance about the exhibit.

Filling in the Visitor Engagement Framework after every observation was a fast and easy process that needed only a tick in each of the behaviours if they were present/absent, whereas filling the eSOCS after each observation required more time considering the proportion of time each behaviour had been presented. Thus VEF is a feasible tool to use during live-observation, although filling it in after an interaction still relies on observer's memory. On the other hand, eSOCS was not a feasible tool to use during live-recording due to the larger number of behaviours it contains, as well as requiring an estimation of the number of times a behaviour was present.

This live observation also emphasized that the context in which data is obtained is susceptible to alterations and last-minute changes, such as interruption of observation by interaction with children when they are not clear on what needs to be done in the exhibit, or by helping colleagues interact with children.

Feasibility during Video recording report

Due to technical difficulties videos were only available for interactions with the Balance Board exhibit. Using a video recording for capturing data with both engagement scales confirmed the potential to capture a richer data, which is less dependent on researcher's memory / ability to code everything happening in the moment. Naturally,

because videos allow observers to pause, rewind and play, video observation was a more time-consuming method than live-observation. Although both scales provided information regarding engagement during a science exhibit, eSOCS captured more instances present of the behaviours, showing more detail about the interaction than when used live, where the VEF only records Present/Absent for a behaviour, while still keeping record of whether these behaviours developed from initiation into breakthrough behaviours. Moreover, in the VEF, some items like “Seek/share information”, relied on conversations happening between the participants, which were not captured in the recording. Although this behaviour could then be coded when the child was interacting, or when they asked the facilitator about clues or hints on how to interact with the exhibit. However, the larger number of details obtained from coding the scales relying on video shows a good trade-off to overturn the sound limitation presented with live-recording. That means that video recording also enabled researcher to pause and code more behaviours while live-observation provided sound and tracking of conversations, however, some information would be missed while observing and coding the behaviours. Both scales used were feasible to capture engagement using a video recording.

Data captured by tools

When VEF was scored, it showed that most of the participants had achieved all three types of possible behaviours (initiation, transition and breakthrough), but unless additional observation notes were being kept, tracking the exact point in which the breakthrough behaviours happened would not be as easy to detect. Moreover, it contained an item called “Engagement” that was opaquely defined, and which could only be scored as present/absent, complicating the coding of engagement and muddling what construct was being measured by this scale. and also provided redundancy with the coding framework since the scale was called “Visitor Engagement Framework” and did not contribute much to the understanding of engagement. On the other hand, eSOCS was more useful during video-observation, while it also captured a wider array of behaviours but also tracking the

intensity of use of each of these behaviours. Regarding children's engagement, most of the children did not show a clear expression of emotions, neither positive nor negative.

D.3 Discussion

This feasibility study evaluated the use of the most prominent engagement assessment tools in a science centre, Visitor Engagement Framework, and compared it to a tool that was designed to capture the three engagement components although not specifically designed to use with science centres (eSOCS). The results showed that when comparing between scenarios (live and recorded) relying solely on the observer memory, to recall all the actions, and have a divided attention while filling in the scale and observing/interacting with the children during the activity impacted on missing information about the children's interaction patterns. Although the video was more time-consuming, it provided the advantage of going back multiple times and register with more detail the interaction and code the engagement scales. Also, important to keep in mind remarks from Sparrman (2005) who has used video recordings with 6-8 years old and acknowledges that it allows to see what children focus on but acknowledging that the observation from a video recording comes out of the interaction of child with technology and researcher.

Even though the VEF tool was practical, easy to learn and use rapidly, it was observed that after five minutes of interaction the VEF tool had already captured the highest engagement levels of the tool (i.e. breakthrough behaviours) both in live and recorded observation. This could potentially be since the tool was designed to capture participants' natural interaction which in an average visit would not typically go above 3 minutes (Tisdal & Perry, 2004; Serrell, 1997). The obtained results between contexts were similar, recording a behaviour as present/absent fails to provide details about the nature of the behaviour observed beyond its achievability, nor about the tracking of any relevant timestamps where there is a fluctuation of the varying levels of engagement. Although the latter could be achieved by using accompanying notes, it would become a combination of methods for data

collection. This tool was developed to be used in the science museum's floors to screen and understand visitors' behaviours throughout a whole-day visit. This result is similar to the reasoning behind the use of timing and tracking of holding time observed in Serrell, 2020, although this tool gathers an additional level of cognitive-behavioural engagement and progression into deeper levels of engagement after initiating contact with the exhibit,

In contrast, the eSOCS tool, which was developed with consideration of the three components of engagement, became more useful when videos were used for attention to detail, which in addition to the option of pausing and going back within the video helps to annotate and keep better track of anything happening and to get a better sense of the levels of the behaviour throughout the interaction. If this tool were to be used in live observation it would become very time consuming and the details dependent on the researcher's impression, memory and inhibitory control to focus only on one of the children or dyads at a time.

In conclusion, while both tools were feasible to use in a videorecording, they did not provide the same level of detail regarding engagement. VEF is not recommended for analysing a longer interaction of an exhibit since it reaches a point of saturation when open-ended exhibits are used. The use of eSOCS or an alternative scale is recommended for video recording.

E - Testing Feasibility of Head-Mounted Camera

In informal science centre environments, having a camera that records the person's point of view has been evaluated by Burris (2017) who placed an action camera to the participant's chest, however, after her findings revealed limited understanding of participant's field of vision she recommends using head-mounted camera to gather better evidence of where exactly a participant might be focusing their attention, since a chest-focused camera does not account for occasions where a visitor might be turning their head. A head-mounted camera has been suggested as a more appropriate tool for recording both looking behaviour and the potential visual environment around, particularly when the activity is related to free-play, where it is more useful to gather information from the surroundings of the child by letting them direct their attention and explore (Schmitow et al., 2013). Therefore using this method can contribute to capturing the process of engagement by providing the child's first-person view of their interaction pattern, along with the third-person view captured by a fixed video recorder, which can be triangulated to show possible triggers within the exhibit or outside the child's point of view that could be influencing their engagement.

However, literature using head-mounted cameras with children is still scarce, thus, the current study aims to test the feasibility of younger children wearing a head-mounted camera while they are exploring an interactive science museum, in particular, the study evaluates children's comfort in wearing it, the amount of information that can be captured in a frame from this type of camera and whether it is a useful proxy for mobile eye-tracking as evidenced by Burris (2017).

E.1 Method

Participants - Two siblings, one 3-year-old female and a 7-year-old male were invited to take part in this study since they covered both the higher and lower end of the chosen age range for the early years population. Both children were recruited using convenience sampling.

Ethics - The study followed ethical guidelines from the British Psychology Society and was approved by the Research Ethics Committee of Moray House School of Education and Sport, University of Edinburgh. Written and verbal assent of each child, and their parents was asked, paying careful attention to any verbal or non-verbal signs of inhibition, reluctance or discomfort.

Tools - An action camera (Yi Action 4k+[®]) was mounted onto the child's head using an adjustable head-band strap. Figure 0.1 illustrates what the camera looks like and how it fits a participant. Two hats were also used to test comfort and camera obstruction, the first one was a cap, which is more commonly used by children, and an explorer-like hat, which is assumed to be more comfortable and has a flatter top. The camera was connected wirelessly to an iPad that displayed the same image as the action camera.

Figure 0.1

Head-Mounted Camera



Note. Head-strap elastic band with camera (*left*).

Example of child wearing it (*right*).

Procedure - This study was conducted on the ground and third floor of the National Museum of Scotland, since these floors have more interactive exhibits in contrast with the other floors that are more object-based. Each participant was asked to wear the head-mounted camera for 5-10 minutes while both of them walked and interacted freely around the museum. First both children were asked to put the head-camera with the head-strap only

to test the comfort and grip of the headband, whether it would slide onto the forehead or to the sides/back of the head. They were also asked to wear it on top of two different hats (an explorer-like hat and a cap) to test both the comfort of the camera and the grip on top of another surface. This would also help to test if any parts of the hat would be visible or interfered with the recording frame.

After the visit to the museum, the video obtained from the head-mounted camera was edited by using a 3x3 grid to mark different locations within the frame. This was also used to look for focus of attention (Figure 0.2).

Figure 0.2

Example Grid Used on First-Perspective Video



Note. Grid was superimposed on video, leaving two clear squares simulating focus of attention

E.2 Results

Wearability of head-mount

The main things evaluated regarding the head-mounted camera were the grip, camera angle and comfort while using it. While both children interacted with the different exhibits, the head-mounted camera with only the head-mount remained in its place. However, when it was placed on top of the explorer-like hat, the head-strap was less stable as the hat would move and slide slightly. When children wore the cap with the head-mount, the head-mount had a better grip, however, the screen of the cap would be within the camera frame.

According to the participants, the camera felt more comfortable when it was on top of a cap as they could not feel the strap around their forehead. However, the size of the cap peak was large and interfered with the image frame. Even if the cap was reversed to avoid the peak this would impact on children's comfort by having the head-strap on their forehead. The camera angle could also change depending on size and type of cap and whether the top involved angling the camera to capture the scene better.

Performance of head-mounted camera

The angle of the camera was also found to be important, requiring an angle towards the floor in order to track where the child was looking at and capture how/where they were interacting. It was ascertained that if the remote-app disconnected between the mobile device and camera, the recording remained intact. The quality of the recording was high, the audio was crisp and clear, potentially not requiring any additional microphones, and the image was highly defined, stable and clear. The camera recorded good quality audio, particularly from the first-person perspective. Albeit was observed that light must be considered when positioning the child, as the camera could project a shadow on top of the activity and interfere with the observation analysis. Additionally, the head-mounted camera did not cause interference from other observing children, some glances were observed, but no interruptions nor pointing to the child nor the head-camera occurred. Figure 4.3 illustrates an example of a scene from the head-mounted camera, particularly highlighting how the first-person perspective provides a better idea of how the child is directly experiencing an exhibit when at the science museum. It can also be observed that the wide-angle lens from

the camera includes more of the environment where the child interacts, as well as showing how the angle of the camera has the potential to observe how children are using their hands.

Figure 0.3

Example Snapshots From Head-Mounted Camera at National Museum of Scotland



Note. Snapshots taken from a video recording whilst child explored ground floor of the National Museum of Scotland.

