Have children’s drawings changed?

A look at children’s human figure drawings and perspective drawing in 2021/2022.

Master’s thesis

Sandra Bencomo
Faculty of Art and Design
University of Lapland

Spring 2022

Lapin yliopisto, taiteiden tiedekunta
Työn nimi: Ovatko lasten piirustukset muuttuneet? Katsaus lasten ihmis- ja perspektiivipiirtämiseen vuonna 2021/2022.
Tekijä: Sandra Bencomo
Työn laji: Pro gradu -tutkielma
Sivumäärä: 115 + 9 (liitteet)
Vuosi: Kevät 2022

Tiivistelmä

Tässä Pro gradu -tutkielmassa tutkin, ovatko lasten piirustukset muuttuneet niistä ajoista, kun taiteellisen kehittymisen vaiheteoriat ensin käsitteellistettiin. Genovese oli tutkinut samanlaista kysymystä aiemmin lasten ihmispiirrostestien kohdalla, ja hän oli havainnut niissä pisteiden nousun. Hän ehdotti, että koska näillä testeillä on yhteys kognitiivisiin ajattelutaitoihin, jos lapset ajattelevat abstraktimmin aiemmin, niin he kenties myös piirtäisivät kuvia perspektiivissä aiemmin.

Tutkin kolmea asiaa tässä Pro gradu -tutkielmassa. Ensin tutkin, oliko ihmispiirrostesteillä yhteyttä perspektiivipiirtämisen. Toiseksi tutkin, olivatko lasten pisteet nousseet ihmispiirrostesteissä. Tätä kysymystä valitsin Harriksen ja Goodenoughin Draw-A-Man-testin (Piirrä mies -testi), joka oli standardisoitu 1950-luvulta saaduilla pisteillä. Vertailin tuloksiani tähän standardisointiin. Kolmanneksi tutkin, piirsivätkö lapset perspektiivissä nuorempana kuin aiemmin, vertaamalla tuloksia Littletonin 1980-luvulla tekemiin perspektiivitesteihin. Vertailin tilastoja kvantitatiivisilla menetelmillä. Lasten iät olivat samat omassa otoksessani ja otoksissa, joihin vertailin. Suurin otokseni sisälsi 5–8-vuotiaita lapsia.

Tulokset eivät tukeneet yhteyttä ihmispiirrostestien ja perspektiivipiirtämisen välillä. Harris-Goodenough Draw-A-Man-testin ja peittävyyttä mittaavan perspektiivitestin välillä oli pieni yhteys. Tämä voi selittyä sillä, että Draw-A-Man-testi mittaa joissakin kohdissa myös peittävyyttä. Siten tämä ei ole vahva todiste siitä, että jokin abstraktimpi piirre, kuten kognitiivinen ajattelutaito, ajaisi tuloksia molemmissa testeissä.

Tulokset Harris-Goodenough Draw-A-Man-testissä eivät olleet nousseet, vaan olivat erittäin samankaltaisia 1950-luvun tuloksien kanssa. Tyttöjen ja poikien tulokset osoittivat kuitenkin suurempaa eriytymistä. Peittävyyttä mittaavien testien välillä löytyi vain keskikokoinen ero 2021/2022 ja 1980-luvun tulosten välillä. Lasten piirustuksissa ei näkynyt mitään piirustustapaa, joka olisi vallannut lasten piirustukset. Tulokset sen sijaan korostavat, että lasten piirustuksissa löytyy monenlaisia ilmaisutapoja.

Avainsanat: Kuvataidekasvatus, lasten piirustukset, peittävyys, viivaperspektiivi, ihmispiirrustus.

Suostun tutkielman luovuttamiseen kirjastossa käytettäväksi.

University of Lapland, Faculty of Art and Design
Title of the thesis: Have children’s drawings changed? A look at childrens human figure drawings and perspective drawing in 2021/2022.
Author: Sandra Bencomo (she/her)
Type of the work: Master’s thesis
Number of pages: 115 + 9 (attachments)
Year: Spring 2022

Abstract

In this master’s thesis, I set out to study whether children’s drawings had changed since the first stage theories of artistic development were conceptualized. Genovese had studied a similar question previously regarding Human Figure Drawing Tests, and he had noticed a rise in points. He suggested that this might be because the tests’ connection to cognitive skills, and that if children master a more abstract form of thinking earlier, then they would also draw in perspective earlier.

I studied three things in this master’s thesis. First, whether there was a relationship between Human Figure Drawing Tests and perspective drawing. The second question was whether children’s scores in Human Figure Drawing Tests had risen. For this question I chose to use the Draw-A-Man test by Goodenough and Harris, which had been standardized with scores obtained in the 1950’s. I compared my results with this standardization. Third, I researched whether children depicted perspective at an earlier age, compared with perspective tests done in the 1980’s by Littleton. I compared results with quantitative methods. The children in my sample and in the samples I compared to were the same age. My larger sample consisted of children aged 5:9 to 8:11 years old.

The results did not support a relationship between Human Figure Drawing Tests and perspective drawing. However, there was a small relationship between the Draw-A-Man Test and a perspective test measuring the use of occlusion. This can be explained by the Draw-A-Man Test measuring occlusion in a few of its items and is therefore not strong evidence of a more abstract feature such as cognitive skill guiding both tests.

Scores in the Harris-Goodenough Draw-A-Man had not risen, but were remarkably similar to those obtained in the 1950’s. However, results between boys and girls had become more polarized. Only a medium difference was found between results in tests measuring use of occlusion as a feature of perspective between 2021/2022 and the 1980’s. There was no one way of drawing that would have overtaken children’s drawings. Rather, results highlight that there is diversity of expression in children’s drawings.

Keywords: Art Education, children’s drawings, occlusion, linear perspective, human figure drawing.

Contents

1. Introduction ... 4

1.1 Direction and concept definition ... 4

1.2 Context within the field of Art Education ... 6

1.3 Research questions ... 6

2.Teoretical background ... 10

2.1 Ways to study drawings ... 10

2.2 Perception and realism ... 12

2.3 Stage theories of artistic maturation ... 14

2.4 Critique of stage theories ... 17

2.5 What children *can* draw ... 19

2.6 What children *prefer* to draw ... 23

3. Approach ... 27

3.1 Quantitative research and its relationship to ontology ... 27

3.2 The basics of statistical tests ... 29

4. Course of the study ... 34

4.1 Order of the tests ... 35

4.2 The Balls test ... 36

4.2.1 Background ... 36

4.2.2 Procedure ... 37

4.2.3 Scoring ... 39

4.2.4 Differences between the Balls test and the original test by Littleton (1991) ... 41

4.3 The Corridor tests ... 41

4.3.1 Background ... 41

4.3.2 Procedure ... 43

4.3.3 The Corridor Selection test ... 44

4.3.4 The Corridor Drawing test ... 45

4.3.5 Scoring ... 46

4.3.6 Differences between the Corridor tests and the original test by Littleton (1991) ... 47

4.4 The Harris Goodenough Draw-A-Man test ... 48

4.4.1 Background ... 48

4.4.2 Procedure ... 49

4.4.3 Scoring ... 51

4.4.4 Consistency of measure in the Harris-Goodenough Draw-A-Man Test: ... 53

4.4.5 Differences between my implementation and the original DAM Test ... 56
4.5 Participants ... 56
4.5.1 Overview ... 56
4.5.2 Full sample ... 57
4.5.3 Partial sample ... 58
5. Results ... 59
5.1 Relationship between results in all three tests ... 59
5.1.1 Statistics ... 59
5.1.2 Interpretation of results ... 76
5.2 The relationship between the DAM Test 1963 standardization and results in 2021/2022 78
5.2.1 Statistics ... 78
5.2.2 Interpretation of results ... 83
5.3 Relationship between Littleton's (1991) results and results in 2021/2022 ... 88
5.3.1 Statistics ... 88
5.3.2 Interpretation ... 96
6. Discussion ... 100
6.1 Implications for Art Education and Art Education research ... 100
6.2 Limitations of the study ... 104
7. Conclusions ... 107
References ... 109
Attachments ... 116

1. Introduction

1.1 Direction and concept definition

I was babysitting a child when I drew with one fast stroke all of the separate fingers on a creature's hand. The child exclaimed: "Whoa!", and in that he summarized how such a small act hides behind it so many complexities. I did not have to count the fingers, or pause my stroke to think, and I could have easily expressed the volume of the creature's hand without putting much thought into it. These are the results of years of growth and practice. I can never go back to being that child who is still learning these little things. Yet, as an Art Education student, I am headed to teach children who do not know the same things I do. I need to be able to understand artmaking from their perspective.

I have already forgotten many of my own learning experiences. My artistic knowledge all blends into one big mass that I find hard to articulate. Fortunately, there are many researchers across history who have studied and pinned down all the little things that people learn as they learn to do various forms of art. Unfortunately, many of these theories of artistic development are very old. There can be even more than a century separating the children studied then, and the children I will teach in the present. Therefore, a relevant question to ask was: *"Can I still rely on the same theories of artistic development that were articulated back then?"*. This formed into my research question: *"Have children's drawings changed?"*.

Art is a wide concept, and different forms of art have been considered in relation to human development (Matthews, 2004, 279). The general interest of psychologists, artists and art educators has been on the broader concept of *artistic development* and its relation to age and maturity. I had to narrow this concept into one that was possible to study in this master's thesis. First, I tried to find out if other researchers had found differences in children's drawings. This led me to a paper by Jeremy E. C. Genovese (2018) in which he had studied children's human figure drawings and observed a change. He related this change to possible changes in perspective drawing. Based off this I chose to replicate a Human Figure Drawing Test by Harris (1963) and two perspective Drawing tests by

Littleton (1991). Therefore, I limit myself only to studying *drawings*, instead of *art* or *artmaking* as a larger concept. By drawings I mean images produced on paper, with pencil, an eraser, and occasionally colored pencils as well.

Children grow, and the art they make changes as they grow. They learn many little things that build on each other and create larger capabilities. It is the aim of this master's thesis to study the pace of this change that occurs in artmaking, and whether its speed has changed in current generations compared to previous ones. This places the drawings I study in the context of *stage theories*.

Stage theories of artmaking outline how children change the way they make art as they grow older and move from one stage to the next. More precisely, this master's thesis will be focusing on the stage theories as first described by Luquet in 1927, and later reiterated by Piaget in the book *The child's conception of space* (Matthews, 2004, 269). These stages are *logical realism*, which is the stage during which the child draws what they know, and *visual realism*, a stage during which the child draws what they see (Salminen & Koskinen, 2005, 43; Cox, 2005, 116). Although I would come to question the validity of these terms and concepts while researching the theoretical background for this master's thesis, they are still a large part of the history of how children's drawings have been studied. They are also what I was taught in my Art Education studies.

Ultimately this master's thesis is driven by my own fascination towards how art is connected to human psychology. I first studied this border area during a university course called Kuvataideopetus I (Art education I). Since I was not able to attend the lectures, I did an exam on a book called *Pääjalkainen (Tadpole drawing)*, by Antero Salminen and compiled by Inkeri Koskinen (2005). This book married many concepts from psychology into the newly forming academic field of Art Education in Finland. It reworked my interest in both art and psychology into a new form that had many interesting paths to study. The topic of this master's thesis in its earliest form was suggested by Professor Timo Jokela of the University of Lapland, in an attempt to steer me toward a more concrete path in this large border area between art and psychology.

1.2 Context within the field of Art Education

It is difficult to think of a child, who has gone through several years of Art Education, and never drawn once. Although many non-depictive forms of artmaking have risen in the field, such as performance-art, depicting something by drawing it is still a meaningful part of Art Education. I have yet to attend or supervise an art class completely devoid of drawing. The better we understand the drawings children make, the better we can support them.

The shift from drawing what one knows to drawing what one sees as a flat plane, brings with it sensitivity and fragility. During the later period, the period of *visual realism*, children are often dissatisfied with what they draw, as it does not look “real” (Salminen, & Koskinen, 2005, 47). The child is prone to face more disappointment than before, and lack of success in creating drawings admired by the society around them may cause the child to slowly stop trying. (Salminen, 2005, 47–49).

Drawings are an innate part of artmaking. The purpose of this master’s thesis is to better understand children’s drawings and how they develop, so that we can provide better Art Education. Both differences and similarities in children’s drawings now and then would provide fruitful starting points for further research. After all, we art educators base our education not only on how well we know each student, but what we know about children and their artmaking overall.

1.3 Research questions

The hypothesis I was testing was modelled after the suggestion made by Genovese (2018, 181) who based his suggestion on a hypothesis he cited from Oesterdiekhoff (2012). First Genovese (2018, 176) found a Flynn-effect in children’s Human Figure Drawing Tests between 1902 and 1968. In these tests children draw the human figure, which is then scored for completeness and detailedness, and this score correlates with other measures of cognitive ability (Genovese, 2018, 177). In other words, subsequent generations were getting higher points for more detailed drawings, and that in turn should reflect better thinking skills at a younger age.

Oesterdiekhoff's (2012, as cited in Genovese, 2018, 180) hypothesis was that increase in formal operational thinking is the reason behind the Flynn effect seen in psychometric and developmental tests that tap into cognitive skills, such as IQ tests. In other words, subsequent generations get better scores in measures of cognitive ability because they master a more abstract form of thinking earlier. Genovese (2018, 181), in turn, suggested that:

"-- if perspective drawing does indeed reflect more abstract reasoning skill, it might be possible to re-examine existing datasets of children's drawings. This could provide a test for Oesterdiekhoff 's (2012) hypothesis. If Oesterdiekhoff is correct we should see children drawing in perspective at younger ages over the course of the 20th century."

Because this is my master's thesis, I wanted to contribute to a dataset rather than just use pre-existing ones. I consider learning how to collect data and set up studies an important part of my education. When conducting my literature review, I started to question the link between abstract reasoning and perspective drawing. The link between thinking, drawing, and appraising drawings is a complicated one. Genovese's (2018, 181) suggestion links perspective drawing and the Human Figure Drawing Tests through the idea that a maturation in thinking skills would reflect in both (Figure 1).

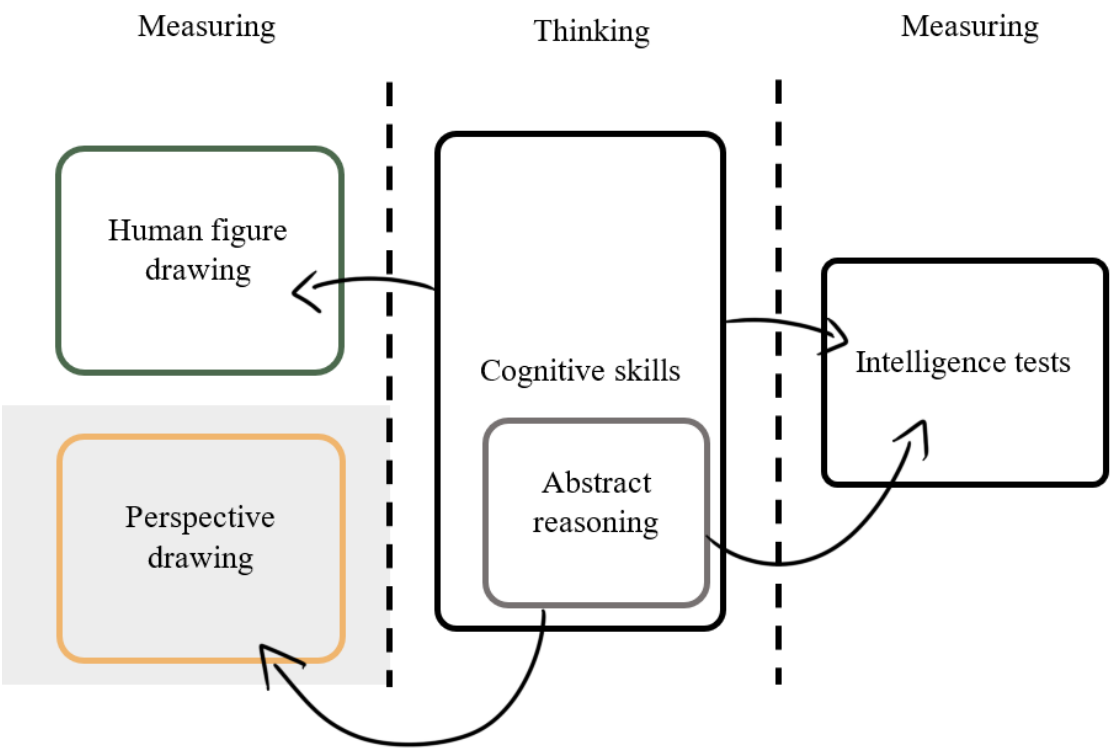


Figure 1: Concepts and their relationships in the hypothesis suggested by Genovese (2018) as interpreted by me.

