Drawing from the past to learn tomorrow

Ann Cronin

A thesis submitted for the degree of

Doctor of Philosophy

At the University of Otago

September 2018
Abstract

Humans are unique in their ability to use tools to produce works which communicate information about 3D objects through 2D drawings. Drawings give us insight into the origins of our species, our uniqueness amongst animals, and our creative ability to produce images from the strategic placement of line and marks. Considerable research has been devoted to the ontogeny of drawing skill and considerable speculation has been devoted to the psychological value of drawings themselves. For example, children’s ability to produce human figure drawings develops in a highly predictable stage-like manner, but the rate at which individual children progress through each stage varies considerably. These two characteristics - a consistent developmental pattern coupled with individual differences in the rate of development - have led many to argue that human figure drawings (HFDs) can be used as a measure of intelligence. One characteristic of traditional pen and paper intelligence tests is that they are highly resistant to instruction. In Study 1, I assessed the effect of instruction on 11- and 12-year-old children’s scores on the most recent HFD test, the DAP:IQ. Children showed significant gains shortly after art instruction, but their scores returned to pre-instruction levels when they were tested 6 months later. These data challenge the view that the DAP: IQ provides a valid measure of intelligence.

Despite the limited value of drawings as measures of intelligence, the possible benefits of drawing per se have been under-explored in contemporary research. Students of art have traditionally drawn from life as a way to improve their drawing skills, but little is known about the cognitive benefits of this practice. In Study 2, I examined the effect of drawing on visitors’ memory for museum exhibits. Three groups completed a tour of a museum exhibit, ‘Wonders of the World’: two of these groups consisted of adults (self-identified artists and non-artists) and the third group consisted of children under the age of 13. Individuals from each group were
accompanied by a researcher who instructed the participant to sketch or to merely observe each object while completing the tour. Adults observed ten objects and drew ten objects and children observed six objects and drew six objects; irrespective of whether they drew or observed, all participants spent the same amount of time with each object. One week later, they were asked to recall the objects they learned about on the tour, and then they were asked a series of specific questions about the visual details of the objects. For adult artists and children, drawing enhanced recall and recognition, but for adult non-artists, it did not. That is, for those participants who were comfortable with the medium, drawing provide an inexpensive opportunity to facilitate learning and memory in museums.

Despite the positive outcomes associated with drawing, drawing as a pastime is decreasing in popularity with children, which some would attribute to the rise of digital culture and increased screen time in young people’s lives. But not all digital technology is without benefit and one digital interface that allows for creative activity in a virtual world is Minecraft. In Study 3, I examined whether using creative digital tools might also aid children’s memory for museum visits. Children took a tour of an industrial museum, either ‘in-game’ using Minecraft or by visiting the physical site and then three of the four groups completed a follow-up worksheet to find additional information. Two groups completed this worksheet ‘in-game’ using Minecraft.

Students who experienced the physical tour and the follow-up worksheet activities in Minecraft had the best retention of facts one week later. Interestingly, students who completed a physical tour followed by the onsite worksheet activities recalled the same amount of information as students who completed all activities ‘in-game’ in their school classroom. That is, Minecraft proved to be an inexpensive opportunity to facilitate learning and memory on field trips.

Taken together, the present research challenges some traditional assumptions about drawing, demonstrating at least one way in which it can be used to enhance
learning and memory, at least for some groups of participants. For children on field trips, this research also extends the learning and memory value of digital gaming, a contemporary pastime which has superseded drawing in children’s day-to-day lives. The evolutionary history of drawing is linked to the emergence of tool use in humans: modern, digital tools build on those that have come before them. The studies in this thesis explore part of this journey.
Acknowledgments

This thesis is dedicated to my mam.

This thesis, as many before it, has been a process of building, discovery, and exploration. This was made possible by the generosity and support of Otago University. There are so many people to thank, I hope this goes some way towards repaying them.

First and foremost, thank you Jules, you have been a constant source of hope, support, and mentorship, when things seemed impossible you just got on with it and brought me with you. Thank you for teaching me not to say sorry so much and all the love and minding. I am so glad I had you as my supervisor and thank you for your unending patience with me. I will never be able to repay what you did for me and how you changed my life.

None of this would be possible without the kindness and brilliance of Harlene. Who managed to stick with me even when I ran after every distraction that came my way. You have made an indelible impact on me.

Thank you to Leon and Emma for being there for me and for each other, you are my light and my hope, I hope this country rewards you both.

Thanks to my mam and dad for believing in me, my sisters and brother, and all the friends, companions, and supports who carried me when I could not walk through the last few years.

And of course, thanks to all of the schools, adults, parents, and children whose participation made this research possible.

This research was funded by Marsden Grants from the Royal Society of New Zealand awarded to Harlene Hayne.
## Table of Contents

Abstract .............................................. i

Acknowledgements ................................. iii

Table of Contents ................................... iv

List of Tables ......................................... v

List of Figures ....................................... viii

List of Appendices ................................. xii

Publications Arising from this Thesis (to date) ............................ xiii

Chapter 1 General Introduction ................. 1

Chapter 2 The Effect of Instruction on Children’s Human Figure Drawing (HFD) Tests: Implications for Measurement ............................ 42

Chapter 3 Put Away Your Camera: Drawing Facilitates Memory of Museum Exhibits ............................ 61

Chapter 4 New Kids on the Block, Learning while Exploring in Minecraft ............................ 84

Chapter 5 Concluding Comments ................. 105

References .......................................... 124

Appendices .......................................... 172
List of Tables

Table 2.1  Individual Participants’ Standard IQ Scores at Times 1, 2, and 3.  
The superscripts refer to the ability classification at subsequent testing points.

Table 2.2  The Distribution of Participants in Each Ability Classification Category at Times 1, 2, and 3.
## List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Horses and rhinoceros drawn on the cave walls of Lascaux cave, France. (35,000 - 40,000) years old</td>
<td>4</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Lion, bear and rhinoceros drawn on cave walls at Chauvet Cave. (35,000 - 40,000) years old</td>
<td>4</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Deer or goat drawn on the wall of a cave in Sulawesi, West of Borneo. (35,000 - 40,000) years old</td>
<td>5</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>Human figures with clothing shown in Bradshaw rock paintings, Kimberley, Australia.</td>
<td>9</td>
</tr>
<tr>
<td>Figure 1.5</td>
<td>Scribble drawings by young children</td>
<td>18</td>
</tr>
<tr>
<td>Figure 1.6</td>
<td>Shows a range of human figure drawings produced by 3- to 5-year-old children. Panels B and C represent classic tadpole drawings</td>
<td>20</td>
</tr>
<tr>
<td>Figure 1.7</td>
<td>Drawings by children showing a variety of inaccuracies common in drawing development.</td>
<td>22</td>
</tr>
<tr>
<td>Figure 1.8</td>
<td>Improvements in drawing human eye between the ages of 15- and 18-years of age.</td>
<td>23</td>
</tr>
<tr>
<td>Figure 1.9</td>
<td>Examples of Koppitz EI drawings</td>
<td>30</td>
</tr>
<tr>
<td>Figure 1.10</td>
<td>Example of HTP drawings completed by adult</td>
<td>32</td>
</tr>
<tr>
<td>Figure 1.11</td>
<td>Example of KFD Family drawing</td>
<td>35</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Examples of human figures drawings created at three time points.</td>
<td>56</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Examples of real objects from the museum and sample drawings from each of the three participant groups</td>
<td>71</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>The proportion of objects that were correctly recalled by the artists in Experiment 1 (left panel) and by and by the non-artists and children in Experiment 2 (right panel) as a function of condition.</td>
<td>73</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>The proportion of objects that were correctly recognized by the artists in Experiment 1 (left panel) and by the non-artists and children in Experiment 2 (right panel) as a function of condition and recognition test.</td>
<td>74</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>The proportion of multiple-choice questions about the objects that were correctly answered by the artists in Experiment 1 (left panel)</td>
<td>75</td>
</tr>
</tbody>
</table>
and by the non-artists and children in Experiment 2 (right panel) as a function of condition

Figure 4.1  A schematic representation of the 4 experimental groups.
Figure 4.2  Top Panel: An aerial view of the Gasworks site and the corresponding aerial view from the Minecraft version.
Figure 4.3  Mean scores (+1SE) on the final worksheet as a function of experimental group.
Figure 4.4  Mean scores (+1SE) on the LOES-S subscales (learning, engagement, and quality) as a function of experimental group.
List of Appendices

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix A</td>
<td>Brief summaries of a selection of past theorists that have considered children’s drawing development as stage-like</td>
<td>172</td>
</tr>
<tr>
<td>Appendix B</td>
<td>Brief summaries of a selection of perceptual- and production-focused theorists that have investigated children’s drawing development</td>
<td>180</td>
</tr>
<tr>
<td>Appendix C</td>
<td>Information contained in onsite and virtual tours of industrial museum</td>
<td>187</td>
</tr>
</tbody>
</table>
Publications Arising From This Thesis

Refereed Journal Articles


I designed the experiment, collected and analysed data, took the lead role in writing and revising the manuscript. Co-authors provided guidance on all drafts and on revising the manuscript.


Cronin, A. P. (2015, November). Drawing as an encoding tool, why not paint a 1000 words

Paper presented at Kiwicam, Otago University. New Zealand.

Chapter 1

General Introduction

A full history of humankind is impossible, but through the messages and signs that have been left behind as art by past civilisations, we can begin to understand something about what makes us human. The art produced by people throughout the millennia of human existence is not only a snapshot of the lives they lived, but it also offers insight into the development of our species. Over time and across continents, the ability to adapt tools and create art is present in all societies and the marks these early humans made on rocks and in caves provides a brief glimpse into the evolution of human culture.

On the basis of carbon dating of cave paintings, we know that humans have drawn since the late Stone Age, but the history of human mark-making is much older. Approximately 164,000 years ago, humans were using pigments and dyes and engraving patterns into bones and beads, and more than 60,000 years ago, they were engraving symbols on surfaces in their surroundings with rock tools. Crayon-like pigment blocks more than 50,000-years old have been found in Australia, and the oldest preserved cave paintings have been carbon dated as 40,000-years old (Bower, 2014). Cave paintings have provided scientists with information about the methods, materials, and tools that were used to create them, as well as information about the natural environment of the creators (Lewin, 1989). The oldest known art created by human hands came from the African continent. Radiocarbon dating of a cave buried between 26,000 and 29,000 BP\(^1\) shows that its inhabitants drew animal figures on stone slabs within the cave before its collapse (Masson, 2006). These paintings provide a window into the lives of the earliest humans, their interests, and their environment (Cutting & Massironi, 1998; Morriss-Kay, 2010).

\(^1\) BP refers to items dated Before Present by carbon dating techniques (present estimated at 1950, the start of affordable carbon dating practices)
From a psychological perspective, the primary value of cave art is what it tells us about the evolution of human cognitive ability. By examining cave drawings, scientists seek to uncover the underlying processes that governed these artistic activities including both motor control and thought. Although the motivation behind the creation of this art will undoubtedly continue to elude us, these works can inform us about the lives of early humans. Soon after the first humans appeared in Europe, they began making marks on surfaces in their surroundings, which included standard graphic elements such as handprints and outlines of figures, dots, and lines.

Humans, like animals, have two major drives: preservation of the individual and preservation of the species. Charles Darwin distinguished between survival of the fittest and sexual selection, arguing that animals’ natural ornaments (e.g., colourful plumage) are explicitly related to the need to attract a mate even though these characteristics sometimes threaten the survival of the individual (Penn, Holyoak, & Povinelli, 2008). It is not surprising then that reproductive themes litter early human artworks. Some evolutionary theorists have even suggested that human art-making may have initially served the additional purpose of mate attraction (Hodgson, 2006). Preservation of the individual (and the social group) is inherent in images related to hunting or to animals (and more rarely, fish) that were important food sources. Preservation of the species, which involves sex and fertility, nurturing, and group protective behaviour, is symbolised in many drawings, both overtly in pubic triangles, and phallus-like objects, and more covertly with images of animals engaged in the behaviour associated with the mating season (Morriss-Kay, 2010).

Many sites in Europe feature images including engraved outlines of animals and humans, representations of both male and female sexual organs, and carved statuettes of animal and human figures (Pettitt, 2008). Fertility is the dominant theme in many early images of humans in European cave art, including human female ‘Venus’ figurines, reliefs,
and engravings. Female sex and/or sexuality in cave art is reduced to a triangle for the mons pubis and an engraved line to represent the vulva. The main fertility attributes of females such as breasts, bellies (presumably pregnant), labia, and vulva feature in these works. A female pubic triangle measuring only 40-56 cm carved in limestone at La Ferrassie, Dordogne, France and dated to around 32,000–34,000 BP is one of the earliest known types of carved human images in Europe (Valladas et al., 1992).

The representations of anatomical features in these early images illustrate an awareness of basic animal husbandry and breeding practices, a precursor to early farming practice. Neolithic man lived in small groups, hunting and gathering in different regions for food sources that were plentiful or available at particular times. The inclusion of drawings about the animals in an area may have served a teaching or instructional function at sites where these groups congregated during different seasons. For example, the analysis by Aujoulat (2005) of the images of animals in Lascaux Cave, Dordogne, France concludes that the images were painted in the following order: horses, aurochs and then deer, and that each species shows physical features characteristic of its breeding season – the horses have the thick coats of late winter/early spring, the aurochs have their summer coats, and the deer have antlers and are represented in groups characteristic of early autumn. Figure 1.1 shows drawings from Lascaux and Chauvet Caves in Europe.
Figure 1.1. Horses and rhinoceros drawn on the cave walls of Lascaux Cave, France (35,000-40,000 years old).

(Image: Aujoulat, 2005)

Figure 1.2. Lions, bears, and rhinoceros drawn on cave walls of Chauvet Cave (35,000-40,000 years old).

(Image: Aujoulat, 2005)
The drawings of early Homo sapiens provide insight into the development of mankind from hunter-gatherer to early farmer, the development of belief systems, and social sharing of information (Vaesen, 2012). The habit of documenting the environment in drawings is further demonstrated by the addition of stylised human figures and companion animals, including dogs and other domesticated species, animals in the locale, and a wide range of geometric shapes.

Analysis of frequency patterns of over 3300 representative images from European Stone Age art panels revealed consistent patterns of presentation of various subjects across sites. Horses, bison, and ibex, were most commonly presented and constituted more than 10% of the total images, followed by Mammoth, aurochs, hinds, and stags. Humans or human-like forms are relatively rare, representing only 4.3% of all drawn images, followed by very rare animals such as reindeer, bear, lion, fish, and rhinoceros. (Sauvet & Wlodarczyk, 2008). In Northern Europe and Russia, the most common element in Stone Age rock art are boats (Kolpakov & Shumkin, 2012), deer and eagles feature in Eurasian art (Omarov,
Baigunakov, & Sabdenova, 2014; Toleubayev, Zhumatayev, & Baimuhamedova, 2014), and babirusas (a tusked pig-like animal) in Indonesia (Aubert et al., 2014).

Animals are the most frequently represented images throughout early art, however, early attempts at the representation of human features emerge across continents and civilisations. Drawing of people is one of the first drawing behaviours observed in human children (Callaghan, 2015) so it is perhaps not surprising that early man also made attempts at representing themselves in drawings. In Siberia, for example, Stone Age humans drew stylised faces, both separate and as part of anthropomorphic figures. Using red mineral dye, they etched, pecked (hit with a tool repeatedly), or painted these images onto hard surfaces. Faces were drawn using a range of styles; circular, semi-circular, heart shaped, and rhomboid forms. Concentric circles and lines around the eyes and mouth exaggerated features. Suggested teeth, beards, and horn-like protrusions also form part of these drawings. Triangular-shaped bodies in silhouette and with details also feature in the art of the late Neolithic period (Karşilaştırma, 2010; Zaika, 2012). These stylised representations of humans can be seen in contemporary hunter-gatherer societies such as with the Mikea of South-western Madagascar, the Okiek of Kenya, as well as Australian Aborigines (Lee & Daly, 1999).

In South Africa, Pleistocene paintings found at the Wonderwerk Cave show both naturalistic representations of animals and geometric finger-painted figures in contrast to other African sites where engravings (petroglyphs) are the prevalent rock art form (Chazan & Horwitz, 2009). Early Brazilian rock art discovered at Lapa do Santo cave, alongside faunal and macro-botanical remains of early hunter-gatherers, reveals figures pecked in the bedrock. Among them is a small anthropomorphic figure with three digits, a C-shaped head, and an oversized phallus (Hayden et al., 2015; Neves, Araujo, Bernardo, Kipnis, & Feathers, 2012).
Carbon dating of cave art from Sulawesi, situated east of Borneo in the Wallacean archipelago, suggest that figurative art was already part of the cultural repertoire of the first modern human populations to reach this region more than 40,000 years ago. Cave paintings at Chauvet Cave in France suggest that rock art emerged independently at around the same time and at roughly both ends of the global spatial distribution of early modern humans (Aubert et al., 2014).

In addition to telling us something about how early humans lived, rock art also allows us to view the progress of at least some of the cognitive skills of early humans. Not only do the animals they drew inform analysis of their surroundings and prey, but they also provide a window into the observation skills of early humans. For example, Horvath, Farkas, Bonec, Blaho, and Kriska (2012) compared the paintings of four-legged animals (quadrupeds) by modern artists and by Upper Palaeolithic Homo sapiens. Quadrupeds have a defined sequence of footfalls during walking, trotting, and running. Horvath et al. assessed the animal paintings to see if the correct sequence of foot positions was displayed in the work. For each painting they calculated the number of feet in the correct position relative to the movement type for each painting. On the basis of their assessment of the artistic rendition of animal gait relative to the actual sequence of footfalls in walking, running and trotting, Horvath et al. concluded that early Homo sapiens depicted these movements more accurately than did modern artists. The error rate for the incorrect depiction of gait and limb movement in cave-art (46.2%) was significantly lower than that of modern artists (83.5%). The high level of accuracy in Palaeolithic cave art suggests that these early humans paid careful attention to animals’ motion which was important to their hunting and their survival.

The cave paintings of early humans also display advanced drawing and painting techniques such as perspective-taking and shading (Blater, 1999). The paintings in the Altamira Cave in Northern Spain and at Grotte Chauvet in France feature over 300 portrayals of animals, both static and in motion, either etched or painted, in a variety of
colours and include shading, perspective, and tone to create realistic representations rather than just outlines. In one portrayal, an artist depicts a bison’s head using a combination of sculpture and drawing. The artist used the surroundings to maximise the representation of the animal including drawing the animal's eyes on natural contours of the rock face in an attempt to mirror the physical dimensions of the animal's skull. These techniques illustrate planning on the part of early artists. Not only did they observe and recall the form and movement of animals, but they also observed and recalled placement, shade, perspective, and size which subsequently form part of their design process and execution.

Evidence of sketching outlines in rock art in the Kimberley region of Australia also provide evidence of planning before the production of art. The drawings in this region span around 40,000 years of human history and document the culture of the Aboriginal people in Australia. These drawings are known as Bradshaw Rock Art, after the pastoralist Joseph Bradshaw who first documented them in the mid-1800's (Bradshaw, 1892). Bradshaw drawings are small, typically only 10-70-cm long and 4-22-cm wide, but despite their small size, they give the first evidence of clothing use by human beings, with humanoid figures shown wearing headdresses, tassels, skirts, and pants (Michaelsen, Ebersole, Smith, & Biro, 2000). Alongside these early clothing representations, Australian Aboriginal shamanistic culture is also evident in Bradshaw drawings. These drawings feature female figures in the role of shaman, indicating particular roles for women. The women in Bradshaw drawings are in central positions within the drawings, they are drawn in ceremonial robes with their heads arched towards the sky with breasts clearly visible either side of the silhouette figure (see Figure 1.4).
It is clear that the neurological potential to create art, and probably the creation of art, was established before Homo sapiens left Africa, but we cannot know whether regional stylistic differences were already established within each emigrating group or were acquired en-route or at their final destinations. The archaeological record is more generous in information on tool-making and it is beyond doubt that all of these Homo sapiens groups left Africa with an accomplished ability to create 3D tools with a great variety of applications (Hodson, 2006). Nearly all of the marks discovered so far have been engraved onto relatively hard surfaces, such as rock, bone, and ochre. No artistic style is static, so the passage of time and generations, with different cultural and environmental influences, changes in climate,
different available materials and technological skills, not to mention specific highly-gifted individuals, must have been important factors influencing stylistic development, culminating in the regional variety apparent in world art today.

The emergence of rock art in the repertoire of early man occurs at the same time as the development of new hand tools, displaying a heightened sophistication of cognitive skills and the ability to balance tools to make them more efficient for hunting and adaptation to the environment. With the ability to create more useful and flexible tools came increased spatial abilities. Improved spatial skills underlie the capacity to create a 2D representation of an object from visual memory. Hodson (2009) suggests that this emergent skill is the result of a cerebral structural change in Homo sapiens. The drawings created by early humankind do not differ significantly from the abilities of contemporary humans, indicating that early human cognition, hand-eye coordination, and the use of symbols has not changed significantly in 40,000 years.

**Drawing as a uniquely human activity**

One unique outcome of human evolution is that as a species, we alone produce drawings and artworks that can be considered as objects themselves as well as representations of other objects. From an early age, children spontaneously attempt to represent their world through drawing (Callaghan, 2015); individuals with special skills in drawing often create products that we celebrate and admire, and that have considerable monetary value. Nonhuman animals, by contrast, do not produce drawings spontaneously, although some captive animals have been encouraged to draw by trainers. The accepted view is that although intense training may result in animals acquiring some mark-making abilities, the works that they produce cannot be judged in the same way as those produced by humans.

Drawing may also serve a social function as art is usually produced for others, either as a communicative medium or for pleasure. While many animals live and grow in
complex social groups, humans have also evolved a unique suite of cognitive skills and motivations—collectively referred to as shared intentionality—for living collaboratively, learning socially, and exchanging information in cultural groups (Tomasello & Herrmann, 2010). Drawings hold information for both viewer and producer, and to a certain extent serve as a communication tool. In this way, our ability to draw, like our other communication skills, may have been unique drivers for evolutionary change within our species.

Researchers have repeatedly tested the ability of animals to create artwork, using a variety of training methods and mediums. These tests have resulted in some examples of mark making, but no conclusive evidence of the intent to communicate ideas, a process that emerges spontaneously in children. Animals in captivity have been trained through a variety of methods to produce marks on surfaces with tools. From elephants to primates, drawing has been used with captive animals to examine perceptual, motor, and psychological processes, by tempting animals to produce drawings or paintings in some way. In one example, an Asian elephant named Siri was observed making marks on the concrete floor of her enclosure with a stone held in her trunk. A keeper subsequently introduced drawing materials to her and permitted her to continue mark making (Gucwa & Ehmann, 1985). Siri’s “artwork” quickly achieved public recognition and other zoos and animal sanctuaries have begun to use art as an enrichment activity for captive elephants. The justification for this approach has been based on the assumption that painting is a therapeutic activity; in the case of the elephants, it has been argued that the paint brush may emulate the natural use of tools in the wild (Hart, Hart, McCoy, & Sarath, 2001).

When these assumptions have been empirically tested, the benefits of painting have been shown to be minimal for the animal (English, Kaplan, & Rogers, 2014), and the interpretation of the works have been shown to be anthropomorphic. For example, although art critics have lauded elephants’ use of form and colour (Gilbert, 1990), the stroke patterns,
which emerge in elephants’ paintings, have been shown to reflect the stereotypical patterns of trunk movements in these animals rather than an attempt to generate shapes or forms (Shoshani, 1997). Furthermore, elephants' colour vision is comparable to that of colour-blind humans; they are unable to see the full spectrum of primary colours, and any suggestion that their “use” of colour is deliberately generated is misleading (Yokoyama, Takenaka, Agnew, & Shoshani, 2005).

In addition to elephants, other captive animals have also been taught to paint. For example, dolphins and sea lions are taught how to apply paint to paper for the entertainment of the public. It is important to note that marine mammals who have been trained to use colours to paint—seals, dolphins and whales—all have monochromatic vision, and are unable to see in colour (Crognale, Levenson, Ponganis, Deegan, & Jacobs, 1998; Fasick, Bischoff, Brennan, Velasquez, & Andrade, 2011; Griebel & Schmid, 2002). Any suggestions of conscious choice or a psychological sense of aesthetics are again unwarranted.

The five great ape species (orangutans, gorillas, chimpanzees, bonobos, and humans) share perceptual, cognitive, and behavioural traits, and descended from a common ancestor over 15 million years ago. All great apes communicate through a range of vocalisations, use tools in their environments, and live in complex social groups. All great apes are capable of learning, causal reasoning, and planning. (Mulcahy & Call, 2006; Tomasello & Call, 1997). Given these similarities, whether drawing is a skill that all apes can utilise has been the subject of intense examination (Tomasello & Herrmann, 2010).

Chimpanzee drawing has been used to examine primates' perceptual organisation and also to uncover the common psychological basis for art making. Chimpanzees and the other great apes are the only animals whose drawing behaviour in some way parallels that of humans, although in the wild, there is no evidence of spontaneous mark-making by great apes (Whiten, McGuigan, Marshall-Pescini, & Hopper, 2009). When provided with drawing materials in captivity, chimpanzees will begin to draw without reinforcement or training.
(Boysen, Berntson, & Prentice, 1987). Their ability to make marks has been studied for decades, starting with Kohts (1935) who drew comparisons between her personal observations of a male chimpanzee (Joni) from the age of 18 months to 4 years and of her son (Roody), born a decade later. In her comparisons, Kohts observed that Joni failed to progress in his use of pencil, but by age 3, Roody could complete elementary sketches of nearby objects and would identify the topics in his drawings.