Head-mounted camera markers for visual attention

The head-mounted video demonstrated how the first-person perspective could captured in a comparable way to eye-tracking tools, particularly those that allow tracking where visual attention is placed and for how long (i.e. fixations) and where and when this attention shifts (i.e. saccades). With the obtained video a 3x3 translucent grid was superposed on top of the video to serve as visuo-spatial referent, a highlight for only the middle and lower centre squares was used to simulate the focal point of attention in what children were observing (Figure 0.4). From this grid particular markers of visual attention were proposed as a proxy to some measures used by eye-tracking measures (Table 0.1).

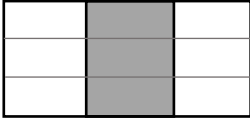
Figure 0.4

Comparison of Head-Mounted Camera Video With and Without Grid



Table 0.1

Head-Mounted Camera Cognitive Markers

<i>Marker Description</i>	<i>Type of measure obtained</i>	<i>Rationale</i>
<p>Focus of Attention Length of the scene directed at task- relevant activities in exhibit.</p> <p>The centre defined as the middle section from a 3x3 grid divided into thirds.</p> 	<p>-Quantitative: Proportion of time looking at different objects in exhibit</p> <p>Qualitative: Descriptive data of objects of interest and actions performed with them.</p>	<p>Duration of focus of attention could provide reference to where visual attention is being guided within an exhibit.</p>
<i>Marker Description</i>	<i>Type of measure obtained</i>	<i>Rationale</i>

Shift of focus of attention

When the centre of the scene is shifted rapidly to left or right third.

Quantitative:

1. Amount of task-relevant* changes
2. Amount of task-irrelevant* changes
3. Length (ms) of changes
4. Number of objects

Qualitative:

Descriptive data of objects that shifted attention.

It is a potential proxy for eye-movements, particularly when focusing on the centre of the scene helps narrowing the focus of attention.

**Task relevant:* Any shifts within the physical space of the exhibit, directed to different objects or agents interacting with the exhibit.

**Task irrelevant:* Any shift that moves away or is not directed at the exhibit or someone interacting with the exhibit.

E.3 Discussion

The head-mounted camera was a feasible tool to use with young children from the age of 3 and 7 years old. This indicated that children within that age range should have no significant difficulties wearing the head-mount while moving freely around a science exhibit. Whilst children reported using a hat as the most comfortable option when wearing the head-mount, this might not be recommended for future use given the reduced awareness of wearing the camera, which could lead to a child removing the hat and accidentally damaging the equipment. Furthermore, a reversed cap would impact the angle of the camera, the stability and visibility of the image, and for sanitary reasons, it would not be possible to share a cap continuously on the same day with different children. However, children still reported wearing the head-mount as comfortable, not applying much pressure to their head and not feeling the head-mount was too intrusive as they interacted, which was behaviourally

evidenced by none of the children touching the camera while they interacted. Furthermore, the superimposed grid helped to separate the scene into different planes that help as a visual scene- referent of where the child's attention is located. The spotlight section of the grid also acts similarly to the 2-5 degrees focal point found in eye-tracking foveal and parafoveal vision measures, outside of which images start losing sharpness and quality (Rayner, 2009).

The video from the head-mounted camera allowed tracking the focus of attention in the centre of the image and the shifts in focus, which was confirmed when the grid was superposed on the video and the clear centre of the video was followed. Although the grid facilitated following the focus of attention over a visit in a science museum, it still should be considered similar to eye-tracking, though not as precise and should be used with purposes that do not require in-depth tracking of the visual scanning pattern of a particular scene (Lai et al., 2013). The video also captured the perspective from the environment where the child was interacting. This aligns with the suggestion from Burris (2017) regarding the potential of using head-mounted cameras to track visual attention.

Overall, this tool is feasible to use with children in the age range of 3-7 years old whilst they freely interact with different science exhibits.

F - Testing Feasibility of Electrodermal Sensor

There is little literature around ambulatory devices with early years children (Hedman, 2008; Malmberg et al., 2019; Sridhar et al., 2018), that would help in evaluating whether the sensor would be comfortable while performing tasks, if the quality of the data recording would be sufficient for analysis, or how much movement would it allow children to do whilst using/playing hands-on items. There is also a lack of research when it comes to electrodermal activity and early years children, as addressed in Section 2.3, for which

there are no established parameters of skin conductivity in children, less so in ambulatory or wearable devices that measure EDA from the wrist in comparison to the fingers or palm.

Therefore, a feasibility study was needed to examine whether an EDA sensor can be used with the youngest age from the early years age group (i.e. 3 years old) as well as the feasibility of use while interacting with hands-on tasks. If this can be possible, then it is assumed that the device would also work well with older children within the early years. The current study aimed to test the feasibility of a child using a wearable psychophysiological sensor while interacting with a science programming game. That is, evaluating fit, comfort and data quality recording and processing.

F.1 Method

Participants – A 3-year-old girl was invited to the University of Edinburgh to take part in this study. She was conveniently sampled from a child of staff within the University of Edinburgh.

Ethics– The feasibility study was approved by the Research Committee Ethics of Moray House School of Education and Sport, University of Edinburgh. Written and verbal assent and consent was sought from both parents and the participants.

Tools – An electrodermal activity sensor and the software needed to process the data were evaluated.

- Empatica E4® Sensor – Wearable wrist-band device. (Figure 0.5) – This device contains four sensors to track 5 different psychophysiological markers: galvanic skin response sensor to measure electrodermal activity, accelerometer to measure movement, photoplethysmography to measure heart-rate variability and blood-volume pulse, infrared thermophile to capture peripheral skin temperature and movement using an accelerometer. The device also contains a button in the centre that when pressed this is recorded it as an event marker.

The E4 sensor can also connect to a mobile device via Bluetooth and stream the real-time data acquisition to the screen.

Figure 0.5

Empatica E4 Sensor



- Ledalab V3.4.9 – MATLAB- based open-source software for skin conductance analysis that allows importing, cleaning and processing of the electrodermal raw activity using a Convolution-Deconvolution method that separates the phasic from the tonic data (Benedek & Kaernbach, 2010b, 2016).

Procedure - The participant was asked to wear the electrodermal sensor for 12 minutes while playing with a hands-on programming robot game for children. First, the sensor was fitted, testing if the band closed comfortably without being too tight nor too loose. To do this the child was asked to move her hand, shake it and close into a fist. To capture a baseline of her electrodermal activity, the child chatted to the researcher for five minutes before playing the game. The sensor was placed from the beginning of the child's participation and taken off at the end of the interview. After the calibration period was finished, the child had to solve a maze using an electronic toy that was programmable by pressing buttons on its back to move in the direction of the arrows. The E4 sensor was also connected by Bluetooth to a remote tablet to test the live-recording screening that showed what the sensor was recording. Finally, after the recording was done, the data was downloaded and the Ledalab software was used to process the electrodermal activity data.

Figure 0.6

Empatica E4 Sensor Fitted on Child's Hand

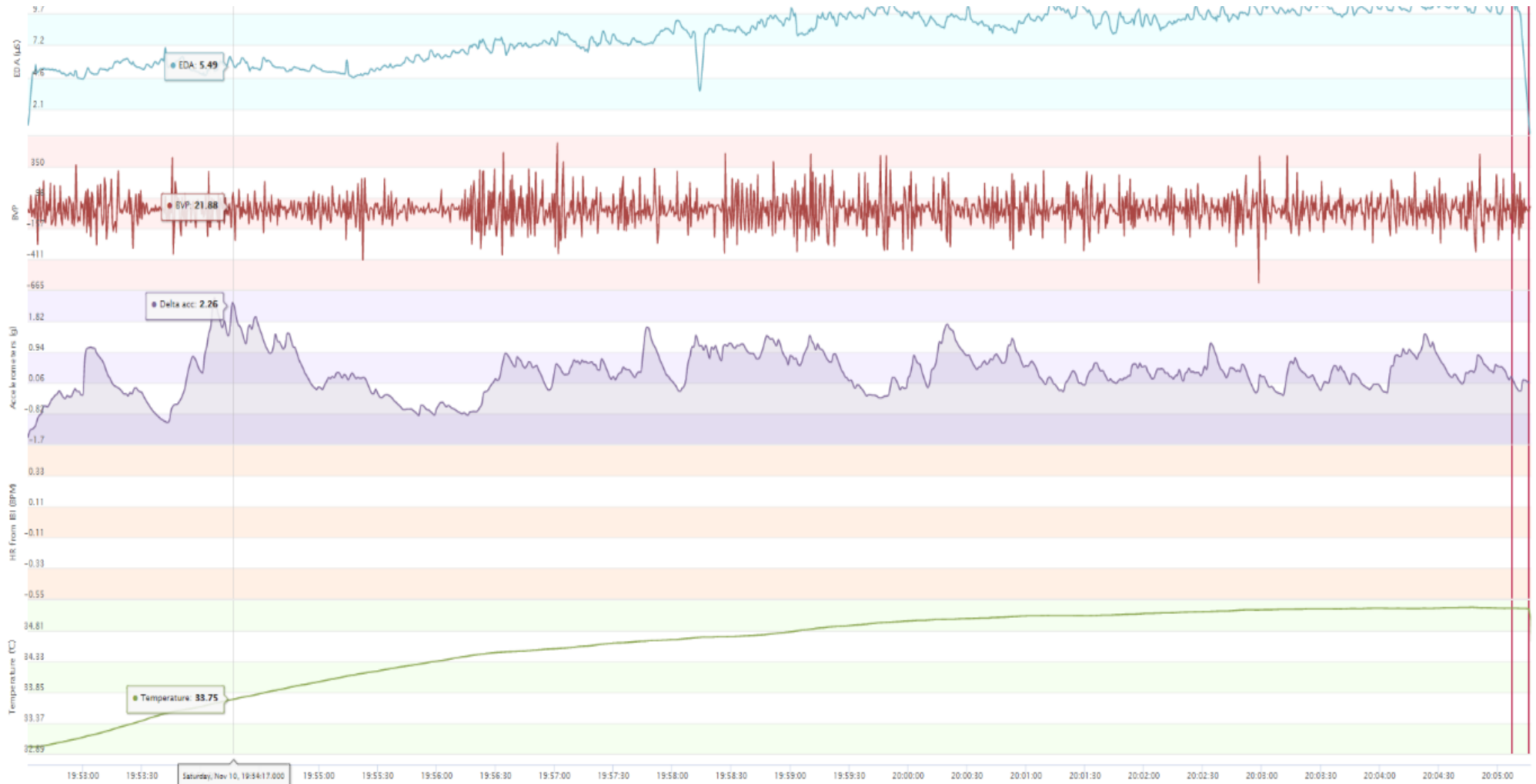


F.2 Results

Wearability of Electrodermal sensor. The E4 Empatica device can be worn by children from 3 years old. The size of the child's wrists did not hinder the sensor's grip, allowing a comfortable fit on the wrist without the device sliding off. When the child was interacting with a hands-on toy robot, the sensor remained on the wrist and did not intrude even when another child came into the room. When removing the sensor, the metal heads although harmless, could leave a small mark on the wrists that fades away over minutes. Figure 4.6 illustrates how the wearable sensor was positioned in the early years' child.