To address this link between perspective drawing and Human Figure drawings I would need to see if they were measuring the "same thing". I cannot guarantee that this "same thing" is any specific cognitive skill within the scope of this master's thesis. I can only evaluate their synchrony when the same child does multiple of these tests. Therefore, my first research question is:

1. Do children get similar results in perspective tests as they do in a Human Figure Drawing Test?

For the Human Figure Drawing test, I chose the Draw-A-Man Test by Dale Harris and Florence Goodenough (Harris, 1963). This is also one of the tests that Genovese (2018, 177) studied for differences in scores compared to earlier datasets in similar tests. I wanted to know whether my scores followed the same pattern of rising scores that

Genovese (2018, 176) had noticed in his own study. This forms my second research question:

2. Do children in 2021/2022 get higher scores on average in the Harris-Goodenough Draw-A-Man Test than they did before?

Genovese (2018, 181) also proposed the idea that children might draw using perspective at earlier ages. This forms my second research question:

3. Do children in 2021/2022 draw with features of perspective at a younger age?

For the perspective tests I chose two tests by Karen Susan Littleton (1991) that measured both occlusion as a feature of perspective and linear perspective. These are called here the Balls test and the Corridor tests respectively. *Linear perspective* means that a view is depicted as seen from one perspective point and represents three-dimensional depth (Littleton, 1991, 197). It is similar to one-point perspective in the setting I chose to pick from Littleton's Experiments. *Occlusion* means that one object is obscured behind the other, and not fully visible from where the observer is looking (Littleton, 1991, 197).

These were features of drawing that also changed as children became adolescents and adults (Littleton 1991, 28, 157). Littleton's tests were also done in the 80's, according to an e-mail she sent me. Therefore, there was enough of a time gap to consider if this change occurred at a different pace in the current generation compared to the older one. I will detail these perspective tests later in this master's thesis.

With all of these three questions I aimed to answer the following main research question. **Have children's drawings changed?** In order to answer this question in a way that could be analyzed with statistical methods, I gathered test results in the previously mentioned tests from 71 child participants from pre-primary and first grades in Finland. Each child completed all three tests. I then observed the results of each participant in all three tests and compared the results to those of Harris (1963) and Littleton (1991).

2.Teoretical background

2.1 Ways to study drawings

Visual art is sometimes defined quite narrowly in the study of *artistic development*. Kindler (2004, 233) puts it simply and clearly in her review of the literature on the topic: *"People who are not experts in art often see artistic development in rather simplistic and unilinear ways—as an ability to progress from pictorial production that "looks like nothing" to creation of images that "look like something"*. Kindler (2004, 233) contests this idea, saying that pictorial realism and technical skills are not necessary in some manifestations of art, and that realism alone cannot be a sufficient condition for good art. The fact that making visual images does not require drawing exactly what is observed has also been noted by some contemporary psychologists studying the areas between art and psychology like Kozbelt (2017, 101). Perspective too, is something that has been remained irrelevant in some artistic traditions (Kindler, 2004, 234).

Psychologists have typically been interested in art as a window to the child's inner workings. In order to measure changes in the child, art and particularly drawings, have offered a medium where the child's growth is reflected. Psychologists have taken note of denotation systems, figure differentiation, segmentation, contouring, volumetric and surface representation and how the child uses graphic symbols (Kindler, 2004, 233). In other words, the most important thing is what and how is something represented on a page. For example, Jean Piaget (Piaget & Inhelder, 1967), whose writing has been central to a lot of the theory presented here, studied drawings to figure out how the child could think of space. Genovese (2018) whose hypothesis I test in this master's thesis, looked at how numerical results of Human Figure Drawing Tests changed over time. These tests were also conceived within the field of psychology, and they consist of giving a certain number of "points" to a drawing depending for example, if the child has drawn all five fingers on the hand.

Central to this master's thesis is the Human Figure Drawing Test by Harris (1963) and Goodenough, that aimed to discover how the drawing's detailedness could relate to intelligence. Littleton (1991) studied drawings to see how children solved the "problem" of representing space in eleven different Experiments. In this thesis I aim to replicate two of the conditions she used in Experiments One through Four, and Experiment Nine. All these researchers have focused on features seen in drawings, which they have then categorized and calculated to test different theories of human development. This master's thesis also draws from this longstanding tradition of looking at children's drawings from an experimental point of view, where the given "exercise" is the same for every participant in a group, and the features of the drawing are operationalized to treat drawings statistically. I will also treat drawings in terms of three features: occlusion, linear perspective, and detail.

Art Education researchers have been interested in what the child can draw and communicate with the drawing. Acquisition of techniques, complexity of drawings, and the expressiveness of their works have been of interest to art educators from certain traditions (Kindler, 2004, 233). However, art educators have not been satisfied with narrow operational definitions of art and its development (Kindler, 2004, 235). For example, Art Education researcher Salminen is someone whose writings are still taught in Art Education studies. Laukka (2005, 13) tells that Salminen observed how his own son drew growing up. This is a research method not unknown to psychologists and can involve a high degree of operationalization. However, in my observation, Salminen's technique is more human-centered. While Piaget and Inhelder (1967, 182) and Littleton (1991, 60–61,150) also took note of the circumstances where the drawing was created (speech, instructions etc.), Salminen's observations tell a story of a growing person who wants to communicate. Kindler (2004, 237) notes, that the tradition of studying Art Education, has not considered the intent and purpose of drawings. According to Laukka (2005,14) Salminen crticized the idea of universal development and the idea that children's drawings were somehow "incomplete" representations of the world. He did all this while still taking note of various theories on development, creativity, perception, and communication. In other words, where he did not see drawings through the lens of tightly controlled experiments, he saw the holistic picture of human growth. It is the aim of this thesis not to reduce drawings just to features, but also to present them as part of something larger.

2.2 Perception and realism

In Salminen's opinion, Art Education's mission and meanings cannot be understood in depth unless we ponder what it means to perceive (Salminen & Koskinen, 2005, 92). Seeing and creating things to be seen go hand in hand. The theories on perception I will present here are contemporary, but they are not the basis of all the theories on drawing that I will present in later chapters. These older theories might have relied on older theories on perception.

As I have studied perceptual psychology and looked at various forms of art, I consider if our perception of the world is defined by our physiology as active omnivorous mammals. Our idea of realistic images, besides being culturally defined, is limited to what spectrum of light we see, how our eyes are positioned and things we do not see, such as polarization of light. If a bee could draw "realistic" images, they would probably be very different from those of humans. Even among humans we see the world differently. An image that might both look "realistic" to someone who is color vision deficient, might look very "wrong" to someone who is not.

A single drawing, no matter how "realistic", can only be an imitation of what is seen through one unmoving eye. This is terminology I borrow from Salminen, (Salminen & Koskinen, 2005, 104–112) who took the theories of James J. Gibson and concluded that thinking about the environmental perception as flat "pictures" at the back of the eye is a historical yet misleading viewpoint. There are many reasons as to why, but I will go over the two most important ones from my perspective as they are described in Yantis and Abrams (2017) coursebook for Perceptual Psychology. First, the two "pictures" reflected to the backs of our eyes (retina) differ in subtle ways. This is called binocular disparity and one obvious example is when we look at our finger and switch between closing the left or the right eye (Yantis & Abrams, 2017, 209). A real view always involves two "images" if you have two working eyes. Second, retinal "pictures" are not sent into our brain like a digital image is represented on a computer pixel by pixel. Our eye has limited resolution outside of a very small area of high resolution called the Fovea, and we even have a blind spot where we do not "see" at all (Yantis & Abrams, 2017, 56). Different layers of the visual system separate, arrange, and rearrange the most important features and changes that occur in the "picture" reflected to our retina (Yantis & Abrams, 2017, 81–105).

Perception is the sum of changes. Salminen, relying of Gibson's theories, proposes that visual information is in the optic array, which is the constantly changing organization of light in the viewer's eye (Salminen & Koskinen, 2005, 104–112). In other words, what we see is like a shifting mass of dots with different hues and brightness, that we learn to interpret as we move around in the world (Salminen & Koskinen, 2005, 111–112). Bumping into a chair follows a very particular arrangement of these dots and learning to interpret that arrangement makes sure we don't bump into that chair again. Salminen emphasizes, that seeing is active, and that it always involves motion, time, and the viewer in some way (Salminen & Koskinen, 2005, 105). The eye is not a camera (Salminen & Koskinen, 2005, 173) This is also in line with contemporary teaching of perception as seen for example in Yantis & Abrams (2017) coursebook.

Normal perception is usually not just limited to vision but works together as a unit with other senses (Salminen & Koskinen, 2005, 118). The optic array always includes our viewpoint, it's movement and our body in relation to other targets in the array (Gibson, as cited in Salminen & Koskinen, 2005, 112). Proprioception, as in the sense of self-movement, is part of our ability to see. Another good example of the merging of the senses is the ventriloquism effect. An example presented by Yantis and Abrams (2017, 384) is that if you were to hear the sound the conversation between a man and a woman from an earbud in your right ear and had your eyes closed, you would interpret all the sound coming from one location to the right side of your head. However, if you saw a video clip of a man and a woman talking at the same time you listened, you would experience the sounds coming from the speakers' mouths in two different locations (Yantis & Abrams, 2017, 384). In this sense vision works together with the very experience of hearing, and cooperation is not limited to only these senses.

Perception is also personal. Two people looking at the same thing will not perceive it exactly the same. Things such as interest and hobbies influence how we perceive the world around us (Salminen & Koskinen, 2005, 135). As an artist, I might pay more attention to the subtle differences between hues than someone who does not have to mix paint. We also recognize familiar and meaningful shapes easier than unfamiliar ones (Yantis & Abrams, 2005, 122). Lines, dots and surfaces are not as important as the meanings embedded in them (Salminen & Koskinen, 2005, 154).

2.3 Stage theories of artistic maturation

The idea that children's drawings change in a systematic way as they grow up is the very essence of stage theories. In these theories, children go through various stages, marked by different features that appear and disappear in their drawings. In this master's thesis I will only be focusing on two stages, commonly known as logical realism and visual realism. These were first described by Luquet in 1927, which then later hybridized with Piaget's interpretation of the stages in the book *The child's conception of space* (Matthews, 2004, 269). They are also the theoretical basis of much of the research I present in this master's thesis. Cox (2005, 116) summarizes the stages concisely in her review and criticism of the stage theories: *"In its simplest form, the argument goes, as the child attends to what they see rather than what they know, they move, over time, to a more accurate representation based on perceptual reality"*.

The stage during which the child draws what they "know" is called *logical realism* (Luquet n.d., as cited in Salminen & Koskinen, 2005, 43). This stage has also been called *object centered* by some researchers (Matthews, 2004, 270) and "*the joy of telling*" by Salminen (Salminen & Koskinen, 2005, 39). This stage has many features, but I will only go over the most relevant ones for this master's thesis. During the stage of logical realism children understand that symbols can represent things that are not present (Salminen & Koskinen, 2005, 39). A ball can be an eye or the sun. They present things in two dimensions and through multiple perspectives at once so that their most recognizable features are seen (Salminen & Koskinen, 2005, 40–45). I personally remember that when I was a child I drew a blanket as a square, as if it was being looked from above, just so I could draw the pattern on it clearly. This did not match the perspective in the rest of the picture, where people were depicted in profile. Sometimes presenting what's important means making things transparent or fully visible, so that all the important features can be seen (Salminen & Koskinen, 2005, 46; Matthews, 2004, 274). After all, they *know* the features are there, even if they cannot see them at the moment.

The stage during which the child draws what they "see" is called *visual realism*. Drawings made during this stage have also been called *viewer centered*(Matthews, 2004, 270). Where during the stage of logical realism children made pictures that seemed "right enough" for their needs, during the stage of visual realism they wish to make pictures that also "look right" (Willats, as cited in Kindler, 2004, 236; Salminen, 2005, 47). During this stage, the child uses occlusion and covers the further object with the nearer one, if the further one cannot be seen completely (Matthews, 2004, 274). There is no longer transparency to show both objects as whole at the same time. The child also starts to practice features of perspective, such as the horizon line and how things further from the viewer look smaller (Salminen, 2005, 47).

This master's thesis relies on the assumption that there are certain ages where children move from one stage to the other. Literature has given various cut-off ages for this shift. These cut-off ages do not change in a systematic way. At least that has not been my impression when looking at studies on this subject. Older studies do not find the change to happen earlier than newer studies or vice versa. For example, Burt (1921), according to Harris (1963, 18) places this shift between logical and visual realism at around the age of 10 to 11, while later Piaget and Inhelder (Piaget & Inhelder, 1967, 51) found the transition to visual realism to start already by the age of 7–8. However, near the same decade Lark-Horowitz and Norton place the start of a "true-to-appearance" stage at age 9–10 (Lark-Horovitz & Norton, 1959, 434–435). I would argue that the cut-off age depends much on the criteria chosen. The general idea is that the shift to visual realism could be said to begin somewhere around age 7 and then be more clearly visible near the age of 10. From this we can get the vague idea that children below the age of 7 are not typically producing visually realistic drawings. This shift from logical to visual realism is by no means definite, and children will often produce both types of pictures in tandem (Lark-Horovitz & Norton, 1959, 434 – 441; Harris, 1963, 19).

Separate claims have been made for children's understanding of perspective, which in this master's thesis means both drawing with linear perspective (for example drawing train tracks that meet at the horizon) and using occlusion (drawing one object in front of the other without transparency). Many writers tend to refer to linear perspective with the word "perspective", but some refer to both of these features or more. This makes it difficult to differentiate when these features should appear according to the literature. Lark-Horovitz and Norton (1959, 435–441) separate the *True-To-Appearance* stage (4th stage) where the child draws occlusion as starting at age 9–10, from the *Perspective stage* (5th stage) which is achieved by only a small number of children. Harris (1963, 154) lists perspective as one of the few and far between details that allow the Draw-A-Man Test to be extended to adolescents as well. In the Draw-A-Man Test scoring manual he considers both transparency and foreshortening (Harris, 1963, 255–263). Use of perspective is what seems to differentiate children's drawings from adolescent's drawings. This was also the result obtained by Littleton (1991, 151–157) in her Experiment Nine, where adults and adolescents drew a real-life road, they saw by drawing lines that converged in the distance, whereas children drew the road with even parallel lines. Piaget notes that children begin to master perspective at age 9 and beyond and specifies that children aged 8,5–9 can draw in linear perspective in the sense that they draw how distance can distort shape (Piaget & Inhelder, 1967, 175; 184)

Freeman and Janikoun (1972, 1118) found that children started omitting parts of the view they did not see as they reached the age of 8. Salminen, summarizing from various studies, concludes that transparency typically appears between the ages of 5 and 9, but that transparency is also considered a rare feature at any age by some researchers (Salminen, & Koskinen, 2005, 46). In Littleton's (1991, 56–84) experiments one to four children only occasionally used occlusion when drawing two balls presented one in front of the other. However, when faces were drawn on the balls, use of occlusion was more frequent (Littleton, 1991, 88). This could be interpreted in relation to another feature of logical realism, animism, in which the child enjoys giving human features to inanimate objects (Salminen & Koskinen, 2005, 46). As a summary, I will estimate that occlusion might be already used at age 5–9, but that linear perspective might never show up in some children's drawings. Results also depend much on how the tasks are structured, as shown by Littleton (1991) in her thesis.