When learning to draw, human children initially scribble, but soon begin to produce simple shapes. This pattern of human development led early researchers to examine whether chimpanzees also produced shapes when they were given the opportunity to draw. Schiller (1951) examined the drawing behaviour of a captive-bred chimpanzee, Alpha. Alpha had been exposed to drawing in captivity for 10 years before Schiller’s systematic assessment of his drawing behaviour over a 6-month period. Schiller was interested in seeing if Alpha would complete incomplete shapes when given stimulus pages with incomplete line drawings of shapes on them. In Schiller’s experiments, the incomplete shapes, such as half-completed triangles, circles, and squares, were located in the centre of test pages. Alpha repeatedly scribbled in the centre of the page. This scribbling was taken as evidence that the ape was systematically trying to complete shapes. The findings were presented as evidence for the principle of closure in a non-human animal.

Subsequent studies by Smith (1973) and Boysen, Berntson, and Prentice (1987) revealed that chimpanzees display distinct preferences for making marks in the centre of pages, regardless of what shape is already presented on the paper. For example, Smith (1973) examined the mark-making tendencies of three captive-bred chimps, Audrey (3 years, 4 months), Leo (5 years), and Walter (4 years). The test was administered to the animals in a setting that they were used to, along with their usual handler present in their enclosure. The animals were presented with 5 different sheets of paper with shapes in different positions on each page. Black and blue crayons were provided as the animals preferred to eat red crayons
rather than to draw with them. The animals produced 100 drawings each although Smith noted that, on average, 20% of these drawings were produced when the animals were distracted. Animals tended to fill in blank spaces in the figures provided on the pages, no evidence of completion of an incomplete shape emerged from the systematic study of any of the chimp’s work. Boysen et al. (1987) found similar patterns to Smith. In their study, three chimpanzees examined under circumstances similar to those used by Smith also tended to mark in the centre of pages regardless of what was already printed on the paper. These findings suggest that animals are not making meaningful attempts to complete figures, and instead merely exhibit a tendency towards mark making in the centre of the page.

Pens or brushes are easily gripped in children’s hands; apes, by comparison, have long curved fingers and short thumbs set well back on their hands. These morphological differences may present a challenge in comparing mark-making across species. In more contemporary studies, researchers have attempted to overcome the possible difficulties that animals experience in using a pen by employing touch-screen technology. For example, chimpanzees have been encouraged to trace lines between points on the screen using their fingertips (Iversen & Matsuzawa, 1996). The animals in this experiment did succeed in closing shapes on a screen, but only after extensive training; their performance on this task did not reflect self-directed representational drawing. Golomb (1992a) has argued that representational drawing depends on the ability to produce enclosed forms such as circles and ovals. Without the capacity to complete simple shapes, ape drawing does not progress in the same way as human drawing. Furthermore, drawing by human children does not require training; by the age of 3, children can produce simple representational drawings, using a variety of shapes such as circles and triangles.

What about other elements of ape drawing? Are they similar to those of humans? Apes have colour vision that is similar to that of humans, so what about their colour choice? Colours have a significant, though not critical, role in visual art. Zeller (2007) investigated
whether colour choice in drawings and paintings by great apes is random or shows conscious decision-making. Zeller collected 396 works by apes in captivity from a range of facilities and catalogued them in terms of the number of colours used, first and last colour preferences, use of novel colours, respect for boundaries, negative space, placement, and pattern. She compared these works to those of 57 paintings produced by human children in their homes. The artworks were created on a variety of blank surfaces using watercolour paints, magic markers, or pencils; paint in these works was applied by brush, stick, pencil, fingers, or tongue. Some were only three or four inches across while others measured several feet.

Zeller (2007) personally observed the production of some of these pieces, but she received others from institutions housing captive apes. While great care was taken by Zeller to allow the animals to familiarise themselves with her and the task at hand before collecting data, a range of methods had to be employed to work with animals. Some animals would paint on pages held outside the bars of their cage; others would only produce artwork in complete privacy with no human observers. Zeller collected 40 pictures from chimpanzees, 153 from gorillas, 146 by orangutans, and 57 from human children under the age of 4. All were given the choice of 13 to 19 different colours and water which could dilute and change hues, and the paintings produced were then analysed for common themes and choices. Zeller’s findings suggest distinct preferences for certain colours, for example, yellow was commonly chosen as a first colour by both humans and apes and blue was the colour used most often in pictures overall. The variety of conditions under which the paintings were produced, as well as differences in the materials that were provided should be considered in relation to these findings. The paintings in Zeller’s analysis came from a variety of sources, and the presentation of palettes and colours to the animals may not be equivalent. Without systematic study, it is difficult to draw any conclusions about the animals’ intent or about possible connections between ape drawing and children’s drawing. A child painting at home
with an adult and an adult orangutan reaching through the bars of his cage to paint are not necessarily comparable.

Despite some similarity in their colour preferences, the act of drawing recognisable objects appears to be a solely human behaviour. There are some measurable differences between humans and other animals which impact drawing ability. Motor control, cognitive development, social facilitation, and language all appear to contribute to humans’ ability to create drawings. For example, humans have greater fine motor control and a larger amount of brain area devoted to hand-eye coordination than do their non-human counterparts. Physical limitations of animals also contribute to their limitations in tool use; apes lack an opposable thumb, which gives humans increased dexterity in holding small tools and completing fine motor movements. A historical review of drawing in human beings clearly illustrates its critical role in the development of symbols and artwork in the species. From 350,000 years ago, we as humans have taken tools and adapted them to produce marks, symbols, and illustrations of a range of topics, both real and imagined. Further examination and comparison confirms that drawing is unique to humans, has been a feature of cultures across the globe, and has changed relatively little during the recorded history of humankind.

**How does drawing develop?**

The drawings produced almost 40,000 years ago at Chauvet’s Cave appear at different levels on the walls, suggesting artists of different heights (Pettitt, 2008). Although it is impossible to know whether these shorter artists were children, we do know that the development of drawing skills in contemporary human populations is a lifelong pursuit. Beginning in early childhood, drawing emerges spontaneously when children are given materials to make marks (Brooks, 2009; Coates & Coates, 2006; Feldman, 2000; Watts, 2010). Drawing skills improve with both age and experience during development, and once learned, these skills are maintained (Andrade, 2010; Davis, 1997a; Oğuz, 2010; Rosenblatt & Winner, 1988).
In addition to its artistic value, drawing gives children an early symbol system which prepares them for the symbol use that formal education will rely upon in mathematics, reading and science (Athey, 1990; Barratt-Pugh & Rohl, 2000; Cox, 1991; Eisner, 1972; Gifford, 1997; Matthews, 1999). The agency afforded by being able to use symbols as representations and the ability to assign meaning to marks has been essential to the evolution of our species as well as contributing to development within our own lifetime.

Most children’s drawing development follows a highly-predictable pattern, changing from unintelligible scribbles on a page to increasingly recognisable, detailed, 2-dimensional representations of objects and ideas. Beginning around 2-years of age, children will enthusiastically create scribbles on surfaces by gripping a crayon or pencil in their hand in a fist or pincer grip. Many theorists have considered that these initial scribbles to be exclusively kinaesthetic—produced purely for motor pleasure (e.g., Arnheim, 1954; Lowenfeld, 1947; Luquet, 1927). Children’s lack of interest in their scribble creations after they have completed them has contributed to the belief that these early mark-making activities are a motor activity and not for the purpose of representing objects graphically (Ring, 2006; Thomas & Silk, 1990). Nonetheless, scribbling results in marks being made, and over time, a range of different kinds of marks are produced such as circles, triangles, crosses, and lines (see Figure 1.5 for examples).

With social support, children soon begin to label or attribute meaning to their scribbles. Drawing, like other childhood activities, does not occur in a vacuum but is influenced by a range of social and cultural expectations. Humans, as previously outlined, are the only animal with complex symbolic symbols to represent the world around them; language is one symbol system and drawings are another. By 2 years of age, children understand that both words and 2-dimensional images can be used to represent people or things (Callaghan, 1999; Hall, 2009; Hope, 2008; Kress, 1997; Matthews, 2003; Ring, 2006). Gradually, children begin to ascribe meaning to their scribbles or will plan to draw specific scribbles to represent actions or
objects. The elements included in early scribbles, such as lines, circles and patterns are subsequently utilised by children to produce drawings that represent a person, place, or thing. Children will start refining their scribbles, and it is often the people and things of interest to the child that will begin to be represented in their drawings (Hope, 2008; Matthews, 2003).

*Figure 1.5. Left panel: Scribble drawing by 2.8-year-old child labelled "my family". Right panel: Scribble drawing by 3.2-year-old child labelled "the playground".*

By the age of 3, children spontaneously ascribe meaning to their own drawings, as well as to the drawings of others even when the drawings per se do not contain sufficient information for adults to identify them accurately (Bloom & Markson, 1998). In one study, for example, Gross and Hayne (1999) asked 3- to 6-year-olds to draw three different events and to talk about the content of their drawings. Each child was asked to draw a picture of a birthday party, a trip to the park, and shopping at the supermarket. Following delays ranging from 1 day to 6 months, children were first asked to select their drawing from a group of 4 drawings—3 of which had been produced by children of the same age. Once they identified their drawing, they were asked to describe the specific things they had drawn. Gross and Hayne found that even children as young as 3 could recognise their own drawing after a delay. In addition, they could also describe the specific elements of the drawing (e.g., “here are the swings,” or “this is the supermarket trolley”) even though many of these elements were
completely uninterpretable by adults. Taken together, these findings suggest that children include symbolic information in their early works which may not be evident to adults because of their highly abstract nature.

By the age of 3 or 4, children also start to produce recognisable drawings of humans. These figures lack substantial detail, but they are nonetheless recognisable to most adults as human figures. Children’s first attempts at representing the human form are commonly referred to as tadpole drawings because they usually consist of a single central shape representing both the head and the body, with some or all limbs projecting from this central figure. Despite the shortcomings of tadpole figures, these early drawings are enthusiastically received by adults as representations of a person. Children’s human figure drawings gradually begin to include more elements and detail as they mature, and by the age of 5 or 6, most children can draw conventional representations of people that include all the main elements (head, body, limbs). Figure 1.6. shows examples of human figure drawings by young children featuring the progression from tadpole figures to recognisable human forms.
A: Human figure drawing by a 3.4-year-old child

B: Human figure drawing by a 3.9-year-old child

C: Human figure drawing by a 4.6-year-old child

D: Human figure drawing a by 5.6-year-old child

*Figure 1.6.* Examples of human figure drawings produced by 3- to 5-year-old children. Panels A and B represent classic tadpole drawings.

From the age of 6 or 7 children show a growing concern with realism in their drawings. In terms of human figure drawings, increasing complexity, differentiation, and elaboration are
marked by the addition of more body parts and features and more sophisticated drawing techniques such as continuous contour lines which allow children to link parts of the body previously drawn as separate segments (Harris, 1963). Figure 1.6 shows a range of drawings which illustrate the development of figure drawing from simple tadpole figures in plates A and B (a head and limbs) to the increasing details and facial features included in Plate C to the continuous outline of the whole body and details in Plate D (hair and fingers and whole figure).

Human figures are not the only source of inspiration for children’s emerging drawings; a number of common themes and schema are also shown in young children’s art. Between 5- and 8-years of age, children will start to include a range of elements in their drawings, such as a baseline for the ground, schematic representations of familiar objects and a range of viewpoints in one image. In addition, young children tend to draw what they know an object looks like, from a canonical viewpoint. For example, a cereal box on a table top will be represented from the front even if the rest of the drawing composition is viewed from above. Plate D in Figure 1.6 shows a family seated for dinner. Items on the table top are shown from the front view, but the scene is clearly from an overhead perspective of the table. Schematic representations of familiar objects are also used in young children’s drawings. For instance, houses are routinely drawn with A-frame roofs and chimneys; this becomes the way of drawing all houses. Figure 1.7 shows a series of drawings produced by children between the ages of 5 and 7 which illustrate a range of these production deficits.
Figure 1.7. Drawings by children showing a variety of inaccuracies common in drawing development.

Seven-to 11-year-old children will start to represent proportion, overlapping, and scale. They include more details in their drawings and increasingly attempt to produce drawings which are photo-realistic. By the time that children reach 8 or 9 years of age new
developments emerge in drawn representations of people. Instead of drawing front-facing people, an interest in drawing figures in profile often emerges. Children eight to 9-years old attempt to harmonise body parts in correct proportions (Cox, 1993). With practice, children’s drawing development will start to show additional features, such as occlusions, where parts of the represented objects or figures may obscure the view of other parts, such as an arm across the body which hides some detail, or a person standing in front of another will hide part of the second figure.

Beginning at the age of 2 until about 11-or 12 years of age, children’s drawing ability consistently improves (Kindler & Darras, 1997; Lowenfeld, 1947; Willats, 2005), but between the ages of 11 and 14, progression abruptly slows and stagnates, usually due to frustration with the inability to represent things satisfactorily (Davis & Gardner, 1992; Gardner, 1980; Rosenblatt & Winner, 1988). Without regular practice, drawing, like many skills, does not improve and will remain the same throughout later life (Kindler & Darras, 1997). Despite this typical stagnation, an individual’s drawings can become more detailed, accurate, and lifelike over the course of their lifetime if sufficient time and attention are devoted to practicing drawing. For example, Figure 1.8 shows the progression of drawing skills through practice over the course of 3 years starting at 15 years of age.

*Figure 1.8. Improvements in the drawing human eye by the same artist between the ages of 15 and 18.*
Research on drawing that was conducted in the first half of the 20th century relied predominantly on large collections of drawings produced under a variety of conditions; these drawings were then collated to identify common themes and patterns of change. Using this approach, we know that most children develop drawing skills over time and that their drawings develop from simple scribbles to more realistic graphic representations, becoming more easily understood and interpreted by others. Given the systematic way in which children’s drawing changes with age and experience, it is hardly surprising that a number of theories have been put forward in an attempt to systematically describe how and why this process takes place (see Appendix A for a summary of these stage theories). The chief concern with the stage models outlined in Appendix A is that they cannot explain why children progress through each stage, abandoning earlier strategies in favour of others. In contrast to the early theorists, empirical researchers working in the latter half of the 20th century brought an experimental methodology to bear on drawing development. In contrast to classifying children into stages, contemporary research has instead examined how children’s drawings are related to the demands of a particular drawing task and how these demands influence children’s drawing production over time (Barrett, Beaumont, & Jennett, 1985; Cox, 1985, 1991; Freeman, 1972, 1980; Gilchrist, 1991; Golomb, 1992; Lansing, 1984; Milbrath, 2012; Smith & Gilchrist, 2005).

The psychological value of children’s art

The uniquely human ability to produce drawings has raised important questions about the value of drawing in many clinical settings. For example, neurologists sometimes use drawings as an adjunct to detailed medical examinations; abnormalities in drawing production can form part of an overall diagnostic assessment for a range of potential problems, including neurological disorders, motor learning problems, psychiatric diseases, and the effects of alcohol or substance abuse (Mergl, Tigges, Schroter, Moller, & Hegerl, 1999). Researchers
have clearly shown that Parkinson’s disease, old age, and neurological insult all result in loss of fine motor control which effects drawing skill and accuracy (Smith, Gilchrist, Butler, & Harvey, 2006; Van Gemmert & Teulings, 2006).

In addition to the role of drawing as a diagnostic tool in some medical conditions, a number of theorists, researchers, and clinicians have suggested that there is also a link between an individual’s drawings and their mental states, arguing that the drawings of mentally-ill people may offer insight into their psychological make-up (Riethmiller & Handler, 2010). Much of this work can be traced back to the work of Freud (1921). Freud’s analysis of the artistic output of Leonardo Da Vinci sparked the argument that artistic output could be used as an insight into personality. In Freud’s book he proposed that the paintings and drawings of Da Vinci could offer insights into the artist’s personality in the absence of any verbal reports. This diagnostic methodology was rooted in Freud’s own theories that drawings and dreams are representations of unconscious drives or repressed conflicts within the individual which may only be brought to the surface in symbolic expression. Freudian theory suggests that drawings, like dreams, are a way of accessing the unconscious.

Freudian theory subsequently gave rise to the development of many drawing tests that are currently used in clinical settings. The use of drawings as a measure of underlying psychological states proliferated in the first half of the 20th century and drawings produced by the mentally ill have been interpreted from existential, psychoanalytic, or phenomenological schools of thought (Kris, 1952, Pickford, 1970, Reitman, 1951). While the artistic output of people with mental illness may include insight into their world view, only 2% of those hospitalised for psychiatric disturbances spontaneously produce art. Despite the low level of artistic output from this population, the belief that drawings include inaccessible information about a person proliferated and a number of drawing tests were developed for use with a range of populations (Golomb, 1992).
Additional drawing tests were also created to assess emotional and cognitive states and traits in the more general population (Buck, 1948; Burkett, Barrett, & Davis, 2003, 2004; Burkett & Newell, 2005; Groth-Marnat & Roberts, 1998; Gulbro-Leavitt, & Schimmel, 1991; Kessler 1994; Machover, 1949). By the mid twentieth century a range of drawing tests were being employed to assess personality, evaluate current or past emotional states, access personally-sensitive information, and for the assessment of intelligence or developmental level. In all cases, drawings were used as a proxy for other forms of information—that is psychological conclusions were derived on the basis of what someone had drawn. A sizeable percentage of clinicians, across a range of contexts, from education to therapeutic settings continue to utilise these projective drawing tests (for a review of studies of drawing assessment usage from 1960 to 1999 see Piotrowski, 1999).

Despite the popularity of the projective drawing tests, psychological science has questioned their use for over 50 years. In particular, serious questions have been raised about their reliability, validity, and utility (Anastasi, 1982; Lilienfeld, Wood & Garb, 2000). The adequacy of any clinical instrument is typically assessed on the basis of reliability, validity, and utility. The reliability of a test refers to the degree to which scores on the test consistently show the same result over time and the result is the same no matter who is using the instrument. The validity of an instrument has two core components, internal validity; which ensures that the instrument is measuring the specific characteristic it claims to measure, and external validity which ensures that the findings are applicable across populations. The utility of an instrument is the practical usefulness of that instrument. For example, a long, detailed questionnaire may provide a fuller picture of an individual’s mental state, but it may be difficult to get people to answer hundreds of questions on such an instrument. Typically, objective instruments used in clinical settings must show high reliability, validity, and utility, in order to be adopted for use. Substantial criticism has been raised concerning these three evaluative criteria with respect to projective tests which use drawings.
The springboard for many contemporary projective drawings tests was Machover’s Draw-A-Person (DAP) personality test (Machover, 1926). In this test, individuals are given a pen and two sheets of paper and are asked to draw a person on both. The drawings produced by the individual are then interpreted by a clinician on the basis of a range of features outlined by Machover, such as size, placement, clothing and various included or omitted details. The figures are assumed to include acceptable impulses on the same-sex figure and unacceptable urges on to the opposite-sex figure. Sexual maladjustment is indicated if an individual draws a person of the opposite sex in the first drawing (Carrera & Honigman, 1955; Fleming, Koocher, & Nathans, 1979). The Machover Draw-A-Person test was initially developed for adolescents and adults and was extended for use with children as young as two, with no allowances made for children’s emerging drawing ability.

From a psychometric standpoint, there are a number of problems with the Machover test. For example, no reliability or validity data were reported by Machover (1926). In subsequent research, the Machover DAP has consistently failed to be reliable or valid. For example, Wanderer (1969) provided 20 experts in Machover’s DAP test, including Machover herself, with pairs of drawings to analyse. These drawings were made by 5 persons who were labelled as mentally-defective, homosexual, schizophrenic, neurotic, or control. The experts were given the list of possible labels and tasked with identifying the person who produced each drawing. With the exception of the mentally-defective person’s drawing, none of the experts could reliably predict the psychological condition from drawings (see Swenson, 1957 for a review of studies that have failed to find support for the use of Machover’s Draw a person as a reliable or valid measure of personality). Despite these findings, the use and popularity of the Machover test persists.

Many projective drawing tests developed in the latter half of the 20th century use one of two approaches to assess and interpret the drawings produced by individuals. One approach has been to attribute specific meaning to isolated elements of drawing; the other approach has
been to use a scoring system for a whole drawing which is then used to establish an individual’s level on a psychological construct (Lilienfeld et al., 2000). Assigning value or meaning to particular aspects of an individual’s drawings was championed by many psychodynamic theorists and has been applied across various populations. One way in which this ‘signs approach’ was utilised was to identify homosexuality, assuming that homosexual men would include particular aspects in their drawings of a person. In practice, however, it has proven difficult to reliably identify sexual identity on the basis of the presence or absence of particular signifiers. For example, Craig and Bovan (2002) examined the drawings of 88 self-identified gay men and 88 heterosexual men for 22 signs that presumably indicated homosexuality (Barker, Mathis, & Powers, 1954; De Martino, 1954; Hammer, 1954; Jolles, 1971; Levy, 1950; Machover, 1951; Schildkrout, Shenker, & Sonnenberg, 1972; Verner, 1952). Craig and Bovan (2002) compiled a list of 22 signs of homosexuality that could be found in human figure drawings according to theorists working from a ‘signs approach’ These include hair emphasis, unusual eyes/eyebrows/ears, eyelashes details, indented waist, phallic protrusion, boots or high heels on males, male effeminate features, and tie emphasis. Participants were asked to draw a person. Two independent raters, unaware of the purpose of the study and the sexual identity of the artist, were given the list of indicators and were asked to assess each drawing for signs of homosexuality. Only 2 variables, hair and hips, were more frequently emphasised in the drawings of gay men, but even these indicators occurred in less than 50% of the drawings. Overall, the raters were unable to reliably distinguish between the two groups based on the details contained in their drawings.

In addition to individual signs in drawings, a variety of other instruments have employed whole scores obtained by assessing a variety of individual elements in drawings. The majority of these tests are modifications or additions to Koppitz’s Emotional Indicators of Human Figure Drawing (1966), Buck’s House-Tree-Person test (1948), and Burns and Kaufman’s Kinetic Family Drawing test (1970). For example, Koppitz (1968) argued that emotional
indicators are contained in various aspects of children’s drawings and that clinicians can use these indicators to identify a child’s underlying attitudes and characteristics. The structure of the drawings produced by children in Koppitz Draw a Person (K-DAP) is determined by the age and maturation level of the child, but the styles used in the drawings are indicators of attitudes, concerns, and anxiety. A total of 30 Emotional Indicators (EI’s) are outlined in the K-DAP scoring system. These EI’s are signs of poor emotional functioning and are derived from studies of children’s drawings produced in clinical settings and from the Goodenough-Harris DAP test of intelligence (see Figure 1.9 for examples). They include the absence of facial features or body parts, over and undersized body parts, crossed eyes, shading in specific areas, legs pressed together, and genitals (Eno, Elliott, & Woehlke, 1981). Each item is classified as expected, common, not common, or exceptional, according to the norms established by Koppitz. The norms for the K-DAP are based on drawings obtained from a sample of 1856 American children between 5 and 12 years of age. The reliability of the Koppitz scoring system was established with 50 children between 6 and 12 years of age and yielded inter-scorer reliabilities of .97, and validity coefficients with the full scale of the Weschler Intelligence Scale for Children WISC Verbal IQ (.54), and WISC Performance IQ (.67) (Gayton, Tavormina, Evans, & Schuh, 1974). Similar, but not quite so high, levels of inter-rater reliability was found by Rae and Hyland (2001) using Koppitz’s scoring scheme (.84-.86), however they report test-retest reliability over a 2-week period of only .32. Why a test of emotional functioning was validated using IQ scores is not clear, but the first version of Koppitz’s emotional indicators were modelled on Goodenough’s Draw a Man test of intellectual ability (Goodenough, 1926). Koppitz (1967) Human Figure Drawing scoring system was a simplified version of the Gooenough-Harris (1963) scoring system. Koppitz’s EIs were initially used to differentiate between the drawings of children in mainstream education and those in psychiatric care (Koppitz, 1966). Koppitz’s EI thus provided a measure of mental maturity (IQ score) to adequately match groups to be compared, along
with a range of indicators which assess emotional functioning with no adequate validation of the emotional components of the test.

Figure 1.9. Examples of Koppitz EI drawings, reprinted from *Journal of Paediatric Nursing*, Jackson & Ott, 1990.

Traditionally, the K-DAP has been used to differentiate between clinical and non-clinical populations on the basis of the presence or absence of emotional indicators, but there is limited evidence for the validity of the test. In a classic study of this kind, Pihl and Nimrod (1976) administered 44 children between 10 and 11 years of age, the K-DAP test and the Children’s Personality Questionnaire (CPQ). Two clinical psychologists rated each drawing on
the personality dimensions outlined in Koppitz’s scoring manual; self-assured/apprehensive, reserved/outgoing, relaxed/tense and happy/depressed. The clinicians agreed about what aspect of the drawings they were rating (according to the 4 dimensions outlined in the scoring methodology), but none of their scores on these measures were correlated. Seven out of 12 correlations were negative and those that were positively correlated \( (r = .26-.29) \) failed to achieve significance. In addition, there was no relation between the clinician’s scores on the K-DAP were compared with neuroticism, anxiety and depression scores obtained from the child personality questionnaire (CPQ), no significant correlations were found (Lessing & Smouse, 1967). More recently, Catte and Cox (1999) failed to distinguish between a group of emotionally-disturbed boys between the ages of 6 and 11 years of age and a matched group of well-adjusted boys using Koppitz’s DAP scoring system.