Figure 0.7

Visualization of Data Obtained from E4 Empatica Sensor



Note. Visualization contains all sensors from E4. Data is shown for EDA sensor in top row. Next row shows PPG sensor. Third row shows accelerometer data. Last row shows temperature

Evaluation of Data Collection. The sensor captured quality data, particularly since the visual inspection did not show missing data, nor pronounced activity peaks with sudden drops to negative values, which would indicate the data was noisy. The results also showed a user-friendly interface to visually inspect the data after capture. Having the device connected via Bluetooth did not interfere with the recording process. There was also no interference in the data recording when the device button was pressed to “mark” an eventful event during her interaction. Figure 0.7 shows the overall data obtained from the interaction and how the events are shown within the interface.

Measures to consider for future use. After visual inspection of the data, the measures presented in Table 0.2 were selected to use for their potential to inform emotional engagement in a larger study.

Table 0.2

Markers for Electrodermal Activity

Measure Obtained	Description	Output	Rationale
<i>Baseline</i>	5 minutes of rest prior to any tasks – sit comfortably while doing a relaxing task.	An average level of Micro-Siemens activity	It is important to have a baseline level to compare to any other events during the activity and evaluate if there were significant changes.
<i>Skin Conductance Responses (phasic activity)</i>	Sudden peaks of electrodermal activity throughout the task. (Highest values every amount of time)	Values of electrical activity peaks (above baseline) in specific periods of time	These tend to be sudden event-related responses that could point to specific important events happening throughout the interaction.

Measure Obtained	Description	Output	Rationale
<i>Skin</i>			
<i>Conductance Level (tonic activity)</i>	Level of skin conductivity in absence of any significant phasic activity	Numeric values of electrical activity over 5 minutes of activity	In the scenario where there were no sudden peaks, but a slow smooth increase, could also be an indicator of total affective engagement.

F.3 Discussion

This feasibility study showed that the E4 sensor is feasible to be used in children as young as 3 years old even if they are using their hands while interacting with a device (i.e. programming robot) and they have to walk around while wearing it. Asking the child to last five minutes without moving or doing much was not easy, however, this task was easier when the child was able to see the live recording of the EDA sensor. For future studies a shorter baseline time is recommended, which could then be similar to the calibration time (~1min) required by the sensor to set an internal baseline. Whilst interacting with their toy the child also did not try to remove or touch the device which could indicate how it was comfortable enough to ignore after the novelty of wearing the sensor has passed. Furthermore, when the data was analysed using the E4 manager visualization, the visualization did not allow phasic and tonic to be easily separated EDA activity, thus the visualization is only recommended for short inspections of the data and more in-depth analysis is recommended. The analysis software Ledalab provided an initial overview that divided both types of EDA activity and allowed for separate inspection. The Ledalab software also provided a list with the amplitude and time-stamp of each of the SCR peaks after adjustment for a significant amplitude threshold where peaks needed to have more than 0.1 to 0.5 μs to be considered (Boucsein et al., 2012) since the sensor is used in a hands-on

setting, the threshold used was higher than the average 0.1 recommended to avoid noise data due to movement. The obtained time-stamp will facilitate triangulation of events with the time-stamp from video recordings in the main study. This software also provides the amplitude scores in a standardised scale, which could help when comparing between participants by homogenising the range of values for each participant. Furthermore, the other three sensors included in the E4 device do not provide information that could be informing the process of engagement, for example, the accelerometer provides information about wrist movement, however, when interacting with a hands-on exhibit it is likely this measure would not uncover any additional information from the movement observed from a fixed camera or tracked during behavioural video-coding. As a conclusion, since a child from the youngest age range was able to comfortably wear an E4 device while interacting with a tangible object, it can be a good indicator of the feasibility of use by children that age and older for the main study.

G - Self-Report Questionnaire – Item Rationale

Pre-exhibit interaction – Before any type of interaction	
<i>To evaluate any type of pre-cursor of engagement: Interest, curiosity, expectations, previous knowledge</i>	
Question	Why ask this?
<i>Have you been here before? Do you come often?</i>	To find out if it is a first visit or a repeated experience
<i>What do you think/know about science?</i>	To know about their understanding/ feelings or opinion about science
<i>What are your favourite things you like learning about?</i>	Try to map their interests
<i>How do you feel about coming/playing in the science centre?</i>	Are they excited or have positive feelings towards the experience?
<i>What do you want to see in the science centre today?</i>	Try to map their interests and expectations of the science centre
<i>What do you think is going to happen?</i>	Know about their expectations
<i>Do you like science?</i>	Know about their feelings towards science

Post-exhibit interaction - After each exhibit.	
<i>This section is to measure “engagement” as perceived by the child. It involves child’s self-perception of their experience during the interaction but under a guided structure.</i>	
Question	Why ask this?
<i>How do you feel about the exhibit?</i>	<i>To capture their self-perception of enjoyment or not towards exhibit</i>
Do you think it was: <ul style="list-style-type: none"> • Interesting or boring? • Easy/hard? 	To record child’s self-perception of own experience similar or relevant to engagement (at least cognitive-behaviourally) as well as to self-perceived challenge
What were you trying to do with the sand?	To detect the intentions behind observed actions
What was your favourite part?	In case they have a particular preference different or in case they were not as specific as previous item

What was your least favourite part?	To detect which parts of the exhibit they are less interested about and see if that correlates with lack of engagement/learning
What would you tell your friends about it?	Knowing what they would say to a 3 rd person reduces agreeability bias of participant's opinion
What do you think your friends would say about the exhibit?	To know what emotions or perception of the exhibit. Again, to reduce agreeability bias, referring to a 3 rd person may capture a more honest answer from participants.

Post-experience - After transfer task test (and after earned sticker)

To evaluate what the outcomes of the experience could have been and if they changed in comparison with their expectations and opinions at the beginning

RELEVANT INFORMATION ABOUT EQUIPMENT

Head strapped camera

Do you feel wearing the hat was...

- comfortable/annoying?
- Light/heavy?

Do you think it made it harder to play?

To evaluate what the child felt about using the equipment and to decide how viable or feasible is to use it.

E4 sensor

Do you feel wearing the sensor was...

- comfortable/annoying?
- Light/heavy?

Do you think it made it harder to play?

To evaluate what the child felt about using the equipment and to decide how viable or feasible is to use it.

OVERALL EXPERIENCE

Question

Why ask this?

Did you enjoy taking part in this study today?

To know their overall experience in the exhibits

After being a part of this, do you think it changed how much you like science?

To evaluate if they developed further interest from their interaction in a longer term

Do you want to know more about what is happening in this exhibit?

To explore how much curiosity is left after interaction with the exhibits

H – Behavioural Codes for Triangulation Feasibility

Behavioural labels					
Measure Obtained	Description	Output	Type of measure	Rationale	Notes (considerations)
<p>Time on task/ off- task (cognitive)</p> <p><i>Distraction /attention - (eSOCS)</i></p> <p><i>On-task behaviour - Learning engagement coding scheme (LECS)</i></p>	<p>eSOCS -Total amount of time children doing task-relevant activities.</p> <p>LECS -The extent to which the child maintains focus* on task and active productive engagement with the task** throughout the session.</p> <p>* <u>Focused</u> → involves visual attention to materials and to the examiner if he/she is talking or if child is requesting help.</p> <p>** <u>Engagement with task</u> → involves manipulation of materials in a way</p>	<p>eSOCS</p> <ul style="list-style-type: none"> • Time stamp when it starts and when it ends. • Length of periods of on- task / off-task • Descriptive label of on/off task matching periods of time stamps <p>LECS</p> <ul style="list-style-type: none"> • Time stamp when period starts and when it ends according to their definition. • Code from 1-5 that matches the description of intensity of activity (see LECS). It will vary on intention of focus and engagement with task relevant or 	<p><i>Quantitative:</i></p> <ol style="list-style-type: none"> 1. Number of events in 5 min for on/off task 2. Length of each event. 3. Time stamp 4. Proportion of time of on/off task behaviour → Level in 1-7 scale for eSOCS. 5. Code from 1-5 matching the description for LECS 	<p>A percentage of how much of the total time they did task-related activities as a proxy for “total engagement”</p> <p>- It is also a behaviour that has been established in two different behavioural engagement scales and commonly used when measuring engagement.</p> <p>The time stamps can help triangulate quantitative information with EDA or qualitative description with Head-camera</p>	<p>*<u>Time on task</u> – observing the exhibit, walking around it, touching the materials in order to understand cause-effect of what can happen in exhibit</p> <p>*<u>Time off task</u> – when child is looking away from exhibit for more than 30 sec, when they are playing with materials with exhibit-irrelevant topics, looking to talk to researcher or someone else, when they verbalise, they want to stop.</p> <p>→ Child does not need to be successful to be on-task. Just working with some level of concentration / effort.</p>

	that would result in completing the task.	irrelevant behaviours.			
<p>Touching task-materials (eSOCS) (behavioural)</p>	<p>Amount of times child touches exhibit and/or task-relevant materials throughout total interaction time (5min).</p>	<ul style="list-style-type: none"> • Time stamp when it starts and when it ends. • Length of touch • Description of object 	<p>Quantitative:</p> <ol style="list-style-type: none"> 1. Number of events in 5 min 2. Length of each event. 3. Time stamp 4. Proportion of time touching task- relevant 5. Level in 1-7 scale <p>Qualitative:</p> <ol style="list-style-type: none"> 1. Description of object/part of exhibit being touched 	<p>-The start/end code “stamp” in video can help triangulate with EDA and head-mounted camera.</p> <p>Even if child is looking away, touching exhibit could provide an indication of child passive engagement with exhibit as spatial “holding place” for attention (Dancstep, 2018)</p>	<p>It will be scored (1-7) which can be assigned depending on the proportion of time children touches task-related materials.</p> <p>*Touching does not necessarily mean interacting with the exhibit or materials, just placing hands.</p> <p>*Touching also does not mean the interaction is categorised as task-relevant or task irrelevant.</p>
<p>Fine motor manipulation (eSOCS) (behavioural)</p>	<p>Fine motor movements involving the device or task materials</p>	<ul style="list-style-type: none"> • Time stamp when it starts and when it ends. • Length of manipulation • Description of object 	<p>Quantitative:</p> <ol style="list-style-type: none"> 1. Number of events in 5 min 2. Length of each event. 3. Time stamp 4. Level in 1-7 scale 	<p>-This item could reflect active interaction with the exhibit.</p> <p>The start/end code “stamp” in video can help triangulate with</p>	<p>It will be scored (1-7) which can be assigned depending on the proportion of time children manipulates task-related materials.</p>