2.4 Critique of stage theories

Looking at the literature (Cox, 2005; Kindler 2004; Matthews 2004; Salminen & Koskinen, 2005) it has become abundantly clear that stage theories are no longer seen as universal, linear, fixed, or necessarily useful concepts. This master's thesis relies on the idea of stage theories since the original studies used them as a basis. In a sense I am testing the assumption of universal development, since I am seeking to see if two similar sets of tests produce similar results across time, space and culture. I can only probe at a very small building block of stage theories and the results can be interpreted in many ways depending on what one's view on stage theories and human development is. My opinion is that stage theories are only useful as a broad overview of what kinds of drawings children can produce. They have been taught to me in my Art Education university training and they have helped me estimate what I might encounter in the field working as an art educator. This is not a viewpoint unique to me, and is reflected, for example, in the writing of Pataky (2020, 77) who has also studied changes in children's drawings. However, actual teaching should always be suited to the individual, time, and place. Teachers raise people, not populations.

Some criticisms are aimed at the false hierarchy implicit in stage theories. Again, we encounter the idea that photorealism equals good art. Luquet, who coined the terms, originally questioned whether intellectual realism was somehow more "primitive" than visual realism (Costall, 1995, as cited in Cox, 2005, 11). However, according to Matthews (2004, 271) the idea that intellectual realism is somehow inferior to visual realism persists today. He sees that most proponents of this idea have it implicitly in their writing, conceiving of visual realism as a way of drawing that shows the "true" shape of objects (Matthews, 2004, 271). Other researchers of the subject have criticized this culturally guided hierarchy as well, such as Kindler (2004, 235) and Cox (2005, 116). Hamblen (1993, 45, as cited in Kindler, 2004, 241) claimed that stage theories conform to "*the values of modernity, to the characteristics of a hierarchical society, to the institutional needs of education*". Harvard's Project Zero documented a U-curve when it comes to children's art getting "good" "worse" and "good" again as children grow (Kindler, 2004, 242). However, this shape derives from the fact that modern artists idolized some of the features found in young children's drawings, and so those features can be seen in the revered works of these artists (Kindler, 2004, 242). Again, there is a tie to what we appreciate culturally, and how we think children should develop.

Criticism has also focused on how the features that account for visual and logical realism are measured and categorized into groups. Matthews (2004, 270) proposes that each drawing is extremely complex, and that drawings are actually difficult to categorize as either intellectually or visually realistic. In a sense, categorizing flattens the vibrant interpretations we might get from a drawing.

Linearity has also been a subject of discussion, though even the earliest theories accounted for undulation between stages and that the separations are not clear-cut (Lark-Horovitz & Norton, 1959, 434 – 441; Harris, 1963, 19). Kindler (2004, 234) tells from her experience teaching various forms of perspective for several years, that there is not a clear progression from one form of perspective to another. In some cases, supposedly simpler orthographic projections were more difficult than linear perspective for some students (Kindler, 2004, 234). This is in line with other researcher's results, such as Hagen's (1985, 1986, as cited in Kindler, 2004, 234). Drawing from Hagen's ideas Kindler (2004, 235) suggests that studying various features of drawing in parallel, instead of grouped stages that follow each other, might be a more fruitful way of looking at drawings.

Research that has focused on the meaning of drawings has emerged to contrast with drawings that are only seen as working toward realism. For example, Matthews (2004, 292), having criticized stage theories earlier, suggests that we see children's drawings as things tied into this world, the moments in which they are made, and activities in which the child explores possibilities just as in pretend play. Cox (2005, 121) sees visual realism as a tool children use which is deemed useful or not by them depending on the context in which the drawing is created, since different drawings reflect different purposes. Children are not working toward some sort of retinal truth. Salminen (Salminen & Koskinen, 2005,

71) saw too, that children's aim is to communicate and express and explore. This is why drawings, especially in what is considered the intellectually realistic stage, often include talk, which interacts with the meanings of the drawing, transforming both (Cox, 2005, 123; Salminen & Koskinen, 2005, 42).

2.5 What children *can* draw

It is evidently clear from the literature that by the age of 6–7 years most children understand the concept of symbols, and exercise quite a bit of control over what kind of marks the pencil leaves on the paper (Salminen & Koskinen, 2005, 38). They can do lines and circles at will and understand that with these shapes can represent a sun, for example. I will not discuss motor control and symbols further in this thesis, as other topics are more relevant. Creativity is also a topic I will leave outside of this thesis, as the instructions in the exercises I give are extremely narrow, with little room for creative solutions.

Piaget and Inhelder popularized the idea that a child's drawings can be used to assess what the child can conceive of. In this sense, children's drawings have been used as a measure of what the child *can* do and think of. They are a window to the child's abilities. Their most relevant publication on the subject is the book *The Child's conception of space*, originally in French (Piaget & Inhelder, 1967). Both Littleton (1991) and Genovese (2018, 176), the main authors this master's thesis' experiments rely on, refer to Piaget and Inhelder's publications when forming and countering hypotheses. Even Harris (1963, 7) related his findings to Piagetian stages of development. Although the formerly mentioned book contains many theories on children's conception of space, I will focus only on two relevant topics in this chapter.

From Kindler's (200, 237) perspective Piaget "*closely linked cognitive development with development in drawing and regarded early pictorial work of children as evidence of their cognitive deficit*" in *The child's conception of space.* In other words, Piaget thought what children could draw is what children could think of, more or less. He does note that children verbally acknowledge things that they do not present in their drawings. For example, when it comes to distinguishing different viewpoints, even children aged 6 realize that perspectives change the appearance of an object, and that their drawings lack some sort of distortion caused by perspective (Piaget & Inhelder, 1967, 180–181). They know that a doll sees a red pencil placed on the table not as a long line, but as a point from the doll's perspective, but they cannot reproduce it on paper, or as Piaget suspects, imagine it (Piaget & Inhelder, 1967, 177)

Piaget wonders why the child does not draw what they see, since for example drawing a dot for a needle facing the child directly should be just as easy as drawing a straight line (Piaget & Inhelder, 1967, 178) he speculates that this is because the child is unaware of the different viewpoints and cannot see them as discontinuous (Piaget & Inhelder, 1967, 178). Piaget (Piaget & Inhelder, 1967, 158) notes in some tests that don't require drawing that the child is able to separate their own viewpoints at age 7. He calls this to *take aim* or *sight*. That is, the child starts acknowledging that particular viewpoints are temporary and different from each other and can co-ordinate between different viewpoints when viewing a model (Piaget & Inhelder, 1967, 165). In the words of a child from one of his studies: "*I can see better from the side*" (Piaget & Inhelder, 1967, 170). The child acknowledges their viewpoint, their own relation to the object they are viewing, and can move themselves into a better position to do the exercises Piaget demands. They become more aware of themselves inside the world.

This was Piaget's way of interpreting the phenomenon that children do not draw what they "see" from their perspective. There are some theories that counter this assumption from the perspective of what a child *can* do. For example, Salminen notes, we rarely have to distinguish our own viewpoints in daily life, as the flow and wholeness of everyday perception allows us to live our daily lives (Salminen & Koskinen, 2005, 102–103). We rarely think "*I am looking from the chair*" and "*I am looking now 2cm from the left*" to ourselves. It could be argued that the ability to distinguish different viewpoints is a skill acquired when this distinction is necessary to a task, such as drawing. Salminen proposes that seeing as flat pictures, as in seeing from different static viewpoints, is a tough skill to learn, because it is so different from an ordinary, multisensory, and continuous way of seeing (Salminen & Koskinen, 2005, 121). It seems that the appearance of the drawing can be framed from the perspective that children have not learned to see the world in a way that is useful for drawing. At least not yet.

Freeman (1980, 1995, as cited in Kindler, 2004, 237), disagreed with the notion that "mistakes" in drawings were a sign of some cognitive deficit as well. Freeman (1980, 1995, as cited in Kindler, 2004, 237), suggested that instead of being cognitively deficient, children had a hard time translating their knowledge of the world onto paper. However, at least when it comes to the idea of acknowledging one's perspective in a task where a ball is placed in front of another, Littleton's (1991, 88) showed that most children between six and seven years *can* translate this view onto paper if they consider it relevant.

A tougher view to translate might be one point perspective. Littleton (1991, 193) hesitantly suggests after failing to elicit one point perspective drawings from children aged 6–7 that one of the explanations might be that children do not notice the convergence of the lines in the horizon. However, this is not her only explanation. When it comes to Piaget's experiments on drawing 1-point perspectives children close to 7 years were unable draw it, but they recognized that the convergence of the lines was there and could choose converging line-drawings to represent it (Piaget & Inhelder, 1967, 175). Children this age sometimes draw the lines parallel until they meet abruptly close to the horizon, like the tip of a pencil (Piaget & Inhelder, 1967, 188). At age 8.5–9 they are able to slowly converge the parallel lines, like a long, gradually narrowing triangle (Piaget & Inhelder, 1967, 175). It is at this age that Piaget suspects they have begun to understand a "law" behind the distortions and are no longer just relying on intuitive impressions of shape (Piaget & Inhelder, 1967, 191). In other words even Piaget did not suspect that children do not understand the idea of converging lines. In Littleton's (1991, 184) experiment eleven where she used a photograph instead of a live model of a road, children did produce drawings where lines converged in the horizon. If we come back to Salminen's (Salminen & Koskinen, 2005, 121) idea of the "skill" of flattening views into two dimensional pictures, the work of flattening was already done for the children in advance in this exercise. It could be that children have a hard time translating what they see in real life, from a moving perspective, into a two-dimensional picture.

There is another curious undercurrent of research that relates to the Flynn effect seen in scorings of drawings. Genovese (2018), it seems, was not the first to study the relationship between rising IQ scores and visual art. Darras (2002, as cited in Kindler, 2004, 243) presents the idea that the general rise in scores on IQ tests over time and generations (the Flynn effect) might be the result of engagement and exposure with the visual world. In Darras' (2002, as cited in Kindler, 2004, 243) study, it was noted that improvements in scores were almost exclusively happening on nonverbal tasks that relied on visual-spatial competences. The study suggested that changes in the visual environment resulted in growing expertise in visual analysis, which in turn, is reflected in IQ test results (Darras, 2002, as cited in Kindler, 2004, 243). However, Kindler (2004, 244) notes that this area of visual-spatial intelligence is far larger than just the concept of drawing. If seen from the seven intelligences identified by Gardner, it is not solely the privilege of the arts, as engineers, surgeons and sailors also rely on visual-spatial intelligence (Kindler, 2004, 248).

In the Draw-A-Man Test, which is the third test in my set, the gist is the following. Harris (1963, 165) concludes from several studies, including later his own, that as children grow older and more mature, their ability to create detailed yet wholly unified drawings becomes better. Harris (1963, 203), in agreement with Eng (1931), considers drawings not to be guided by meaningfulness as much as by how the child condenses their limited conception of the model into the forms they have learned to draw on a two-dimensional surface. In other words, children learn "formulas" of drawing and fit whatever limited amount they can conceive of an object/animal/person into this formula to the best of their abilities. These formulas or condensations have also been called *schemata* (Harris, 163, 202) A good example is found in Salminen's (Salminen & Koskinen, 2005) book as he discusses the same topic. There he shows a drawing of a hamster and a bison which are virtually identical, however the bison has a tail and is shown to be bigger than the hamster with some mountains for scale (Salminen & Koskinen, 2005, 235). They have the same formula but different details.

In the Draw-A-Man and Draw-A-Woman Tests, children are asked to draw a picture of a man and woman and lastly participants are asked to draw a picture of themselves (Harris, 1963, 239–311). These drawings are scored according to a manual and this score is related to intelligence but is not in itself an intelligence score such as the IQ (Harris 1963, 246).

However, the ability of the Draw-A-Man Test to probe directly into children's conceptions varies if its instructions are changed. In Mott's (1945, as cited in Harris, 1963, 23) experiments children were asked to move the body part they were about to draw, which elicited more detailed drawings of the body parts from the children. This puts extra emphasis on repeating the tests I have done exactly as the preceding researchers did them. After all, it seems a trend that different instructions give rise to different kinds of drawings.

In conclusion, there are many theories to what children *can* draw, and these theories often interact. First, some theories pose that children cannot imagine, conceive of or think of everything they see in a model. Second, they have not yet gained the skill and experience to translate what they see, or conceive of, onto paper. It is not the case that children do not know how to draw the features in a physical sense, it is that they have not learned where these forms might be useful and to relate these shapes directly to the model. They are still away from that moment of "*Aha! This can be drawn like this, and now it looks more real*" which comes with the next stage of visual realism. In this sense, the shift between stages is one of ability to depict and the ability to think of many features at once. This argument, however, relies on the idea that realism is somehow the desired result. It would seem that certain aids help reach this realistic result, such as a change in instructions. I cannot completely rule out that my tone, infliction, setting and language might have changed the drawings presented in this master's thesis.

2.6 What children *prefer* to draw

Depending on the context in which the drawing is produced, certain representations are more advantageous than others. Even researchers who view drawing from the point of cognitive development recognized this. Harris (1963, 200) summarizes past studies, and states that the child does not draw all that they know, just the bits that are meaningful to them at that moment. He quotes Goodenough (1926, as cited in Harris, 163, 200), who noted that children's drawings may lack hair, even if the child can easily point at their own hair. In her thesis (Littleton 1991, 199) rejects the Piagetian notion that children are too conceptually immature to draw perspective or occlusion and concludes that her experiments give support to Luquetian arguments that children are simply not that interested in drawing in a visually realistic way. Her experiment's support, however, that this interest can be piqued in different kinds of situations with different instructions. Kindler (2004, 239) also considers situations, even those that happen inside the drawing itself, to guide preference.

Communication is often posited as something guiding preferences for depiction. Salminen (Salminen & Koskinen, 2005, 45) notes that children know that their drawings lack some features present in the real model, but that they have simplified their drawings to express an idea. This is also reflected in the name Salminen gave to intellectual realism: *The joy of telling* (Salminen & Koskinen, 2005, 39). This theory of communication is supported by the fact that children in the logical realism stage often accompany their drawings with explanations as to why they look the way they look (Cox, 2005, 121). Not everything has to be shown visually. "*The bare essence of the story*" matters (Kindler, 2004, 240). Children in the intellectual realism stage draw in an economical way, depicting only what is needed (Cox, 2005, 120). If the drawing needs to put emphasis on how something looks, then realism is expressed (Kindler, 2004, 240). This idea that the scene needs to be depicted "*as it looks from where you are*" is included in the instructions for some of the tests in my study.

Communication does not happen in a vacuum. Just as many theories have considered the western obsession with "realism", many researchers have proposed that children draw from other cultural models of drawing as well (Kindler, 2004, 237). Darras (2002, as cited in Kindler, 237) notes that we live in a world that is not solely composed of visually realistic images, but also pictographic/schematic images, and mixes of realistic and non-realistic. Cox (2005, 121) proposes that children in the logical realism stage are less constrained by the conventions of their culture, and in their freedom, they draw what they want to. I noticed these conventions in myself, when in my preliminary studies I often dismissed the child's worries of whether a depiction was "right" with saying "*don't worry it does not matter*" almost automatically. This response does not come automatically from me when I talk with adolescents who worry if their drawings are realistic.

Golomb (2002, 90–91), in her book *Child art In Context*, showcases an interesting study by Belo (1955) where children from Bali drew their very first pictures in an environment enriched with Balinese art. This tradition includes shadow plays, in which puppets depicting humans in a distinctly elongated way with their heads shown in profile entertain and educate at festivals (Golomb, 2002, 90–91). However, children aged 3–4 who clearly saw and were familiar with these Balinese types of depiction started out by drawing the human form much in the way all children start, with round heads, lumpy bodies and stick legs. Only at ages 5–7 did the humans drawn by children start to develop features typical to Balinese art, such as the shadow-puppets elongated forms and profile-view (Golomb, 2002, 91). In other words, they started drawing in a way appreciated by their culture, which often starts by looking at the drawings of peers (Golomb, 2002, 91).

Interest in cultural models seems to come with age, and when children get interested, they want to copy. Even 5-year-old children themselves acknowledge that they learn to draw from others, by being taught or by copying (Kindler & Darras 1995, as cited in Kindler, 2004, 243). Wilson and Wilson (1977, 1982, 1985, cited in Kindler, 2004, 243) argue that every image a child creates can be traced back to sociocultural schemata.