Koppitz’s EI norms were developed over 50 years ago in the USA. In an attempt to address any differences in drawing that may have emerged over time, Catte and Cox (1999) examined the percentage of children producing Koppitz emotional indicators at a range of ages. Catte and Cox administered the DAP to 1598 children between 5 and 11 years of age in the UK. They found that seven of the 30 original indicators were no longer present in the drawings produced by their sample. Of the 23 remaining indicators, six of these appeared at different ages in Catte and Cox’s sample than predicted by Koppitz’s scoring system. When they re-evaluated their original sample of emotionally-disturbed and control boys using these revised indicators, Catte and Cox still failed to differentiate between them on the basis of indicators included in their drawings.

Another projective drawing technique is Buck’s (1948) House-Tree-Person test (HTP), designed to assess personality; each drawing in the test corresponds to a different aspect of the artist’s personality. In the test, a person is given 3 pieces of paper and is asked to draw a House, a Tree, and a Person on separate sheets. The House is thought to reflect the persons’ familial relationships, the Tree is thought to reflect deep rooted self-concept, and the Person is
believed to represent the artist’s ideal self. To assess personality, a range of aspects of each drawing are interpreted by a clinician using the HTP guide. For example, when interpreting a drawing of a house, doors and windows are used to interpret the artist’s relationship with the surrounding world. Many windows and doors indicate neediness, but large windows or doors indicate openness. Other aspects of the drawing such as sidewalks, roof styles, chimneys, brickwork and placement on the page are also considered to reveal underlying aspects of the artist’s personal relationships (see Figure 1.10 for example). Details included in any of the three drawings from the HTP can be interpreted as indicating trauma, discord, or unhappiness (Hammer, 1969). The presumption in these tests is that drawing directly reflects the artist’s own body image, self-esteem, or some other underlying state (Merriman & Guerin, 2006).

Figure 1.10. Examples of HTP drawings completed by adults. The interpretation of these drawings using Bucks’ HTP scoring system would suggest that the absence of feet on the person figure and the absence of root structures visible on the tree figure indicate a lack of autonomy or powerlessness. The level of detail on the roof of the house figure would indicate that this person focuses on or indulges in fantasies, while the inclusion of bushes
and foliage around the house could indicate that this person is hesitant to open themselves to others (reprinted from Merriman & Guerin, 2006).

Buck’s original HTP test was developed and normed on the basis of drawings produced by a sample of 140 adults from the general population (Buck, 1981). The sample was broken down into groups of 20 based on evaluations of their intelligence; imbecile, moron, borderline, below average, average, above average, and superior. These classifications are not based on standardised intelligence testing and no explanation of why intelligence was used as a grouping mechanism for personality was made. No attempt was made to balance gender within these groups and the age range of the entire sample ranged from 13.6 to 29 years of age. In terms of reliability and validity, the original 1948 manual for the HTP test contains no reliability or validity information. In the 1981 revised version of the HTP scoring manual, Buck himself states that the validity of the qualitative scoring system fails to show statistical validity or reliability but is a ‘mature clinical instrument’.

Despite these limitations, Buck’s HTP test was adapted to be used as a measure for other underlying states. One use was a modification of the test to assess abuse in children. A 1991 survey of 212 mental health professionals in the US indicated that 87% of them used some form of free-drawing exercise in their evaluations of abused children (Conte, Sorenson, Fogarty, & Rosa, 1991). Recently, Palmer et al. (2000) investigated the clinical validity of the HTP test to discriminate between children who were sexually abused and a non-abused comparison group. According to the HTP scoring booklet (Van Hooton, 1994), these groups should differ on aspects contained in their drawings. Each drawing was scored by 2 independent raters who had completed extensive training in HTP scoring. Despite extensive training in the scoring system, Palmer et al. reported reliability co-efficients of only .60 and .39 for the sub-scales of the HTP test. Palmer et al. employed 6 raters in total rather than the usual method of using 2 people to establish inter-rater reliability, which may give more insight
into the variety of individual interpretations of elements within a drawing by clinicians in the field who all come from a variety of training and educational backgrounds. Additionally, in Palmer et al.’s experiment, none of the sexual-abuse indicators in the children’s drawings reliably predicted which group the drawing came from. The indicators included emphasis on the bedroom, large open windows, and figures drawn more or less mature than the child’s age. The results of Palmer et al.’s experiment study suggests that HTP drawings do not reliably contribute to assessments of child sexual abuse. While drawing offers a low-stress facilitative technique for clinicians to establish rapport and allow a child to engage in an enjoyable activity during interview sessions, in particular in cases of highly-distressing abuse situations, the claims of projective drawing techniques as diagnostic tools remain unsupported (Cohen & Phelps, 1985; Joiner, Schmidt, & Barnett, 1996; King, 1960; Koppitz, 1948; Lindgren, 1971; Thomas & Silk, 1991).

A different approach to projective drawing techniques was championed by Burns and Kaufman (1970), who insisted that dynamic, action-orientated drawings, provided key insights about a patient. In their Kinetic Family Drawing (KFD) test, children are asked to “draw your family doing something” with an emphasis on drawing a dynamic action rather than a stationary image of the family. According to Burns (1982), this instruction allows the child to place him or herself in a functioning active unit and allows the clinician gain insight into the interactions amongst the people in the drawing. When administering the KFD, a child is provided with a plain piece of paper and a pencil and asked to “Draw a picture of everyone in your family, including you, doing something. Try to draw whole people not stick people, Remember, make everyone doing something—some kind of action”. The child completes the picture and is then asked to describe and identify each element in the picture whose symbols and actions are later interpreted by a clinician according to the KFD scoring manual. According to Burns and Kaufman (1972), various aspects of a familial relationship can be determined from careful assessment of the individual features, placement, and interaction of individual figures in the
drawing and additional items and symbols included in the artwork. For example, according to the KFD scoring system, the omission of body parts indicates conflict or anxiety, inaction of a figure indicates negative attitude towards that person, inclusion of throw-able objects such as a ball indicate competition, jealousy, or aggression (for another example, see Figure 1.11).

![Figure 1.11. KFD Family drawing with figures showing compartmentalisation, which the KFD scoring manual would interpret as feelings of isolation or alienation by the artist (reprinted from Handler & Habernicht, 1994).](image)

Normative data for the KFD is scant. Burns (1982) provided normative data for 44 5-to 6-year-old children using the KFD, Jacobsen (1973) provided normative data for 146 children 6-to 9-year old children, and Thompson (1995) established KFD norms for adolescents using 196 KFD drawings by 13- to 18-year-old adolescents. Comparing scores compiled by independent raters yields good inter-rater reliabilities using the KFD scoring system (.87 - .95, Handler & Habernicht, 1994), test-retest reliabilities are more variable, with reliabilities ranging from 46% to 90% (Acosta, 1989). Proponents of projective measures argue that these re-test changes offer insight into the ever-changing dynamics of a child’s
perspective on relationships, however, a reliable instrument should remain relatively stable over time, while still being nuanced enough to pick up changes in an individual.

The KFD test has a scoring system which features four categories of drawing elements; actions, distance, barriers, and positions. Each part of a drawing is allotted a value which delivers a total to identify possible psychopathologies in the artist. The reliability of the KFD has been disputed. For example, Mostkoff and Lazarus (1983) administered the KFD to 50 elementary school children between the ages of 9 and 12 years of age on two occasions. Children’s drawings were evaluated based on 20 individual variables; number of people in the family, self in picture, style (compartmentalisation, edging, lining the top or bottom of page, etc.), barriers between figures, or drawings on the back of the page (see Mostkoff & Lazarus, 1983 for full list of scored variables). These variables were scored in both pictures and if they remained stable (present/absent in both) they were scored as the ‘same’ but if they did not remain stable that is they were not included in both drawings they were scored as ‘different’. All drawings were scored by two independent raters to assess inter-rater reliability (97% agreement). Less than 50% of the 20 variables scored appeared in both drawings (46-90% remained the same). Only 9 of the 20 variables from the KFD scoring manual appeared in both drawings, therefore the authors recommended that clinicians exercise caution in the use of all indicators included in the KFD. In many studies that have examined the validity of the KFD the researchers have utilised modified scoring systems making it difficult to compare and contrast studies (for a full review, see Handler & Habernicht, 1994). For example, Schacker (1983) compared only one variable (lining at the bottom of the paper) with marital adjustment of the parents of the artist. No significant relation was found between these two variables. Although the KFD was not designed to be used with single variables, this has not stopped many researchers and clinicians cherry picking elements of drawings to support projective hypotheses.
The fundamental assumption of any drawing measure is that it is possible for a clinician or a researcher to infer something about the artist’s inner mental life on the basis of what they have drawn. The key question, however, is whether these projective measures provide a valid measure of psychological constructs? When the KFD’s validity has been tested in field studies, it has also failed to adequately distinguish between controls and target groups. For example, Layton (1983) compared the drawings of two groups of children between the ages of 6 and 12 years of age. One group consisted of children with behaviour, social-emotional, and learning difficulties. The control group was matched for age and background. When the drawings from both groups were compared using 133 KFD signs, no significant differences were found. In fact, signs of ‘pathology’ were seen more frequently in the control group. Similar findings which question the ability of the KFD to distinguish accurately between normal and abnormal populations have been reported for adolescents (Sobel & Sobel, 1976), psychiatric patients (Joiner, Schmidt, & Barnett, 1996), and abused children (Hackbarth, Murphy, & McQuaid, 1988).

Despite well-documented problems with validity, many clinicians continue to attest to the clinical utility of projective drawing measures, arguing that their validity is not always revealed by empirical testing, and that an experienced clinician will be able to utilise the tests more effectively. To test the assertion that clinicians are better at interpreting projective drawings, Levenberg (1975) compared the ability of clinicians, pre-doctoral interns, hospital secretaries, and KFD experts to differentiate between abnormal and normal children based on their drawings. To do this, Levenberg examined 18 KFD drawings by 8- to 11-year old children with known psychopathology and 18 from a matched sample of children from a local elementary school who had no known behavioural problems. Behavioural problems in the control sample were rated by classroom teachers using a Child Behaviour Rating Checklist. The 36 KFD drawings were randomly ordered and the identity of the child artist was masked.
Levenberg (1975) used three discreet levels of clinical expertise were used to segment the judging groups; 5 per judging group were randomly selected from the following categories; hospital secretaries, post-doctoral level psychology graduates, and pre-doctoral interns. A KFD expert, the author of two books on the KFD, was also given the 36 drawings to assess. Each judge was given the drawings and KFD scoring system and was asked to assess whether the drawing came from a normal or clinical sample. The interns ($M = 61\%$) and post-doctoral level clinicians ($M = 72\%$) distinguished normal from disturbed children at a rate slightly better than chance, and they significantly outperformed the KFD expert ($M = 47\%$).

In studies conducted by Roe, Bridges, Dunn, and O’Connor (2006) and Wagner, Mills-Koonce, Willoughby, Zvara, and Cox (2015), researchers examined children’s family drawings in an attempt to predict family relationships. Roe et al. analysed the drawings of 166 children living in different family settings at two time points (when children were aged 5 and 7 years). The exclusion or inclusion of immediate family members and the grouping of figures in these children’s drawings were examined to see if changes in the pictures over time reflected the child’s view of who their family was. Patterns of association between these representations were also examined to see if they were predictive of behavioural disturbance. Only the child’s drawings of his or her parents and siblings (full and half siblings) were coded; drawings of other relatives such as grandparents, aunts, uncles, etc. were excluded from coding. At both test administration points, children drew and excluded the same people in their family drawings. Roe et al. also measured children’s adjustment using the Child Behaviour Checklist (CBCL) (Achenbach, 1991) and found that the mean differences in CBCL scores between the children who excluded family members and those who did not were extremely small, only 0.10 of a point difference on a 3-point scale. The degree of overlap in these samples would make it impossible to detect, on the basis of one child’s drawing, whether he or she were at risk of problems or not. These are similar to the findings noted by Imuta et al. (2013) who found poor predictive power of human figure drawings for classifying children’s intelligence. For
example, Imuta et al. found high false-negative and false-positive rates in screening, with an overall hit rate of only 38%, indicating that the majority of children were not correctly classified according to their intellectual level (for a full discussion of Imuta et al.’s findings, see Chapter 2).

Wagner et al. (2015) examined the family drawings of 1239 children when they were aged 5 as well as measuring conduct problems and callous unemotional behaviours. They also assessed parenting when the children were aged 24, 36, and 58 months. Although Wagner et al. reported significant correlations between scores on the family drawings and measures of parenting, conduct problems, and callous unemotional behaviours, the correlations were extremely small and Wagner et al. acknowledged that the statistical significance of the correlations was undoubtedly due to the very large sample size. Taken together, the studies by Roe et al. (2006) and Wagner et al. (2015) provide further evidence that it is impossible to learn anything from a child’s drawing in the absence of additional information about that child.

In conclusion, interrater reliability for projective drawing techniques is generally high, making them attractive across a range of settings, both in research and in clinical practice. The reliability of scores available from these techniques is frequently given as a reason for their continued use; Kahill (1984) reported that the interrater reliabilities of two-thirds of figure-drawing indicators was above .70. While these reliability scores relate to the whole drawing score, scoring of individual elements within drawings can vary to the extent that they are indistinguishable from chance. Projective drawing tests have been robustly criticised by many researchers who argue that it is impossible to use drawings to reliably assess underlying psychological states, and there is no way of reliably predicting when a child or adult will include or exclude important significant information in a drawing (Cohen & Phelps, 1985; Koppitz, 1966; Van Hutton, 1994). In addition, a range of researchers have established that diagnoses based on drawings are a poor substitute for empirically-validated instruments
(Peterson & Batsche, 1983; Sherman 1958). Questions have also been raised about the use of inappropriate statistical tests with projective drawing tests (Elin & Nucho, 1979) small or biased sample sizes, and low replicability (Kelley, 1984; Neale, Rosal, & Rosal, 1993) as well as vague research criteria (King, 1960). Despite extensive critique, repeated modifications to projective drawing scoring methodologies continue to appear, and the use of these tests across psychological practice continue, convincing some clinicians and researchers that drawings do speak louder than words.

**The Current Research**

The persistent use of children’s drawings as a source of information about internal mental states can be traced to one researcher in particular - Florence Goodenough. Florence Goodenough’s (1927) contention that children’s human figure drawings (HFDs) can be used as a measure of intelligence launched a range of subsequent tests discussed in this opening chapter. The use of HFDs as measures of intelligence continues; the most recent version based on Goodenough’s work is the DAP:IQ (Naglieri, 1998). The first empirical chapter of this thesis will highlight ongoing concerns about the validity of HFDs as measures of IQ. In Study 1, I will examine the effect of instruction on 11- and 12-year-old children’s scores on the DAP: IQ. In the second empirical chapter, I will focus not on the end product, the picture drawn, but on the process of drawing and its potential benefits for learners during field trips by evaluating the memory outcomes associated with drawing in museum galleries. Life drawing in galleries and museums is a common practice used by artists to develop their drawing skills, but drawing may also serve an additional educational purpose beyond the development of the skill to draw, per se. To date, no researchers have measured specific memory outcomes associated with drawing in museum settings. In the final chapter of this thesis I will further explore the possibilities for learning on field trips using digital tools—tools that are beginning to replace traditional pastimes like drawing during childhood. Despite an increase in dedicated screen time, few researchers have explored the memory outcomes
associated with using digital technology to learn. In the final empirical chapter, I will explore contemporary creative technology and its possible benefits in real and virtual field trips.
Chapter 2

The Effect of Instruction on Children’s Human Figure Drawing (HFD) Tests: Implications for Measurement

The empirical study of children’s drawing ability has a long history in psychology. Over 100 years ago, psychologists noted that children’s drawings follow a highly consistent pattern of development (for review, see Lansing, 1984). Across Western cultures, children’s representation of the human form begins with abstract marks on a page. These initial scribbles gradually shift to *tadpole figures* in which children represent the head and the trunk as a single shape, adding a single set of limbs (sometimes only arms and sometimes only legs) to the sides of this central figure. Over time, children’s drawings begin to represent more realistic human forms as they differentiate the head from the trunk and add increasingly more detail to their drawings, including elements such as eyelashes, clothing, and jewellery (Klepsch & Logie, 1982). By the time children reach adolescence, most are able to produce a recognizable human figure with no support (Maloney & Glasser, 1982).

Despite the uniformity of the general stages of drawing development, the speed with which individual children transition through each stage varies dramatically. For example, in a given sample of 4-year-old children, some only scribble, others produce tadpole drawings, and still others produce human figures with a separate head, torso, and four limbs. The highly consistent, stage-like development of children’s human figure drawings (HFDs), coupled with the large individual differences in children’s progression through these stages, has led some psychologists to argue that children’s drawings of the human figure might provide a surrogate measure of their intelligence (e.g., Bardos, Softas, & Petrogiannis, 1989;

---

2 This chapter is published as Cronin, A., Gross, J., & Hayne, H. (2017). The effects of instruction on children’s human figure drawing (HFD) tests. *Psychology of Aesthetics, Creativity, and the Arts*, 11, 179-186. DOI:10.1037/aca0000097. As the paper is reproduced in its full published form, there is some duplication with the introduction material presented in Chapter 1.

The first formal HFD test designed to measure intelligence was developed by Florence Goodenough in 1926. The basic premise of Goodenough’s Draw-a-Man Test (DAMT) was that the intelligence of early-school-age children could be established on the basis of their drawing of a human figure. In the DAMT, children were asked to produce their best possible head-to-toe drawing of a man. Children were told to include as much detail as possible. Each drawing was scored by assigning points for details of parts of the body that were included; the scores were summed to produce an overall HFD score. Estimates of a child’s cognitive ability were then made on the basis of the score obtained for his or her drawing. Harris (1963) subsequently extended the DAMT to include three HFDs; a man, a woman, and the self. Harris found that children’s scores on the new \( r = .47 \) Goodenough-Harris Drawing Test (GHDT) were moderately correlated \( r = .47-.74 \) with scores on standardized IQ tests, including the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949; \( r = .47 \), the Stanford-Binet Intelligence Scale (Terman & Merrill, 1960; \( r = .65-.74 \), and the Wechsler Adult Intelligence Scale (Wechsler, 1958; \( r = .56 \); Aikman, Belter, & Finch, 1992; Kamphaus & Pleiss, 1991; Scott, 1981).

Since the original publication of the DAMT and the GHDT, the procedure and the scoring systems for these tests have been revised multiple times to reflect modern trends (e.g., Draw-A-Person: A Quantitative Scoring System (DAP:QSS): Naglieri, 1988; Draw-A-Person Intellectual Ability Test for Children, Adolescents, and Adults (DAP:IQ): Reynolds & Hickman, 2004). For example, in Reynolds and Hickman’s (2004) most recent scoring system, the DAP: IQ, an IQ score is calculated based on a single drawing of the self. In addition, the scoring systems have been updated so that the influence of current hair and clothing styles is reduced (e.g., gender-neutral clothing). Because of its time- and cost-effectiveness and ease of administration relative to more traditional IQ tests, the GHDT and
subsequent HFD tests are extremely popular with school psychologists and clinicians (Anastasi & Urbina, 1997; Camara, Nathan, & Puente, 2000; Cashel, 2002; Maloney & Glasser, 1982; Skybo, Ryan-Wenger, & Su, 2007).

Although HFD tests are used widely by practitioners in clinical and educational settings, the critical factor in assessing the value of any psychometric measure is establishing its reliability and validity. With few exceptions, there is considerable evidence that HFD tests are highly reliable (e.g., Bardos et al., 1989; Gross & Hayne, 1998, 1999; Imuta, Scarf, Pharo, & Hayne, 2013; Kamphaus & Pleiss, 1991; Motta, Little, & Tobin, 1993; Naglieri, 1988; Nasvytiene, 2007; Prewett, Bardos, & Naglieri, 1989; Reynolds & Hickman, 2004; Willcock, Imuta, & Hayne, 2011; Williams, Fall, Eaves, & Woods-Groves, 2006). The scoring manual for the most recent HFD test, the DAP:IQ (Reynolds & Hickman, 2004), reports internal consistency (alpha coefficients) ranging from .74 to .87, along with test–retest reliability of .86, and intra-rater and interrater reliability of .95 and .91, respectively.

Researchers who have used these drawing tests have reported equally high levels of reliability. For example, Williams et al. (2006) reported interrater and intra-rater reliability scores of .83 and .92, respectively. Similarly, across four studies, Hayne and her colleagues reported interrater reliability scores of .95 and higher (Gross & Hayne, 1998, 1999; Imuta et al., 2013; Willcock et al., 2011). Equally high levels of reliability have also been reported for earlier HFD tests (Abell, Wood, & Liebman, 2001; Kamphaus & Pleiss, 1991; Lassiter & Bardos, 1995; Motta et al., 1993; Nasvytiene, 2007; Prewett et al., 1989; Scott, 1981; Wisniewski & Naglieri, 1989).

Despite their high levels of reliability, there has been considerable concern about the validity of HFD tests. Most of the evidence used to support their validity has come from correlational data. In general, scores on HFD tests are moderately correlated with scores on other standardized measures of intellectual ability (e.g., Abell et al., 2001; Kamphaus & Pleiss, 1991; Lassiter & Bardos, 1995; Motta et al., 1993; Nasvytiene, 2007; Prewett et al.,
1989; Scott, 1981). For example, significant correlations have been found between children’s scores on the DAP:QSS and their scores on the full-scale version of the Wechsler Intelligence Scale for Children - Revised (WISC-R; Wechsler, 1974; \( r = .42 - .51 \)), and on single subscales of the WISC–R (Wisniewski & Naglieri, 1989). Significant correlations have also been found between scores on the DAP:QSS and scores on the Wechsler Preschool and Primary Scale of Intelligence-Revised (WPPSI-R; Wechsler, 1989; \( r = .30 \); Lassiter & Bardos, 1995), the WISC-R \( (r = -.46 - .55) \), and the WISC-III \( (Wechsler, 1991; r = -.29 - .36; Abell et al., 2001) \).

Moderate correlations of this kind have been used to argue that HFD tests provide a valid measure of IQ, but the problem with correlational data is that they are based on group means obtained from large samples. In practical terms, intelligence tests must accurately screen individuals, and under these conditions, HFD tests consistently fail. For example, Dykens (1996) tested individuals from atypical populations and found that scores on the Draw-A-Person test (DAP; Naglieri, 1988) failed to accurately differentiate between those individuals who would be classified as gifted or as learning disabled on the basis of more traditional IQ tests. Similarly, although Willcock et al. (2011) found significant, albeit modest, correlations between 5- and 6-year-old children’s scores on the DAP:QSS and their scores on the WPPSI-R or the Wechsler Abbreviated Scale of Intelligence (WASI), when it came to classifying individuals, 82% of the children were misidentified as having borderline intellectual functioning, and only 25% of children were correctly identified as having borderline intellectual functioning as estimated by traditional IQ tests. In a similar kind of study, Imuta et al. (2013) tested 4- to 5-year-old children and adults with the DAP:IQ, and with the WPPSI-III (children) and WASI FSIQ-2 (adults). Imuta et al. also found high false-negative and false-positive rates in screening, with an overall hit rate of only 38%, indicating that the majority of children were not correctly classified according to their intellectual level.

In addition to their inability to screen individuals, HFD-generated IQ scores are often lower than traditional intelligence test scores. Scores on the GDHT, DAP:QSS, and
DAP:IQ have all been found to underestimate IQ when compared with traditional pen-and-paper IQ tests (e.g., Abell et al., 2001; Aikman et al., 1992; Haddad & Giuliano, 1991; Imuta et al., 2013; Lassiter & Bardos, 1995; Mendoza & Soto, 2011; Wisniewski & Naglieri, 1989). HFD-generated IQ scores are also lower than scores on alternative tests such as the Ravens Color Progressive Matrix test (Troncone, 2014).

One critical feature of a valid intelligence test is that scores on that test remain relatively stable over time and are not substantially influenced by practice. For example, scores on traditional IQ tests taken decades apart have been shown to vary by only one quarter to one fifth of a standard deviation (e.g., Deary, Whiteman, Starr, Whalley, & Fox, 2004). Furthermore, IQ scores assessed using traditional pen-and-paper tests have been shown to exhibit only small gains through repeated exposure to the test. When a change in score does occur, the magnitude of the change is dependent on a number of factors.