				EDA and head-mounted camera.	
<p>Persistence (eSOCS) (LECS) (behavioural)</p>	<p>eSOCS</p> <p>-Begins and completes task without breaking away, stopping or giving up.</p> <p>LECS</p> <p>-This code comes into play when the child has some difficulty with a task – it can be an easy part, but the child has trouble with it or a more difficult component of the task.</p>	<p>eSOCS</p> <ul style="list-style-type: none"> • Time stamp of when persistence behaviour starts and when it ends. • Length of persistence periods <p>LECS</p> <ul style="list-style-type: none"> • Time stamp when period starts and when it ends according to their definition. <p>Code from 1-5 that matches the description of intensity of activity (see LECS).</p>	<p>Quantitative: eSOCS</p> <ol style="list-style-type: none"> 1. Number of events in 5 min 2. Length of each event. 3. Time stamp 4. Level in 1-7 scale <p>Qualitative: LECS</p> <ul style="list-style-type: none"> - Code description level (1-5). - Description of behaviour when child was being persistence or not. 	<p>- An item that illustrates how many tries is a child going to have before changing to a different activity of the exhibit/activity.</p> <p>- The start/end code “stamp” in video can help triangulate with EDA and head-mounted camera.</p>	<p>eSOCS</p> <p>It will be scored (1-7) which can be assigned depending on the proportion of time children persists on the task.</p> <p>- As part of LECS, in a 1-5 scale, each level has particular criterions.</p>
<p>Enthusiasm/ Energy (LECS) (emotional)</p>	<p>Quality of child’s involvement with the task → appearing eager and enthusiastic to try each problem presented (regardless of success).</p>	<ul style="list-style-type: none"> • Time stamp when detected • Total amount of times detected • Code from 1-5 that matches the description of 	<p>Qualitative : LECS code description level (1-5).</p> <p>Quantitative:</p> <ol style="list-style-type: none"> 1. Time-stamps of when enthusiasm is detected. 2. Total amount of times detected 	<p>-Item that can highlight affective engagement and differentiate from “just doing” to “doing with motivation”.</p> <p>-The start/end code “stamp” in video can help triangulate with</p>	<p>- As part of LECS, in a 1-5 scale, each level has particular criterions.</p>

		intensity of activity (see LECS).	3. Duration of enthusiasm	EDA and head-mounted camera.	
Negative affect (LECS) (emotional)	<p>Shown through facial expressions, involving frowning, narrowing of eyes or a dejected look.</p> <p>May include frustration, anger, annoyance, sadness and boredom.</p>	<p>*Time stamp of events.</p> <p>* Duration of event.</p> <p>* Description of expression.</p>	<p>Qualitative:</p> <ol style="list-style-type: none"> 1. LECS code description level (1-5). <p>Quantitative:</p> <ol style="list-style-type: none"> 1. Time-stamps of when enthusiasm is detected. 2. Total amount of times detected 	<p>Useful to code other aspects of an emotion, such as facial expression, voice tone.</p> <p>The start/end code “stamp” in video can help triangulate with EDA and head-mounted camera</p>	
Monitoring progress-Strategy use (LECS) (Cognitive)	<p>The extent to which the child is aware of their own progress solving the task and/or can recognise a problem preventing a correct solution and use problem-solving strategies.</p>	<p>* Time stamp of events.</p> <p>* Duration of event.</p> <p>* Description of strategy.</p>	<p>Qualitative:</p> <ol style="list-style-type: none"> 1. LECS code description level (1-5). 2. Description of particular behaviour <p>Quantitative:</p> <ol style="list-style-type: none"> 1. Time-stamps of when enthusiasm is detected. 2. Total amount of times strategy use is detected 	<p>Provides further insight in cognitive processing behind the observed interaction with exhibit.</p> <p>The start/end code “stamp” in video can help triangulate with EDA and head-mounted camera</p>	<p>This behaviour could be changing depending on the instruction and the facilitation→ This could be relevant for later studies.</p>

I -Self-Report Questionnaire





THE UNIVERSITY of EDINBURGH

Questionnaire "Understanding engagement"

Before playing with the exhibit.



Ask children the following questions, write and record the answers they provide.

Have you been to the Science Centre before?	Yes / No
(if answer Yes):	-Do you come often? -How many times?
What do you think about science?	
What are your 3 favourite things you like learning about?	
How do you feel about coming to play in the science centre today?	 Awful Not very good So-so Really good Amazing
What do you want to see in the science centre today?	
Do you like science?	 Really don't like it Don't like it Not sure Like it Really like it

Participant code:

After playing with exhibit



Ask children the following questions and record the answers they provide.

<p>How do you feel about the exhibit?</p>	 <p>Really don't like it Don't like it Not sure Like it Really like it</p>
<p>Do you think it was:</p>	<p>Interesting / Boring / Fun / Other? + Easy / Hard</p>
<p>What were you trying to do with the sand?</p>	
<p>What was your favourite part?</p>	
<p>What was your not so favourite part?</p>	
<p>What would you tell your friends about it?</p>	
<p>What do you think your friends would say about the exhibit?</p>	 <p>Awful Not very good So-so Really good Amazing</p>

Participant code:

After complete experience

Ask children the following questions and record the answers they provide.

<i>Head strapped camera</i>	
Do you feel wearing the hat was...?	Comfortable / Annoying + Light/ Heavy
Do you think it made it harder to play with the exhibit?	Yes / No
<i>EA sensor</i>	
Do you feel wearing the sensor was...?	Comfortable / Annoying + Light/ Heavy
Do you think it made it harder to play with the exhibit?	Yes / No
<i>Overall experience</i>	
Did you enjoy taking part in this today?	 Really don't like it Don't like it Not sure Like it Really like it
<i>Attitude/Interest change</i>	
After being a part of this, do you think it changed how much you like science?	 Really don't like it Don't like it Not sure Like it Really like it
<ul style="list-style-type: none"> Do you want to know more about what is happening in this exhibit? Yes / No Is there any other thing you would like to know or ask about this exhibit or about why I am here? 	

Participant code:

J - Theory Behind the Codes for Engagement Codebook

ELAN tier	How do they support definition of engagement	Assumptions I have	What do I expect from the data?
Touching	Touching an object can contribute to the understanding of sustained attention and some of the indicators behind it.	<i>Touch = directed attention.</i> That is, there is a volition and a decision from the child to pose their attention on certain objects, while revealing a certain intentionality to act with/on them.	Duration of episodes of interaction with exhibit
Objects/ Items being touched (different independent tiers after pilot)	Specific Components: Behavioural and cognitive engagement. Because they reflect voluntary movement and probably an intention to act, respectively.	<i>Object touched = directed attention + intention to act</i> The objects used might reveal more of the situational thinking of children by deciding to choose certain objects to act in a specific way and not others in a different way.	Provide a list of objects where child places direct attention and for how long.
FMM (Fine Motor Manipulation)	Actions reveal the purpose/intention a child is giving to a certain object by the way it is using. Specific Components: Behavioural and cognitive engagement Because they are both observable, reflect an intention and a purpose to act. Also, it can reflect motor planning, observing cause/effect of action and use of motor schemas.	<i>Fine motor movement = related to the developmental term of fine motricity / planned action</i> FMM reveals knowledge of skills and dominance of certain actions and use of objects.	I expect it to provide labels easier to generalize for each exhibit as main actions performed in it and whether these are similar or not between children. Also if these actions are particularly related (in order of appearance) to persistence or strategy use.
Strategy Use	Supports situational thinking by revealing either a problem-solving solution, exploring and testing different action-consequence (by modifying slightly actions from previous ones, either in the movement or in the object used) Specific Components: Mainly Behavioural, Cognitive	Strategy Use = problem solving, testing hypothesis and adapt/modify actions or objects from previous actions.	I expect a list of strategies and how much they are used throughout interaction. I expect it to increase as time goes by, as well as the last sections of it to relate to EDA and emotional attachment/enthusiasm.

	<p>engagement. They reflect a voluntary action as well as a problem-solving and experimentation behind this action.</p> <p>*If strategy achieves goal emotional engagement could be involved too.</p>
<p>Apparent Goal</p>	<p>This could help to keep track whether an open-ended exhibit is engaging because it allows for different goal-setting or the exploration of self-imposed problems that need to be explored.</p> <p>Reflects situational thinking from child, as well as goal-oriented behaviour</p> <p>Not sure yet.</p>
<p>Attention off-task</p>	<p>This code directly to contributes to the component of sustained attention and to unentangle whether there could be bottom-up or top-down attentional demands that are driving attention off task.</p> <p>Specific Component: Behavioural and Cognitive because through their behaviours they could be reflecting the decision to focus back after possible attentional intrusions or whether to investigate somewhere else.</p> <p>Attention = Process through which child shows/maintains interest with the task.</p> <p>This also includes re-engaging with the task despite the possible drivers off-task.</p> <p>To show if an exhibit is interesting enough to capture children's attention for longer periods, which would depend on amount of times child paces attention elsewhere off-task.</p> <p>Also, a list of possible intrusions elements that could point to disruptive events.</p>

**Persistence
(Meta-coding)**

The fact that the child is persisting on an action could reflect that the intended result has not been satisfactory or the encountered a similar problem, yet they are still willing to try.

Specific Components:

Behavioural, Cognitive and emotional engagement.

Observed voluntary actions, while reflecting problem-solving and goal-oriented behaviour at the spot. Different emotions could be associated with being persistent, as some children might be enthusiastic and other could persevere in a more negative manner.