Cohn (2012) looks at drawings from the perspective of language and emphasizes the importance of copying cultural models. The ideas of “formulas” that have been present in previous literature (Harris, 1963, 202; Salminen & Koskinen, 2005, 235) are recontextualized in his work as a syntax. Cohn (2012, 175) proposes that children learn graphemes (such as circles or lines), a syntax (such as how to put these in the shape of a stick figure) and a production script which dictates the order in which they should be put together to form a stick figure. The goal is not necessarily realism, but to learn the language that others speak in their drawings. Cohn (2021, 187) however, does see this learning as having a critical developmental period that ends during puberty. During this window of time, imitation of other people’s drawings is critical in order to learn the language of drawing (Cohn, 2012, 188). Curiously Cohn (2012,188) proposes that if children miss this window their drawing abilities stagnate for the reminder of their lives. These notions tie drawing back again to cognitive ability though they do account for preference and relevance as well.

Another viewpoint that traces preferences to abilities is that for children to be able to depict something that others understand, they must be aware of what the viewer needs to see (Matthews, 2004, 275). For example, if the child wants to depict a coffee cup so that other people understand that it is a coffee cup, it is reasonable to assume other people will understand it faster if a handle is included even if it's not visible from the child's viewpoint. Light (1985, as cited in Cox, 2005, 123) argues that drawings should be looked at as trying to convey something, not just as containing certain features common to stages.

In conclusion, both the perspectives of abilities and preferences are interesting and have potential ramifications for Art Education. Most literature accounts for both, although they might be presented as opposites. It seems likely that abilities and preferences interact, which is a view that has also been acknowledged all along the literature. The results of this thesis can be interpreted from many viewpoints, although it originally set out to answer a question presented only in the context of cognitive abilities, as in the skill of thinking. It is not for me to say if it is abilities or preferences that have changed or stayed the same. My scope is much more limited. My answer simply looks at whether results in certain tests have changed.

3. Approach

3.1 Quantitative research and its relationship to ontology

My research questions are full of terms such as *higher, average,* and *younger*. In order to truly know these relationships, I would need to investigate things quantitatively. I also decided to do something akin to a replication of tests, so that I would not be comparing apples to oranges. My study demanded that I compare quantities of the same things. My study was also tied to the quantitative research paradigm in the sense that the original studies I picked from the 1950's and 1980's were done quantitatively as well.

"*Quantitative research is the process of collecting and analyzing numerical data. It can be used to find patterns and averages, make predictions, test causal relationships, and generalize results to wider populations*" (Bhandari, 2021). In the field of Art Education, it seems to be less used, except when it comes to quantitative handling of content analysis data. In psychology, it has been used a lot historically, but current textbooks on psychological research such as Coolican's (2019, 29) hold that human experience cannot be summarized with just mathematical laws. In the field of education, a similar pattern can be seen, where historically quantitative studies were more popular in the 50's and 60's, but qualitative research is more common presently (Rautopuro & Malin, 2008, 108–113).

One of the reasons why quantitative research has fallen out of favor is because qualitative research can be seen as providing deeper and nuanced information, whereas quantitative research has provided narrower information (Atjonen, 2001, as cited in Rautopuro & Malin, 2007, 114). Quantitative research still holds a place where it is needed. Rautopuro and Malin (2001, 109) argue that it is still much needed in education research. Coolican (2019, 59), in his textbook on research in psychology, expresses that "*quantitative methods should be used as appropriate to the subject matter and research aims*". In this case, I deem my quantitative approach appropriate, but I emphasize that the subject should be studied from many different perspectives and with many different research methods. My desire to use quantitative methods is also fueled by the desire to diversify research methods in Art Education. Using different methods to study phenomena helps us see them from multiple perspectives (Metsämuuronen, 2011, 265).

Historically quantitative research methods have been tied to the ontological idea that an ultimate truth can be reached if we are just careful enough with our methods (Coolican, 2019, 349). This is the philosophy of *realism* (Coolican, 2019, 268). However, the more common view in current science is that quantitative studies are not free of subjectivity (Coolican, 2019, 349). A *relativist* position to science poses that all knowledge is relative and uniquely different to each (Coolican, 2019, 268). I do believe that even the numbers and charts of a carefully constructed quantitative study get twisted in our human heads into a position that's continuous with our own experiences and biases. I also believe that information also means different things in different contexts. I would like you to imagine quantitative analysis in this thesis like a writing partner who points to things that I might not have noticed with the naked eye or just by evaluating and sorting them in my head. Despite statistical standards it is ultimately the human expert, or in this case the human Art Education student, who can determine what the meaning of these observed numbers is in real life (Kalakoski, 2014).

However, I cannot deny that the way I am looking at drawings in this master's thesis is rooted in *objective drawing* as outlined by Riley (2008, 154). Riley (2008, 154 – 156) considers that when we look at drawings through the lens of accuracies and viewer-centeredness we are participating in the same view that rationalism and empiricism hold. This view is that there is one absolute reality that can be revealed by reasoning and empirical observation (Riley, 2008, 154). Though I consider the concept of "drawing" through different cultural and social contexts, I am almost erasing human intent when it comes to how I observe drawings in this master's thesis. I am only looking at superficial features, in order to search for a very narrow truth.

This study is not experimental, although it derives from experimental research done by Harris (1963) and Littleton (1991). It is a between groups comparison study, as determined by Coolican (2019, 136–137), that does not have almost any control over external variables, does not have clear conditions, and can only determine relationships, not causes or consequences. Time and culture are also not something I can precisely administer (Coolican, 2019, 136–137). However, because it is the nature of time that 6 and 7-year-olds will grow up, culture will change, and I am a sea away from the places of the original studies, this study could not have been done any other way. Genovese (2018, 177) also examines Human Figure Drawing Tests done with different cultural groups in Brazil and Turkey, when comparing the results of his own research to a wider scope. I can only aim to widen this horizon of research done on children's drawings in different places and times.

3.2 The basics of statistical tests

In order to conduct a good quantitative study, I had to pay attention to various things. Outlined here are my knowledge of statistical methods, operationalization, consistency of measures, validity and the demands placed by doing a replication. I will go over each of these in further detail when I present my concrete methods and tests and when I analyze statistical significances.

For statistical methods I learned the basics of SPSS and looked at differences in percentages and whether any differences might have been just caused by chance. For the later, I was advised by my statistics teacher, Marianne Silén to use the Pearson Chi-squared and Pearson correlation. Correlation tells if two datasets are related to each other in a linear way (Coolican, 2019, 525). I will explain here in colloquial terms what these statistical tests are, as statistics is not a standard part of Art Education studies at the University of Lapland, and I intend this publication to reach as many audiences as possible. It is also a way for me to express and remember what I know.

Positive correlation occurs when variable A increases when variable B increases, whereas if variable A increases when the variable B decreases, the correlation is negative (Coolican, 2019, 554). A correlation closer to –1 expresses a stronger negative correlation, whereas a correlation closer to +1 expresses a stronger positive correlation, with values close to 0 indicating no relationship or a relationship that cannot be expressed with a semi-straight like, like a U-curve or just a random splatter (Coolican, 2019, 554 – 560). Pearson's correlation demands that at least one variable is continuous (1,2,3… and so on) or at least interval (20,60,80…) and the other variable is continuous or interval as well, or dichotomous (yes/no) (Coolican, 2019, 580). The latter case of one continuous variable and one dichotomous variable is typically called a Point-Biseral Correlation and it is a special case of Pearson correlation (Lund Research Ltd., 2018b). For two dichotomous variables a Chi-squared test is optimal, and it expresses whether two datasets are associated by comparing frequencies of yes/no answers with expected frequencies of these answers (Coolican, 2019, 521 – 524). The expected frequencies are what the frequencies *would* be if there was no association between the datasets (Coolican, 2019, 524). In short, the Pearson correlation and the Pearson Chi-squared tell us whether two variables are "*doing their own thing*" or if they are changing hand in hand. I also ended up doing a *Z*-test for two normally distributed variables I had. The *Z-test* tells whether doing a test twice or more gives results that are similar enough to each other (Glen, 2022).

One of the most relevant numbers these some tests can give us when we compare them to certain charts is the *p* value. The *p* value tells us if we should reject the null hypothesis (H0) in favor of the alternative hypothesis (H1) (Coolican, 2019, 437). Different statistical tests have different null and alternate hypotheses, for example in the Chi-squared test H0 is that two variables are not distributed in a meaningful linear pattern that would indicate one variables dependency on the other's distribution (Coolican, 2019, 525). In the *Z*-test, H0 is that the two means of two variables are the same (Glen, 2022). In correlation tests H0 usually means that correlation is 0 (Coolican, 2019, 437). There isn't a strict rule for *p* values in Art Education that I would have encountered. However, it is generally accepted in the traditions of psychology, that *p* has to be less than .05 for the null hypothesis to be rejected (Coolican, 2019, 437). Since the studies I repeat here come from the tradition of psychology, I follow this guideline as well. When $p < .05$ it means that there was only a probability of less than 5% that this pattern we observed in our results would show up just by luck or chance (Coolican, 2019, 437). In other words, if $p < .05$, then it's likely we are observing some real relationship between the two things we measured.

The Kappa statistic also tells us if two things are working hand in hand, but instead of datasets it predicts whether two measurements, measurers or judges agree on the results of something, when we consider how much they'd agree by accident (Läärä & Aro, 1998, 49–50). A Kappa statistic approaching 1 tells us of agreement beyond happenstance, and –1 of disagreement beyond happenstance, with values closer to 0 indicating that measures were more up to chance (Läärä & Aro, 1998, 49).

For testing significance values a one-sided significance value was used where I was testing if the difference between two sets of results was meaningful in one direction (higher than, lower than) and the two-sided significance was used to determine if there was a general difference without direction (different than) (Coolican, 2019, 438). For the Chi-squared test I always used the two-sided significance as Coolican's (2019, 537) book tells this must *always* be the case except in very specific conditions.

I would have to divide one continuous variable (Harris-Goodenough Draw-A-Man Test scores) into a categorical one, so I could compare it with two categorical variables (Balls test and Corridor tests) using the Chi-squared test. Silén advised me to have at least thirty participants in each of the categories I would split, so that it would hold some statistical credibility. Therefore, I aimed to gather at least sixty children for my study, so I could split it into groups of 30. The final number of participants in my study was 71. My sample is not massive, but even smaller samples than this can be analyzed statistically so that conclusions about significance can be made from it (Metsämuuronen, 2011, 262).

On top of these measures, I observed effect sizes as suggested by Silén. The *p* values cannot be compared without considering sample size, the type of statistical test and the effect size (Coolican, 2019, 437). An effect size considers the distortion that a sample size can cause (Coolican, 2019, 458). An effect size also tells us whether the relationship between two variables is small or large (Coolican, 2019, 458; Statistics Solutions, 2022a). Measures of effect size used in this master's thesis are Cohen's *d* for two continuous variables, and φ (phi) and Cramer's V are for dichotomous variables, although the first is optimal for 2x2 tables (Statistics Solutions, 2022a; PsychED, 2016; Coolican, 2019, 583).

Cramer’s V varies between 0 and +1, with +1 indicating the strongest association and 0 no association (van den Berg, 2022a) φ varies from –1 to +1, with 0 indicating no relationship (Glen, 2016).

In order to get numerical data, I had to operationalize the abstract act of drawing into numbers. Fortunately, some of my work was cut out for me when it came to operationalization and measurement. The original studies had detailed how the data was obtained and converted into numerical form. Much of the material was visual and easy to understand. If I was not sure about an aspect of the original tests, I asked for advice from my master’s thesis advisor Maria Huhmarniemi, or from the maker of the original test if they were available and alive. I will explain the concrete way in which I turned drawings into numbers with each test in the following chapters.

I also wanted to know how well the results of the tests I used maintained their consistency. If researchers got wildly different results each time they used the test on the same participant, I would have to consider the effects of chance or external influences a lot more. I will not be evaluating the consistency of Littleton’s (1991) perspective tests as the tests are not so well established as to have other researchers evaluating their consistency. However, Littleton’s (1991) thesis itself constitutes as something that evaluates in which conditions the results of a similar test are repeated. I have taken into consideration the effects that changing the stimuli or wording had on the results in her thesis by aiming to repeat her conditions to the best of my abilities. I will be reviewing some of the studies evaluating the consistency of the Harris-Goodenough Draw-A-Man Test, as it is well established and researched.

Although I chose not to operate from a strictly empirical, objective nor realist point of view, I was still concerned whether I had done all the tests “right”. This is the measure of *validity*. In maintaining internal validity, the researcher aims to eradicate any systematic errors (Läärä & Aro, 1988, 40). This could be caused for example if I always did one test first, causing participants to be always more tired in the other test. In this case I would be measuring more the effect of drawing when tired than the drawing outcomes themselves.

I assumed I would not manage to perfectly replicate the quantitative studies I chose. Firstly, the Harris-Godoenough Draw-A-Man Test manual was in English, and I needed to translate it for Finnish students. There were other differences as well, over which I will go as I describe my tests. Doing replication studies where the original study is not replicated to the dot is not unheard of. For example, the Open Science Collaboration (OSC) was a large replication project in the field of psychology, which was later found to have included many studies that used samples and methods that could be considered different from the ones used in the studies they were replicating (Gilbert, King, Pettigrew & Wilson, 2016, 1037). This of course, is not the best science, and has been criticized (Gilbert, King, Pettigres & Wilson, 2016, 1037). I will not call my study a replication, but it is close to the term. I have done everything in my power to keep the samples and procedures as similar as possible. Restrictions have mostly come from the fact that I am a lone student doing research in a different country, in a different time. I also did some modifications on an ethical basis.

4. Course of the study

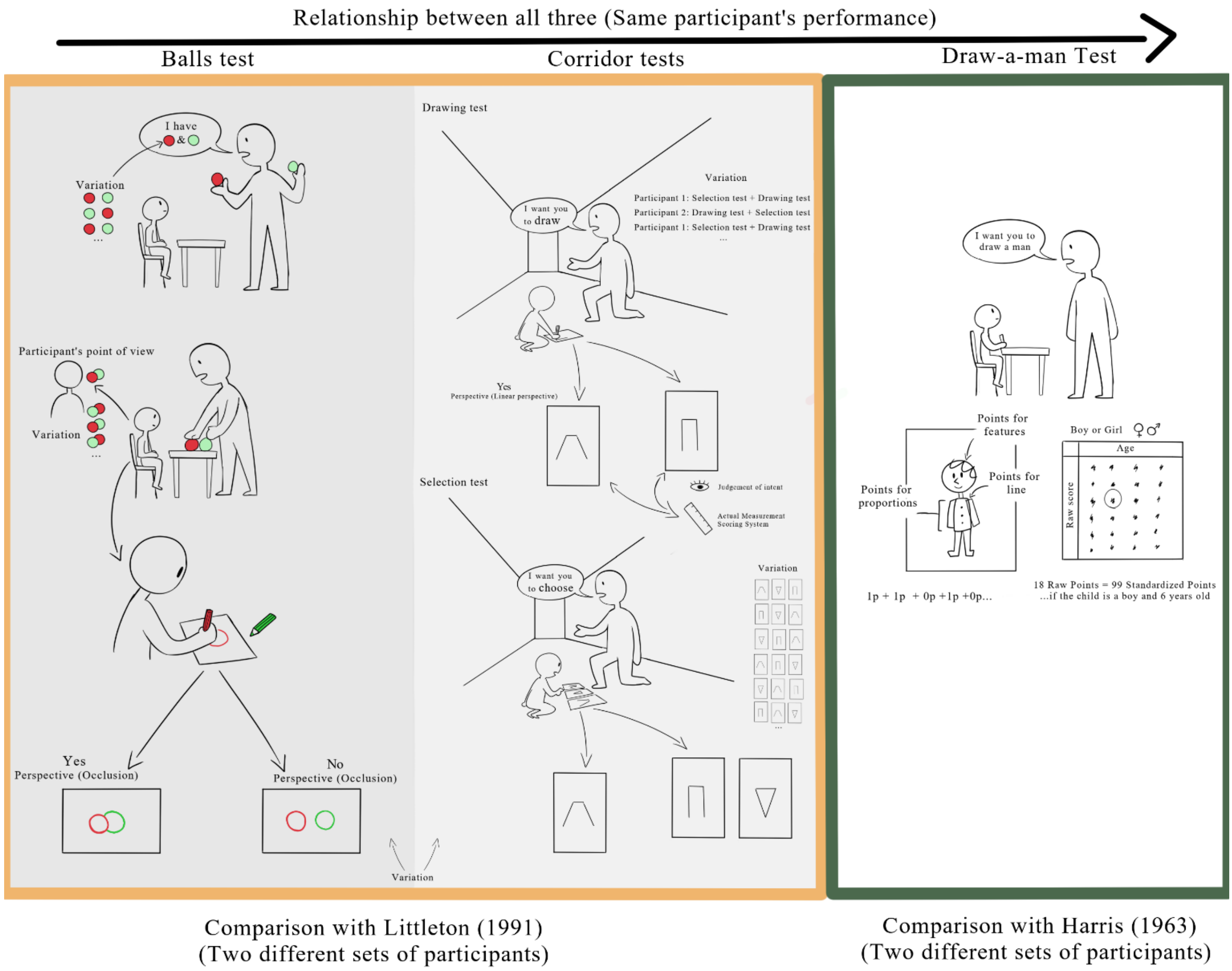


Figure 2: Outline of the course study. Perspective tests are outlined in orange and the Human Figure Drawing test in dark green. On the participant's first day, they completed one of the tests outlined in orange and the Human Figure Drawing test. On the second day the completed the other test outlined in orange.