Extensive coaching on aptitude and intelligence tests, for example, has been found to result in small gains in scores. Kulik, Kulik, and Bangert (1984) conducted a meta-analysis to determine the effects of coaching on aptitude and IQ test scores. In 24 out of the 38 studies in the meta-analysis, there was evidence that coaching raised scores on aptitude and intelligence tests. This, however, was related to the relative complexity of the intelligence test; coaching on IQ tests with low to medium levels of cognitive complexity could result in gains in scores of up to half a standard deviation. In addition, the time commitment required to improve test scores is substantial. Focused instruction for 20 hours on mathematical and verbal subtests of intelligence resulted in improvements of about 20% of a standard deviation. Doubling these improvements required in the region of 120 hours instruction in mathematics and 250 hours for the verbal subtest (Messick & Jungblut, 1981; Powers, 1993). Gains in IQ scores over time are also more pronounced for tests examining very specialized aspects of intelligence, rather than for tests that measure global IQ. Moreover, improvements to an individual’s IQ score on particular aspects of IQ
tests are also influenced by the individual’s level of intelligence to begin with (Arendasy & Sommer, 2013, Kulik et al., 1984); increases in IQ scores are more likely to occur in people of average to above-average intelligence rather than in people with below-average intelligence. Although education, socioeconomic status, and environmental factors all contribute to increases in IQ scores over time, no intervention has been shown to substantially alter the IQ scores of children near the bottom of the scale (e.g., >71–75; Cassidy, Roche, & Hayes, 2011; Deary, Penke, & Johnson, 2010; Heil, Rösler, Link, & Bajric, 1998; Kanaya & Ceci, 2007; Youn & Youn, 1991).

Given that practice has little effect on performance on traditional IQ tests, what is the effect of repeated testing and practice on HFD scores? Lange-Küttner, Küttner, and Chromekova (2014) investigated whether merely practicing human figure drawing would lead to changes in children’s test scores. They asked 6- to 12-year-old children to complete a series of six HFDs over the course of an hour. Lange-Küttner et al. found that, overall, there was no effect of practice on children’s HFD scores. Instead, children revised their drawings during each attempt, sometimes including additional details, but omitting other details at the same time—the net result was a similar score on the test over repeated trials.

To the best of our knowledge, only one study has examined the effect of more general drawing practice on the children’s scores on HFD tests. In that study, Burns and Velicer (1977) found that when 11-year-old children were given specific art instruction in human figure drawing in the classroom, their GHDP-estimated IQ scores increased by more than 19 points. This finding raises important concerns about the validity of HFD tests as a measure of intelligence, suggesting that the test may provide a better measure of artistic skill than of IQ (see also Imuta et al., 2013; Troncone, 2014; Willcock et al., 2011). Given the continued popularity of drawing tests in educational settings, it is striking that the effect of instruction on children’s scores on more contemporary HFD tests has never been assessed.
In the present study, we examined the effect of drawing instruction during a classroom lesson on 11- and 12-year-old children’s scores on the DAP:IQ - the most current version of an HFD IQ test. The DAP:IQ is designed to estimate intellectual ability in the general population between the ages of 4 and 90. We selected 11- to 12-year-olds for a number of reasons: (a) We wanted to compare our findings with the only other study examining the effect of drawing practice on the children’s HFD scores (Burns & Velicer, 1977), (b) our drawing instruction could easily be integrated into the students’ biology curriculum at school, and (c) 11- and 12-year-olds are still at an age at which they have not yet become self-conscious about their drawing ability, so will attempt to draw without much encouragement (Cohn, 2012). During their normal day at school, children received three classes on sensory information processing that also included instruction on drawing accurate figures for scientific illustrations. To measure potential changes in their HFD scores, the DAP:IQ was administered three times: prior to the instruction, shortly after the final instruction session, and, finally, 6 months later.
Method

Participants

Forty-four 11- and 12-year-old children (24 girls and 20 boys; \( M \) age = 12.14 years, \( SD = .39 \)) were recruited from a local middle school in a small city in New Zealand. The school was assigned a decile rating of 8 (maximum = 10) by the New Zealand Ministry of Education. This rating is based on socioeconomic indicators from national census data obtained from households with school-age children in the areas from which the school draws students. In general, a decile rating of 8 roughly corresponds to a socioeconomic status of upper-middle class; as such, the student roll included children from throughout the city and from a wide range of socioeconomic backgrounds. The children were predominantly of European descent, and all had written parental consent to participate. They were drawn from intermediate-level science classes that were not streamed for ability—that is, assignment to class groups was random. The research was reviewed and approved by the University of Otago’s Human Ethics Committee, which is approved by the New Zealand Health Research Council and whose guidelines are consistent with those of the American Psychological Association.

Procedure

Participants were administered the DAP:IQ in their class group during their weekly science class. They were given a pencil and an eraser and, following the instructions outlined in the DAP:IQ manual, were asked to “draw yourself, the best drawing you can do, put in as much detail as you can into it.” The picture was drawn on a piece of white construction paper (21 cm x 30 cm), and a maximum of 10 min was allowed to complete the drawing.

Six months later, the experimenter returned to the class and provided a novel science lesson in which students learned a variety of things about the senses over the course of two 1-hr classes. We chose a 6-month delay because students had transitioned from one school.
year to the next, and the 6-month delay allowed for maturation of students who may have been close to cut-off points for IQ estimations given their age.

In the first class, the experimenter encouraged students to draw the eyes, nose, mouth, ears, and eyebrows while learning about the sensory information that each of these body parts conveys to the brain. Students were given support in drawing each feature and were encouraged to practice these at home. In the second class, the experimenter taught students about the central nervous system and how neurons transfer information from the periphery (e.g., from the arms and legs) to the brain. Students were supported in drawing a correctly proportioned human figure by learning the classical proportions of the human frame (the head fitting in the body 8 times). They were then supported to produce their own drawing by folding a sheet of paper (21 cm x 30 cm) in half, then in half twice more to produce eight fold-lines on the page. Students were able to use the physical landmarks of the folds to help correctly proportion the body parts that make up a human figure. Two days after the last class, students were re-administered the DAP:IQ test in their class group. The experimenter gave students the same materials (pencil, eraser, and a piece of unfolded paper) and instructions as in the original DAP:IQ test 6 months earlier.

Six months after the second DAP:IQ administration, or 12 months after initial testing, all of the original participants completed a third drawing under the same testing conditions that were used at Times 1 and 2. During the 6-month delay, students received no further support in learning to draw; art was not part of their normal curriculum. As before, participants were tested in school, but in mixed groups, as they had changed classes at the beginning of the new academic year.

All of the drawings were collected and scored in accordance with the DAP:IQ scoring system (Reynolds & Hickman, 2004). Each drawing yielded a raw score out of 49. One experimenter independently scored all of the drawings and a second experimenter
scored 25% of the drawings to assess interrater reliability. Consistent with prior studies that have employed HFD tests, intra-class interrater reliability was very high ($r = .92$).
Results

We used a one-way repeated measures analysis of variance (ANOVA) to compare children’s DAP:IQ raw scores across the three testing points. This analysis revealed a main effect of test point, $F(2, 72) = 75.63, p < .0005$, partial $\eta^2 = .68$. Pairwise comparisons using the Bonferroni correction revealed 3 important findings. First, children’s DAP raw scores increased significantly between Time 1 ($M$ score = 32.81, $SD$ = 5.84) and Time 2 ($M$ score = 42.32, $SD$ = 4.33), $p < .0005$. Second, children’s DAP raw scores decreased significantly between Time 2 ($M$ score = 42.32, $SD$ = 4.33) and Time 3 ($M$ score = 32.89, $SD$ = 6.39), $p < .0005$. Finally, there was no difference in DAP raw scores between Time 1 and Time 3, $p = .938$.

To account for maturation over the testing period, we converted the DAP:IQ raw scores to IQ standard scores using the method outlined in the DAP:IQ manual. As shown in Table 2.1, at Time 1, children’s standard scores on the DAP:IQ ranged from 84 to 132 ($M$ score = 113.52, $SD$ = 12.98), at Time 2, they ranged from 102 to 153 ($M$ score = 133.43, $SD$ = 11.59), and at Time 3, they ranged from 77 to 137 ($M$ score = 110.84, $SD$ = 12.86).

Table 2.1

<table>
<thead>
<tr>
<th>IQ Classification</th>
<th>Time 1 IQ</th>
<th>Time 2 IQ</th>
<th>Time 3 IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3) Below Average</td>
<td>84</td>
<td>1317</td>
<td>964</td>
</tr>
<tr>
<td></td>
<td>86</td>
<td>1074</td>
<td>1044</td>
</tr>
<tr>
<td></td>
<td>92</td>
<td>1124</td>
<td>1054</td>
</tr>
<tr>
<td></td>
<td>96</td>
<td>1327</td>
<td>984</td>
</tr>
<tr>
<td></td>
<td>98</td>
<td>1374</td>
<td>1014</td>
</tr>
<tr>
<td></td>
<td>99</td>
<td>1296</td>
<td>914</td>
</tr>
</tbody>
</table>

3 To meaningfully compare children’s IQ scores across all time points, we report both raw and standard scores.
<table>
<thead>
<tr>
<th>Grade</th>
<th>Number</th>
<th>Code 1</th>
<th>Code 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td></td>
<td>99</td>
<td>137</td>
</tr>
<tr>
<td></td>
<td></td>
<td>101</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td></td>
<td>103</td>
<td>102</td>
</tr>
<tr>
<td></td>
<td></td>
<td>103</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td></td>
<td>103</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td></td>
<td>107</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td></td>
<td>107</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td></td>
<td>107</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>109</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td></td>
<td>109</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td></td>
<td>111</td>
<td>141</td>
</tr>
<tr>
<td></td>
<td></td>
<td>111</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td></td>
<td>113</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td></td>
<td>113</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td></td>
<td>113</td>
<td>141</td>
</tr>
<tr>
<td></td>
<td></td>
<td>114</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td></td>
<td>116</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>116</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td></td>
<td>116</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>116</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td></td>
<td>118</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td></td>
<td>118</td>
<td>141</td>
</tr>
<tr>
<td></td>
<td></td>
<td>122</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>122</td>
<td>134</td>
</tr>
<tr>
<td></td>
<td></td>
<td>122</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td></td>
<td>126</td>
<td>149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>127</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>127</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td></td>
<td>128</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td></td>
<td>129</td>
<td>141</td>
</tr>
<tr>
<td></td>
<td></td>
<td>131</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>132</td>
<td>152</td>
</tr>
<tr>
<td></td>
<td></td>
<td>132</td>
<td>144</td>
</tr>
</tbody>
</table>

(4) Average

(5) High Average

(6) Superior

(7) Very Superior
As with children’s raw scores on the DAP:IQ, a one-way repeated measures ANOVA across children’s standard scores at the three testing points revealed a main effect of test point, $F(2, 86) = 90.79, p < .0005$, partial $\eta^2 = .68$. Between Time 1 and Time 2, there was a statistically significant mean increase of 19.91 points, 95% [15.59, 24.23], in children’s DAP standard scores, $p < .0005$. Although nine children failed to increase their IQ group classification, seven of these children were classified in the highest 2 groups, Superior and Very Superior, to begin with and remained in these two groups after the drawing instruction. Between Time 2 and Time 3, on the other hand, children’s DAP standard scores decreased by an average of 22.59 points, 95% [18.12, 27.06], a statistically significant decrease, $p < .0005$. In fact, every child’s scores declined from Time 2 to Time 3. Finally, there was virtually no change in children’s standard scores between Time 1 and Time 3, the mean decrease of 2.68 points was not significant, $p = .535$.

Next, we used the ability classifications listed in the DAP:IQ manual to assign children to categories on the basis of their IQ standard score at Times 1, 2, and 3. To illustrate the extent to which children’s standard scores changed between each time point, Table 2.1 shows the individual score for each participant at Times 1, 2, and 3 according to his or her ability level at Time 1. Table 2.2 shows the distribution of participants in each ability classification category at the three testing points. As shown in Table 2.2, 41% of children were classified in the Below Average and Average categories at Time 1. However, following a brief drawing intervention, only 7% of children were classified in the Average category or lower at Time 2; 93% of children were classified in the three top categories. At Time 3, however, the distribution of participants in each ability classification category returned to pre-drawing
instruction levels; 43% of children were classified in the Average category or lower at Time 3. To illustrate these findings another way, Figure 2.1 shows the side-by-side comparisons of a selection of representative drawings produced by the participants.

Table 2.2

The Distribution of Participants in Each Ability Classification Category at Times 1, 2, and 3.

<table>
<thead>
<tr>
<th>Ability Classification</th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Significantly Impaired</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2: Mildly Impaired</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3: Below Average</td>
<td>2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4: Average</td>
<td>16</td>
<td>3</td>
<td>17</td>
</tr>
<tr>
<td>5: High Average</td>
<td>12</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>6: Superior</td>
<td>8</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>7: High Superior</td>
<td>6</td>
<td>30</td>
<td>2</td>
</tr>
</tbody>
</table>

Although we found that children’s raw scores and standard scores at Time 3 were not significantly different from their raw scores and standard scores at Time 1, we did notice that the quality of the drawings appeared better at Time 3 than at Time 1 (see Figure 1.1). To examine this further, we asked 20 naive adults (15 females; $M_{age} = 33.30$ years, $SD = 12.10$) to rank each child’s set of three DAP drawings from 1 (worst) to 3 (best). The individual rankings were then averaged, and each drawing was assigned a mean ranking. A one-way repeated measures ANOVA indicated that there was a significant effect of time point, $F(1.54, 29.17) = 214.99, p < .0005$, partial $\eta^2 = .92$. Pairwise comparisons using the Bonferroni correction revealed that children’s Time 2 drawings were ranked significantly higher ($M = 2.64, SE = .03$) than were their drawings at Time 1 ($M = 1.51, SE = .03$), $p < .0005$; children’s Time 3 drawings were ranked between these two extremes ($M = 1.85, SE = .02$) and were
significantly different from both, $p's < .0005.$

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Figure" /></td>
<td><img src="image2.png" alt="Figure" /></td>
<td><img src="image3.png" alt="Figure" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Above Average</th>
<th>Very Superior</th>
<th>Above Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>(score: 111)</td>
<td>(score: 149)</td>
<td>(score: 112)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Average</th>
<th>Very Superior</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>(score: 109)</td>
<td>(score: 145)</td>
<td>(score: 102)</td>
</tr>
</tbody>
</table>

*Figure 2.1.* Examples of human figures drawings created at three time points over a 12-month period by two children. Each row in the figure shows the DAP:IQ classification at each time point and corresponding standard IQ score.
Discussion

In the present study, we examined the effect of specific drawing instruction on middle-school-aged children’s scores on a standard HFD test that is commonly thought to provide a measure of IQ. We found that even a limited amount of drawing instruction, provided in a different academic context, had a substantial impact on the HFD test results. When they were tested shortly after the instruction, children’s scores increased by an average of 16 points compared to their test scores obtained 6 months earlier. When converted into standardized IQ scores, the increase in DAP: IQ scores over time resulted in an average IQ gain of 19 points. When the scores were interpreted relative to standardized IQ categories (e.g., below average, average, above average, superior, and very superior), 22.7% of the children moved up one category, 36.6% moved up two categories, 15.9% moved up three categories, and 4.5% experienced a category increase of four. Of the 20.4% of children who failed to change categories, 7 out of 9 of these children were in the superior or very superior classification at Time 1 and remained in the same classification after the drawing instruction. Despite these major changes over the short term, when they were tested 6 months later, many of the children’s scores returned to their pre-training levels. How long the results of training would persist after drawing instruction remains a question for future research.

The present findings stand in stark contrast to prior research where practice has been shown to have little effect on individuals’ scores on traditional pen and paper intelligence tests (Bangert, 2015; Casey et al., 1928; Catron, 1979; Slaughter, 1976), but they are highly consistent with prior research on the effect of instruction on other HFD tests of intelligence. In an earlier study, for example, Burns and Velicer (1977) reported increases of up to 20 IQ points when 11-year-old children completed the Goodenough-Harris Drawing Test following drawing instruction; here, we recorded increases of up 47 points on IQ scores obtained using the more modern DAP: IQ. Taken together, these data indicate that a consistent feature across various versions of HFD measures of intelligence is that they are highly susceptible to the
effects of even brief instruction--a characteristic that sets them apart from other validated measures of IQ, where the time requirement to improve test scores is significant (e.g., te Nijenhuis, Voskuijl, & Schijve, 2001).

Why might drawing instruction have had such an effect on children’s IQ scores measured using a HFD test of intelligence? In HFD tests, the degree of differentiation or sophistication in a child’s drawing of the human figure is viewed as a reflection of intelligence. By including more detail in the drawing, children score more points on the test. For example, drawing the eyes allows children to obtain 1 point on the test, but drawing the eyes in proportion, in the correct oval shape, and roughly the same size scores 3 points. With respect to the drawing instruction in the present study, children participated in two science classes where, as part of learning about the senses, they were taught how to draw parts of the body. When learning about the visual system, for example, students were encouraged to draw and include details such as the iris, pupils, eyelashes, and laugh lines. Even after relatively minimal drawing instruction, children were equipped, very quickly, with enough information to draw more complex or detailed human figures. Thus, when re-tested on the HFD test, the majority of children had acquired the drawing skills that allowed them to improve their HFD scores.

In the past, researchers have argued that HFD tests may provide a better measure of artistic skill than of global IQ (Butler, Gross, & Hayne, 1995; Gross & Hayne, 1998; Imuta et al., 2013; Kamphaus & Pleiss, 1991; Troncone, 2014; Willcock et al., 2011). For example, in one study, the researchers examined the relation between 5- to 6-year-old children’s scores on the DAP-QSS, the WPPSI-R and WASI tests, and their verbal performance during an interview about a visit to a local firestation (Willcock, 2004; see also Willcock et al., 2011). Some of the children were asked to draw about the experience while they were talking while other children were simply asked to tell the experimenter about the experience. For children in the draw and tell group, both their DAP:QSS score and their WPPSI score predicted the amount of information that they reported in the interview (DAP: $r = .30$; WPPSI: $r = .29$;
both $p's < .05$). For children in the tell group, on the other hand, only their score on the WPPSI predicted the amount of information that they reported in the interview ($r = .51, p < .01$), their score on the DAP:QSS did not ($r = .17, ns$). Given that IQ predicted the amount reported by children who had the opportunity to draw as well as children who merely told, but DAP:QSS score only predicted the amount of information reported by children in the draw group, it was hypothesised that although scores on the DAP:QSS do not reflect individual differences in IQ, they might reflect individual differences in drawing ability. The present findings refute that hypothesis.

In the present study, although children’s DAP:IQ scores returned to pre-instruction levels over time, the quality of their drawings did not. As illustrated in Figure 2.1, there were visible improvements in how children structured their human figure drawing, paying more attention to underlying body shapes and proportions. This difference was most notable shortly after instruction (Time 2), but it continued to occur 6 months later (Time 3). Despite this long-term improvement in the quality of their drawings, the DAP:IQ scoring method was not detailed or nuanced enough to pick up on these differences. Taken together, the present findings suggest that not only does the DAP:IQ fail to adequately assess IQ, but may also fail to adequately assess drawing quality.

Despite growing evidence that challenges the validity of HFD tests, many clinical and educational psychologists continue to use children’s drawings as a surrogate measure of intellectual ability. IQ scores based on variations of the DAP are still one of the most widely used screening tests of children’s intelligence (Anastasi & Urbina, 1994; Skybo, Ryan-Wenger, & Su, 2007). In both the clinic and the classroom, the ease of administration of HFD tests, along with the reliability of the scoring systems continue to outweigh serious, evidence-based concerns about their validity. More recently, HFD tests have also begun to creep into research as a surrogate measure of intellectual function. In a recent study of behavioural genetics, for example, researchers used HFD tests as a proxy for IQ in their work on the genetic basis of
intelligence across 7,252 twin pairs (Arden, Trzaskowski, Garfield, & Plomin, 2014). In that study, the authors found the typical, modest correlation between IQ ($g$) and HFD scores at 4 and 14 years of age (i.e., $r_s = .20-.33$); they also found that the correlations between $g$ and HFD scores were higher for MZ twins than they were for DZ twins. From these data, Arden et al. argued for a genetic link between intelligence and drawing ability - once again, assuming that the latter provided an estimate of the former.

In our view, two lines of research challenge Arden et al.’s (2014) conclusion. First, as shown in the present study, HFD scores are highly influenced by the environment. On the basis of just a few drawing sessions, the children in the present study produced drawings that shifted estimates of their IQ dramatically - shifts that are not common when using traditional measures of $g$. Second, in their enthusiasm for HFD tests, Arden et al. may have overlooked more parsimonious genetic links between drawing and fine motor skills (Missitzi et al., 2013) or between drawing and artistic talent and aptitude (Hur, Jeong, & Piffer, 2014; Vinkhuyzen, van der Sluis, Posthuma, Boomsma, 2009). Based on current estimates from twin studies, explanations based on the heritability of fine motor skill or on talent and aptitude provide substantially greater explanatory power (even across much smaller samples) than does an explanation based on any direct relation between HFDs and IQ.

In summary, the current study was conducted with an unselected group of children from normal classrooms. Over the short term, drawing instruction had a substantial effect on performance on the DAP:IQ test, leading to major changes in estimates of IQ. Over the longer time, however, and in the absence of further drawing instruction, IQ scores on the DAP:IQ test returned to pre-instruction levels. This finding, coupled with prior research on the validity of these tests, once again indicate that they should not be used as a measure of intelligence in the clinic, in the classroom, or in the context of research.
Chapter 3

Put Away Your Camera: Drawing Facilitates Memory of Museum Exhibits

Taken together, contemporary research has raised serious questions about the projective value of drawing. To the best of our knowledge, there is no empirical evidence that researchers or clinicians can glean psychologically-valid information from a drawing alone. Despite these concerns about projective drawing tests, there are applied settings in which drawing can be extremely useful. In a series of studies by Hayne and her students (Butler et al., 1995; Gross & Hayne, 1998, 1999; Gross, Hayne, & Drury, 2009; Macleod, Gross, & Hayne, 2011; Patterson & Hayne, 2007; Woolford, Patterson, MacLeod, Hobbs, & Hayne, 2015), for example, giving children the opportunity to draw while they talk increases the amount of information that they report about a prior event without decreasing accuracy (see also Davison & Thomas, 2001, Drucker, Greco-Vigorito, Moore-Russell, Alvatroni, & Ryan, 1997; Rae, 1991; Rollins, 2005; Salmon, Roncolato, & Gleitzman, 2003; Wesson & Salmon, 2001).

In the first study of this kind, Butler et al. (1995) examined the effect of drawing on 5- to 6-year-old children’s reports of a field trip to a fire station. Children were taken on a tour of a fire station and interviewed about their visit 1 day or 1 month later. Children were interviewed in one of two ways, children in the draw group were allowed to draw while reporting all that they could remember about the event; children in the tell group were interviewed using the same methodology but did not have access to drawing materials. Children who were asked to draw what happened during their trip to the fire station reported more information compared to children who were asked to tell what happened. In a similar study, Gross and Hayne (1999) examined the factual and autobiographical memory outcomes for 5- and 6 year-old children who experienced a field trip to a chocolate factory. When tested over long and short delays, drawing was found to enhance the reporting of both narrative and factual information over a short delay (1-2 days) but to only enhance the reporting of narrative
information 7 months later (see also, Gross et al., 2009). This same basic finding has been replicated in a number of different laboratories around the world (Eschenfelder, Gravalas, Perdergast, & Gorman, 2018; Faccio, Costa, Losasso, Barrucci, & Ricci, 2016; Piotrowski, 2016; Salmon et al., 2003).

The research by Hayne and others has clearly shown that drawing facilitates children’s reports of their prior experiences, but what about the potential impact of drawing on encoding per se? Does drawing at the time of an event increase the amount of information that children (or adults) encode in the first place? The research outlined in the current chapter addresses this experimental question. Museums are repositories for cultural knowledge, not only preserving and presenting ancient artefacts, but also providing multiple opportunities for visitors to learn and explore (Burton & Scott, 2016; Smith, 2014). For visitors of all ages, a trip to a museum can be both enjoyable and educational (Ahmad, Abbas, Yusof, & Mohd.Taib, 2015; Charitonos, Blake, Scanlon, & Jones, 2012; De Backer et al., 2015; Sienkiewicz, 2015). For children in particular, museum visits are commonly used to supplement classroom experiences (Farmer, 1995; Hackett, 2012). One key issue for museum curators is to design displays that are engaging and informative (Sienkiewicz, 2015), but are there also other ways to enhance visitors’ experiences, particularly when it comes to what they learn and remember over time? In the present study, we examined the role of drawing in facilitating visitors’ memory for museum exhibits.

Museums must meet the learning needs of a diverse range of populations including families (Borun, Chambers, & Cleghorn, 1996; Moussouri, 1997), school children (Birney, 1988; Griffin, 1998), and adults (Falk & Dierking, 1997; Falk, Moussouri, & Coulson, 1998; McManus, 1993). Each of these groups visits the museum with different sets of expectations, motivations, and learning needs (Jensen, 1994). Given this, museums employ a range of strategies to facilitate learning; these learning interventions are as diverse as the audiences they serve, and include a range of tools from interactive displays and communicators within
galleries, to treasure hunts and information panels (Anderson, Piscitelli, Weier, Everett, & Tayler, 1997; Doering 1999; Falk & Dierking 2000; Hein 1998). The key question is which strategies work in maximising what visitors learn and remember?