Persistence = Appropriate level of challenge.

Testing success/failure of a strategy. Also, situational problem solving if they

To show a pattern of general goals to be achieved-larger actions that can be achieved throughout the exhibit.

I expect a pattern to emerge between persistence, use of strategy and type of FMM.

I expect the emergence of a pattern by the second half of the task.

Negative Affect

This code directly contributes to the component of sustained attention and to untangle whether there could be bottom-up or top-down attentional demands that are driving attention off task.

Specific Component:

Behavioural and Cognitive because through their behaviours they could be reflecting the decision to focus back after possible attentional intrusions or whether to investigate somewhere else.

Attention = Process through which child shows/maintains interest with the task.

This also includes re-engaging with the task despite the possible drivers off-task.

To show if an exhibit is interesting enough to capture children's attention for longer periods, which would depend on amount of times child paces attention elsewhere off-task.

Also, a list of possible intrusions elements that could point to disruptive events.

**Enthusiasm/
Positive Affect**

It contributes to understanding of the emotional attachment or reaction when interacting with the exhibit.

Specific Component:

Emotional engagement (and behavioural engagement)

Enthusiasm = positive emotional arousal. Reflects some of the feelings within the inner world

Expecting these results to directly correlate with the EDA data.

Intrusiveness

They directly might not support engagement but could be useful to know what events are interfering with children's interaction.

This can be a marker to events out of child's control

Periods of time and list of events of external events that might be affecting child's interaction

GMM (e)

It contributes to understand if and how the whole body might be moving in the exhibit could have any impact in how the child interacts with exhibit

- Voluntary Body movement = Motor planning
- Involuntary movement (or unintentional) = could be indicator of emotional distress/excitement.

I expect to see the duration of movement of whole body throughout the interaction.

Verbalizations

They can provide insight into child's situational thinking

Specific Component:
Cognitive engagement

Verbalizations provide a window to the thought process of the child.

I expect these moments to mostly match with what is happening with the FMM, Strategy or Persistence.

Visual Mismatch

This can provide an insight into child's thoughts, expectations, sequence and action planning.

Specific Component:
Cognitive engagement

Visual-touch mismatch = intention of action, corroboration of information used to act, expectation of a result.

To show if this mismatch is presented before/after any FMM and if a relationship could be found between them.

Interesting Moments

Unknown yet.

Some of these events could play a relevant role with the interaction as a whole

List of interesting moments throughout the interaction

K - Engagement Codebook Items

ELAN tier	Touching	Objects/Items being touched (different independent tiers <i>after</i> pilot)
Definition/ Description	Whenever children's hands (either one or both) are positioned on any relevant part of the exhibit and/or touching task-relevant materials.	
Direct evidence they give	1. A list of objects were child places direct attention. 2. Frequency and duration of episodes where child touches relevant objects of exhibit	1. Frequency and duration of episodes where child touches each object of exhibit
Data entry	There is no data entry here, just time selection.	1. Scraper: Red/Green/Blue/Yellow 2. Funnel: Red-Small / Transparent-Large/Both hands 3. Spinning Disc 4. Red Button 5. Sand 6. Flask 7. Graduated Cylinder: Small/Large/Both hands
Timing/ Coding	From the moment the hand is positioned on object's surface until it is not in contact with any surfaces anymore.	From the moment the hand is positioned on object's surface until it is not in contact anymore.
Examples	*When a child touches the red button in sand table. This counts whether they push/rub or just touch object. *They can be touching two different objects at the same time. *If they touched one and then another without taking their hand off first, it will all be coded until none of the hands is touching anything.	When touching a funnel with one hand and flask with other: Both are coded independently for each independent duration.
Notes	This tier has sub-dependents tiers with the particular names of the objects you can interact with at the sand exhibit. They follow same rules as this tier.	If objects are combined and being moved /touched simultaneously (e.g. funnel inside grad cylinder) and another object is also being used (e.g. flask or sand) they will ALL get coded separately and coincide in time. On the other hand, if objects are combined at some point (e.g. Funnel on top of flask) but are not being touched (e.g. sand in being poured with one hand and the other hand is not doing anything) then only object in hands (i.e. sand) is coded.

Action -FMM (Fine Motor Manipulation)
(Modified from eSOCS)

Definition/	Actions performed throughout the task that involve manipulation of task-materials.	
Description	Action: A voluntary movement in the interaction using an object which "suggests" the intention of changing something in the environment, e.g. moving a hand in the sand (Ghazali-Mohammed et al., 2020)	
Direct evidence they give	<ol style="list-style-type: none"> 1. List of actions performed throughout the interactions and objects associated with these. 2. Duration and frequency of actions performed 	
Data Entry	<p>The description will always start with a verb representing the action, then how the object(s) were used.</p> <p>Verb+Object</p>	
Timing / Coding	<p>From the moment the hand(s) in the object starts to move with the object in a particular way or direction (it has to last at least 150ms) until it stops:</p> <ul style="list-style-type: none"> • In space/time for more than 0.5-1 sec., • Because momentum/inertia of the movement has finished <p>*If one hand is doing something and the other hand starts doing the same but slightly delayed, it still counts as one action.</p> <p>4 different FMM</p>	<p>Exceptions:</p> <ul style="list-style-type: none"> • If FMM is mainly done with one hand and the other starts doing an action with a different object or an independent action from the first one --> They will be coded as 2 actions (FMM) • If there are clearly two actions in one single movement with less than 0.5 pause in between.
Examples	<ol style="list-style-type: none"> 1. Touches red button 2. Grabs Funnel 3. Pours sand through funnel 4. Puts funnel in flask and observes it. <p>* Wipes sand off of spinning disc (starts with one hand, then uses both hands)</p> <p>When NOT:</p> <p>* Changes item from one hand to other, fidgets with object in hand, when objects are ONLY touched but not manipulated</p>	<p>Exceptions:</p> <ul style="list-style-type: none"> • One hand is grabbing a container and is directly collecting sand from table (1) and using other hand pushes sand into the container too (2). • Without stopping: Collect sand with hand (1) and pour into container (2).
Notes	Fine Motor manipulation is not the same as Fine Motor Movement, thus excluded from this code is fine motor movement which does not involve an object/item (e.g. clicking fingers)	

Whenever the child is performing one or more actions that appears to be trying to solve a particular problem in the moment.

Or

That child appears to be reaching a particular goal/outcome

**Definition/
Description**

(e.g. Goal might be 'to see what happens when I do X after I have used Z' or it might be 'fill container'):

Involves the extent to which the child seems to be aware of their own progress solving the task and/or can recognise a problem preventing a correct solution and use problem-solving strategies.

**Direct evidence
they give**

1. List of types of strategies used.
2. Frequency and duration of episodes that the child uses strategy.
3. Objects or actions involved in the strategy.
4. Frequency and duration of particular strategies used by children.

Data entry

It can be one of the following:

- **New** - An action that has not been performed previously in the interaction
- **Adapted** - Whenever there is a change to the previously used - new (or adapted) - strategy that still appears to have the same goal.
- **Previous or going back** - When returning to an action that has been used before at some point in the past of the interaction, but not immediately before.
- **Repeat** (previous action) - An action that has been used immediately before with no changes or actions in between at all.

Linked to FMM, but it can encompass more than 1 FMM. Describe the use of strategy, rather than repeating FMM.

- **New** - From the moment the action (s) that seem to have a clear observable goal start until they finish.
- **Adapted** - Any change to a **new/adapted/previous** strategy in the following areas:

Timing/ Coding

* **action** in one of the movements of the action- (e.g. change in direction of action - from up/down to left/right).

* **object** from previously used in a similar action (if an object is added, it would fall here)

* **Location** - From previous location where the action took place into a new location.

- **Previous or going back** -
 - When an action is used in the exact same way as when it was coded as "new or adapted".
 - In the exact same way as before performing a change/new action (no more than 5 sec in between).
 - previously through the interaction but has not been used in the last 10 s
- **Repeat** - It uses exact same object as before
 - * Cannot be labelled this if the previous action is not the exact same (same action, same object).

- **New** - First time a child has brushed sand off table
- **Adapted:**
 - *Adapt (action): goes from pour sand into container --> to pour sand from container.
 - *Adapt (object): goes from big funnel--> to smaller funnel.
 - *Adapt (location): goes from pouring sand into general table --> to pouring sand on spinning disc.
- **Previous or going back** - Goes back to using container to fill directly from table
- **Repeat** - Pours sand into container (graduate cylinder)

- Important to keep in mind:**
- **It is highly likely that this will be coded retrospectively after FMM.**
 - Not all FMM will have a Strategy Code associated with it, as not all of them might seem to have a purpose.
 - It might still be possible that the purpose observed and the child's intent to be different, however, since they are not think-alouds, this is the best and closest guess.

<i>ELAN tier</i>	Apparent Goal* (New tier)
Definition/ Description	Higher-order category of Strategy - Describe whether the strategies used seem to be following a same goal/solution, a new goal/solution or are going back to an "unfinished" goal/solution.
Direct evidence they give	Frequency and duration of episodes when a child seems to establish and/or maintain a same goal to solve problem or a self-imposed goal.
Data entry	<ul style="list-style-type: none"> • New Goal - Whenever it seems to be a new goal/problem to solve • Same Goal - Whenever the strategy seems to be following the exact same goal as the strategy used before (new or previous) • Previous Goal - Whenever the strategy seems to be going back to a strategy used in a previous goal (left incomplete, or perceived as finished) during the interaction but not immediately before.