4.1 Order of the tests

I have depicted a visual summary of the entire course of the study in Figure 2. Each participant completed all three tests on two separate days in each school. The separation between these two testing days ranged from one day to fourteen days. Most Draw-A-Man Tests were produced on the first day, some in a group, some individually. In his manual, Harris (1963, 239) explains that the test can be done individually or in groups while using the same instructions, scoring and standardization.

Participants were haphazardly assigned to either do the Balls or the Corridor tests on the first day, with half doing the Balls test first and half the Corridor tests. Decision was influenced by five things. First, participants who were not engaged in play or homework at the moment did the tests before those who seemed to be doing something that needed their attention at the moment. Second, some of the participants were chosen by the teacher to go first. Third, some of the participants were chosen by me, while I was reading the names in the consent forms I had. It is likely this was influenced by alphabetical order. Fourth, I tried to take participants who had done the first test late on the first day early on the second day of testing. This way they would not have to wait as long on the second day too. Fifth, participants who had not yet done their first test were more likely do the Balls test first if the corridor was occupied. On the other hand, if the corridor was empty participants were more likely to be asked to do the Corridor tests first.

Before administering the tests, I presented myself to the class or to the individual students. I told them my name, where I was from and what I was here for. I also told them that there would be no right or wrong answers and that I wanted to know how *they* draw. In accordance with the times, I maintained COVID-19 regulations in our interactions and wore a mask for example.

4.2 The Balls test

4.2.1 Background

The Balls test was similar to Littleton's (1991, 61,70, 78, 83) Experiments one, two, three and four. The main difference between her tests and mine is that I used only one of the conditions. This condition was called *Task 2: The Ball Behind a Ball Task (Occluding)* by Littleton (1991, 60–61) and it was repeated in Experiments two, three and four. It was a test where just the balls were presented behind one another, with no attempt at anthropomorphizing them and so that one of the balls was only partially visible (Littleton, 1991, 60–61). The aim was to see if children would occlude the further ball behind the closer one (Littleton, 1991, 66). Results from this condition in Littleton's experiments compose the dataset I will be comparing to, dubbed *Littleton (1991)* by me. Littleton (1991, 60, 68, 78, 82) studied children aged 6:6 to 7:6 in her experiments one to four. My participants were 5:9–8:11 years old, but I also grouped them into the same range as Littleton (1991) when I compared my results with hers.

I chose this test condition because I assume it to separate adults from children. Littleton (1991, 28) herself describes partial occlusion as an "adult" technique, and as I reviewed earlier, it is something that starts appearing in children's drawing around ages 5–9. Occlusion is also a feature of perspective. After all, if you want to hide objects behind objects, they are only hidden from certain viewpoints. When not bound by perspective we can peek around the corner and see that the two balls are indeed whole. I will use this test to investigate if there are changes in how children draw occlusion as a feature of perspective, with the assumption that depictions of this feature would change as children grow older. Littleton's (1991, 58) original goal with this test was to see if altering the instructions and presentation of the balls would alter the children's drawings in a way that they would produce partial occlusion. Task 2:The Ball Behind Ball Task (Occluding) was just a condition among many in Littleton's (1991) various experiments. In other conditions, such as when the balls had faces drawn on then, children produced far more occlusion (Littleton, 1991, 75). That is to say, they drew one ball behind the other, so that the further ball was only partially visible.

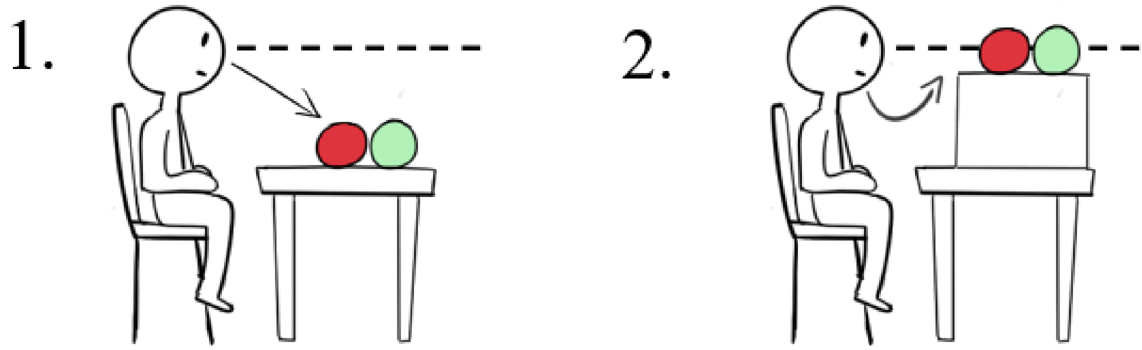


Figure 3: Cleaner versions of two images I sent Littleton to verify how the Balls test should be set up

I corresponded with Littleton via e-mail after I had some confusion of the height at which the balls were viewed. I sent her an image (Figure 3), and she told me that the first setup (1.) was correct. The original image I sent was cruder, and I have refined it here for purely aesthetic reasons. In the pretests, the balls were positioned incorrectly as in the second setup (2.) and this is why I rejected all the results from the pretests as they were not comparable with Littleton's (1991) experiments. Littleton also told me in the e-mail that her tests were done in the 1980's.

4.2.2 Procedure

I invited the participant to the place where I had set up the materials. Other children might have heard what was said during the experiment, but they had no line of vision to the participant doing the test unless they went for a bathroom break and peeked at the setting. In some cases, I had left the previous model on the table, so I detached the balls from the tape with the participant present. The setting was laid out as in figure one. Table and chair heights varied from school to school. The closest ball was always placed 45cm away from the edge of the table. When it came time to put the balls on the table, they were placed close but not touching. I made some small talk with the participant before the experiment. I emphasized that this was voluntary, and that if the participant felt too nervous or if this wasn't fun, they were allowed to say: "*I don't want to do this anymore*". I also expressed I was going to read my "script", and that it might sound clunky at times. The original text by Littleton (1991, 60–61) went as follows:

"I've got a blue ball and a red ball". Then, moving each ball in turn towards the child she said: "Here is the blue ball" and "Here is the red ball". The blue and red balls were then placed simultaneously on the table (the first ball was positioned about 45 cm from the edge of the table) in front of the child. As the experimenter did this she said: "Oh look, I am putting the blue ball behind the red one", and then, pointing to the array: "Now, I want you to draw a picture of the blue ball behind the red one."

I kept the same gestures, but the color of the balls was changed. Half of the participants saw the red ball behind the green one, and half of the participants were presented the red ball before the green one. The order was varied systematically in the way shown in Figure 4. I later checked if I had varied the order systematically. The systematic switching of the red and green ball in both speech and presentation, just as Littleton (1991, 61) had done, was successful. In 49% (*n* 32) of the cases I presented the red ball first, and 52% (*n* 34) of the participants got to hear about the green ball first. In 47% (*n* 31) of the models the red ball occluded the green one and in 53% (*n* 35) of the cases the green ball occluded the red one.

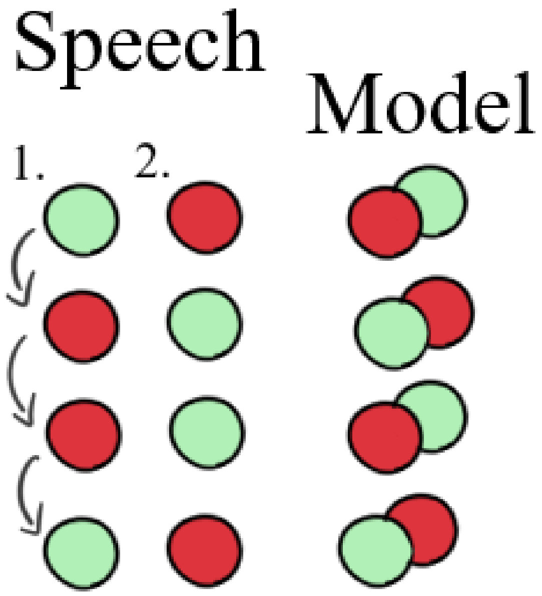


Figure 4: The order of the presentation and occlusion were varied systematically, though I made some order errors occasionally

I translated Littleton’s (1991, 60–61) text as: “*Minulla on punainen/vihreä pallo ja vihreä/punainen pallo. Tässä on punainen/vihreä pallo, tässä on vihreä/punainen pallo. Katsos! Laitan (takana olevan pallon väri) (edessä olevan pallon väri):n taakse. Nyt haluan, että piirrät kuvan (takana olevan pallon väri):sta pallosta (edessä olevan pallon väri):n pallon takana.*”

Materials that were used were a red and a green coloring pencil, a red ball and a green ball. The balls were from the same manufacturer and with the same dimensions. The diameter of the balls was approximately 5cm. White A4 paper for the drawing was presented in landscape orientation. Each ball was fixed to the desk with tape.

4.2.3 Scoring

Balls were marked as occluding or not occluding according to Littleton’s (1991, 63) original criteria. The depictions seen below were *all* marked as having occlusion (Figure 5). Any other way of depicting the balls was marked as not having occlusion. This includes depiction f) if the closer ball is not colored in. The lower row represents drawings I came across that were not depicted by Littleton (1991, 63). With these depictions I, in consultation with my classmates and my advisor Huhmarniemi, deemed the drawings as including occlusion. When I presented the drawings to others, I blacked out the name of the participant. In one ambiguous case I relied on the color of the balls to tell me whether occlusion was present.

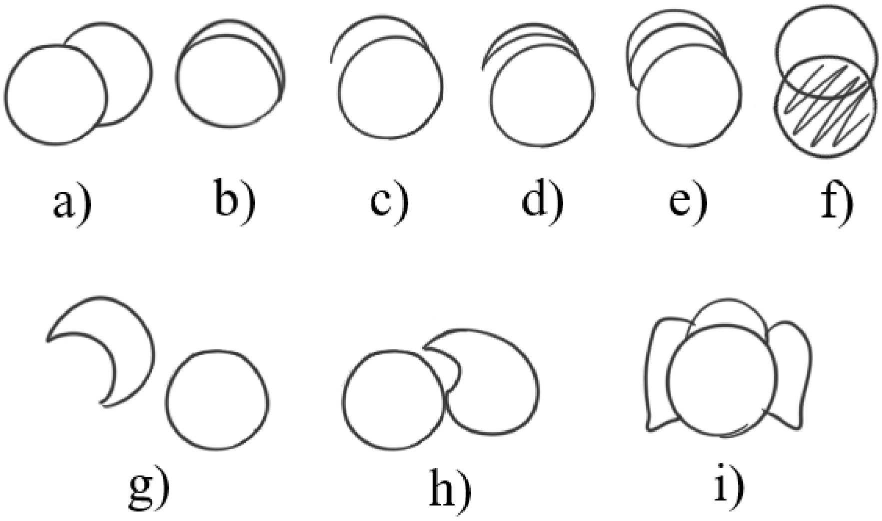


Figure 5: Depictions marked as occluding. The higher row is modelled after Littleton's (1991, 63) criteria. The lower row presents drawings I encountered and deemed as representing occlusion.

Apart from whether the balls were drawn as occluding I noted the following things:

- Which ball was placed in front?
- Which ball was presented first?
- Was there an interruption during the test?
- Did the participant look at the model when drawing?
- Did the participant have to be asked to draw both balls?
- Did the participant draw other details besides the balls? (such as: shadows, seams etc.)

I ended up not looking at most of this data, since I tried to keep the scope of my master's thesis reasonable. However, I felt it was important to note this information as it addressed some of the concerns Littleton (1991) presented in her thesis. For example, she pointed out that drawing the front or back ball first might have an influence on the result (Littleton, 1991, 76). A concern I did end up addressing was interruption of the task. Interruption was something Littleton (1991, 28) mentioned because in other studies interruption during drawing had led to more depictions of occlusion.

4.2.4 Differences between the Balls test and the original test by Littleton (1991)

There are four main differences to the original test by Littleton (1991) I will note here. First, my test was done in Finnish, to Finnish students. There might be some translation issues, as many participants asked whether they should draw the ball in front as well. Second, my balls were slightly smaller in diameter, and had a visible seam, which clearly influenced student's depiction. The seam was included in some drawings. Third, I made human errors in my order, so it was not 100% systematic. I was not as consistent with giving the paper and pens after or before the script, as Littleton (1991, 60–61) had been. Fourth, I was explicit about the fact that the participants could stop their participation at any time and that I was reading a "script".

Before corresponding with Littleton via e-mail, I did some tests where the balls were at the wrong height. I also did some tests where I did not mark some features, such as the order in which the balls were drawn. In these tests I did not vary the order of the balls systematically with a chart, but by trying to remember if I had used this combination before. I have excluded all the data from these tests from the final data, but I will refer to these tests as pretests in this master's thesis. These students also did the Draw-A-Man Test, and Corridor tests, but I did not include the results as they were not comparable with each other in the way that I aimed.

4.3 The Corridor tests

4.3.1 Background

The Corridor tests in my study was as similar as possible to Littleton's (1991, 144) Experiment nine. Results from this Experiment compose the dataset I will be comparing to, dubbed *Littleton (1991)* by me. Instead of a bridge from where a road could be seen like in Littleton's (1991, 144) Experiment nine, participants saw a corridor, or a mat placed on the corridor. This would be seen as going off into the distance in a similar fashion to the road. Participants were asked to draw this model and to select a premade line drawing that looked as much as possible like the road.

Child participants in Littleton's (1991, 149) Experiment Nine were aged between 6:6 and 7:11, but some of her participants were adults and adolescents. My participants were the same as those who did the Balls test. They were 5–8 years old, grouped in two different ways which I will detail later.

Results in Littleton's Experiment Nine (Littleton, 1991, 157) clearly differentiated adults and adolescents from children, and that is the reason why I chose this test. Whereas children drew and selected roads where the sides were parallel (i.e without linear perspective, not narrowing in the horizon) adults and adolescents selected and drew the road in perspective (Littleton, 1991, 157). Linear perspective is obviously a feature of perspective, and it is also often mentioned as a feature of visual realism in that it appears in drawings as the child grows older (Salminen, 2005, 47; Piaget & Inhelder, 1967, 52). If we are not bound by one viewpoint, we can easily change where our horizon line is, by looking around and realizing that a straight corridor going into the distance is rectangular, instead of triangular. The object does not narrow into the distance, it just looks that way if we stare at it from one point of view. I will use this test as well to investigate if there are changes in how children draw linear perspective as a feature of perspective, with the assumption that depiction of this feature would change as children grow older. This assumption is supported by Littleton's (1991, 157) results in Experiment nine, as depictions of linear perspective were far more common in the drawings of adolescents and adults.

Littleton's original goal with this test was to see if altering the model and wording of the exercise would make children produce drawings with linear perspective and she included many more combinations of conditions than just this one (Littleton, 1991, 160, 174). Littleton concluded that she failed in facilitating the production of visually realistic drawings by children when using a three-dimensional model (Littleton 1991, 192). Linear perspective did appear in children's drawings and selections in some conditions, such as in Experiment 11 conditions, where the model was a photograph instead of a three-dimensional model (Littleton, 1991, 184).