To understand what children learn from educational events it is essential to have an agreed-upon way of assessing the learning that takes place. Formal assessment of learning and memory from museum exhibits is still an emerging field, and researchers have employed a range of dependent measures to conduct evaluations. For example, engagement, enjoyment, and learning are sometimes ambiguously defined, and often conflated and ratings of enjoyment and engagement are often used as proxy measures of learning. For example, in one study, Zoldosova and Prokop (2006) examined motivation, enjoyment, and attitudes, which they argued impact learning on field trips to museums (see also Flexer & Borun, 1984). Their study involved 363 11- and 12-year-old children. Half of the children attended a field trip to a science education centre. The other half acted as controls; these children received science education in their regular classroom. Participants then completed the Motivational Orientations Diagnostic Test (MODT: Johnstone & Al-Naeme, 1995), which measures four different motivational orientations: social-oriented (social), consciousness-oriented (conscientious), effort-oriented (achiever), and curiosity-oriented (curious). Zoldosova and Prokop found no significant differences between the motivations to learn in the classroom and field trip samples, however, there was no control over the learning content that the groups were exposed to, and there were no measurable learning outcomes, making it difficult to estimate how a concept like motivation may quantifiably impact learning.

In another experiment examining learning in a museum setting, Smedley (2015) assessed the learning outcomes for 11- and 12-year-old children who completed a 90-min tour of an art museum. On the tour, students were given blank postcards which they could fill with written or drawn information during the tour; they were also given a list of prompts to encourage them to write on or fill in their cards. One month later, the postcards that were
completed on the tour were returned to the students immediately prior to the completion of the follow-up reflection piece. Using their postcard as a cue, students completed a reflective writing activity during which they described their experience during the field trip and explained what, if anything, they learned as a result of the field trip. Students reported a deeper ‘connection’ to and detailed memory for the objects that were included in their postcard records. However, as there was no control group included in this study, it is difficult to determine if prompting from the participants’ postcards was responsible for the recall of items from the tour. Similarly, exactly how much information they recalled reflected autobiographical memory, as opposed to factual memory for the trip and pieces encountered on the tour, is unclear. While participants in Smedley’s study reported being highly engaged, this study does not empirically measure learning outcomes on site.

We know that museum visits are a rich source of autobiographical memories which are retained over time (Fivush, Hudson, & Nelson, 1984; Gross et al., 2009; Henry, 1992; Stevenson, 1991). Fivush et al. (1984), for example, found that 5-year-old children recalled specific elements of a museum field trip, including making clay models and digging for artefacts after one and six year delays. While memory for the field trip did decrease over time, when cued with visual aids, children could accurately report a substantial amount of information about the event. Similarly, Henry (1992) assessed the memory outcomes of a field trip to an art museum and found that after an 18-month delay, children between the ages of 10 and 12 remembered the trip and details about it. However, recalling what was encountered on the tour and location of various points of interest, such as the positioning of guards, and where lunch was eaten, but retaining autobiographical memories are not the same as recalling factual information presented which forms part of the museum learning experience.

In one study that did differentiate between participants’ autobiographical and conceptual recollections for a museum event, Stevenson (1991) interviewed participants ranging in age from 6 years to adults, 6 months after their visit to the Launch Pad centre at the
London Science Museum. The Launch Pad centre contains interactive science exhibits that are designed for both children and adults. Stevenson found that only 14% of the responses given during the interview indicated that participants had an accurate understanding of the concepts that were presented during the visit. The majority of responses that participants gave were autobiographical in nature; that is, participants described what occurred during the visit, rather than reporting new information that they had learned at the centre. Because Stevenson did not assess participants’ immediate recall of the trip, the small number of conceptual responses reported 6 months following the event may have been due to a failure to learn the information in the first place, but it could also have been due to forgetting over the delay.

In another study, Gross et al. (2009) examined 5- and 6-year-old children’s memory for a field trip to an ecological learning centre. Children were interviewed either 1–2 days or 7 months after the field trip. To assess how much children learned and remembered about the field trip, first they were asked to provide a comprehensive verbal report of the trip and then they were asked 12 comprehension questions that covered material that the museum staff considered to be most relevant to the visit. When interviewed after the short delay, children answered 50% of the questions correctly. After a 7-month delay, children answered less than 20% of the comprehension questions correctly. This method enabled the researchers to identify what types of information, factual or autobiographical, could be recalled over delays.

In a handful of studies, researchers have measured specific memory outcomes for children on field trips. For example, Benjamin, Haden, and Wilkerson (2010) examined the use of conversation instruction techniques by parents and children visiting an engineering exhibit at a children’s museum. In their study, 121 dyads consisting of children aged between 4- and 8-years old and their adult caregiver were assigned to one of five experimental groups. Some dyads received instructions about building and/or conversation techniques (build and talk instruction, build-only instruction, talk-only instruction), some dyads watched videos of previous visitor dyads interacting with the exhibit (model group), and the remaining dyads were
in a control group that received no instruction at all.

Immediately before touring the exhibit, dyads in the build groups received information on engineering concepts, while dyads in the talk groups received instructions on elaborative conversational techniques (open-ended *wb*-questions that add new information about the event under discussion). Immediately after the dyads’ visit to the exhibit, children completed a picture comparison task, consisting of six pairs of photos showing 2D or 3D representations of structures with cross-bracing shown in one of each picture pair. Children were asked to choose the stronger of the two examples and received one point for each correct answer that contained the triangle braces, for a maximum score of 6. Children in the build + talk (M = 4.52) and build-only (M = 4.55) groups performed significantly better than did those in the talk-only (M = 2.65), model (M = 2.74), and control (M = 2.74) groups.

In addition, 35 dyads were given a tape recorder and instruction sheet and were asked to record two subsequent conversations at home. One day and 2 weeks later, caregivers asked children about their experience at the children’s museum using *wb*-questions. Conversations were analysed for the amount of engineering related information recalled. Analysis of recordings made after the 1-day (T1) and 2-week delay (T2), showed that compared with children in the talk-only (T1: M = 0.00; T2: M = 1.25) and control group (T1: M = 2.31, T2: M = 1.79), children in the build and talk (T1: M = 25.00; T2: M = 16.54) and build only (T1: M = 14.29; T2: M = 4.76) groups reported a greater amount of engineering information in response to *wb*-questions from caregivers.

In a similar study, Haden, Jant, Uttal, and Babcock (2014) examined the influence of conversation and hands-on learning for 78 adult-child dyads consisting of parents and children aged between 2.9 and 6.6 years of age during a natural history museum visit. The procedure involved a pre-exhibit activity, a visit to the first exhibit, a visit to a second exhibit, completing a family questionnaire and follow-up at-home memory conversations. Dyads visited both exhibits; for one exhibit, dyads were supported with learning materials by the experimenter
and for the second exhibit, dyads received no additional learning materials. Immediately before visiting the first exhibit, each dyad participated in a 10-minute session in a room in the museum’s educational department. During this pre-exhibit activity, each dyad was assigned to one of four experimental conditions: conversation cards only, conversation cards and objects, objects only, and control. The conversation cards showed 6 target objects together with 3 to 5 'wh'-question suggestions. Dyads who received these cards were asked to examine and discuss the object on the card using the suggested ‘wh’ questions provided. A second group received cards and physical examples of the objects shown on one side of the conversation cards. The third group received just the physical objects and no instructions on conversations. Dyads who received neither cards nor objects served as the control group; during the pre-visit activity, they were shown fossils unrelated to exhibit’s one and two. Dyads’ conversations were recorded and analysed to compare verbal and non-verbal behaviour at the first and second exhibit. While at the museum, children’s spontaneous talk and joint engagement with target objects in both exhibits was highest for those who received conversation cards. By facilitating use of ‘wh’-questions about the first exhibit, dyads subsequently applied this methodology without the need for cues in the second exhibit.

One day and 2 weeks after their visit to the museum, 30 dyads recorded follow-up reminiscing conversations at home. Children’s spontaneous reporting about the exhibits made after a 1-day (T1) and 2 week delay (T2), showed that compared with children in the control (T1: $M = 12.10$; T2: $M = 8.11$), and children in the cards only (T1: $M = 7.87$; T2: $M = 7.33$) conditions, children who received objects (T1: $M = 18.11$; T2: $M = 10.33$) and objects and cards (T1: $M = 10.27$; T2: $M = 14.10$) reported more information about their museum experiences over time. Over the delay, children who received the pre-exhibit activity cards were able to access and report more about their museum experiences even without their parents questioning. That is, providing support to enhance parent-child conversations in a museum resulted in greater recall of information about the exhibits.
The research reported by Haden and her colleagues clearly shows that some


techniques can facilitate what children learn in a museum, but many of those techniques are
time intensive and are unlikely to be adopted by casual visitors to a museum. Are there easier
ways to enhance what visitors learn and remember about objects in a museum? Photography
is one tool that museums have recommended that their visitors use to enhance their memory
of the visit. This recommendation is based on the assumption that taking photographs is an
easy way of making and preserving lasting memories. In a study by Henkel (2014), however,
museum visitors who took photographs actually recalled less, not more, information about the
objects they encountered, compared to visitors who only viewed the items.

Taking another tack, galleries such as The National Gallery in Scotland and The
Rijksmuseum in the Netherlands have gone back to basics, encouraging their visitors to draw
while at the museum. Traditionally, it has been students of art who have drawn in galleries as a
means of learning techniques from past and contemporary masters, but now it is common for
sketch books and pencils to be made available for all visitors to use to record what they see in
the museum. While drawing is an enjoyable activity for many visitors, virtually no research has
examined the potential learning outcomes of drawing in this context. Given that drawing
when recalling a past event increases the amount of information reported by children and
adolescents (e.g., Butler et al., 1995; Gross & Hayne, 1999; Gross et al., 2009; Salmon et al.,
2003; Wesson & Salmon, 2001), can drawing also facilitate the formation of memories in the
first place? Furthermore, does the effect of drawing vary as a function of the age or artistic
ability of the museum visitor? With these issues in mind, the aim of the present study was to
determine whether simple pencil and paper drawing would enhance children’s and adults’
memory of objects that they encountered in a museum. Participants were led on a guided tour
of a natural history museum, during which they were asked to view particular objects of
interest and were directed to draw some of the objects and merely observe others. In
Experiment 1, we tested adult artists and in Experiment 2, non-artist adults and children.
Experiment 1

Method

Participants. A total of 25 self-identified artists (18 female; aged between 18 and 65) were recruited in the local community by snowball sampling through college art departments, artist groups, and art shops. Each participant had extensive experience in drawing from life and scored their drawing ability as above average (\(M = 7.48, SD = 1.29\)) on a 10-point scale (0 = very poor drawing skills, 10 = very good drawing skills). Although the artists used a variety of artistic mediums in their current work, all had a foundation in drawing. All of the participants had been to the participating museum before, but reported that they had not been there in the past 6 months or longer. The research was reviewed and approved by the University of Otago’s Human Ethics Committee, which is approved by the New Zealand Health Research Council and whose guidelines are consistent with those of the American Psychological Association. Written informed consent was obtained from all participants. Participants received a small gift in return for their participation.

Procedure. Each participant was taken individually on a tour of a gallery at the Otago museum. Before the tour began, participants were told that they would be asked to draw some objects and to simply observe other objects; they were asked to pay attention to the objects because they would be asked to recall them later, as well as some details about each one. Participants were given an A5 drawing pad and a 2B pencil and no further direction on how or what to draw, only that their sketch was strictly for the purpose of helping them recall the items and would not be used to evaluate or compare their work with that of others.

On the tour, participants visited 20 objects. The objects spanned over 2000 years of human history from Egyptian pottery to Corinthian artefacts and Asian sculpture and included sculptures, pottery, antiquities, costumes, and curiosities (see Figure 3.1, left panel, for

---

4 We based our sample size (25 participants in Experiment 1 and 30 participants in each age group in Experiment 2) on the sample size of 27 used by Henkel (2014).
examples). To counterbalance the objects across the sketching and observed conditions, participants were randomly assigned to one of two object orders. Of the 20 objects, 10 were sketched and 10 were observed. Each participant was given a list of the names of the 20 objects. To begin with, the participant read the name of the first object out loud to the experimenter, who then took him or her to that object. For the observed objects, the participant was directed to look at the object while the experimenter sketched it for 2 minutes; for the sketched objects, the participant was given 2 minutes to sketch the object him or herself (see Figure 3.1, right three panels for examples from each of the participant groups). During this time, the participant was also free to talk about the object with the experimenter. After 2 minutes had elapsed, the participant read out the name of the next object on the list, and the same procedure was repeated, one after another for all 20 objects. Participants always sketched one object and then observed one object and the tour was organised such that participants viewed each object once without passing by it later.
Figure 3.1. Examples of real objects from the museum and sample drawings from each of the three participant groups. **Top:** A carved relief of grave Stele featuring a young girl holding doves and three drawings of the same object by an artist, a non-artist adult, and a child. The visual detail question about this object was, “What was the girl holding in her hands? 1) A lily; 2) A dove; 3) An offering box; or 4) 2 doves.” **Bottom:** An Egyptian pot and three drawings of the same object by an artist, a non-artist adult, and a child. The visual detail question about this object was, “The Egyptian Pot featured? 1) Slaves building a pyramid; 2) The king being crowned; 3) A boat carrying souls to the after world; or 4) A farmer in his fields.
One week after the tour, participants came to the laboratory to take part in a memory assessment. All of the testing was completed on a computer using Medialab (Jarvis, 2012). Memory was assessed in 4 ways: First, during free recall, participants were asked to list all of the objects that they remembered from the tour of the museum. After completing the free-recall test, participants were given a name-recognition test that consisted of the names of the 20 objects from the tour randomly intermixed with the names of 10 other objects that were not part of the tour but were objects that participants could plausibly have seen at the museum. In the name-recognition test, participants were seated at a computer and the names of the objects appeared one at a time on the computer screen; participants were asked to indicate for each object if it had been one of the items that they had seen on the tour with the experimenter, or was new. Following the name-recognition test, participants were given a photo-recognition test on the computer and were asked to indicate which of 30 photographs of objects from the museum had been part of the tour and which were new. Finally, for each of the 20 objects on the tour that had been sketched or observed, participants were asked about a visual detail of that object and given a choice of four multiple-choice response options. For example, for the two-faced Grecian oil bottle, participants were asked, “Which of the following statements is descriptive of this piece?” They could select from the following answers: 1) A clean-shaven man, 2) A woman without eyebrows, 3) A bearded man, or 4) A man with a moustache.
Results

To standardise the data so we could later compare the performance of all of our participant groups, we first converted participants’ scores on each component of the memory assessment to proportions. On the free-recall test, we divided the number of objects that each participant correctly recalled by 20 (the total number of objects that participants visited on the tour). As shown in Figure 3.2 (left panel), participants recalled a significantly greater proportion of the objects that they had sketched ($M = .55, SE = .03$) than objects that they had merely observed ($M = .27, SE = .02$), $t(24) = 6.26, p < .001, d = 1.92$

Figure 3.2. The proportion of objects that were correctly recalled by the artists in Experiment 1 (left panel) and by and by the non-artists and children in Experiment 2 (right panel) as a function of condition.

The left panel of Figure 3.3 shows participants’ scores on the two recognition components of the memory assessment. These data were subjected to a 2
(Condition: sketched, observed) x 2 (Test Phase: name recognition, photo recognition) analysis of variance (ANOVA) with repeated measures over condition and test phase. Overall, recognition accuracy was significantly higher for the sketched objects (74%) than for the observed objects (58%), $F(1, 24) = 24.80, p < .001, \eta^2_p = .51$, and was significantly higher when participants were presented with photos of the objects (77%) compared to when they read the names in the name-recognition test (55%), $F(1, 24) = 89.35, p < .001, \eta^2_p = .79$. There was no interaction, $F(1, 24) = .287, p = .60, \eta^2_p = .01$.

Figure 3.3. The proportion of objects that were correctly recognised by the artists in Experiment 1 (left panel) and by the non-artists and children in Experiment 2 (right panel) as a function of condition and test phase (i.e. recognition test).
In the final phase of the memory assessment, participants were asked to answer specific questions about the specific visual details of objects (see Figure 3.4, left panel). Again, we found that participants answered a greater proportion of the questions correctly about the objects that they had sketched ($M = .65, SE = .03$) than about the objects that they had only observed ($M = .40, SE = .04$), $t(24) = 3.68$, $p = .001$, $d = 1.23$.

![Figure 3.4.](image)

*Figure 3.4.* The proportion of multiple-choice questions about the objects that were correctly answered by the artists in Experiment 1 (left panel) and by the non-artists and children in Experiment 2 (right panel) as a function of condition.
Discussion

Taken together, the results of Experiment 1 revealed that drawing enhanced adult artists’ memory for objects they encountered in the museum; that is, participants recalled and recognised more objects and more details about those objects when they had sketched those objects rather than merely observing them. The effect of drawing on memory was not due to the total amount of time spent viewing the objects; in both conditions, participants spent an equivalent amount of time with each object on the tour. Apparently, something about drawing the objects, made them more memorable. Given that the participants in Experiment 1 were self-identified artists who had scored their drawing ability as above average, it is possible that their advanced drawing skills allowed them to gain particular benefits from sketching objects. Would the same positive effects of drawing be incurred in a group of adults who were not artists? What about children? To answer these questions, in Experiment 2, we tested whether the memory of adult non-artists and of children would also benefit from the opportunity to draw.
Experiment 2

Method

Participants. A total of 30 adults (19 females; aged between 18 and 65) with no acknowledged drawing ability (M self-rated score for drawing ability on a 10-point scale = 2.40, SD = 1.30) were recruited in the local community by snowball sampling in social media groups. A total of 30 children (16 females; M age = 10.20 years, SD = 1.32, range = 8 to 12) were recruited from a database of families who had expressed interest in participating in research. Parents were contacted by phone and were asked if their child would like to take part in a drawing experiment at the museum. We selected children in the 8- to 12-year-old age range because the objects in the museum were feasibly been part of their school curriculum and they were still of an age where they had not yet become self-conscious about their drawing abilities (Malchiodi, 1998).

The research was reviewed and approved by the University of Otago’s Human Ethics Committee, whose guidelines are consistent with those of the American Psychological Association. Written informed consent was obtained from all participants. Participants received a small gift in return for their participation.

Procedure. The procedure was similar to that used in Experiment 1 except that to reduce fatigue and frustration, the number of objects that participants were asked to sketch or to observe was decreased from 20 to 18 for the non-artist adults and from 20 to 12 for the children.
Results

As in Experiment 1, all of the scores were converted to proportions prior to analysis. The free-recall scores of adults and children are shown in the right panel of Figure 3.2. These data were subjected to a 2 (Age) x 2 (Condition: sketched, observed) ANOVA with repeated measures over condition. There was no effect of age, $F(1, 58) = 0.325, p = 0.57, \eta^2_p = 0.01$, but, there was a main effect of condition, $F(1, 58) = 4.98, p < 0.05, \eta^2_p = 0.08$, and a significant Age x Condition interaction, $F(1, 58) = 17.61, p < 0.001, \eta^2_p = 0.23$. Adults recalled the same (albeit small) proportion of objects regardless of whether they were sketched ($M = 0.35, SE = 0.02$) or merely observed ($M = 0.39, SE = 0.02$), $t(29) = -1.65, p = 0.11, d = 0.30$. Children, on the other hand, recalled a greater proportion of the objects that they sketched ($M = 0.45, SE = 0.02$) compared to the objects that they had only observed ($M = 0.32, SE = 0.02$), $t(29) = 4.00, p < 0.001, d = 0.93$.

The right panel of Figure 3.3 shows participants’ scores on the two recognition components of the memory assessment. A 2 (Age) X 2 (Condition) X 2 (Test Phase) ANOVA with repeated measures over condition and test phase revealed main effects of age, $F(1, 58) = 13.83, p < 0.001, \eta^2_p = 0.19$, condition, $F(1, 58) = 65.73, p < 0.001, \eta^2_p = 0.53$, and test phase, $F(1, 58) = 4.02, p = 0.05, \eta^2_p = 0.07$. These main effects were qualified by significant interactions between Age and Condition, $F(1, 58) = 56.65, p < 0.001, \eta^2_p = 0.49$, Age and Test Phase, $F(1, 58) = 42.01, p < 0.001, \eta^2_p = 0.42$, and a three-way interaction between Age, Condition, and Test Phase, $F(1, 58) = 4.43, p < 0.05, \eta^2_p = 0.07$.

To evaluate the three-way interaction, we conducted two 2 (Condition) X 2 (Test Phase) ANOVAs with repeated measure over condition and test phase for each age group separately. As shown in Figure 3.3 (right panel), adults’ recognition
accuracy was not enhanced by the opportunity to sketch the objects, $F(1, 29) = 0.16, p = .69, \eta^2_p = .006$. Recognition accuracy for sketched objects (72%) and observed objects (71%) was high and was virtually identical. Not surprisingly, recognition accuracy was significantly higher when participants were presented with photos of the objects (82%) compared to when they read the names in the name-recognition test (61%), $F(1, 29) = 22.50, p < .001, \eta^2_p = .44$. There was no interaction, $F(1, 29) = 1.58, p = .22, \eta^2_p = .05$. Children’s recognition accuracy, on the other hand, was enhanced by the opportunity to sketch the objects; their recognition accuracy for the sketched objects (75%) was significantly greater than for the observed objects (46%), $F(1, 29) = 130.77, p < .001, \eta^2_p = .82$. Children also recognised a greater proportion of objects when presented with photos of the objects ($M = .66, SE = .02$) compared to when they read the names in the name-recognition test, ($M = .54, SE = .02$), $F(1, 29) = 25.01, p < .001, \eta^2_p = .46$. There was no interaction, $F(1, 29) = 2.93, p = .10, \eta^2_p = .09$.

Finally, we examined participants’ performance on the multiple-choice questions about the visual details of the objects (see Figure 3.4, right panel). A 2 (Age) x 2 (Condition) ANOVA with repeated measures over condition revealed main effects of age, $F(1, 58) = 69.26, p < .001, \eta^2_p = .54$, and condition, $F(1, 58) = 60.73, p < .001, \eta^2_p = .51$, and a significant interaction, $F(1, 58) = 49.11, p < .001, \eta^2_p = .46$. Again, adults did not benefit from the opportunity to sketch the objects; they correctly answered the same (albeit small) proportion of questions regardless of whether the objects were sketched ($M = .41, SE = .03$) or merely observed ($M = .38, SE = .03$), $t(29) = 0.62, p = .54, d = 0.20$. Children, on the other hand, did benefit from sketching the objects: they correctly answered a higher proportion of questions.
about the objects that they had sketched ($M = .88, SE = .13$) than about the objects that they had observed ($M = .39, SE = .23$), $t(29) = 9.56, p = .001, d = 2.62$. 
General Discussion

The overarching goal of the present study was to assess the effect of drawing on children and adults’ memory of objects that they encountered in a museum. In contrast to the negative effects of taking photographs (Henkel, 2014), drawing objects in the museum did not impair participants’ memory, and for some participant groups, drawing increased their memory. Adult artists who drew objects in the museum not only remembered more of those objects, but they also recalled more specific information about each one. Although drawing was not effective for adults who were not artists, children, like adult artists, recalled more objects that they had drawn and remembered more information about each one. For children, for whom learning in museums is a key outcome of such visits, it is particularly noteworthy that children answered more than twice as many questions correctly about objects that they drew relative to objects they only observed. In fact, when they were allowed to draw the objects, children’s performance on the free-recall measure and on the multiple-choice questions about the details of the objects exceeded the performance of adults. Overall, when drawing was effective, it was very effective—the effect sizes in this study were consistently large. Furthermore, the effect of drawing on memory by adult artists and by children could not be attributed to increased time with each object; the amount of time that participants spent with each object was exactly the same.

Some of the present findings were possibly more predictable than others. For example, it is probably not surprising that adult artists remembered more about the objects that they drew than objects that they merely observed. Research has shown that, compared to non-artists, when artists are drawing, they make more eye movements on and around the objects to be drawn (Cohen & Bennett, 1997), which might facilitate their encoding of the objects’ specific features. What we do know is
that the artists’ advantage was not due to a general memory advantage. Although artists have been shown to exhibit superior fine motor skills, and heightened perceptual and feature extraction abilities relative to non-artists (Arnheim, 1954; Chatterjee, 2004; Kozbelt, 2001; Mitchell, Ropar, Ackroyd, & Rajendran, 2005; Ramachandran & Hirstein, 1999; Zeki & Lamb, 1994), and to perform better on tests of visual memory, visual spatial processing, and mental rotation (Kozbelt & Seeley, 2007), these general skills did not help them on the non-drawing version of the task. When they did not draw the objects in the museum, their performance was equivalent to that of non-artists.

While the adult artists and the children used drawing to enhance their memory of objects in the museum, adults who identified themselves as non-artists did not. Can these differences be attributed to differences in drawing skill? As illustrated in Figure 3.1, the adult artists clearly exhibited superior drawing ability relative to the other two groups, but the drawing skills of the non-artist adults and the children were not that different. In fact, if anything, the drawings that were produced by the non-artist adults were of better quality than those produced by children. Despite this, children benefit from drawing but non-artist adults did not. We suspect that differences in their attitude towards drawing rather than drawing skill or expertise per se may have contributed to the differential effectiveness of drawing in these two groups. Children draw spontaneously from early childhood and continue to improve their drawing skills until early adolescence (Allen, Bloom, & Hodgson, 2010; Callaghan, 2015; Davis, 1997a; Golomb, 1987). By the age of 7, children value the ability to make convincing observational and representational drawings and will continue to refine and improve their drawing skills for several years (Watts, 2005). Despite this early, universal interest in drawing, the majority of people cease to improve their drawing skills beyond adolescence; instead, the self-imposed
pressure to produce photorealistic representations of objects leads to inhibition (Davis, 1997b; Rosenblatt & Winner, 1988). Thus although drawing, unlike taking photographs (Henkel, 2014), did not impede memory by non-artist adults, their unease and beliefs about their drawing skill may have prevented them from enjoying the same benefits of drawing that were seen in artists and children.