Timing/ Coding	Will be linked to the coding of Strategy Use and FMM
	<i>New goal:</i> Fill in container
Examples	<i>Same Goal:</i> Fill in container <i>Previous:</i> Brush off sand from spinning disc
Notes	This label will link both Strategy Use and Persistence together
ELAN tier	Attention off-task (adaptation from eSOCS + LECS)
Definition/ Description	The period when the child deviates attention (focus) from the task or task-relevant items. This could be in a form of visual attention, touch or verbalizations.
Direct evidence they give	Frequency and duration of episodes where child focuses attention somewhere else.
Data entry	<ul style="list-style-type: none"> • Verbal (hands are on-task, but speech is not task relevant) • Visual (hands are on-task, but visual attention is not task-relevant) • Touch (vision is on task, but hands are touching task-irrelevant items). • Mixed (when more than one modality is off-task).
Timing/ Coding	It will align with Visuo-tactile Mismatch, and verbalizations. *Coded from period when visual/verbal cue is placed/mentioned off-task.
Examples	Examples 1. Someone approaches exhibit 2. They are taking with parents about unicorns.
Notes	None
ELAN tier	Enthusiasm/ Positive Affect (Adapted from eSOCS + LECS)
Definition/ Description	An expression of a positive emotion while interacting with the exhibit.
Direct evidence they give	1. They can reflect whether a child is experiencing any positive feelings towards the task. 2. Number of times of these episodes and duration of them
Data entry	Describe any expression of enthusiasm: <ul style="list-style-type: none"> • Smiling, laughing, excitement, enjoying the task; vocalization and/or facial expression. • Appearing eager and enthusiastic to try each problem presented
Timing/ Coding	From the initial facial or body movement/sound that reflects the positive emotion until it moves back to neutral or a negative emotion.
Examples	1. Short glimpse of a small smile (?) 2. Smile while playing 3. Excited vocalization after successful action

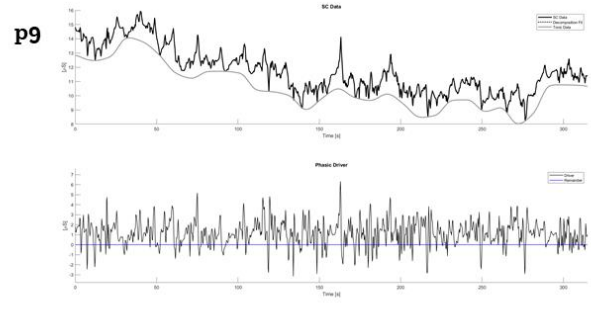
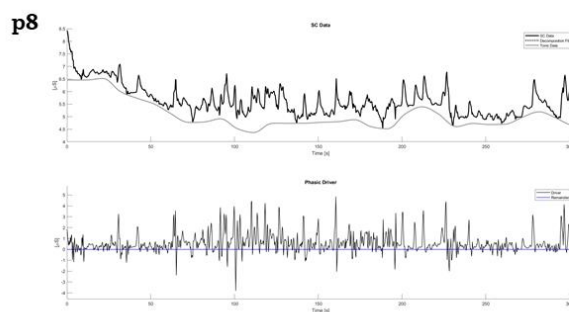
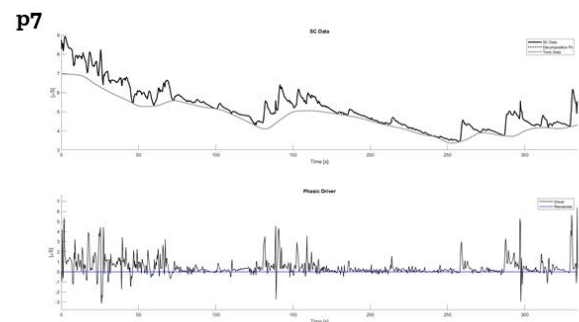
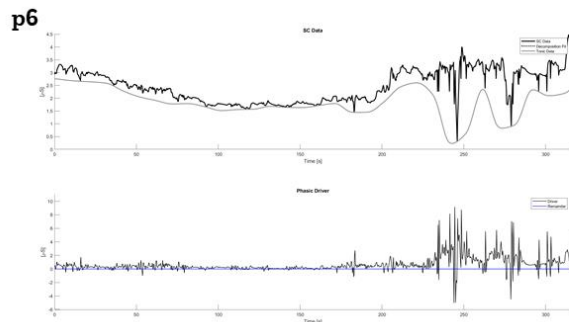
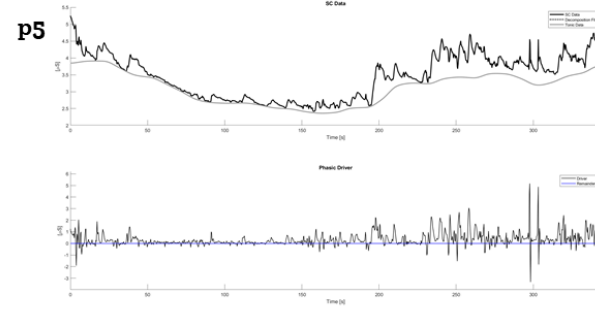
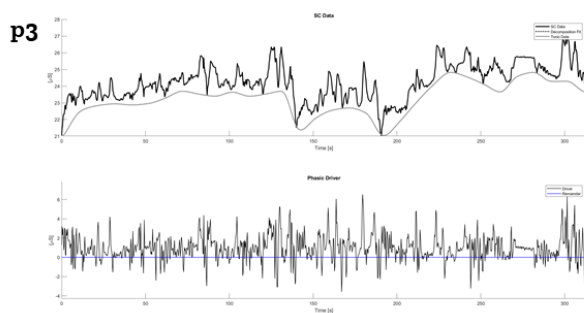
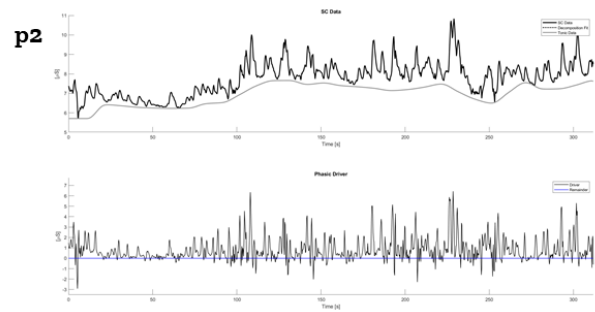
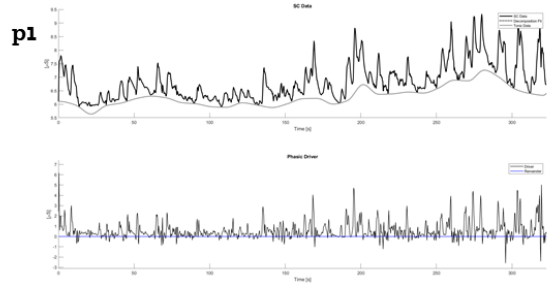
Notes	This code is harder to code as children do not always express intense facial expression.
ELAN tier	Negative Affect (Adapted from eSOCS + LECS)
Definition/Description	An expression of a negative emotion while interacting with the exhibit.
Direct evidence they give	<ol style="list-style-type: none"> 1. They can reflect whether a child is experiencing any negative feelings towards the task. 2. Number of times of these episodes and duration of them
Data entry	<p>Describe any expression of a negative emotion (frustration, sadness, annoyance, boredom, anger):</p> <ul style="list-style-type: none"> • Yelling or tossing elements from the exhibits. • Loud vocalization with negative emotion • Expressing feeling a negative emotion • Tears
Timing/ Coding	From the initial facial or body movement/sound that reflects the positive emotion until it moves back to neutral or a negative emotion.
Examples	<ol style="list-style-type: none"> 1. Frustration when disc won't stop turning 2. Forcefully tries to stop disc
Notes	This code is harder to code as children do not always express intense facial expression.
ELAN tier	Intrusiveness (Taken from eSOCS)
Definition/Description	<p>Physical and/or verbal events (whether voluntarily or involuntarily) that disrupt, manipulate, or control child.</p> <p>(Added here whenever there were external factors that disrupted child's interaction)</p>
Direct evidence they give	Direct cause of intrusion in child's activity
Data entry	Describe event (person/item/situation) that caused intrusion
Timing/ Coding	Code will start at the moment where another person or event disrupt the task, until they are done.
Examples	<i>Uninvited child walks towards exhibit</i>
Notes	This does not happen very often
ELAN tier	Body Movement – GMM (Adapted from eSOCS)