4.3.2 Procedure

I guided the participants individually to a spot that I had agreed with the teacher would suit the needs of the study. This was either a corridor that was at least over one meter long, or an open space with a narrow dark mat that was at least over one meter long. Because this was a place everyone used, I often had to wait with the participant until students had gone to recess or back to class, so that they would not be in the way. Sometimes students came in during the test. I tried to make sure students who would participate in the test later would not overhear or see the test procedure. If they were on the corridor during the time of the test, I stopped it until they went into a different space. If the people in the corridor would not be participating in the test later, I asked the participant if they wanted to stop and wait. If yes, we waited, if not, we continued with the test. We often waited when adults walked across the corridor as well.

There were two parts in the Corridor Test. These were the Corridor Selection test and the Corridor Drawing test. Half of the participants did the Selection test first. Half of the participants did the Drawing test first. I checked later to see if I had varied this systematically. 47,8% (*n* 32) of the participants did the Drawing test first and then did the Selection test, whereas 52,2% (*n* 35) did the tests in the opposite order. All participants did both tests on the same day due to time restrictions. I emphasized their option to quit at any time like in the Balls test. I also emphasized I would be reading a “script” that would sound very much the same the two times I read it, but that the two different readings were different exercises that everyone did. This was because Littleton (1991, 186) suspected that asking the children to do two similar tasks might lead them to believe that they did it “wrong” the first time. If they knew ahead of time there would be two of the same tests for everyone, they would presumably not suspect that the second test was some form of correction.

The materials used were three line-drawings of a corridor (Attachments 1, 2 & 3). The space used had either straight corridor or a mat that was at least over one meter long. Participants drew with a pencil and were also allowed to use an eraser.

4.3.3 The Corridor Selection test

In the selection test the participants were presented the following script, originally by Littleton (1991, 150), when we were both sitting with our knees toward the corridor, looking straight at it going into the distance:

"Here is a road (experimenter pointed at the view from the bridge). Here is the road (experimenter pointed at the road only) and here is the pavement (experimenter pointed at pavement). Now if you were going to draw a picture of the road, not the pavement, just the road... this bit here (experimenter pointed' at the road only) and you were going to draw the road exactly as it looks from where you are kneeling which of these three pictures would your road look like (experimenter pointed at the pre-drawn pictures, taking care not to point to any one particular picture)?"

I translated this into Finnish as: "*Tässä on käytävän lattia/matto ja tässä on sen seinät (ja tuolla on vaatteita/ovi/kasvi/vaatteet). Jos (sinä) nyt piirtäisit kuvan käytävän lattiasta, ei siis seinistä (tai vaatteista/ovesta/kasveista/vaatteista), vaan vain käytävän lattiasta. Tämä osa tässä. Ja sinä piirtäisit käytävän lattian juuri sellaisena, miltä se sinulle juuri tästä polvistus/istumapaikasta näyttää: Mikä näistä kuvista näyttäisi samalta kuin piirustuksesi käytävän lattiasta?*"

I kept the gestures the same. I replaced the words road and pavement with corridor/mat and whatever was at the sides of the corridor/mat (art, walls, doors, clothes, etc.). If the participant was sitting on the floor instead of kneeling, I used the word sit instead of kneeling. I chose to present the line drawings after this, since I had to make sure I would not even glance at one drawing in particular more than the others during the procedure. I varied the order of the Selection Test line drawings in the way depicted in Figure 6. I later checked if my variation had been systematic, however the first and sixth order were more frequently presented with a frequency of 19% (*n* 13 out of 67 presentations), despite my attempts to vary the order. The least represented combinations were the fourth and the fifth, with their frequency being 13% (*n* 9). You can find the original line drawings in the attachments section of this thesis.

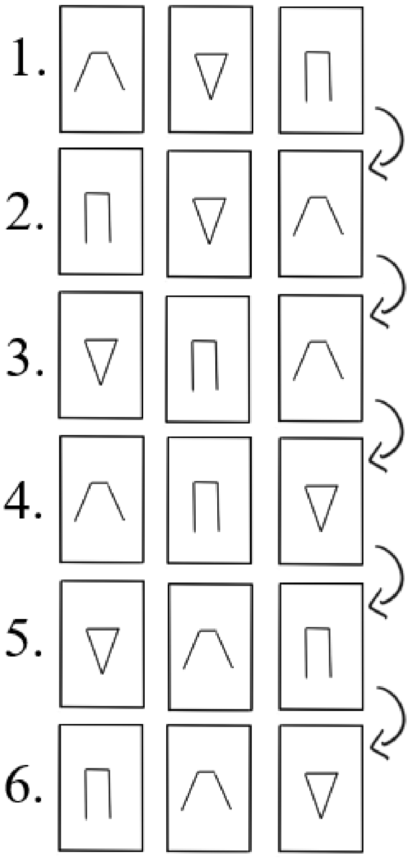


Figure 6: The order in which the Selection test line drawings were presented was varied systematically, though I did make some order errors occasionally.

4.3.4 The Corridor Drawing test

The Drawing Test was very similar to the Selection test. We sat in the same general position as in the Selection test and I primed the test in the same way. Littleton's (1991, 150) script for this task is as follows:

"Here is a road (experimenter pointed at the view from the bridge). Here is the road (experimenter pointed at the road only) and here is the pavement (experimenter pointed at the pavement). Now I want you to draw a picture of the road. Not the pavement, just the road... this bit here (experimenter pointed at the road only). Draw exactly what you can see. Draw the road exactly as it looks from where you are kneeling".

I translated this into Finnish as: *"Tässä on käytävän lattia/matto ja tässä on sen seinät (ja tuolla on vaatteita/ovi/kasvi/vaatteet). Nyt haluan, että piirrät kuvan käytävän lattiasta/matosta. Ei seinistä (tai vaatteista/ovesta/kasveista/vaatteista), vaan vain käytävän lattiasta. Tämä osa tässä. Piirrä juuri se, minkä näet. Piirrä käytävän lattia juuri sellaisena, miltä se sinulle näyttää, tästä polvistus/istumapaikasta."* I kept the same gestures as Littleton (1991, 150), and I used the same word replacements as in the Selection test translation.

4.3.5 Scoring

For the Drawing Test, the drawing was marked as having either linear perspective or no perspective at all. For linear perspective to be marked in the drawings there were two measurements. I marked the results for both measurements separately.

1. Just as Littleton (1991, 99) had had a set of judges to evaluate what the participant's intent might have been, I did a *judgement* if the participant had intended to draw the corridor in linear perspective.
2. Based off Littleton's (1991, 99) Actual Measurement Scoring System the convergence of the sides of the corridor was measured with a ruler. Even a convergence of 1–2mm was enough for the drawing to be marked as having linear perspective if both sides converged.

The Kappa coefficient for these two "judges" (me and the ruler) was .228. This is only a minimal agreement. I will also note that the ruler and me are technically not independent judges since my general judgement of the shape of the lines as converging was always confirmed by the ruler.

Likewise, the Selection Test also had a dichotomous result. As shown in Figure 7, drawings b) and c) where marked as having no linear perspective, whereas drawing a) was marked as having linear perspective.

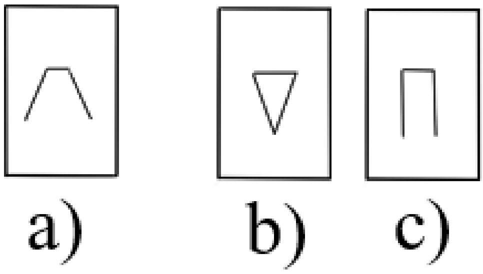


Figure 7: Line drawing a) depicted a corridor in linear perspective, b) and c) did not

Additionally, I noted the following information:

- Did the participant do the selection or the drawing test first?
- What order were the selection test stimuli presented in?
- Did the participant draw any additional details to the corridor? (texture, doorways etc.)

Again, I did not use most of this data, but I considered it important to collect.

4.3.6 Differences between the Corridor tests and the original test by Littleton (1991)

Aside from the obvious difference in stimuli (road vs corridor) there are four main differences to the original test by Littleton (1991, 144–157). First, Selection and Drawing tests were conducted one right after the other in order to save time. I chose to be very clear from the start that everyone would take two very similar tests one right after the other. I also told them that normally these two tests would be done on different days, hence the repetition in instructions, but that I was in a hurry, so I had to do both today. I also emphasized in this test, as I did in others, that I would be reading a "script" so that all my tests would sound the same to everyone, and that is why I would sound janky. Secondly a straight dark mat against a light floor acted in place of a corridor in one of the schools. Third, my line drawings were not the exact same as Littleton's (1991, 146–148), although they had the same general shape. I made my line drawings based on the first corridor I tested at, so they would be closer to a corridor than to a road. If I had to replace them, I redrew them so that I would see the old drawings through the new paper and they would align. At one school my drawings were stolen, and I had to draw them again from memory. Fourth, in the Actual Measurement Scoring System I had to evaluate some of the drawings by the photos I had taken, as I had left some (*n* 7) of the drawings at my other residence in the south when I was categorizing data while living in the north. I do not believe this to have massively affected results.

In one set of tests, which I did not include in the final data, participants did the corridor Selection test only. In this set of tests, they also did the Selection Test in a space where they could hear each other's answers. I will refer to these as pretests as well. These students also did the Draw-A-Man Test, and Balls test, but I did not include the results as they were not comparable with each other in the way that I aimed.

4.4 The Harris Goodenough Draw-A-Man test

4.4.1 Background

The Harris-Goodenough Draw-A-Man (DAM) test is part of a large history of human Drawing tests. Per its name, it is a revision and extension of Goodenough's Draw-A-Man Test (Harris, 1963, vii). The set of tests was originally undertaken in collaboration with Goodenough until her death, and it aimed to "*put in order certain uncompleted aspects of the Goodenough research*", to extend the applicable scale, and to bring more contrasting and even contending viewpoints offered by research into the making of the test (Harris, 1963, vii). The Harris-Goodenough DAM's most obvious use is as a measure of intelligence, stated clearly in the book's title *Children's Drawings as Measures of Intellectual Maturity* (Harris, 1963). This book includes the manual, as well as a review of the research leading up to the making of the new version of the test. Its general idea is that details, proportions and inclusion of relevant features in drawings of the man and the woman reflect the child's maturity in drawing (Harris, 1963, 247).

The Harris-Goodenough DAM is only a part of a battery of several tests in the Goodenohg-Harris Drawing Test, and the manual also includes instructions for the Draw-A-Woman test and a drawing of the oneself (Harris, 1963, 276 – 315). It is important to note that most studies I refer to have administered the whole battery of drawings and not just the drawing of a man, like I have.

The Harris-Goodenough Draw-A-Man Test has been standardized for children aged 3-15-years old and can be done either individually or in groups (Harris, 1963, 239; 294).

4.4.2 Procedure

In contrast to Harris' (1963, 240) instructions for gathering background information I did not ask participants to supply their exact birth date as I had already gathered that information in the consent forms. I did ask them to write their name and age. I considered information about the school they went to a risk for anonymity, and therefore I did not mark it down other than for initial grouping purposes and I do not present it here or in my final dataset. Instead of asking them to circle *Boy* or *Girl*, I asked them to write P if they were a boy, T if they were a girl and M if they were of another gender. This proved problematic when the children didn't know how to write yet, so I usually had to draw examples of the letters in the air or on a whiteboard. In a curious case, a child asked what a boy is, which might mean that my instructions were not clear enough, although the child wasn't a native Finnish speaker. I also did not ask information on the father's occupation, since it had no relevance to counting the points. I tried my best to make a space with as little visual material as possible and where participants did not talk to each other about their drawings, as per Harris' (1963, 241) instructions. I also emphasized I was going to read off a "script", and therefore I would ask even a participant usually gendered as a girl for their gender. Harris (1963, 241) encouraged also praising slower students, telling participants questioning what they should draw to: "*Do it whatever way you think is best*", and avoiding giving suggestions or Yes/No answers. I complied with these instructions as well.

The DAM procedure in the manual, after sitting the participants down and gathering the background information such as age, goes as follows:

"I am going to ask you to make three pictures for me today. We will make them one at a time. On the first page I want you to make a picture of a man. Make the very best picture that you can; take your time and work very carefully. I want to see whether the boys and girls in ___ School can do as well as those in other schools. Try very hard, and see what good pictures you can make. Be sure to make the whole man, not just his head and shoulders." (Harris, 1963, 240).

In Finnish, my script was the following*: "Pyydän sinua tänään piirtämään kuvan. Haluan, että piirrät kuvan miehestä. Piirrä niin hyvä kuva, kuin suinkin voit. Työskentele rauhallisesti ja huolella. Opettaja on kertonut minulle, että piirrät hyvin / täällä on hyviä piirtäjiä. Muista piirtää varmasti kokonainen mies, ei vain hänen päätään ja olkapäitä."*

I opted to not place children in a position where they were competing with another school, as I found this morally dubious. Instead, to encourage them by saying that "*the teacher has told me that there are some very good drawers here*", or that "*the teacher has told me you are very good at drawing*". I made this decision with my advisor Huhmarniemi. Instead of three pictures I told them I was going to ask them for "*a picture*" and excluded the part about making them one at a time and about this being the first page. I pointed at my own body when I asked participants not to draw just the head and shoulders but the whole body. I also told them they could start, after some participants had been questioning whether they could start after the instructions. After the drawing was done, I clarified some uncertain parts with the participants, without trying to make assumptions or suggestions about what they depicted, in accordance with Harris' (1963, 239–241) instructions.

The materials used were a pencil, an eraser and A4 paper. The original test would be done on a test booklet, but since I did not have access to it, I ended up using simple materials that could collect the same information.

4.4.3 Scoring

As I was the only researcher in this master's thesis, I did all of my scoring alone. Initially I practiced with some of the drawings provided by Harris (1963) on pages 264–274 until I only got a maximum of +5/-5 points incorrect in each drawing. I estimated this to be a good enough agreement with Harris' (1963, 264–274) scores. Harris (1963, 112) had originally obtained an inter-rater correlation of +.71–+.91 when raters were using the simpler quality scale. I had the most trouble scoring complex drawings. One of the judges in Harris' (1963, 112), evaluation of the quality scale, a secretary, can be considered just as psychologically naive as me, although I have done some psychology courses as a minor subject.

I scored the Draw-A-Man Test with the scoring manual presented by Harris (1963) on pages 248–263, and raw scores were standardized according to the tables he presented on pages 294– 296. As an example, higher scores are given if the drawing has all five fingers, or if the feet are proportional to the legs (Table 1).

Table 1: Short list of the items measured (Harris,1963, 275) with some specifications I provided based on the more detailed guide on pages 248–263.