The present findings add to a small, but growing, body of research on the educational benefits of drawing. Although drawing did not facilitate memory in the non-artist adults in this study, there is evidence that adults may benefit from drawing in other educational contexts. Within the area of medical education, for example, researchers have shown that adult medical students remember individual items more often and for longer when life drawing is used as a learning tool in lessons involving histology and neurosurgery (Balemans, Kooloos, Donders, & Van der Zee, 2015; Clavert, Bouchaib, Duparc, & Kahn, 2012; Elsharky & Hernesniemi, 2014). At a time in history when governments worldwide are seeking ways to encourage academic success in science, technology, engineering and math (STEM), these results underscore the potential value of integrating art into the science curriculum (STEAM; Jon & Chung, 2010).

In conclusion, drawing represents an inexpensive and untapped asset to facilitate learning and memory in museums, classrooms, and other environments. Given that sketches have the potential to facilitate learning and memory for complex information, we suggest that this technique may be further utilised by a range of learners in a wide variety of settings.
Chapter 4

New Kids on the Block, Learning while Exploring in Minecraft

The results of the research conducted in Chapter 3 point to the possible advantage of integrating drawing into learning experiences in and out of school contexts. Despite the potential learning benefits of drawing, recent evidence suggests a decline in children’s drawings skills as the push for digitization increases. For example, a recent (2015) survey by the Dutch Schools Inspectorate (2017) compared 11- and 12-year-old children’s ability on a range of sketching and illustration tasks with those of children who were surveyed in 1996. They found significant declines in children’s ability to add detail to their sketches. They also found that the more recent sample found it difficult to illustrate a whole story or incident and that they were unable to include all elements of a story in one drawing. The authors of the Dutch report hypothesise that an increase in digital activities is one possible source of the decline of drawing skills in this age group.

Engaging with digital technology is increasingly an ordinary aspect of childhood. Livingstone, Bober, and Helsper (2005) note that 82% of 9-19-year-olds are spending as much time playing computer games as they are on homework. Despite gaming being a normal part of life outside of school, there is considerable reluctance to integrating gaming technology into pedagogy (Bösche & Kattner, 2011; Crowe & Flynn, 2014). With this in mind, in Chapter 4, we examined the potential of new creative technologies to enhance or support learning.

The first decade of the 21st century spurred massive increases in research into human–computer interaction. The development of a broad range of platforms, games, and applications has resulted in technology becoming more accessible and

---

5 This chapter is under review as Cronin A., Gross, J., & Hayne, H. New kids on the block: Learning while exploring in Minecraft. Journal of Experimental Child Psychology.
open (Martin, 2015). In most Western countries, children are immersed in technology earlier, and over increasing periods of time, and technology has now become a common tool in the classroom (Loh & Kanai, 2015). Educators now make use of Educational Technology (Ed Tech), computer games, and virtual worlds as learning tools for children and young adults (Beavis, O'Mara, & MacNeice, 2012; Short, 2012; Wang & Towey, 2013). Most schools are now expected to provide education that integrates new technologies into curricula, which are not only relevant to users, but also facilitate interest and cultivate future career trajectories in technology.

The penetration of digital products into so many spheres of activity brings with it a host of questions about the impact of technology on children (Boulos, Hetherington, & Wheeler, 2007). In addition to the possible benefits of technology-assisted learning, there are also some legitimate concerns about the use of new technologies in educational settings. For example, it has been argued that new technologies may serve as a distraction from established, developmentally-appropriate activities (Hein, 2006; Rajala, Kumpulainen, Jaakko, Paananen, & Lipponen, 2016). Furthermore, the introduction of technology in the classroom has not always been an evidence-based decision and has often been the result of an industry-led push towards technology for technology’s sake. For example, many schools use iPads as digital books in place of physical books with little or no evidence for improved learning outcomes. In addition, many educators have voiced concerns that much of educational technology does not reflect children’s motor, perceptual, cognitive, or social development (Antle, 2013).

Despite these valid concerns, engaging students in the learning process remains one of the most significant challenges to education and it is clear that gaming technology and immersive virtual environments can increase engagement
because they are fun. It has been well established that play is essential to children’s development, and digital games offer a new and stimulating environment in which to play. In this way, play within virtual environments might offer a range of learning possibilities, from social skill development to individual problem solving (Marsh, 2010). In fact, Plowman, McPake, and Stephen (2012) have argued that digital technologies in educational contexts may be the key to engagement and better learning outcomes because effective learning is mediated by situated, active, experiential, experimental experiences which provide opportunities to solve problems and receive immediate feedback (Connolly et al., 2010).

But the unanswered question for educational researchers is: what do children learn from technology-enhanced platforms? It turns out that this question is rarely addressed. For example, Connolly, Boyle, Hainey, and Boyle (2010) systematically reviewed over 7392 papers purporting to show learning using game technology and identified only 129 papers in which the researchers specifically assessed learning outcomes. The majority of the papers in Connolly et al.’s review documented the engagement potential of games, rather than knowledge acquisition or the perceptual, cognitive, behavioural, affective, motivational, physiological, or social outcomes associated with the interface.

One computer game which has been adapted for educational purposes is Minecraft in which the core gameplay revolves around construction—in essence, it provides a 3D game play experience that is comparable to traditional block-building play. Wildly popular, Minecraft is the most downloaded game in history and it can be played across different gaming platforms, from handheld to console devices. Minecraft is not educational software—it is a game designed for entertainment, but it may also offer new opportunities for learning (Bos, Wilder, Cook, & O’Donnell, 2014; Gauquier, & Schneider, 2013; Riordan, & Scarf, 2017). In the game, players
can build or modify the existing structures as they please. Players can gather materials, and either place them in the world to create various structures or use them to create other items. The virtual world of Minecraft consists of square blocks that can be altered and manipulated to build and create landscapes, using a variety of different in-world resources, such as timber or iron. Also, Minecraft has the capacity for multiple players to collaborate within the same worlds at the same time. It is possible for players to collaboratively design and construct a unique place, building, society, or world using the tools provided in the game.

Some schools have begun using Minecraft as a teaching tool on a variety of topics from humanities to chemistry (Minecraft.edu, 2014), for a review, see Nebel et al., 2016. When used in this way, online communities build diverse Minecraft worlds that are available for others to explore from the comfort of the classroom. For example, Minecraft users can take a virtual field trip to the Coliseum at the height of the Roman Empire, or take a tour of Dublin, Ireland at the turn of last century. Many schools incorporate field trips to enrich children’s learning including the acquisition of scientific facts and concepts (National Research Council, 2009). Using Minecraft to create ‘virtual field trips’ has some obvious advantages, children can explore sites that may be geographically remote, inaccessible to them, or no longer in existence.

If a computer game like Minecraft could provide educational benefits similar to those provided by a physical field trip, it paves the way for significant benefits for many students and schools. To the best of our knowledge, no study has examined whether virtual field trips have an educational benefit akin to that following actual visits to educational sites. Here, we assessed the educational value of Minecraft as an exploration and learning tool, by comparing it with the learning and engagement outcomes that occur via an on-site field trip. To do this, 11- to 12-year-old children
were taken on a 30-minute tour of an Industrial Heritage Museum or they toured a
Minecraft version of the same museum. The museum lent itself well to developing a
model in Minecraft as many of the main features of the game are evident in the
museum, such as the use of coal, basic physics, and buildings with specific functions
and interesting shapes. The instruction on the tour was deliverable across platforms
and was of interest to intermediate level school-aged children.
Method

Participants

Ninety-nine 11- and 12-year-old children (42 girls, M age = 11.30 years, SD = 0.45) were recruited from three local primary or intermediate schools in Dunedin via a letter to principals offering class groups a free tour of an industrial heritage museum housed in a closed gas production plant. The schools were assigned decile ratings of six to eight (maximum = 10) by the New Zealand Ministry of Education. This rating is based on socioeconomic indicators from national census data obtained from households with school-age children in the areas from which the school draws students. In general, a decile rating of eight roughly corresponds to a socioeconomic status of upper middle-class, but the actual student roll included children from throughout the city and from a wide range of socio-economic backgrounds. The children were predominantly of European descent and participated with written consent from their parents. The research was reviewed and approved by the University of Otago’s Human Ethics Committee, which is accredited by the New Zealand Health Research Council and whose guidelines are consistent with those of the American Psychological Association.

We selected 11- and 12-year-olds because children in this age group are the target of many science education programs in museum settings, particularly in the context of science learning with the process of extraction of gas from coal being relevant to both chemistry and biology as well as to history topics. It was also important that the children had a degree of computer literacy skills allowing them to carry out a range of functions in a virtual environment.
Field Trip

Students participated in a novel educational event during which they toured a former gas production plant where the students learned about the history of industrial gas production. None of the students had previously visited the museum site. In the live tour a female guide took students on a 30-minute tour of the site that included 6 points of interest: the Engine, Chimney, Boiler Room, Fitting Room, Library, and Forge. At each point of interest, the guide provided students with a small amount of information (see Appendix A). The information was also on signage at each point of interest. The tour was specifically designed to be appropriate for students aged 11-12 years, and the tour guide was experienced in interacting with students of this age. During the in-school version, the same female experimenter guided the students around a Minecraft version of the museum. Students were allowed 5 min. to become familiar with the controls for the Minecraft desktop version, and were instructed to assemble at the same point that other groups would have started their tour. The script for the real and virtual tours were identical and students listened at each of the points of interest to the guide explaining the function of various parts of the plant.

Procedure

Schools were randomly assigned to one of four experimental groups (see Figure 4.1). Students from three of these school groups travelled by bus to the museum site where they were met by the guide who explained the health and safety protocol for the site and then took students on the tour. At the end of the tour, students in the Onsite Tour only group (n = 22; see Figure 4.1A) were given 30 mins to freely explore the site before being asked to evaluate their learning experience using a

---

6 NB. Because participants were run in classroom groups, the sample size for each experimental group was uneven.
modified version of the Learning Object Evaluation Scale for Students (LOES-S; Kay & Knaack, 2009). On the LOES-S, students used a 5-point scale (1 = Strongly Disagree, 5 = Strongly Agree) to rate their level of agreement with 14 statements under the categories of Learning (5 statements, e.g., “Visiting the museum helped me learn.”), Quality (4 statements, e.g., “The museum tour was easy to understand.”), and Engagement (5 statements, e.g., “I want to find out more about this place.”). After completing the evaluation, students returned to school.

For students in the Onsite Tour + Worksheet group (n = 31; see Figure 4.1B), at the end of their tour, they were given a pencil and a worksheet containing 19 questions about information that they had been told or read on the signage at each of the points of interest during the tour and were told that they had 30 mins to explore the site to find the answers to the questions:

1. Why was it dangerous after dark in Dunedin before the Gasworks was built?
2. What century was the Dunedin Gasworks chimney built in?
3. What is an anvil used for?
4. Name 3 things that were in the forge?
5. Where did you see an anvil?
6. What do you call a person who works with iron?
7. What was gas made from in the Dunedin Gasworks?
8. Why was the chimney built so high in the Dunedin Gasworks?
9. Did people enjoy living near the Gasworks?
10. Why?
11. Stephen Stamp Hutchinson worked in 2 other cities before Dunedin, can you name either of them?
12. What year did the Fitting Room blow up?
13. What colour does iron turn when it is ready to forge?
14. What caused the Fitting Room to blow up?
15. If you were working in the Dunedin Gasworks, where would you find information on how to repair a machine that had broken?
16. How is coke made from coal?
17. Coal gas is made up of three chemicals, can you name one?
18. What are some of the by-products produced when making gas from coal?
19. What does coalesce mean?

Finally, students evaluated their experience using the LOES-S, before returning to school.

Students in the Onsite Tour + Minecraft Worksheet group (n = 22; see Figure 4.1C) also completed the worksheet at the end of the tour, however, they were instead given the opportunity to explore the site in a Minecraft world to find the answers. To do this, students were taken to a room on-site that housed 22 iMacs connected to a server running Minecraft.edu (Mojang, 2014) over a private LAN. The built version of the museum site was a low detail Minecraft version of the physical site, albeit on a smaller scale (see Figure 4.2). Students were given instructions on how to operate Minecraft and then had 30 mins to explore and complete the worksheet, before completing the LOES-S and then returning to school.

Students in the final experimental group did not visit the museum site in person. Instead, students in the Minecraft Tour + Minecraft Worksheet group (n = 24; see Figure 4.1D) logged on to a private server running Minecraft.edu in their school classroom and after being given instructions on how to operate Minecraft, completed an in-game tour of the museum site. The in-game tour was identical to the on-site tour and the same guide who had conducted the on-site tours also conducted the in-game tour. At the end of the tour, students were given 30 min to explore the
museum site in the *Minecraft* world to find the answers to the questions on the worksheet, before completing the LOES-S.

![Diagram of experimental groups](image)

*Figure 4.1. A schematic representation of the 4 experimental groups.*

One week after each experimental group had experienced the on-site of Minecraft tour, the same experimenter who had conducted the tour went to each group’s classroom at school and tested their memory for the tour using the same worksheet that she has used in the first session. Students also completed a 10-item spatial awareness quiz to assess their ability to recall the locations or functions associated with the six points of interest that they had encountered during the tour.
Figure 4.2. Top Panel: An aerial view of the Gasworks site (Google Maps, 2017) and the corresponding aerial view from the Minecraft version. Bottom Panel: Side-by-side view of inside one of the buildings. Left: the onsite library space; Right: the library in the Minecraft version.
Results

Learning Outcomes

First, we examined students’ ability to complete a worksheet when they were given the opportunity to search for the answers immediately after the tour, either on site or in game. Overall, students in the three experimental groups who completed the worksheet immediately after the tour found this task relatively easy, answering on average more than 15 of the 19 questions correctly. A one-way analysis of variance (ANOVA) across the experimental groups and a post hoc Tukey’s test ($p < .05$) revealed that students in the On-site Tour + Minecraft Worksheet group answered significantly more questions ($M = 17.14, SE = 0.46$) than did students in the Minecraft Tour + Minecraft Worksheet group ($M = 16.25, SE = 0.44$) and students in the On-site Tour + Worksheet group ($M = 15.13, SE = 0.39$), $F(2, 74) = 5.78, p < .005$, partial $\eta^2 = .134$.

Next, we examined students’ memory performance when they were tested one week after the tour (see Figure 4.3). Recall that after the one-week delay, students in all four experimental groups were tested on the worksheet in their classroom at school. A one-way ANOVA across the experimental groups and a post hoc Tukey’s test ($p < .05$) revealed that students in the On-site Tour + Minecraft Worksheet group answered significantly more questions on the final worksheet ($M = 13.18, SE = 0.58$) than did students in the On-site Tour + Worksheet group ($M = 10.51, SE = 0.49$) and students in the Minecraft Tour + Minecraft Worksheet group ($M = 10.16, SE = 0.56$), who in turn answered significantly more questions than did students in the On-site Tour only group who only took part in the tour of the museum and did not complete a worksheet immediately after the tour ($M = 4.45, SE = 0.58$), $F(3, 95) = 40.28, p < .001$, partial $\eta^2 = .560$. 
Figure 4.3. Mean scores (+1SE) on the final worksheet as a function of experimental group.

Finally, we examined whether students’ performance on the worksheet decreased over the 1-week delay by conducting a series of paired $t$-tests ($p < .05$)
between students’ scores at Time 1 and Time 2. The performance of all of the groups who were tested on the worksheet at both time points decreased significantly over the 1-week delay (smallest $t(df) = 4.40, p < .001, d = 0.94$).

In the last part of our analysis of students’ learning outcomes, we examined their performance on the 10-item spatial awareness quiz administered following the one-week delay. There was an effect of experimental group, $F(3, 95) = 9.26, p < .001$, partial $\eta^2 = .226$. Post-hoc Tukey’s tests ($p < .05$) indicated that students in the On-site Tour + Minecraft Worksheet group performed significantly better ($M = 6.04, SE = 1.70$) than did students in the other 3 groups (On-site Tour only: $M = 3.31, SE = 1.58$; Minecraft Tour + Minecraft Worksheet: $M = 3.50, SE = 1.74$; On-site Tour + Worksheet: $M = 3.93, SE = 2.33$), whose performance did not differ.

**Perceptions of the Learning Experience**

To examine how students perceived their learning experience, we analysed their ratings on the learning evaluation scale (the LOES-S) that they completed after the tour (see Figure 4.4). Recall that the LOES-S consists of three constructs; Learning, Quality, and Engagement; students rate the learning object (in our case, the museum tour) on all three constructs. Because the three constructs are on different scales, we analysed students’ scores on each construct separately using one-way ANOVAs across groups. Any significant effects were examined using post hoc Tukey’s tests ($p < .05$).

As shown in Figure 4.4, overall, students rated their learning experience highly. On the Learning construct, students in the On-site Tour Only and Minecraft Tour + Minecraft Worksheet groups rated their experience more highly than did students in the other two groups. On the Quality construct, students in the On-site Tour Only, On-site Tour + Worksheet, and the Minecraft Tour + Minecraft Worksheet groups rated their
experience more highly than did students in the Minecraft Tour + Minecraft Worksheet group. Finally, on the Engagement construct, students in the On-site Tour Only and Minecraft Tour + Minecraft Worksheet groups rated their experience more highly than did students in the other two groups.

![Figure 4.4. Mean scores (+1SE) on the LOES-S subscales (learning, engagement, and quality) as a function of experimental group.](image)
Discussion

The overarching goal of the present experiment was to investigate whether a field trip in a virtual environment provided comparable educational benefits to that of an onsite field trip. We evaluated students’ memory for information presented during the field trip and their perceptions of the learning experience. We found that children who completed additional activities in Minecraft after an onsite tour recalled 68% of target information about the event one week later. Children who toured and completed worksheets at the museum and those who did the same from their own classroom in Minecraft recalled 52% of the facts presented. Children who toured the museum and did not complete any follow-up activities recalled only 26% of the target information. All groups reported enjoying their learning experience, however, their level of enjoyment did not necessarily relate to how much information they remembered about the experience.

Researchers have shown that museum visits range in educational effectiveness depending on the additional learning tasks carried out to support them (Gilbert & Priest, 1997; Griffin & Symington, 1998; Hofstein & Orion, 1994). One method of increasing engagement with information presented during out-of-school visits are worksheets which force visitors to search for information which is presented on site. Our findings support the value of including on-site activities, the group who completed an On-site Tour with no additional learning activities had the poorest retention of information about the site and its functions. This finding is highly consistent with prior research that shows that educational visits to museums that include no preparation, onsite, or follow-up activities result in poor learning outcomes (Anderson & Lucas, 1997).

Our findings also illustrate that not all post-visit learning activities are created equal. We found that that those participants who took part in an On-site Tour and
then completed the *Minecraft* worksheet remembered significantly more of the target information than did students in the other three groups. Being able to physically explore the site and then engage with a version of the site in *Minecraft* gave students a unique and novel way to engage with the material we asked them to learn. Participants in the *On-site Tour + Worksheet*, and *Minecraft Tour + Minecraft Worksheet* groups were presented with the same information, albeit one group onsite at the museum and the other in their classroom. These groups performed equally as well at test, suggesting that *Minecraft* alone can be used to introduce and consolidate new information in certain educational settings.

Given the differences between a low-detail virtual model of a site and the highly detailed real-world experience of visiting a place of interest (see Figure 5.2), we also examined whether there were differences in children’s ability to map or recall the location of items which formed part of the tour. Spatial memory is learned automatically and may result in additional learning, which supports the recall of information during follow up testing (Hess, Detweiler, & Ellis, 1999). It is also possible that information learned through encoding of locations may be used as a reference guide to other information (Pylyshyn, 1994). For example, Richardson and Spivey (2000) showed that, when tested, people looked towards the original location of an item to recall specific information about it, despite the fact that the answer was not physically present.

In the present experiment, we examined students’ ability to correctly locate items on an aerial map of the museum site. Consistent with their performance on the worksheet, the *On-site Tour + Minecraft Worksheet* group again recalled more information about the location of items on site compared to the other groups. The ability to physically engage with the site and contrast that with the low-detail *Minecraft* model may have advantaged these students. Overall, the best learning
outcomes resulted from a physical experience and an in-game activity suggesting that *Minecraft* can be a beneficial learning tool to consolidate a range of learning and spatial information learned at a site of interest.

One advantage of virtual environments is that they minimize the range of distractions and external influences that are present in the real environment while still being able to facilitate learning (Bitgood, 1991). Gross, Hayne, and Drury (2009), for example, found that children who went on an educational field trip to a local nature museum recalled significantly more autobiographical information about the field trip (e.g., how they got there and what they ate for lunch) than they recalled educational information (e.g., that albatross feed their young by vomiting food into the baby’s tummy). Moreover, Imuta, Carson, Scarf, and Hayne (2017) found that children who received a nature lesson on the Royal Albatross in their own classroom exhibited significantly greater recall of scientific information than did children who received the same information during a field trip to the Royal Albatross Colony. Virtual environments used judiciously may minimise distractions and allow students to focus on the learning points of interest. Online or *in-game* field trips can, similar to their real-world equivalents, expose children to new environments and stimulate interest in a range of science topics (Anderson, Piscitelli, Weier, Everett, & Tayler, 2002; Henry, 2016; Rajala, 2016). With improvements in virtual world environments, the amount of exposure time to the on-site information can be lengthened or augmented in a classroom.

In addition to measuring the learning outcomes associated with these field trips, we also examined students’ evaluations of their experience—the outcome that is most often measured in studies of this kind. Previous studies in different topic areas have found high levels of enjoyment using *Minecraft* in educational settings. Zorn, Wingrave, Charbonneau, and LaViola (2013), for example, found that
students were motivated to engage with programming with *Minecraft*. Similarly high engagement is reported for students using *Minecraft* to learn about biology, ecology, physics, and social interaction, across a range of published studies, but as no learning outcome measures accompany these studies it is difficult to conclusively support its use in the classroom (Short, 2012). In one of the few studies to measure learning outcomes, Wang and Towey (2013) found that students who used *Minecraft* to learn physics concepts performed better on follow up tests compared with a control group who learned the same material during normal classroom instructions.

In the present experiment, we asked students to evaluate the experience (the tour and/or associated activities) using the LOES-S which consisted of 3 subscales; Learning, Quality and Engagement. All groups rated their experience above average, however, the groups who evaluated the experience the highest did not have the best learning outcomes. Interestingly, the *On-site tour* group estimated that they learned the most from the tour and rated it high in engagement and quality, but they performed significantly poorer on the tests of knowledge, compared to the children in the other three groups. Based on our findings, it is clear that simply asking children if they enjoyed themselves, or how engaged they were, is not a proxy for discovering whether they actually learned anything.

Children, like adults, may overestimate their actual learning or underestimate it depending on a range of expectations and prior experiences. Even adult learners are poor judges of their own learning, and that judgments can be influenced by beliefs (Bjork & Bjork, 2014; Jacoby & Kelley 1987). For example, Bjork et al. (2009) found that students evaluated the success of a learning session on how easy it was to understand at the time of learning, but this does not necessarily mean that the information is committed to long-term memory. Bjork, Dunlosky, and Kornell (2013) have argued that what is easy to learn is often thought of as easily ‘absorbed’
into long term memory compared to the more challenging experience of when the learner must actively engage with material to understand it. It is reasonable to expect that children’s evaluations of the learning potential of experiences suffer from similar biases.

In contrast, although children who completed an On-site Tour + Minecraft Worksheet had the best recall of facts and spatial information, these children also gave the lowest ratings on the learning, quality, and engagement, sub-scales of the LOES-S. It is possible that for these children, learning via Minecraft on a field trip did not fit their schema of what a learning experience should be. Other researchers have noted that on museum visits, teachers and students regard having fun to be a major objective for their visit although whether school work should be fun is debated by many educators (Dierking, 1991; Lucas, 1999, 2000). Clearly there are a range of factors that may impact evaluations of learning experiences which do not relate to actual learning. This may be particularly evident when evaluating learning experiences that involve game technology, because the ‘fun element’ may result in mixed evaluations similar to the effects we note here. Our findings provide further evidence that caution should be taken when evaluating learning outcomes solely on the basis of students’ self-evaluation of their own learning.