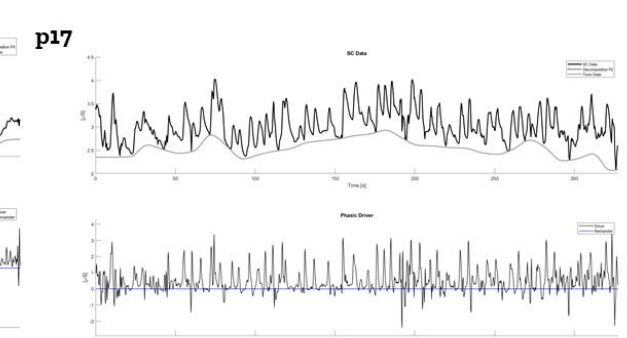
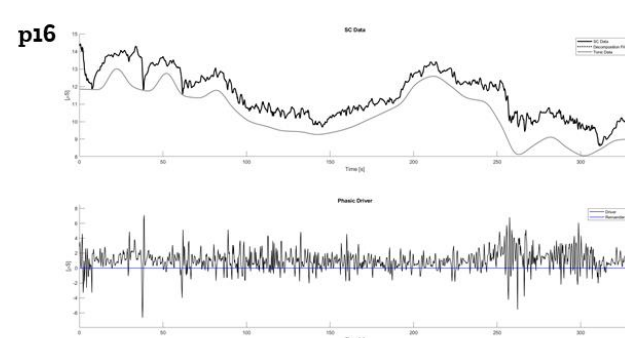
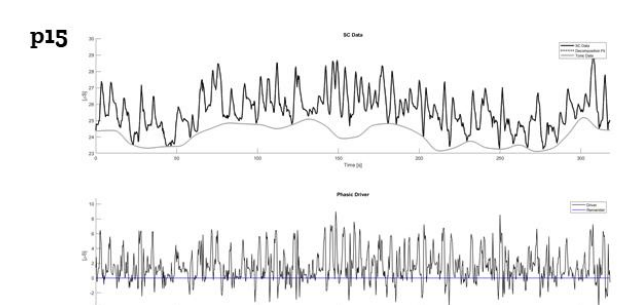
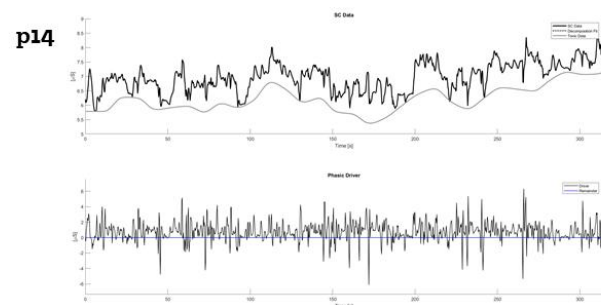
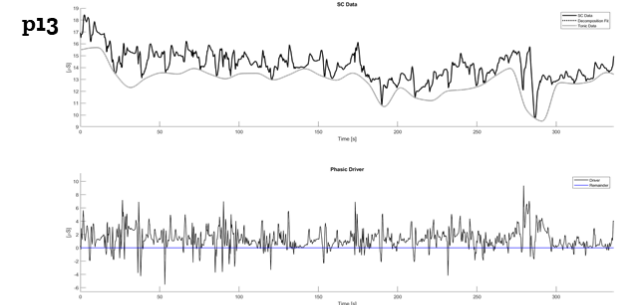
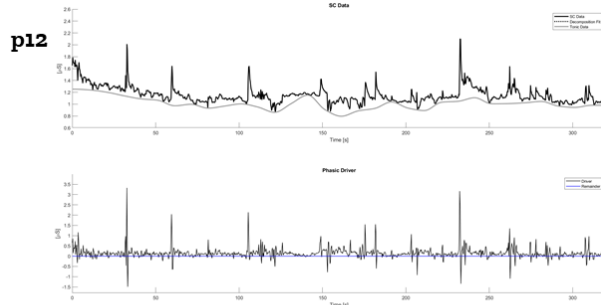
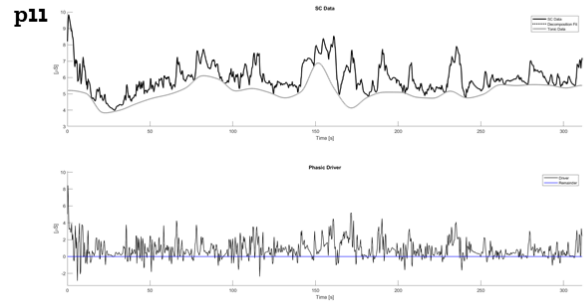
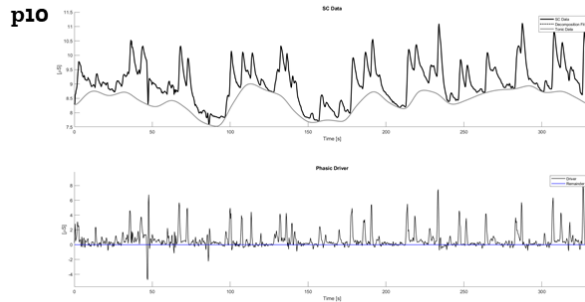
Definition/Description	This definition is grounded on developmental definition of gross motricity, but since Action- FMM always entails movement of arms, for the sake of simplicity in labelling and considering relevance to engagement, it will only track whenever child is moving around the exhibit or in their same place but changing position significantly (e.g. standing --> walking).
Direct evidence they give	It can provide types of gross motor movements that could be associated with the use of this exhibit as well as frequency and duration of these movements.
Data entry	<ul style="list-style-type: none"> • Walk •Whole body - Whenever a child is crouching/sitting/ standing up
Timing/ Coding	From the moment the body moving until it stops
Examples	<ul style="list-style-type: none"> • Walking - When child circles around the exhibit) •Whole body - when child sits down on stool
Notes	This might be a good code to counteract for EDA activity peaks when needed
ELAN tier	Verbalizations (adapted from eSOCS)
Definition/Description	Task-related speech, either self-directed or directed at surrounding members.
Direct evidence they give	<ol style="list-style-type: none"> 1. Episodes when child verbalizes either for themselves or as evidence of interaction with someone else. 2.Frequency and duration of these moments
Data entry	Transcription of their speech
Timing/ Coding	<ol style="list-style-type: none"> 1.From the moment the sound starts until it stops. 2. When the word/phrase seems to have finished. 3. When there is more than 1.5 sec of silence. 4. Also, add if the speech is directed to someone else at the beginning of transcript
Examples	<p>*Child mutters: "will this be full?"</p> <p>*Child speaks to adult: "Whoops, it fell down!"</p>
Notes	
ELAN tier	Visuo-Tactile Attentional Mismatch (New Tier)

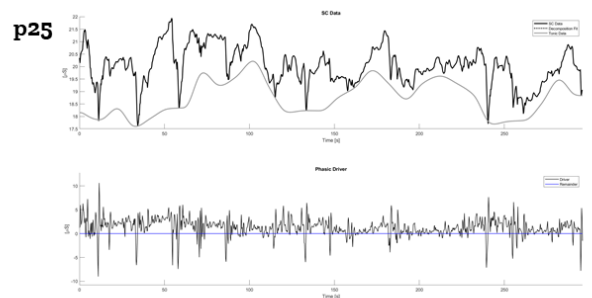
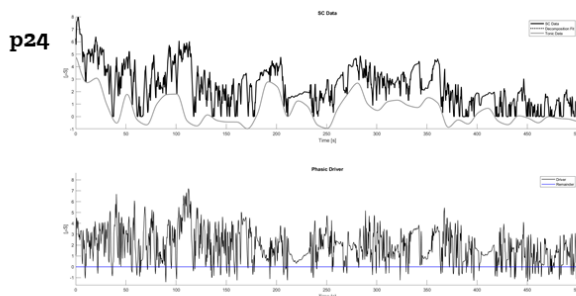
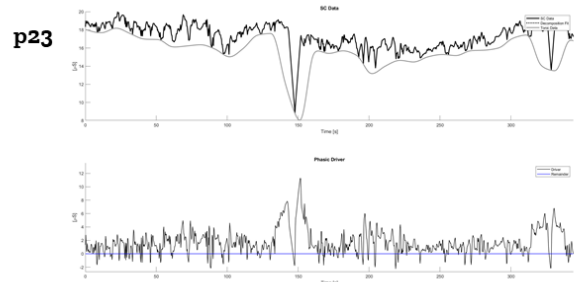
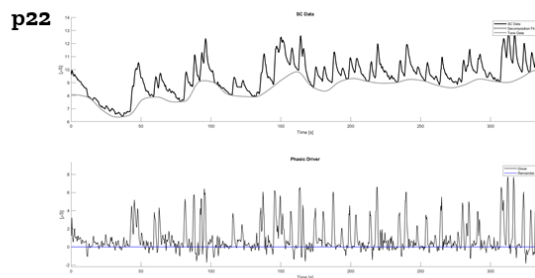
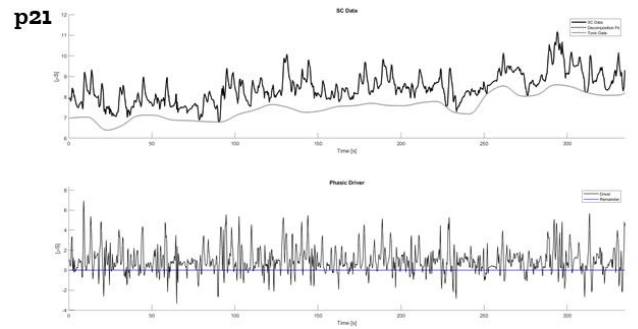
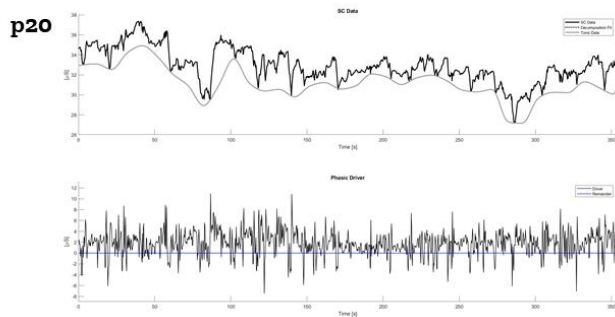
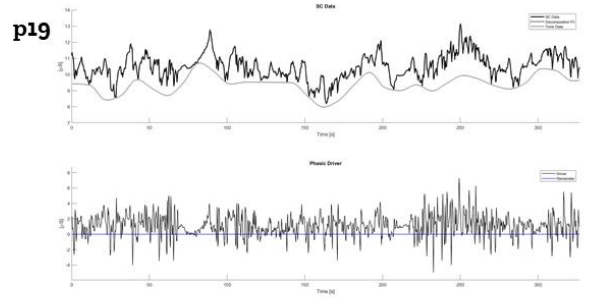
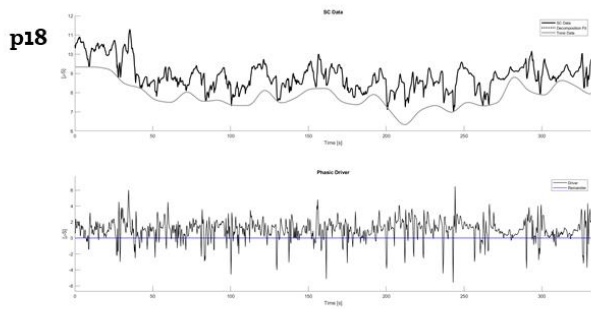
Definition/Description	The periods of time when the head-camera does not match the location of where the hands are being placed, yet it is still task related.
Direct evidence they give	Tangible evidence of when children might be anticipating an action or observing the result of a previously performed action. Frequency and duration of these episodes throughout the interaction
Data entry	Description of what seems to be at the centre of the video.
Timing/Coding	From the moment the head-camera video moves away from what the hands are doing until the hands return to occupy central third or 2/3 of the screen
Examples	*Gaze at spinning disc while having hands on sand. *Looking away because he is listening to container.
Notes	none