1. Head present
2. Neck present
3. Neck, two dimensions
4. Eyes present
5. Eye detail: brow or lashes
6. Eye detail: pupil
7. Eye detail: proportion
8. Eye detail: glance
9. Nose present
10. Nose, two dimensions
11. Mouth present
12. Lips, two dimensions
13. Both nose and lips in two dimensions
14. Both chin and forehead shown
15. Projection of chin shown; chin clearly differentiated from lower lip
16. Line of jaw indicated
17. Bridge of nose
18. Hair I (any indication of hair)
19. Hair II (substance)
20. Hair III (styling)
21. Hair IV (directed lines)
22. Ears present
23. Ears present: proportion and position
24. Fingers present
25. Correct number of fingers shown
26. Detail of fingers correct
27. Opposition of thumb shown
28. Hands present
29. Wrist or ankle shown
30. Arms present
31. Shoulders I (change in the direction of the outline)
32. Shoulders II (more outlined shoulders)
33. Arms at side or engaged in activity
34. Elbow joint shown
35. Legs present
36. Hip I (crotch)
37. Hip II (more outlined crotch)
38. Knee joint shown
39. Feet I: any indication
40. Feet II: proportion
41. Feet II: heel
42. Feet IV: perspective
43. Feet V: detail
44. Attachment of arms and legs I (approximate)
45. Attachment of arms and legs II (strict)
46. Trunk present
47. Trunk in proportion, two dimensions
48. Proportion: head I
49. Proportion: head II
50. Proportion: face
51. Proportion: arms I
52. Proportion: arms II
53. Proportion: legs
54. Proportion: limbs in two dimensions
55. Clothing I (any clear representation of clothing)
56. Clothing II (two articles of clothing)
57. Clothing III (no transparencies)
58. Clothing IV (four articles of clothing)
59. Clothing V (costume)
60. Profile I (head, trunk and feet in profile)
61. Profile II (true profile)
62. Full face (all major body parts in proper location and joined, all essential items such as legs, eyes, ears)
63. Motor coordination: lines
64. Motor coordination: junctures
65. Superior motor coordination
66. Directed lines and form: head outline
67. Directed lines and form: trunk outline
68. Directed lines and form: arms and legs
69. Directed lines and form: facial features
70. “Sketching” technique
71. “Modeling” technique
72. Arm movement
73. Leg movement

4.4.4 Consistency of measure in the Harris-Goodenough Draw-A-Man Test:

I will refer to *reliability* as a concept in the studies that use the term. However, Läärä and Aro (1988, 43) have suggested that *consistency* is a much more precise term for the idea that a test's results can be repeated if the tests are done properly. Unlike my two previous tests, the Harris-Goodenough DAM and other Human Figure Drawing Tests have been studied extensively. Validity of it as a measure of human intelligence is not relevant for this thesis, as I noted before this connection between drawings and intelligence was too multifaceted and debated to be covered in the scope of this study. Ganesh (2012, 217), in his review, states that overall "*there is considerable controversy about the use of children's human figure drawing as a measure of ability or intelligence*". For example Aikman, Belter and Finch (1992, 119) found that the Harris-Goodenough Drawing tests correlated poorly with IQ scores. For me to gauge differences in the performance of children in the Harris-Goodenough Draw-A-Man Test it does not matter what it measures, as long as it measures consistently.

Administration of the test may have its effects. Reichenberg-Hackett (1953) as cited by Rubin, Schachter and Ragins (1982, 663) found that Goodenough scores (not Harris-Goodenough scores) were higher in general when teachers were supportive and permissive. Naglieri (1988, as cited by Ganesh, 2012, 217) modified the Harris version of the test so that "*subjects' language, verbal skills, and cultural differences would not become confounding factors*". They are then, presumably, confounding factors in the Harris-Goodenough Draw-A-Man Test. This emphasizes that any differences or similarities found in this study may very well be related to the translation and Finnish cultural context. Harris (1963, 130) himself deemed experience with print and pictures to be one of the factors influencing results that warranted more detailed research. It can be argued that children today have a vastly different visual world than children in the 60's and 80's, but this study cannot probe it further.

Harris (1963, 130) also noted that sex differences and cultural differences were complexly related, which brings into question how applicable his gender-separated scale may be in a contemporary Finnish cultural context. I use the word gender here, as participants were asked whether they were a boy or a girl, and thus it relates more to social identity.

Culturally significant things may also be reflected in the drawings, such as mukluk-boots in drawings made by Inuit children (Harris, 1963, 132). As a counterpoint to the whole idea of this master's thesis, Harris (1963, 133) debates whether the scale should be used at all for cross-cultural studies if we cannot find out which specific cultural factors influence scoring and which features actually reflect intelligence. This is the case in the context of comparing IQ, which is not the purpose of this study.

There are also a variety of factors that should not have an effect on the scores, some of which were already addressed by Harris (1963, 239). Rubin, Schachter and Ragins (1982, 662) concluded from a variety of other studies that the medium, person, time of the day, and whether the examiner was present should not have any major effect on the scores. Group and individual administration were also found to produce similar results in' Scott's (1981, 501) review. Inter-rater reliability for scoring of the human drawings with the quicker and rougher quality scale was found to be +.71–+.91 by Harris (1963, 112). In a separate study done by Rubin, Schachter and Ragins (1982, 655) inter-rater reliability for the Draw-A-Person test, of which the DAM Test is part of, was +.88. In Scott's (1981, 502) review of the topic of intraindividual variability in human figure drawings several decades later, inter-rater reliability was between +.80 and +.90.

Harris (1963, 90) himself presented that the reliability of the original Goodenough test was fairly high and had been tested numerous times. He tested Goodenough's original scale with children on ten consecutive days and concluded that even though there was a fair amount of individual and sex-related variation, the order of the drawings only accounted for a very small proportion in the variation of the performances (Harris, 1963, 91). In other words, Harris deemed the original Goodenough fairly sound in reliability, through individuals may have provided different kinds of drawings on different days. On page 92 Harris (1963) marks Intra-child variation as nonsignificant.

In her review, Scott (1981, 490) concluded that test-retest reliability was approximately +.70 as a measure of correlation. In a study where the same children did multiple Harris-Goodenough Drawing test drawings on different days, Rubin, Schachter and Ragins (1982, 659) found that in general, drawings done with less of a time interval correlated more than drawings done further apart in time. Collectively measured correlation was +.65 when drawings were done on successive days, but when drawings were set six to eight days apart, correlation dropped slightly to +.54–+.58 (Rubin et al., 1982, 656). This was in line with other studies indicating that successive-day and same-day drawings showed "*dramatically high*" reliability, whereas correlations were lower when more time had passed between drawings (Rubinet al., 1982, 659). Yet they also note, that in their study, successive-day correlations for age groups varied a lot, between +.39 to +.89 and for drawigs done six to eight days apart age-group variations were also between +.31 and +.89 (Rubin et al., 1982, 659). The same applied to order effects regarding scoring, as first-day drawings of each week were scored higher overall, but individual patterns varied wildly (Rubin et al., 1982, 660). In general, although patterns were found, variation was also high when age groups or individuals were analyzed. Rubin, Schachter and Ragins (1982, 660) advise against the dangers of looking only at group means.

Variability in scores between drawings done by the same child measured within age groups rose between ages 6–7, but it did not show a steady increase or decrease, but instead, a zig-zag pattern with variability being lowest at five and ten, and rising between, before, and after (Rubin et al., 1982, 657). In other words, during the ages of 6 and 7, the drawings the children produced varied considerably more from day to day than at age 5 or 10. For the 6-year-olds the mean r (arithmetic mean) for correlations between drawings done on different days was +.52, and the mean standard deviation was 8.72 (Rubin et al., 1982, 656) For 7-year-olds the mean r (arithmetic mean) for correlations between drawings done on different days was .42 and the mean standard deviation as 9.52 (Rubin et al., 1982, 656). The lowest correlations between pairs of drawings in these age groups were +.31 and +.32 correspondingly and the highest were .68 and .45 (Rubin, Schachter & Ragins, 1982, 656). They also found boys drawings to show more variability than girls at all ages except nine and ten (Rubin, Schachter & Ragins, 1982, 657). Overall, Rubin, Schachter and Ragins (1982, 664) conclude that intraindividual variability is probably higher than what was previously recognized.

4.4.5 Differences between my implementation and the original DAM Test

I did not have access to an official booklet for the DAM Test. Therefore the test was conducted on paper. I chose to add the option that participants could identify outside of the gender binary and chose not to compare their performance with a fictional school. The test was translated. I also had to take some freedom in scoring some of the symbols that were not present in Harris' (163, 248–263) manual. I decided that an eye where there was a highlight on a black background counted as the same level of detail as a pupil being depicted separately from the eyeball. A bald head with a beard was counted as an indication of hair. A T-shirt with sleeves separating from the arm (not just separated with a line across the arm but with the shape of the sleeves) counted as longer-sleeved shirts would have. A tank top did not count as a separate item of clothing if it did not separate itself from the outline of the trunk.

4.5 Participants

4.5.1 Overview

A total of 71 children from southern and northern Finland participated in this study. Participants were gathered from three pre-primary schools, two 1-st grades and one joint 1-2-grade class. Finnish children attend pre-primary school's last year, named *esikoulu* in Finnish, the year they turn 6. 1st graders in Finland start school the year they turn 7. All schools were in either southern or northern Finland, with the study spanning four different cities. The schools and pre-primary schools were obtained by looking at lists of schools and pre-primary schools of cities close enough to where I lived at the time (one location north, another location south). Permission was obtained from the city, the school, parents and participants themselves. This participant pool is grouped in two different ways.

4.5.2 Full sample

Table 2: Distribution of boys and girls in the full age range (5:9 to 8:11 years).

	Boys	Girls
n	29	42

First, when I compared how the same participants performed in all three tests, I used a dataset with participants aged 5:9–8:11. I used this same dataset for comparison with Harris' (1963, 294 – 297) standardized averages of the Draw-A-Man Test (DAM). The youngest participant in this grouping was 5 and 9 months old, and the oldest 8 years and 11 months. The mean age was 7 years, with the median and mode being 7. The standard deviation was .66, which reflected that the data was gathered mostly in pre-primary schools and 1st grades. Out of the 5- to 8-year-olds 59% (*n* 42) were girls and 41 % (*n* 29) were boys (Table 2). No participant chose another way of expressing their gender.

Table 3: Sample sizes in the full age range (5:9 to 8:11 years).

	Balls test	Corridor tests	DAM Test
Sample	66	67	69
Missing	5	4	2
Total	71	71	71

The total amount of participants in the full sample was 71. Out of those 71, 4 were not present to do the Balls Test, and 5 were not present to do the Corridor Tests. Because of this the sample for the Balls Test is 66 and the sample for the Corridor Tests is 67 (Table 3). All participants completed the Draw-A-Man Test except for two whose data I excluded, the first being an outlier going beyond the standardization scale and the second being a participant with their hand in a cast. The final number of participants in the DAM Test was 69 (Table 3).

4.5.3 Partial sample

Secondly, when I compared the results in the perspective tests with those of Littleton (1991), I left out of the dataset children who were younger than 6 and six months or older than 7 and eleven months. This was also the age range that (1991,60 ,68 ,78 ,82) used in the Experiments that had conditions I tried to replicate. The youngest participant in this grouping was 6 years and 6 months old, and the oldest 7 years and 10 months. The oldest differs from Littleton's as there wasn't a child 7 years and 11 months of age in my sample. The mean age was 7 years and 2 months, with the median and mode being 7. The standard deviation was .34, which is in line with this sample being narrower, as a smaller standard deviation indicates that ages were closer together. In this partial sample 40% (*n* 19) were boys and 60% (*n* 29) were girls (Table 4).

Table 4: Distribution of boys and girls in the narrower sample (6:6 to 7:10 years).

	Boys	Girls
n	19	29

The total amount of participants in this narrower sample was 48. Out of those 48, 3 were not present to do the Balls Test, and 4 were not present to do the Corridor Test. Because of this the sample for the Balls Test is 45 and the sample for the Corridor tests is 44 (Table 5). All participants completed the Draw-A-Man Test except for two whose data I excluded, the first being an outlier going beyond the standardization scale and the second being a participant with their hand in a cast. The final number of participants in the DAM Test was 46 (Table 5).

Table 5: Sample sizes with the narrower age range (from 6:6 to 7:10 years)

	Balls test	Corridor tests	DAM Test
Sample	45	44	46
Missing	3	4	2
Total	48	48	48

5. Results

5.1 Relationship between results in all three tests

5.1.1 Statistics

First, I set out to answer the following research question:

1. Do children get similar results in perspective tests as they do in a Human Figure Drawing Test?

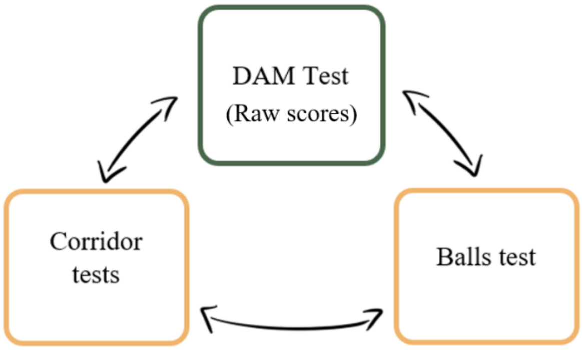


Figure 8: To answer my first research question I compared performances in all three types of tests

For this I needed to analyze results in the Balls test, the Corridor tests (both Drawing and Selection) and the Draw-A-Man Test (Figure 9). For these statistical tests I used the grouping with participants aged 5 to 8 and did not include any of Littleton's (1991) test results.

Table 6: Results in the perspective tests (full sample size, 5:9 to 8:11 years).

	Balls test	Corridor Selection test	Corridor Drawing test Judgement	Corridor Drawing test AM
No perspective	27	53	61	38
Yes, perspective	39	14	6	29
Total	66	67	67	67

Perspective use was measured with the Balls test showing use of occlusion as a feature of perspective, and the Corridor showing use of linear perspective as a feature of perspective. In the Corridor tests, there was the Selection test, which had participants choose a representation of the hallway from three line drawings. The Corridor tests also had a Drawing test, which had the participants draw the hallway. The drawing was measured in two ways, first I judged whether the child had intended to portray linear perspective (Judgement score) and then I measured the convergence of the sides with a ruler (Actual Measurement Scoring System, AM).

5.1.1.1 Ungrouped results

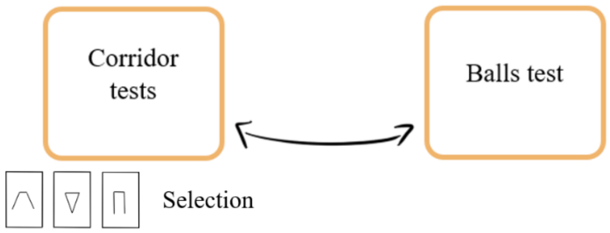


Figure 9: First I investigated the relationship between the Corridor Selection test and the Balls test

Table 7: Crosstabulation of the results between the Balls test and the Corridor Selection test (full sample size, 5:9 to 8:11 years).

		Corridor Selection test		
		No linear perspective	Yes, linear perspective	Total
Balls test	No occlusion	21	4	25
	Yes, occlusion	27	10	37
Total		48	14	62

Before grouping the continuous variables, I set out to analyze the more fine-grained data I had. First, I investigated the relationship between the Corridor Selection test and the Balls test (Table 7). A Chi-squared test between the Corridor Selection test and the Balls test gave a significance of $p > .05$ (two-sided, $p = .308$). This means that there was not a statistically significant association between the Corridor Selection test and the Balls test.

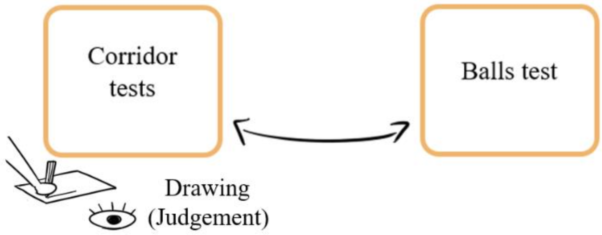


Figure 10: Second, I investigated the relationship between the Corridor Drawing test as measured by judgement and the Balls test

Table 8: Crosstabulation of the results between the Balls test and the Corridor Drawing test as measured by my own judgement of the participant's intent (full sample size, 5:9 to 8:11 years).

		Corridor Drawing test (Judgement)		
		No linear perspective	Yes, linear perspective	Total
Balls test	No occlusion	24	1	25
	Yes, occlusion	32	5	37
Total		56	6	62

Second, I investigated the relationship between the Corridor Drawing test as measured by judgement and the Balls test (Table 8). For this comparison there was not enough observations for me to perform a Chi-squared test. Littleton (1991, 75, 112, 114, 116) had used Fisher's exact test when comparing her own small datasets. Fisher's exact is a test that is often used for small samples with dichotomous variables (Statistics Solutions, 2022b).

A Fisher's exact test between the Corridor Drawing test as measured by judgement and the Balls test gave a significance of $p > .05$ (two-sided, $p = .39$). This means that there was not a statistically significant association between the Corridor Drawing test as measured by judgement and the Balls test.

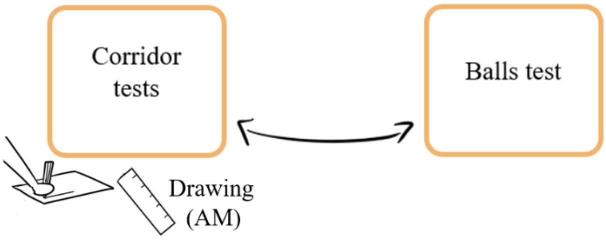


Figure 11: Third, I investigated the relationship between the Corridor Drawing test as measured by the Actual Measurement Scoring System and the Balls test.