Engaging with gaming technology is increasingly an ordinary aspect of childhood. Despite gaming being a normal part of life outside of school, considerable reluctance exists to its integration into pedagogy (Bösche & Kattner, 2011; Crowe & Flynn, 2015). Livingstone and Bober (2005) reported that 82% of 9-19-year-olds spend as much time playing computer games as they do on homework. While we would not suggest that teachers or museum educators abandon completely the sensory experience of on-site field trips in favour of virtual tours, a virtual tour provides considerable benefits for schools who wish to engage students in what is a
highly enjoyable activity with minimal cost. Although in the present research, we controlled in-game activities in Minecraft (adding or removing blocks, in-game talking, ability to fly) that participants could engage in, there is huge potential for educators and children to use Minecraft to edit, develop, and create new learning opportunities. While not all games offer the scope and appeal of Minecraft as a learning tool, this form of activity offers at least some benefits to learners as well as offering an egalitarian solution to access and enjoyment.
Chapter 5

Concluding Comments

The overarching goals of this thesis were threefold. The first goal was to investigate whether children’s drawing is affected by instruction and if there are implications for established psychological tests that derive measures from these artworks. The second goal was to explore drawing-enhanced learning and memory in adults and children at museums, where drawing is utilised as a recreational activity rather than as an explicit learning opportunity. The final goal was to explore children’s contemporary creative practices, which have migrated from pen and paper to digital technologies, and the possible benefits of these virtual experiences for learning in museums. To achieve these goals, in Chapter 2, I examined time-related and experience-related changes in children’s human figure drawings. In Chapter 3, I explored how drawing from life in a museum could enhance adults’ and children’s memory for objects in the museum. Finally, in Chapter 4, I examined how new digital tools might be utilised to enhance memory for information presented during both real and virtual museum visits. In this final chapter, I will review the main findings of each experiment contained in the thesis, discuss their implications, and propose new avenues for future research.

Drawings serve dual purposes: they can be used to express ideas and concepts by the artist as well as to create objects that can be viewed by others. Drawings fulfil many roles across their lifespan and are judged by viewers on a range of factors, from technical accuracy to aesthetic quality, depending on the contextual function of the drawing. In particular, young children’s drawings are not judged for photorealism; instead, they are appreciated for their unconventional nature. Preschool children produce drawings that are considered spirited, original, and aesthetically appealing to adults, although they are unrealistic and often unplanned. Viewers’ interpretation of
children’s drawings has formed the basis for much psychological research for the last 100 years - partly inspired by the fascination with the ‘untamed’ nature of young children’s drawings. For example, young children’s drawings (under 5 years of age) and those of mature artists have been evaluated by viewers as being similar in aesthetic quality (Gardner & Winner, 1982). However, as children progress through primary school, they draw in much more predictable (stage-like) and standard (canonical and schematic representations) ways that conform to a range of cultural conventions. Fascination with the produced work at either end of this ‘aesthetically pleasing spectrum’ is justified; however, interpretation of the drawings themselves presents a range of challenges to scientific research.

One of the primary requirements of scientific research is to reduce complexity to measurable, comparable units. For reasons unclear, various branches of psychological sciences’ study of children’s drawings are underpinned by psychodynamic theory. Psychodynamic theory has held back the goal of understanding the uses of drawing by focusing almost exclusively on the possible meanings behind the drawings that people create (Hargreaves, 1978, Neale, et al. 1993; Peterson, & Batsche, 1983; Skybo, et al., 2007; Trowbridge, 1995). Later cognitive approaches adopted some elements of these early theorists work by adopting the Human Figure Drawing (HFD) as measurement unit. HFDs have remained the basic unit of measurement in a range of psychological studies since the early 1900s, because children are socialised to produce drawings of themselves and family members. However, at no stage has anyone ever justified why a HFD is the most suitable drawing output to use as the primary measurable unit of artistic ability – perhaps, as outlined in Chapter 1, the number of individual elements in the human figure which could be quantified formed the basis of this corner stone selection along with HFDs’ stage-like development. HFDs produce a quantifiable unit of
measurement: there are a range of tests which have been used to infer different psychological constructs based on the presence or absence of different elements of a child’s figure drawing. Psychologists have subsequently drawn conclusions about the child’s internal states in these children based on the presence or absence of these elements. These assumptions formed part of the opening chapter’s work and informed the development of the research described in Chapter 2. The results of Chapter 2 point to a high degree of malleability in HFDs but they also suggest these static images may not adequately reflect IQ.

The justification for the use of HFDs by so many researchers and clinicians is motivated in part, by the availability of a range of scoring systems, the uniformity of expectations, and the ease of administration of a figure drawing test. One aspect that is rarely mentioned by researchers using HFDs as proxy measures is the influence of the environment, particularly practice, on the development of drawing. The extent to which children practice drawing figures is difficult to pinpoint, but it is reasonable to assume that some children have greater access to, and interest in drawing than others, which in turn affects the quality of their HFDs. In addition, children are influenced in their development of HFDs by peers and society, once again, to different extents (Toku, 2001a). As a consequence, there are issues with using HFDs as a valid measure of anything, until the effect of practice and environmental factors are fully accounted for.

On the surface, the development of drawing ability from simplicity to complexity appears to mirror the development of intelligence (IQ). While IQ has been shown to be attributable to a complex interaction of genetic and environmental factors, no firm evidence exists that drawing ability is similarly affected by genes or environment. Contemporary research into the basis of human intelligence has determined the extent to which a range of environmental factors can lead to variance
in intelligence development. For example, Olson, Bates, and Bayles (1984) suggest that maternal IQ is the most critical predictor of a child's subsequent cognitive functioning. Similarly, extensive research has concentrated on the development of cognitive (Landry, Smith, Miller-Loncar, & Swank, 1998) and linguistic (Bornstein & Tamis-LeMonda, 1989) skills based on the amount of and quality of interactions between parents and children. These works highlight the need for a thorough understanding of the role of environment and its influence on the development of skill in any domain.

As Chapter 2 highlights, the development of intelligence in young children and the development of drawing skills were proposed as equivalent in the early 20th century, and almost 100 years later, the myth that HFDs reflect intelligence continues. Given that we understand that parent-child interactions have a meaningful effect on the development of cognitive and linguistic development, it is somewhat surprising that nobody has questioned how family influence may have a role in the development of drawing. In other aspects of children’s skills development, such as musical prowess, parental influence is responsible for stimulating and arousing children's interest and ability (Bloom, 1985; Chadwick, 2011; Gagné, 1993, 1995; Sloane, 1985). The same is true of parents or caregivers encouraging and developing drawing ability. However, no extensive investigation has examined the impact of environmental factors such as practice, encouragement, and instruction on children’s drawing development.

In Chapter 2 of the present thesis, I examined the role of practice and instruction on children’s production of human figures. Considering the effect of instruction in isolation, the results of Chapter 2 point to a high degree of malleability in the drawing skills of children. Few people teach children how to draw - a ‘hands off’ approach to drawing is adopted in most Western cultures, resulting in wide
variations in drawing ability within age groups. Unlike a skill such as musicianship, which also requires hours of practice, instruction, and encouragement, drawing is not usually afforded the same time in a child’s life. The results of Chapter 2 show dramatic increase in human figure drawing scores on HFDs as a result of only 2 hours’ drawing instruction, indicating that even with minimal intervention, children’s drawing skills can improve. At the same time, the results of Chapter 2 also show that without regular practice, children’s drawing skills regress to previous ability levels quickly. In the future researchers could examine the effect of more sustained drawing practice with instruction to better understand the dynamics of the development of this skill.

The second outcome of the research in Chapter 2 addressed the underlying assumption related to HFDs: that they contain information about a child’s intelligence. As outlined in Chapter 1, numerous researchers have attributed psychological meaning to aspects of children’s drawings; these tests have a common ancestry - the Goodenough-Harris Draw-A-Person (DAP) test of intelligence. Indeed HFD-IQ tests based on the Goodenough-Harris DAP have proven to be a highly reliable measure, but of what? In the DAP test, HFDs were utilised as a measure of intelligence as they followed a step-like progression over children’s development. A few studies have found that IQ scores derived from HFDs’ scores are not valid (Dykens, 1996; Imuta et al., 2013; Willcock et al., 2011), yet their popularity persists. One critical feature of a valid intelligence test is that scores on that test remain relatively stable over time and would not be changed as a result of only one hour of instruction. However, drawing is a skill which can be taught or enhanced in isolation from other abilities very quickly.

In Chapter 2, I tested if minimal instruction in drawing would alter scores generated from the most recent HFD IQ test. During the intervention, basic
anatomy and proportions were taught as part of an intermediate-level science class. The subsequent increase in children’s DAP scores resulted in IQ scores increasing 19 points on average. The small drawing intervention yielded a universal improvement in IQ scores derived from children’s HFDs; a similar change in scores on traditional pen and paper IQ tests cannot be achieved using a short intervention like this.

What then, if anything, are HFDs good for? As mentioned, HFD scoring systems are a highly reliable way of measuring how many parts and details a person includes in a human figure drawing. Whether this is of any use in contemporary science is another matter: HFDs are residual from Psychology’s attempt to create a metric for drawing ability and uncover underlying psychological states. The investigation of psychological constructs through drawings has mainly been debunked, but even the use of HFDs as measurements of drawing ability should be called into question. What precisely do scores generated from HFDs reflect? It is common practice to say that the inclusion or omission of detail in a human figure tell us about the artist's ability relative to their peers. Even this metric is of concern, as inclusion or omission of details is not equivalent to greater drawing ability: artists routinely communicate complex structures with minimal line drawings.

As outlined, there are numerous reasons why HFDs may have become the proxy measure for artistic ability, but the fact that HFDs have remained part of psychological toolkits for over 100 years by so many without question is disappointing. An easily administered test and scoring system is a tempting prospect to any researcher, and indeed despite extensive critique, HFDs are still widely used in contemporary research and clinical practice. There are few other psychological tests still in use which originated in the 1920’s. Psychology has experienced a rapid rate of change in the last 100 years, as the discussion of drawing in the opening chapter highlights. The Draw-a-Person test and its many subsequent forms are still used as
diagnostic tools where other early attempts at discovering underlying states through projective techniques such as Rorschach tests have lost favour. Drawing, as outlined in the introduction, has been a tool for humans for millennia. It is a skill which few master but many could, and this disparity between what some people can draw and what others believe they cannot led to the study described in Chapter 3.

The use of pictures to aid comprehension and retention of knowledge in a range of disciplines has been extensively studied (Alesandrini, 1984). The chief finding from a range of investigations is that the addition of a picture that represents the information to be learned has a positive effect on learning outcomes. For example, in a study by Mayer (1989), students were asked to learn about the mechanical braking system: one group of students received information in a text-only format, while the other group received the same information accompanied by illustrations of the process. The groups were tested on a range of follow-up measures to assess comprehension and recall. Performance of the text-only group did not differ from the performance of the text-and-illustration group on the word recognition tests, but the text-and-illustration group outperformed the text-only group on tests of problem solving and explanatory recall. Mayer concluded that the addition of pictures significantly increased the learning and retention of knowledge. However, the amount of attention paid to pictures may vary and thus reduce these positive effects. Finding ways of engaging further with images is one way of increasing the attention devoted to images. For example, Schwamborn (2011) found that comparing, tracing, labelling, and drawing increased the effect of picture viewing and had a positive effect on learning outcomes.

Researchers who have focused on investigating the role of creating drawings to learn, rather than merely observing illustrations, have generated some support for using drawing as a learning tool. For example, Hall, Bailey, and Tillman (1997)
compared the learning outcomes for three groups of learners. All groups learned about the functioning of a hand pump: Group 1 through text only, Group 2 through text and illustration, and Group 3 via text and participants' self-generated illustrations. Participants in Group 3 performed significantly better on post-tests of recall and comprehension than the other two groups, and the accuracy of their drawings was significantly correlated with test scores. Hall et al.’s findings combined with comments by Van Meter (2001) who argued that the post-tests employed in measuring learning outcomes with illustrations and drawings should include higher-order tests (open-ended questions) informed how I structured evaluation of learning outcomes for the study reported in Chapter 3.

Researchers in medical science have also examined the possible uses of drawing as a learning aid. Commentary from medical schools, in particular, is of interest, as the move towards digital images has been felt by some to lessen the learning impact of drawing diagrams for students. For example, Clavert, Bouchaib, Duparc, and Khan (2012) suggest that in-class drawings of human anatomic structures during early medical training is both engaging and actively develops students’ understanding of the relationships between structures. Further research into medical students’ study strategies and their learning outcomes has revealed that creating pencil and paper drawings improved retention of knowledge, even over long periods. Balemans, Kooloos, Donders, and van der Zee (2015), for example, examined the long-term retention of histological knowledge at three different time points, using a randomised cross-over design. In the first part of the experiment, 348 medical students were randomly assigned to a drawing or non-drawing group to study throat anatomy and histology. In the second part of the study, the groups swapped learning method (i.e., the drawing group swapped to non-drawing) and then studied the muscle cells of the heart. When tested 1-, 4-, and 6-weeks later, Balemans
et al. found that more information was retained when it had been drawn than when it had not been drawn.

In a preliminary study conducted as part of this thesis, I examined the differences in learning outcomes for two groups of 11- to 13-year-olds in a drawing and non-drawing intervention. Over the course of 4 weeks during their regular intermediate biology classes, one group illustrated their biology notes, adding pictures of the function of different parts of the brain to their workbooks. Another class, matched for age and academic ability (mixed ability general science class), took notes in their workbook without drawing. Participants were tested 1- and 6-weeks later. There were no significant differences in learning outcomes between the groups; however, after viewing the notes that the learners had created in their workbooks, there may have been a simple reason why the drawing intervention may not have been successful. Participants were given no direction as to how to structure their study or drawings: being naive learners, their notes and drawings were sparse and unstructured. Kornell and Bjork (2007) also noted that in the absence of specific instructions, self-regulated study is a confounding factor when attempting to examine study techniques and their outcomes. The students in my intermediate-level biology class were given workbooks with blank diagrams (to add either labels or drawings to), within the class or as homework, which explained the function of each lobe of the brain. What I did not control for was the amount of time that was spent on these tasks, the attention devoted, or the decisions that led the learners to apply themselves to studying this topic. It is reasonable to expect that the study habits of medical students in Balemans et al. (2015) would differ considerably from those of intermediate-level children, and future research could benefit from creating a set of study strategies to accompany interventions which seek to evaluate the effectiveness of study strategies with this age group. In addition to these concerns, the class
teacher in my experiment had previously focused on participation and engagement in science class rather than exam outcomes; this historical effect may have been at odds with the focus of this experiment on formal learning objectives. With these insights, I moved the focus of our subsequent experiments to direct attention and focus, as well as controlling the time devoted to drawing more rigorously in subsequent experiments.

The experiment in Chapter 3 was the result of two converging events in museum studies. The first is Henkel's work on the use of cameras in museums and their effect on memory: her work suggests that using cameras as 'proxy' memory devices has an adverse effect on recalled information. This effect can be mitigated by increasing attention devoted to taking pictures (focusing, details, camera settings manipulation, etc.). The second source of inspiration for this experiment was the recent attempts by major galleries and museums to encourage people to spend time with their collected works through formal and informally welcoming drawing in their galleries.

Taking a different approach from studies of what people draw and the contents of those artworks, I began to consider drawing outside of formal learning environments and what possible cognitive benefits it may hold. We know that drawing after an event can increase the amount and accuracy of reported information with a variety of populations. However, what about drawing at the time of an event - what memory benefits might this activity have? The possible benefits of drawing on memory in a non-formal learning context are not well understood. With this in mind, in Chapter 3, I examined the relations between drawing and recall of items in a museum gallery, a setting which is used routinely as a venue for artists to practice. I examined the recall of objects and details about them with three distinct groups: adult artists, intermediate-school-aged children, and adults who considered
themselves to have no artistic ability. Participants were asked to draw about half of
the objects on a tour of a museum gallery, and merely observe the other half of the
objects. I began by testing adult artists who were comfortable with drawing; I did not
assess drawing ability as the results of Chapter 2 highlighted the ambiguous nature of
this practice. In addition, I was conscious that others might misinterpret drawings
and the drawings themselves were for the participant to observe and record what he
or she saw. In particular, the work by Gross and Hayne (1999) highlights that others
can poorly understand abstract or very minimal drawings by even young children. I
briefed artists before starting their tour of the museum that their drawings were for
their reference and use and would not be subject to interpretation in the experiment.
I found that adult artists recalled more of the objects that they sketched as well
as specific details about them when compared to their recall of objects that they had
merely observed. Of course, artists may have extensive experience in studying the
form of objects and translating them into 2D sketches and all of the artists in this
experiment rated themselves as above average in drawing ability. This led me to the
question of whether drawing may also be of use to people who do not consider
themselves artistic. As discussed in Chapter 1, drawing skill is generally viewed as the
result of individual skill or innate talent rather than simply as a matter of practice
(Davis, 1997; Gardner, 1980, 1990; Kindler & Darras, 1997; Lowenfeld, 1947;
Rosenblatt & Winner, 1998). Many people abruptly stop increasing the accuracy and
detail they include in their drawings between 11 and 14 years of age. This is due in
part to a lack of practice, as well as self-imposed limitations: without practice, their
drawing ability will remain at the level of a pre-teen child or deteriorate even further
throughout later life (Kindler & Darras, 1997). The adults who participated in the
second experiment in Chapter 3 all rated their drawing ability below average,
conforming to the general trend in stalled artistic development in adulthood. I made
a conscious decision not to ask the children who participated in this experiment to rate their drawing ability as they were at the stage in their development where their drawing could be abandoned completely, and I did not want to contribute to this effect in any way. I also re-iterated to all participants that drawing skill was not being assessed in this experiment and that the drawings were for the artists themselves to help them remember the objects.

I found that adults who self-reported that they were not artistically talented showed no differences in recall for the objects that they sketched versus objects that they merely observed. Intermediate-level children aged between 8 and 12 years, on the other hand, recalled more objects that they had drawn on the tour as well as more details about them compared to the objects they observed. However, for adults, their own self-imposed beliefs around their drawing ability may have affected their performance on the task. Adults may have been overly concerned about the accuracy of the drawings, causing them to look more at their emerging drawing and then attempting to correct or steer it towards a ‘realistic’ drawing, whereas children may have been less concerned about the accuracy of their drawings. As I mentioned earlier, drawing from life involves careful attention paid to the object and the translation of that visual information into a series of strokes in a drawing. When attention is focused on the drawing rather than the object, which may have occurred with less confident artists, less time might have been devoted to the object and more time attempting to produce a schematic drawing. This may account for the effects noted here with non-artists.

Taken together, the results of Chapters 2 and 3 builds on what we already know about drawing and its uses. That is, I found further evidence to support the use of drawing as a memory aid for some groups in informal learning settings, as well as highlighting the benefits of encouraging practice and instruction in drawing for
children between the ages of 11 and 13, demonstrated by the finding that drawing skill improved dramatically with only minimal instruction in this age group. With minimal drawing instruction and support, drawing could be used and developed throughout people's lifespan.

The studies completed during Chapter 2 and 3, as well as the unpublished studies which formed part of my learning journey throughout this thesis, were synthesised into the final study, designed with my findings from the opening chapters in mind. For example, in the Minecraft study described in Chapter 4, I focused on formal (school) and informal (museum) learning spaces, combining the findings of my studies of drawing in the galleries with Imuta et al.'s (2013) study into the learning outcomes from school tours compared with the same information presented within the classroom. In my final study, I also focused my efforts on contemporary children’s interests, specifically computer games. Rather than measure engagement, I focused on changes in measurable learning outcomes which could be compared across groups. The focus on one particular computer game in Chapter 4 was the result of my experiences in working with children over the course of this thesis study. Through the course of my drawing studies, I was increasingly exposed to the inner lives of the children who shared what their passions and interests were during my work with them. While many seemed to enjoy drawing, their talk of contemporary pursuits, Minecraft in particular, forced me to think of learning strategies to use the tools they prefer - namely, computer games.

Learning from Chapter 2 and 3's studies, I focused my research on how contemporary digital technology can be employed as a tool for learning, just as drawing may be used in some contexts to facilitate memory. Humans have always developed and adapted tools to enhance their environment; both the relevance and the importance of these tools are based on current needs. Every few decades or
centuries, a new set of skills and adaptive tools become crucial for interacting with the world around us. As highlighted in the Dutch Schools Inspectorate (2015) report, there is a noted decline in intermediate-level children’s drawing skills, which appears to be related to a rise in other activities such as digital gaming. The parallels between drawing and games are clear: both are considered a pastime, or an adjunct to formal learning; and both are seen as an activity for children, a form of play, and should be abandoned over time in favour of other, more ‘serious’ pursuits (Cox, 2005; Piaget, 1962; Rubin, 1977).

We do not know why early man painted pictures on the walls of caves - was it merely a work of art produced to pass the time in a darkened cave, a communication tool, or was some element of playful behaviour at work? Play is how children acquire essential skills to interact with and understand their world, as well as being an essential element in developing social and cognitive capabilities (Pellegrini, 2009). For example, drawings are often completed as a play activity by young children in educational settings (Coates & Coates, 2006; Kukkonen & Chang-Kredl, 2018; Wood & Hall, 2011). Play is changing - or rather the tools of play are changing - and of course parents, caregivers, and educators want to facilitate this.

Digital expansion into schools has seen a range of initiatives incorporated into teaching and learning. However, the impact that technology has on learning in schools is still a hotly contested topic (Georgina & Olson, 2008; Selwyn, 2007). Using digital tools is fast becoming a normalised part of childhood. For example, 12- to 15-year-olds are spending more time gaming in their free time in 2017 (12.2 hours per week) compared with in 2013 (10.7 hours per week) (Connolly, Boyle, Hainey, & Boyle, 2010; Statista, 2018). Traditional games have been adapted into digital formats, and handheld or desktop computer gaming is now part of most children’s lives, with 97% of young people in the United States playing for at least one hour per day.
All games vary regarding rules, goals, player skills, and tasks to complete, and most contain elements of challenge and risk that the player must overcome. These elements can be combined and modified into an array of experiences which can be classified as games, which may be played on a variety of platforms, either solo or with others (Baranauskas, Neto, & Borges, 1999; Gredler, 1996). While traditional physical games are not a source of controversy, the same elements delivered through a digital platform often are. For example, children playing Capture the Flag in the schoolyard is not controversial, but ‘Capture the Flag matches’ in computer game format are often listed as contributors to school shootings in the United States (Ferguson, 2007, 2013).

Digital gaming continues to be a source of controversy in school environments for many reasons. Since the computer game Logo blended math and programming in the 1970's, video gaming technology has been utilised for educational purposes. This has not been a straightforward process however, initial studies into video game technology were focused on its adverse effects. Similarly to the implications that drawing and art are the expressions of inner workings of the mind and that mental illness and artistic ability are related, the association between mental health problems, aggression, and computer games persists (Toker & Baturay, 2016). Researchers have shown that excessive playing of computer games leads to physical problems, such as joint and muscle problems, auditory hallucinations, neck pain, and hand-arm vibration syndrome. Other correlative data suggests associations between excessive gaming and comorbid conditions such as hypomania, dysthymia, obsessive compulsive disorder, ADHD, borderline personality disorder, as well as low self-esteem and impulse dysregulation (Caplan, 2007; Kwon, Chung, & Lee, 2009; Niemz, Griffiths, & Banyard, 2005; te Wildt, Putzig, Zedler, & Ohlmeier, 2007; Weinstein & Lejoyeux, 2010). While the Diagnostic and Statistical Manual of
Mental Disorders (DSM-V), conclude that there is insufficient evidence to include computer gaming addiction as a psychiatric disorder, the implication that computer gaming may be a risky behaviour has persisted (American Psychiatric Association, 2014, Pies; 2009; Shaffer, Halverson, Squire, & Gee, 2005; Spekman, Konijn, Roelofsma, & Griffiths, 2013).

With concerns about health and behaviour effects, it is understandable that there has been a reluctance by many educators to adopt gaming technology in the classroom, and teacher attitudes have been identified as a major factor in the lack of adoption of digital technology in the classroom (Ertmer, 2005; Hew & Brush, 2007). The ability of games to add enjoyment to previously dull tasks, as well as being an obvious source of enjoyment and an engaging tool which could focus attention, has been noted in a range of studies (Dias, 2017; Kim, 2012; Sailer, Hense, Mayr, & Mandl, 2017). However, the use of computers in formal learning environments has changed the dynamics of teacher and pupil patterns and procedures: the learner now has more freedom to plan and complete their work independently, but also more freedom to be distracted or diverted from the task at hand (Cuban, 2009; Long & Ehrmann, 1995; Yang, Huang, & Li, 2013).

Despite these concerns, computer games are making their presence felt in classrooms - the use of technology in classrooms has increased exponentially over the last 30 to 40 years (Mouza & Lavigne, 2013). Despite this, much of the research on computer games focuses on fun and engagement, or the potential negative aspects associated with increased screen time, rather than asking the question of whether computer games facilitate learning outcomes (Aldritch, 2005; Prensky, 2007). Engagement is not equivalent to learning, and it is worth noting that many experiments using gaming technology in educational contexts measure engagement or enjoyment rather than learning outcomes. Connolly, Boyle, and Hainey's (2012)
review of game-based learning research confirmed that published studies on gaming displayed multiple theoretical models and methodological approaches without measuring learning outcomes specifically. Despite the growing use of game technology in the classroom, little empirical work has included measures of memory for in-game facts and events.