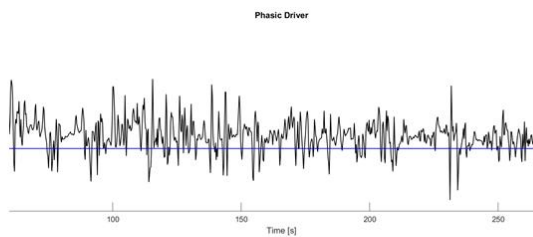
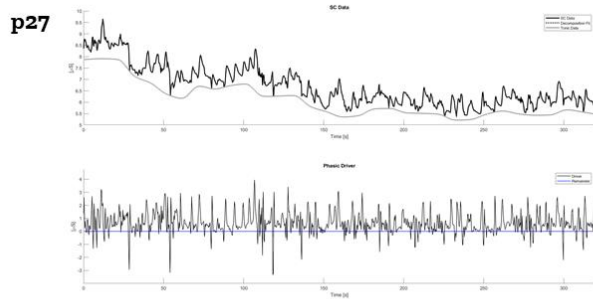
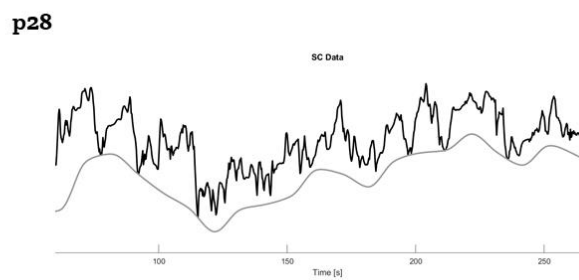
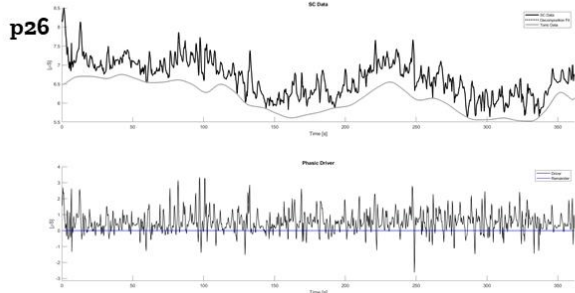
ELAN tier	Interesting Moments (New Tier)
Definition/Description	Any other behaviours or events that are not considered but that they seem interesting or relevant for the interaction.
Direct evidence they give	Possible relevant behaviours/events that have not been considered but could still be related to the other codes provided
Data entry	Description of moment
Timing/ Coding	Will depend on coder, the duration of the moment that has been considered relevant/interesting
Examples	*Tries to set up the cylinder and can't because of sand on spinning disc ---> Uses his hand to swipe sand off spinning disc *Knows what part of container is easier to pour from
Notes	

L – Electrodermal Activity Decomposition per Participant





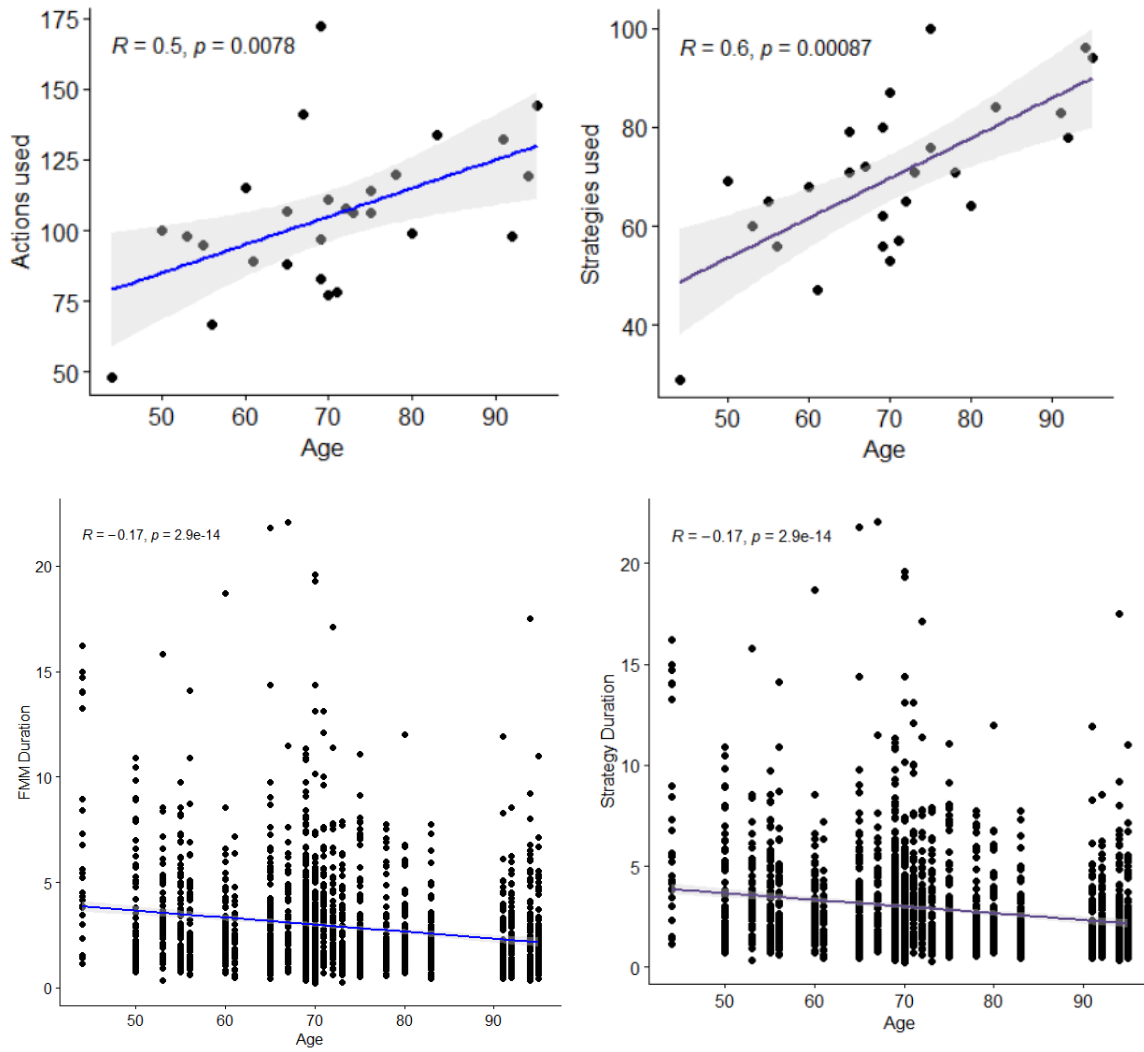




M – Age Analysis

The age of participants was evaluated to ascertain if it had a relationship with any of the engagement variables. No correlation was found between children's age, and the EDA amplitude, nor with the NS-SCRs, which could be reflecting no physiological age effect driving the results observed from the EDA measures and that these had a similar pattern within the selected age range of 4 to 7-year-olds. On the other hand, Age had a positive effect in some children's behaviour, particularly on the frequency of actions performed ($R = .50, p < 0.01$), and with the frequency of the Strategy-Goal driven behaviours ($R = .61, p < .001$), however, this effect did not appear when proportional time spent was considered. Thus the relationship of Age with the duration of the behaviours was explored, where a small negative effect was found between Age and both Actions [FMM] and Strategy-Goal driven actions ($R = -.17, p < 0.001$) which indicates that older children spent less time in each of the performed actions and instead increased the number of actions performed during the exhibit.

Correlation plots between Age-Actions, and Age-Strategies



Note. Age is shown in months. Duration is shown in seconds. Left: Duration of actions [FMM]. Right: Duration of Strategies.

N- Additional Illustrative Examples

Strategy- Goals

These examples show different items from the exhibit used to collect and transport sand showing a range of diverse actions.

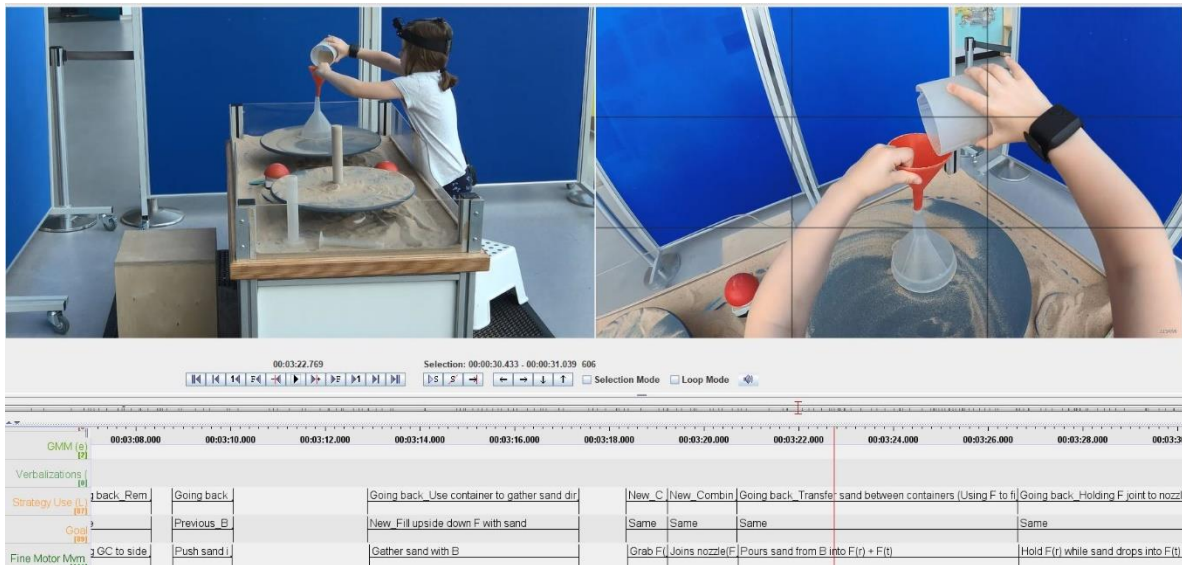
Using Scraper to Push Sand into container

00:01:26.501 Selection: 00:03:08.366 - 00:03:09.699 1333

Enthusiasm/PA (R)							
Negative Affect (R)							
GMM (G) (R)							
Verbalizations (R)							
Strategy Use (L) (R)	New Choose a new tool to help her go	Going back Prepare GC for new	New Uses S t	Going back Check	Previous	New Combines pushing sand	Going back Fill container directly from tabl
Goal (R)	Same	Same	Same	Same	Same	Same	Same
Fine Motor Mv (R)	Grab S from far-end of table	Positions GC horizontally in the t	Push sand into	Inspect inside of G	Position	Fill up GC with sand while push	Filling up GC with sand from table

Child filling in container using one of the tools (scraper) from sand exhibit. Left image shows fixed camera, whilst right image shows the perspective from head-mounted camera. Lower part of image shows codes used throughout the interaction

Using combination of Funnels to pour sand onto disc



Note: Child transferring sand from container, through funnels and into the disc. This was one of the most unusual uses of the two funnels together. Left image shows fixed camera, whilst right image shows the perspective from head-mounted camera. Lower part of image shows codes used throughout the interaction

Using funnels to distribute sand onto spinning disc



Note: Child pouring sand onto spinning disc through funnels until disc was fully covered. Right image shows fixed camera, whilst left image shows the perspective from head-mounted camera. Lower part of image shows codes used throughout the interaction