In Chapter 4, I investigated a computer game, *Minecraft*, which is primarily an open world game where players can interact with a range of challenges or create a range of ‘worlds’ themselves or with others for pleasure. Minecraft can be used as a creative tool as it encourages players to build, modify, and create whatever they wish using various materials all presented as blocks of the same size. *Minecraft* has been lauded as a learning tool by some, but it was not originally designed to be a learning interface - the open world style and an infinite number of ‘ways’ of playing this game have facilitated its use in a range of learning contexts. Communities have emerged online who create together massive, immersive worlds using *Minecraft* as a venue to tackle projects that would overwhelm individuals. For example, groups working on shared servers have created scale models of cities, the Seven Wonders of the Ancient World, Hogwarts, and the Taj Mahal, which other players can then add to, admire or explore. Other additions to the game have included in-game chemistry, geometry, programming and physics modifications that allow players to engage with elements of these topics in-game. With such a vast variety of ways to play and topics to explore in this one game, I considered the potential learning advantages of just one way of playing with *Minecraft*. If children could explore *Minecraft*-created worlds and learn in the process, this would build on prior research in our laboratory which has examined the impact of in-school versus on-site tours of informal learning centres (Imuta et al., 2013).
As highlighted in Chapter 3, there are differences between the number of facts recalled after off-site learning experiences and in-class experiences: I explored using *Minecraft* as a proxy for touring a physical site to overcome this. During the *Minecraft* study, I investigated whether experiencing a school tour virtually, through a computer game, could affect learning compared with traditional school tours. I found that children who experienced the real-world tours and completed follow-up tasks in *Minecraft* benefitted the most regarding the recall of facts presented on the tour compared with others who did not complete follow-up tasks in this way. Including questions that assessed participants’ perceptions of their learning experience was an important differentiator between our and other investigations into game-based learning, where often only engagement and perceived learning are measured rather than recall of facts which form part of the essential information that the tour contains. My results showed that there were varying perceptions of learning, quality, and engagement between groups, which had little predictive ability when compared with actual learning outcomes with the highest performing group - those who toured the site physically with a guide and followed up by completing the worksheet activity in *Minecraft*. As highlighted, this group completed their learning evaluation before boarding the bus back to school, which may have dampened their enthusiasm for filling in the survey. The evaluation results from the group who completed the on-site tour and explored in *Minecraft* at the museum was in contrast to the first group who completed a tour and were allowed a half hour on site to do as they pleased. Interestingly, this group of students who only completed a tour site evaluated their learning and engagement the highest of all groups but showed the least recall in follow-up testing. It seems as if students’ perceptions of their learning outcomes, much like their adult counterparts (Bjork & Lansing, 2015), are of limited use when they attempt to evaluate their own learning. The primary focus of the
Minecraft investigation was to evaluate the learning uses of touring maps in Minecraft, albeit in a limited, ‘access and find’ way. As noted, my study did not allow students the ability to modify, build, or create anything new on the map of the museum that they explored. In a similar way to initial research into the use of pictures to improve learning, in the Minecraft experiment, I controlled the Minecraft world, and directed the participants’ attention. Additional benefits to playing in these virtual worlds could be investigated, allowing children to fully immerse themselves in the game to build and create. Given the infinite uses of Minecraft, there are infinite ways to explore the learning outcomes and the results of Chapter 4 will contribute to the growing body of work exploring child computer interaction and learning. The game itself lends itself to far more advanced classroom uses: with its building, programming, and adaptive uses in a controlled environment, it offers science an affordable platform to study diverse topics from group dynamics to spatial learning. These more creative aspects of the game encouraged me to investigate it alongside my research into drawing in order to increase empirical knowledge about how people may exploit the learning possibilities of these tools.

Taken together, the findings of this thesis demonstrate that drawing and Minecraft are easily accessed tools which may offer learning advantages in the classroom and beyond. I have contributed to the growing body of research which challenges the use of children’s drawings as diagnostic tools and begun new lines of research into tools which are being used in educational settings, namely recreational drawing in museum galleries and a building game, Minecraft. Both these foci were the result of exploration and questions that emerged throughout the course of the research. The results from these studies of tool use show that measurable differences in learning can be achieved with practices usually considered pastimes.
References


preschematic, schematic, and gang age stages. American Journal of Art Therapy,
40(3), 170-192.


Prentice Hall.

Anderson, D., & Lucas, K.B. (1997). The effectiveness of orienting students to the
physical features of a science museum prior to visitation, Research in Science
Education, 27(4), 485-495.

museum experiences: Identifying powerful mediators of learning. Curator, 40,

doi:10.1002/acp.1561

McGraw-Hill Education.

doi:10.1016/j.ijcci.2012.08.001

young children’s human figure drawings and their association with intelligence
doi:10.1177/0956797614540686


doi:10.3102/0028312021002435


doi:10.1177/0143034389103006


doi:10.1080/0300443830120103


doi:10.1002/1097-4679(194804)


doi:10.1080/0144341042000211670

doi:10.1002/15206807(197701)


Why is tool use important in the study of human evolution? In A. E. Russon & D. R. Begun (Eds.), Evolutionary origins of great ape intelligence. Cambridge, UK: Cambridge University Press. doi:10.2277/0521783356


English, M., Kaplan, G., & Rogers, L. J. (2014). Is painting by elephants in zoos as enriching as we are led to believe? *Peer J*, 2, 468-471. doi:10.7717/peerj.471


Horvath, G., Farkas, E., Boncz, I., Blaho, M., & Kriska, G. (2012). Cavemen were better at depicting quadruped walking than modern artists: Erroneous walking illustrations in the fine arts from prehistory to today. *PLoS ONE, 7*, 1–10. doi:10.1371/journal.pone.0049786


Levy, S. (1950). Figure drawing as a projective test. In L. Abt & L. Bellak (Eds.), *Projective psychology clinical approaches to the total personality*. (pp. 257-297). New York, NY: Grove.


http://dx.doi.org/10.1037/h0088273


167


Appendices

Appendix A

Brief Summaries of a Selection of Past Theorists that have Considered Children’s Drawing Development as Stage-Like

<table>
<thead>
<tr>
<th>THEORIST</th>
<th>AGE</th>
<th>FOCUS /RESEARCH</th>
<th>THEORETICAL OUTCOMES AND MAIN FINDINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1896</td>
<td></td>
<td>Lukens analysed 3400 drawings produced by children, collected from parents. Lukens presented series of drawings done by the same children over the course of many years to illustrate their linear development.</td>
<td>No stages:</td>
</tr>
<tr>
<td>Herman Lukens</td>
<td>2</td>
<td>No stages:</td>
<td>The ontogeny of human figure drawings</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>No stages:</td>
<td>Drawings categorised by frequency of representation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No stages:</td>
<td>Noted realism emerging and a preference for drawing observed objects by 9 years of age.</td>
</tr>
<tr>
<td>1912</td>
<td></td>
<td>Ballard, a school’s inspector in inner city London, collected 20,000 drawings by children. Children completed drawings in classrooms which were examined to establish preferences and trends. Human figures were the most popular choice to draw in until age 6 but decreased after until the age of 9.</td>
<td>No stages:</td>
</tr>
<tr>
<td>P.B. Ballard</td>
<td>3</td>
<td>No stages:</td>
<td>Drawings categorised by frequency of representation</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>No stages:</td>
<td>Noted realism emerging and a preference for drawing observed objects by 9 years of age.</td>
</tr>
<tr>
<td>1921</td>
<td></td>
<td>Burt included a section on drawing in his publication ‘Mental and Scholastic Tests’. This chapter was based upon observations generated from artworks collected while he was writing.</td>
<td>7 stages:</td>
</tr>
<tr>
<td>Cyril Burt</td>
<td>3</td>
<td>No stages:</td>
<td>Stage of Scribble</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>No stages:</td>
<td>Stage of Line (by 4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>No stages:</td>
<td>Period of Descriptive Symbolism (5-6 years)</td>
</tr>
<tr>
<td>Year</td>
<td>Author(s)</td>
<td>Methodology</td>
<td>Stages/Phases</td>
</tr>
<tr>
<td>------</td>
<td>-----------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
</tbody>
</table>
| 1927 | Georges-Henry Lucquet | Developed his stage theory based upon a systematic study of the progression of drawings completed by his daughter. | 4 Stages:  
Scribbling stage (2-4 years)  
- fortuitous realism  
Preschematic stage (4-7 years)  
- failed realism  
- intellectual realism  
Schematic Stage (7-9 years)  
Visual realism (9-10+ years) |
| 1927 | Florence Goodenough | Studied the development of drawing and writing skills in children. She collected children’s Human Figure Drawings at different ages in order to identify common themes and classify the development of this skill into stages. Proposed a link between intelligence and human figure drawings. | 2 Stages of Human Figure Drawing:  
Preliminary stage of drawing (scribbling)  
- Unrecognisable as a human figure  
Human figure drawing (assessed under the following headings)  
- Gross Detail  
- Attachments  
- Head Detail  
- Clothing  
- Hand Detail |
Lowenfeld's theory of drawing development consisted of six stages that linked children's age with characteristics of their artwork. Lowenfeld collecting and studying thousands of children’s drawings (Smith, 1987) in Austria and the United States. However, no exact numbers for children studied are available.

6 Stages:

1. Scribble Stage (2 to 4 years)
   - Consisting of four sub-stages.
     - Disordered - uncontrolled markings that could be bold or light depending upon the personality of the child. At this age the child has little or no control over motor activity.
     - Longitudinal - controlled repetitions of motions.
     - Circular - controlled motions to produce more complex forms.
     - Naming - the child tells stories about the scribble.

2. Preschematic Stage (4 to 6 years)

3. Schematic Stage (7 to 9 years)

4. Dawning realism Stage (9 to 11 years)
   - The artist becomes extremely self-critical.

5 Pseudorealistic Stage/Pseudo-Naturalistic Stage (11 to 13 years)
   - The finished product becomes most important factor in drawing.
     - Success is determined by the level of realism achieved in the drawing.
     - Frustration is a common occurrence.

6. Decision Stage (13-16 years old)
   - The artist at this stage will decide to continue drawing or view it as an activity without merit.
<table>
<thead>
<tr>
<th>Year</th>
<th>Author(s)</th>
<th>Age Range</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1943</td>
<td>Herbert Read</td>
<td>2-adult</td>
<td>Collected and collated the drawings of a range of children (undetailed) in the early 20th century. The artworks were classified according to a range of common themes and abilities. No details of exact numbers or studies included in work. Read considered child art a creative output and contemplated the educational benefits of art integration in pedagogy. Established the International Society of Education through Art under UNESCO.</td>
</tr>
<tr>
<td>1953</td>
<td>Daniel Marcus Mendelowitz</td>
<td>2-adult</td>
<td>Drawings by children collected, compared and contrasted, and grouped according to common themes emerging in the art at each age. No details of exact numbers or studies included in work.</td>
</tr>
</tbody>
</table>

**12 classifications for children’s drawings:**
- Theorized child art as falling under the following classifications, children may swap between which:
  - Organic
  - Lyrical
  - Impressionist
  - Rhythmical Pattern
  - Structural Form
  - Geometric
  - Habitic
  - Expressionist
  - Enumerative
  - Decorative
  - Romantic
  - Literacy

**6 stages:**
- The Scribbler Stage (2-4)
- Mark Making or Scribbling Stage. (4-6 years)
- Human Figure's Drawing Stage. (6-8 years)
- Geometric Drawing Stage. (9-12 years).
- Story Drawing/Graphic Narrative Stage. (12+)
Piaget observed the cognitive and artistic development of children, as well as his own three children. Piaget and Inhelder did not focus on drawing development directly, but drawing is part of children's over-all cognitive development which Piagetian theory can account for.

**1955**

Jean Piaget & Barbel Inhelder

2-13

4 stages:

1. **Sensorimotor (up to 2)**
2. **Preoperational (2-7)**
3. **Concrete operations (7-11)**
4. **Formal Operations (11-15)**

Moreover each stage is mediated by the following processes:

- Equilibrium
- Structure
- Scheme

1967

Richard A. Salome

2 -17

Salome, an art education academic, collected the works of thousands of children in formal and out of school learning environments. No details of exact numbers or studies included in work.

5 stages:

1. **Scribble Stage (2-4)**
   - Not drawing symbols for objects.
   - Random scribbling
   - Controlled scribbling
   - Naming scribbles
2. **Pre-Schematic Stage (3-7)**
   - Begins when child begins making representative symbols for objects in
the environment

- Formed with circles, squares, & lines
- Symbols change frequently

3. Schematic Stage  (6-11)

- Representation of symbols for familiar objects, and the use of base line.
- Schematic representations used repeatedly
- Figures are flat and stiff
- Multiple baselines as organisational device

Top view, side view, raised base lines, x-ray or transparent view

4. Transitional Stage  (9-12)

- Attempt to produce works that meet adult standards
- Produces works that contain schematic characteristics.
- Baseline replaced with receding ground plane
- Intentional use of overlapping
- Attention to details, gender roles, and clothing differences
- Linear perspective

5. Realism Stage  (12-17)

- Child produces art work in the manner of adult artists.
- Control over medium, content, organisation
- Figures naturalistic in appearance or intentionally stylised
- Consistently use many organisational devices:
  - Overlapping
  - Diminished size
  - Placement on picture plane
  - Linear and Aerial perspective

| 1967  | 2 - 4     | Kellogg analysed 100,000 drawings 300 (2 to 4-year-old children. She considered the scribbles produced by children to be the building blocks of drawing which could subsequently be assembled to produce more complex forms and representations. |

1 Stage:

20 types of scribbles were categorised under the following headings
  - The cross
  - The square
  - The circle (and oval)
  - The triangle
  - Areas of unique forms (odd form)
  - The diametrical cross
## Appendix B

*Brief Summaries of a Selection of Perceptual and Production focused Theorists that have Investigated Children’s Drawing Development*

<table>
<thead>
<tr>
<th>Theorist</th>
<th>AGE</th>
<th>THEORY/FOCUS</th>
<th>THEORETICAL OUTCOMES AND MAIN FINDINGS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rudolf Arnheim</td>
<td>2-adult</td>
<td>Perceptual Theory</td>
<td>Visual image making and Gestalt principles could be defined under the following headings</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Arnheim (1969) argued that drawing development involves a distinct symbol system that is guided by its own graphic logic and cannot be considered as a mirror of a child’s cognitive abilities. He suggests that children draw what they see (or perceive) and at first do not perceive objects as the sum of observed parts. Rather, children see wholes or total images. For Arnheim perception is learned, and can be improved, through training in visual discrimination and this improvement in perceptual abilities can lead to an improvement in drawing ability.</td>
<td>Balance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>June Mcfee</td>
<td>2-adult</td>
<td>Perceptual Delineation theory</td>
<td>3 Factors which impact the child’s ability to draw.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>McFee (1970) suggests children draw as they do, not because of any one factor, but several which interact to produce drawing effects.</td>
<td>1: Readiness</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Perceptual Ability

#### Culture

---

#### 2: Psychological Environment

- Threats
- Support
- Rewards
- Punishments

---

#### 3: Information Processing

- Organisation
- Categorisation
- Ability to plan and create with materials

---

<table>
<thead>
<tr>
<th>Year</th>
<th>Age</th>
<th>Study</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1972</td>
<td>2-</td>
<td>Art Hochburg</td>
<td>Canonical Viewpoints&lt;br&gt;Children hold a mental image of objects from a canonical viewpoint with characteristic features. In drawings children utilise these views to produce drawings.</td>
</tr>
<tr>
<td>1980</td>
<td>2-</td>
<td>Norman Freeman</td>
<td>Cue-dependency Model&lt;br&gt;Freeman argues that children’s perception and cognitive ability is not reflected in their drawings; instead production deficits and biases contribute to children’s early drawings. Children’s drawings develop from an object centred to view centred perspectives.</td>
</tr>
</tbody>
</table>

---

- Canonical representation
- Characteristic features
- Task demands continuously change throughout the drawing process
- Canonical views convey basic structural information
- Preference for viewpoints (side-on for cars and animals/frontal view for people)
<table>
<thead>
<tr>
<th>Year</th>
<th>Theorist</th>
<th>Theoretical/Factual Aspect</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1982</td>
<td>David Marr</td>
<td>Information processing theory</td>
<td>Marr proposed a computational theory for understanding information processing systems, primarily of visual perception. Marr’s contribution is to the understanding of the breakdown of visual information and its increased detailing during analysis of an object for reproduction.</td>
</tr>
<tr>
<td>1982</td>
<td>Howard Gardner and Ellen Winner</td>
<td>U-Shaped Curve</td>
<td>Artistic development described in terms of a U-shaped curve. Instead of linear stage-like development, Gardner and Winner suggest that young children’s and adult artists’ work shares common features. Empirically supported by Davis (1991), Pariser &amp; van den Berg (1997), and Kindler (2000) who compared the judgment of art professionals of children’s and adults works and found them to be judged equal aesthetically depending on cultural background of the judges.</td>
</tr>
<tr>
<td>1982</td>
<td>Maureen Cox</td>
<td>An extension of Freeman’s theory</td>
<td>Cox sees children’s drawings vary over time depending on mood, intention, and cognitive development. These factors in turn affect performance on drawing tasks.</td>
</tr>
</tbody>
</table>

- Input received from senses (visual)
- Intensities perceived
- Primal sketch of the object – edges, contrast, groups, curves
- $2^{1/2}$ D Sketch – surfaces, orientations, depth
- 3D Model representation
- Aesthetic features of drawings similar at different developmental stages for children and artists:
  - Authenticity
  - Directness
  - Formal inventiveness
  - Expressive force

- Internal mental model of objects mediated by:
  - Immediate perception
  - Prior knowledge (schema)
  - The drawing as it emerges on the page
<table>
<thead>
<tr>
<th>Year</th>
<th>Age</th>
<th>Author(s)</th>
<th>Main Idea</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>2-18</td>
<td>Van Sommers</td>
<td>Drawing is the result of two hierarchical systems: Marr’s model of visual perception and a graphic production system. The latter comprises four hierarchically organized components: depiction decisions, production strategy, contingent planning, and articulatory and economic constraints.</td>
</tr>
<tr>
<td>1992</td>
<td>4-18</td>
<td>Dziuraweic and Jeregowski</td>
<td>Key contours of the object will be incorporated into drawings regardless of viewpoint. Object-centred representation – Contour Mapping. Object surface or canonical view is described as a typical outline.</td>
</tr>
<tr>
<td>1995</td>
<td>2-13</td>
<td>Al Hurwitz and Michael Day</td>
<td>Development from scribbling to manipulation includes a movement towards greater control in each stage. 3 Stages: The Manipulative Stage (Ages 2-5), Linear and Circular patterns, Greater control and choice.</td>
</tr>
</tbody>
</table>
Shape making and manipulation
Naming of a Scribble – Visualisation
Identification of real world objects in scribbles
Identification of the intent to draw an object

The Symbol-Making Stage (Ages 6-9)
Connection between the image drawn and an idea.
Drawing what they know rather than what they see
Details emerging in their work
Symbols joined to form complex representations
Placement of symbols on the page improves.
Natural color choices
X-Ray views of objects within other objects
Fold over representations running off the page

The Preadolescent Stage (Ages 10-13)
Caution and self criticism emerge in child's art work
Interest in detail, perspective, subtle use of color, and new techniques
<table>
<thead>
<tr>
<th>Name</th>
<th>Year</th>
<th>Context</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constance Milbraith &amp; John Willats</td>
<td></td>
<td>Argue that individual children differ in their ability to draw realistically, and these differences can be seen in very early childhood, prior to any kind of drawing instruction. Children who are gifted in realistic drawing and who are able to create life-like representations create drawings that differ in many respects from the drawings of typical children. Differences in drawing ability can be assessed under the following headings: Graphic Representation not action representation</td>
<td>Key findings: Individual graphic development is exploratory Multiple strategies can be utilized to acquire graphic symbolism Distinct approaches to drawing from age 5 onward Copying is a common drawing strategy Increase in pictorial repertoires from experimentation Drawing development is not a linear growth path Visual art and graphic communication are different outcomes of drawing.</td>
</tr>
<tr>
<td>Anna Kindler</td>
<td>1998</td>
<td>Suggests that the multiple uses of drawing during development mean that it is both stage-like and that at each point multiple graphic repertoires may be utilized to communicate with drawings. Graphic development involves a range of exploratory pathways in the production of images. Children can begin to decline in their artistic output from the age of 6 due to frustrations with their lack of drawing ability. Drawing is not the product of inborn creative ability exclusively, but is socially and culturally mediated.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Golomb argues that children’s cognitive and perceptual abilities are not reflected in their artwork. Adult interpretations, she argues, ascribe value to children’s artistic output which does not reflect the unity, balance, expressiveness and intention of children. Children strive towards improvements in their drawing and work to produce more complex drawings. Drawing is not necessarily an easy task for children, however they accept these struggles in drawing production as challenges.

<table>
<thead>
<tr>
<th>Key findings:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task demands effect drawing output, even in scribble stage</td>
</tr>
<tr>
<td>Preferences for complex/differentiated drawings.</td>
</tr>
<tr>
<td>Vocabulary and shape differentiation</td>
</tr>
<tr>
<td>Drawings do not reflect emotional states accurately.</td>
</tr>
<tr>
<td>Drawing content not related to life circumstances</td>
</tr>
<tr>
<td>Personal affect-related content more expressive</td>
</tr>
</tbody>
</table>
Appendix C

Information contained in onsite and virtual tours of the industrial museum.

The Event

1. The Chimney

The first stop here is the oldest surviving structure in the Gasworks museum. In the late 19th century, around 1895, this chimney was constructed - does anyone know why it was built so high? It was built high to make sure the surrounding area was not covered in coal smoke. Despite this, the Gasworks used to stink; the smell of coal, gases, and other products used to cause complaints from some corners of the South City - these high chimneys were a way to disperse some of the coal and soot. Burning fossil fuels is a way of producing heat, but it is also damaging to the environment. The oldest structure in the Museum is the chimney the adjoining buildings on the site were constructed in or after 1900.

2. The Boiler Room

If you want to turn on a light switch you have instant power! Back in the 19th century, you had to light a candle and try and complete work or read books at night in low light conditions. If you walked the streets outside at night they were very dark, and if you wanted to cook you had to burn a fire in a range to heat and cook your food. This is the first step towards the production of gas to use in street lights or home lighting. Now we are at the next important place in the production of power - we've seen the massive chimney outside, now this is the boiler in which coal was burnt to heat the coal in the chambers above. The Gasworks carried out 4 main functions: Gas Manufacture, Gas treatment, Gas pumping and Gas storage. Gas is created when coal is heated in a sealed vessel. Coal gas is created when coal is heated in the absence of air in an enclosed chamber. The coal is heated to about 400 degrees Celsius where it softens and coalesces (comes together to form one mass or whole). The temperature is then raised to 1000 degrees Celsius and the remaining volatile matter (hydrogen) is extracted from the coal, leaving coke. Coal gas is made of hydrogen, carbon monoxides and
methane. When it was created here it also contained stuff like tar and ammonia which had to be removed by processing it further in the Gasworks.

3. The Engine House

For the first European settlers, street lighting was almost non-existent, making travelling after dark hazardous unless guided by moonlight. There were few if any lights on public streets. Citizens regularly fell into streams and open sewers or banged into wandering stock and other obstacles. Lighting was limited to a few candle lanterns over the entrances to hotels, which produced only feeble, flickering light. Kerosene lamps were introduced from the 1860s in many cities, often placed beside bridges, drains and other night-time hazards. Though brighter than candle lamps, they were still dim and were not used on many streets. Dunedin, like many cities in the new world, did not have adequate street lighting, and after dark, the city could be a dangerous place. So when Stephen Stamp Hutchinson moved here, he set about solving that problem. The Dunedin Gas Light and coke company was formed in 1862 by Stephen Stamp Hutchinson. Stephen had previously worked in London and Melbourne and came to Dunedin intending to start a gas light company to provide street lights for the city.

The machines in this section are gas pumping machines which pumped the gas into sections of the Gasworks which treated them to remove impurities before being stored in the storage tanks.

4. The Library

Nowadays when we need an answer to a question we simply Google it; not so 100 years ago when these books were a Wikipedia of information about everything that an engineer may encounter during the day's work at the Gasworks. A lot of the skills needed here were learned on the job, book knowledge was mixed with hands-on, on-the-job skills. This is the library which holds all the books that were needed for people who worked in the Gasworks. These books are full of information about how coal can be turned into gas, how gas can be stored, how it can be used as a source of heat, light and power. These books contain all the information on the process of making gas.
5. The Fitting Room

This room is called the fitting room; now it has some computers in it but 100 years ago it had large cylinders which were part of the water gas purification process where they removed impurities from gas before pumping it to storage. In 1903, a leak in the plant ignited and the entire building exploded. Nobody was killed or badly injured but the building was destroyed. It was rebuilt afterwards and lately has been used as a place to teach kids about computers and technology.

6. The Forge

Where would you go if you needed a nut or bolt nowadays? A tool shop or online perhaps. Back when the Gasworks was operating, you made your own nuts and bolts if you needed them. If something broke, you fixed it, and there was no ordering of parts from anywhere - unless you were willing to wait a very long time. When it was a working Gasworks it was necessary to make and repair things as they broke on site; everything from bolts and nails to beams and machine parts were repaired on site. The Gasworks uses a lot of things made from iron. In order to make iron hot enough to melt you have to use coal which has had all of the impurities removed in a furnace to produce coke. Coke burns much hotter than coal but needs to have air pumped into the fire to keep it lit. A blacksmith is a person who works with iron to make iron products; they heat the iron in the fire until it is very hot, then they use a hammer and the anvil to move the iron around into a shape they need. The temperature here is between 1800 degrees Celsius and 2500 Celsius. The blacksmith knows how hot the iron by the colour. Iron will become yellow when it is hot enough to forge. It then cools to orange, red, then back to black. The hotter the iron is the easier it is to work with, which is where the phrase 'strike while the iron is hot' comes from.