Table 9: Crosstabulation of the results between the Balls test and the Corridor Drawing test as measured by the Actual Measurement Scoring System (full sample size, 5:9 to 8:11 years).

		Corridor Drawing test (AM)		
		No linear perspective	Yes, linear perspective	Total
Balls test	No occlusion	16	9	25
	Yes, occlusion	19	18	37
Total		35	27	62

Third, I investigated the relationship between the Corridor Drawing test as measured by the Actual Measurement Scoring System and the Balls test (Table 9). For this comparison a Chi-squared test between the Corridor Drawing test as measured by the Actual Measurement System gave a significance level of $p > .05$ (two-sided, $p = .32$). This means that there was not a statistically significant association between the Corridor Drawing test as measured by the Actual Measurement System and the Balls test.

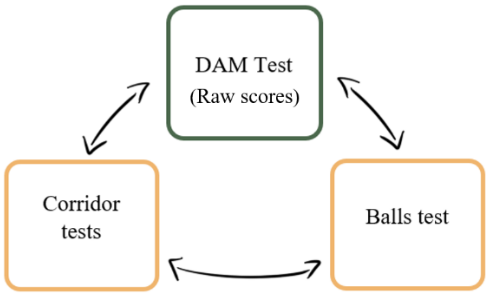


Figure 12: I chose the raw DAM Test scores to compare with the perspective test results since raw DAM Test scores were not distorted by age or gender

I chose to compare perspective test results to the raw scores instead of the standardized scores of the Draw-A-Man Test because like in the Balls and Corridor tests, the gender and age of the participant does not modify the raw score. A child will get the same raw score from eyes, number of fingers et cetera regardless of age or gender. On the other hand, passing this score through the standardization table would yield different results were the child a boy, a girl, eight or six. The mean for the raw scores was 22, the median 22 and there were several modes (15, 23, 25, 26 and 28). The standard deviation was 6.70. The lowest score was 9 and the highest score that did not go off the scale for the standardization was 40.

In order to compare a dichotomous variable with a continuous one I would need to use the Point-Biserial Correlation. There are five assumptions that my data would need to be in accordance with, in order for me to use this test (Lund Research Ltd., 2018b). First, one of my variables was continuous (DAM Test raw scores). Second, another variable was dichotomous (Balls test and Corridor test, each compared separately to the DAM Test raw scores). Third, I had removed outliers. In this case I removed one score that went off the scale and one from a participant who had a broken arm. Fourth, my continuous variable was approximately normally distributed. Lastly, my continuous variable had equal variance for each category of the dichotomous variable. These criteria are all from a website detailing statistic methods (Lund Research Ltd., 2018b).

For the fourth assumption the DAM raw scores gave a significance of $p > .05$ ($p = .37$) when I examined them with the Saphiro-Wilk test of normality. This means that there was a statistically significant likelihood that the data is normally distributed and didn't pool in one end or in the other, but in the middle (Lund Research Ltd., 2018b). A histogram of the results gives us a more visual idea, as the highest frequencies are close enough to the raw DAM Test score mean of 22 (Figure 14). There is also not a clear rise in frequencies at either end. For the last assumption on the list, it is generally accepted that roughly equal sample sizes should have roughly homogenous variances (van den Berg, 2022b).

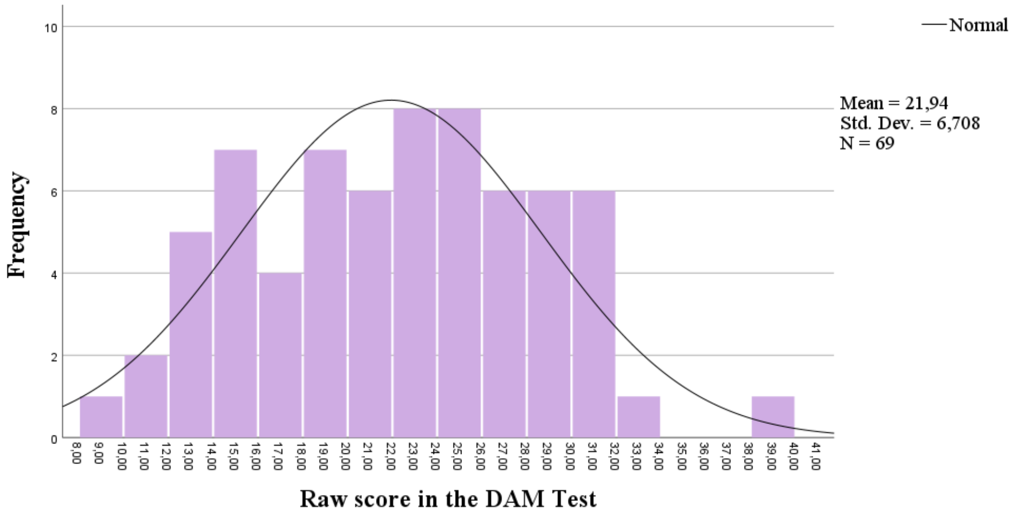


Figure 13: Histogram of the DAM Test raw scores, which were approximately normally distributed (full sample size, 5:9 to 8:11 years).

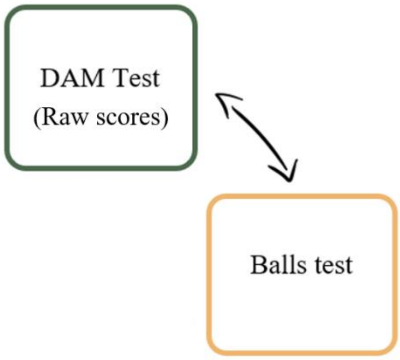


Figure 14: First, I set out to study the relationship between the DAM Test raw scores and the Balls test.

First, I set out to study the relationship between the DAM Test raw scores and the Balls test. Running a Point-Biserial Correlation between these two gave a Pearson correlation coefficient of +.39, in which $p < .02$. This means that the two variables are linearly related in a highly statistically significant way. As it is above .3 the effect size is moderate (Psychology.emory.edu, n.d.).

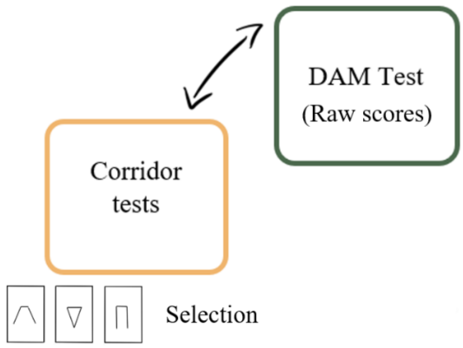


Figure 15: Second, I set out to study the relationship between the DAM Test and the Corridor Selection test

Second, I set out to study the relationship between the DAM Test raw scores and the Corridor Selection test. Running the Point-Biserial Correlation between the Corridor Selection test and the DAM raw scores gave a Pearson correlation coefficient of +.009. This is very close to 0, meaning that these two variables did not show a linear relationship.

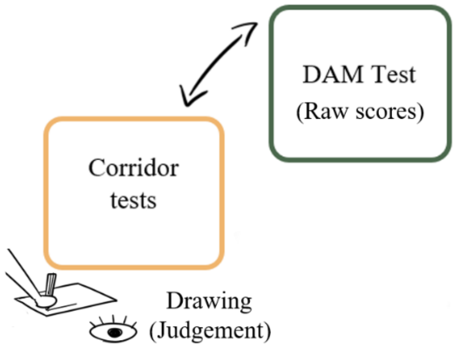


Figure 16: Third, I set out to study the relationship between the DAM Test raw scores and the Corridor Drawing test as measured by my judgement of the participant's intent.

Third, I set out to study the relationship between the DAM Test raw scores and the Corridor Drawing test as measured by my judgement of the participant's intent. Running a Point-Biserial Correlation between these two gave a Pearson correlation coefficient of –.19 This is not a significant correlation as it is close to 0. More curiously it is negative. This set me to explore the low and high ends of the DAM scores later in this thesis.

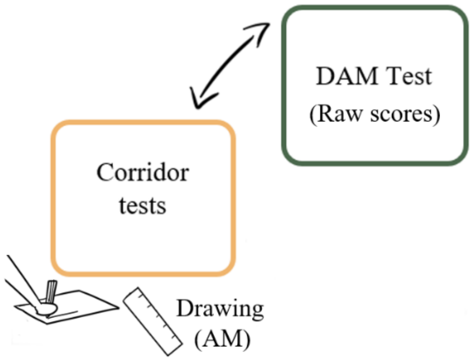


Figure 17: Fourth, I set out to study the relationship between the DAM Test raw scores and the Corridor Drawing test as measured by the Actual Measurement Scoring System.

Fourth, I set out to study the relationship between the DAM Test raw scores and the Corridor Drawing test as measured by the Actual Measurement Scoring System. Running a Point-Biserial Correlation between these two gave a Pearson correlation coefficient of –.03. It is very close to 0 and therefore not statistically significant. It is also negative, which is to be expected after the judgement score was negative too, since the drawings being judged were the same as in the previous paragraph and therefore flow in the same "direction". If I judged a drawing to include linear perspective, it was always also judged to have linear perspective in the Actual Measurement Scoring System.

5.1.1.2 Grouped results

In order to compare the continuous variable of the Harris-Goodenough DAM Test raw score with two categorical variables with the Chi-squared test and other categorical tests, I first split it in two in the middle, so that I could do statistical analysis with groups of around 30. I will also be looking at the highest performing and the lowest performing later in this chapter. The cutoff point for the dividing of the raw DAM score into the *Split DAM Test raw score* variable was 22, with those scoring 22 or lower getting the label *Lower raw score* and those scoring 23 or higher the label *Higher raw score*. This splits my data into 51% (Lower raw score, *n* 35) and 49% (Higher raw score, *n* 34).

For the Corridor test, I first categorized the data so that the Selection Test and the two judgements of the drawing (Actual Measurement System and Judgement by me), were collapsed into one. Only those who selected the perspective representation *and* drew it and were judged to have drawn it by both measures scored the label of *Yes, Total Perspective Representation*. Those who did not meet all of these criteria were labeled *No Total Perspective Representation (TPR)*. I will also be looking later in this chapter at those who selected the perspective representation in the Selection test and either did or did not draw it, as there were very few who drew the perspective representation. This made it so that only 5 participants had a Total Perspective Representation score that was positive, indicating that they had used linear perspective according to all measures.

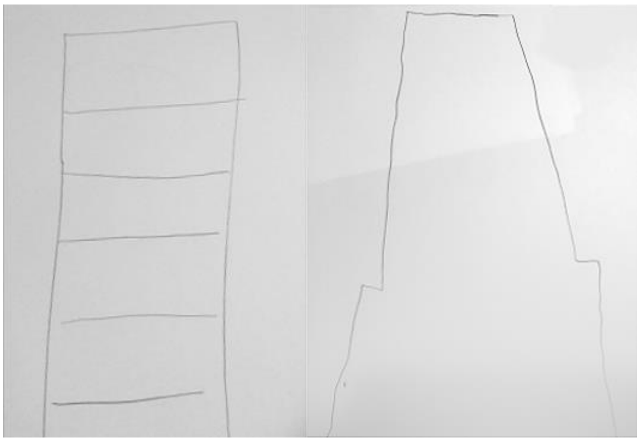


Figure 18: An example of a drawing without linear perspective and a drawing with linear perspective where the judgement score and the AM score were in agreement. Drawings are by participants who gave permission for their drawings to be shown anonymously.

The Balls test did not need any categorization. The test results were only measured in one dichotomous variable. The drawing of the balls either had occlusion or it did not.

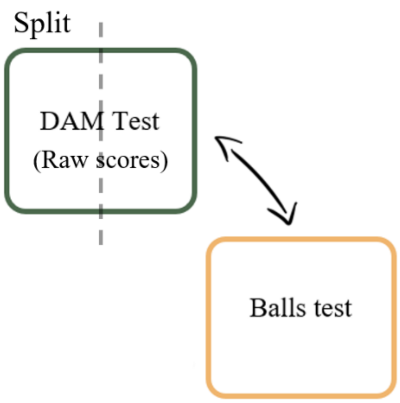


Figure 19: First, I set out to study the relationship between the Split DAM Test raw scores and the Balls test.

Table 10: Crosstabulation between Split DAM raw score data and the Balls test results (full sample size, 5:9 to 8:11 years).

		Split DAM Test raw scores		
		Lower raw score	Higher raw score	Total
Balls test	No occlusion	18	9	27
	Yes, occlusion	15	22	37
Total		33	31	64

First, I set out to study the relationship between the Split DAM Test raw scores and the Balls test (Table 10). Comparing these two with the Chi-squared test yielded a significance of $p < .05$ (two-sided, $p = .04$). The two scores are dependent of each other at a statistically significant level.

I also tested what the Kappa statistic would be between the Balls and the Split DAM raw scores. If we go by the assumption that both the Balls and the DAM Test are measuring cognitive skill, then they could be considered different judges. The Kappa statistic yielded a kappa coefficient of .25. Being between 0.21–0.40, it can be rated as a "fair" amount of agreement.

Another way of measuring strength of association for categorical data are Cramer's V and φ (phi). Cramer's V is used more for any categorical data but φ (phi) is optimal for dichotomous data (Coolican, 2019, 582; PsychED, 2016). SPSS usually calculates both from the same button. Both Cramer's V and φ for the Balls test and the Split DAM raw scores had a value of .26. A Cramer's V lower than .3 when the smallest row or column is 2 meaning that the degrees of freedom are 1, is considered small as by Cohen's original interpretation (Coolican, 2019, 530). Likewise the φ, being below .3, is considered to represent a small effect as well (Zaiontz, 2022).

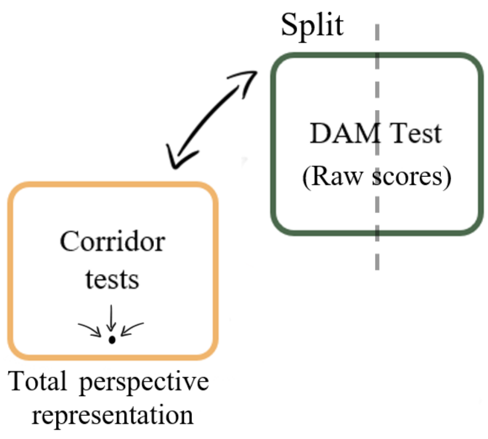


Figure 20: Second, I set out to study the relationship between the Split DAM Test raw scores and Total Perspective Representation in the Corridor tests

Table 11: Crosstabulation between Split DAM raw score data and the Total Perspective Representation from the Corridor tests (full sample, 5:9 to 8:11 years).

		Split DAM Test raw scores		
		Lower raw score	Higher raw score	Total
Corridor tests	No TPR	29	31	60
	Yes TPR	3	2	5
Total		32	33	65

Second, I set out to study the relationship between the Split DAM Test raw scores and Total Perspective Representation in the Corridor tests (Table 11). Doing a Fisher's exact test between these two yielded a significance level of $p > .05$ (two-sided, $p = .67$). This means that there was not a statistically significant association.

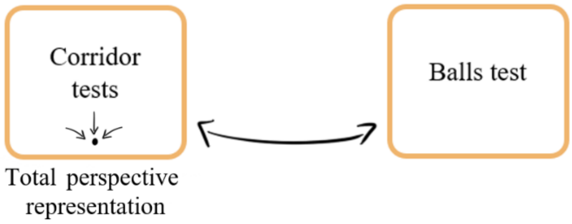


Figure 21: Third, I set out to study the relationship between the Balls test and Total Perspective Representation in the Corridor tests

Table 12: Crosstabulation between the Balls test and the Total Perspective Representation from the Corridor tests (full sample, 5:9 to 8:11 years).

		Corridor tests		
		No TPR	Yes, TPR	Total
Balls test	No occlusion	25	32	57
	Yes, occlusion	0	5	5
Total		25	37	62

Third, I also looked at the Total Perspective Representation in the Corridor tests results and the results from the Balls tests (Table 12). A Fisher's exact test gave a significance of $p > .05$ (two-sided, $p = .08$). This means that there was not a statistically significant association between the Corridor Drawing test Total Perspective Representation and the Balls test.

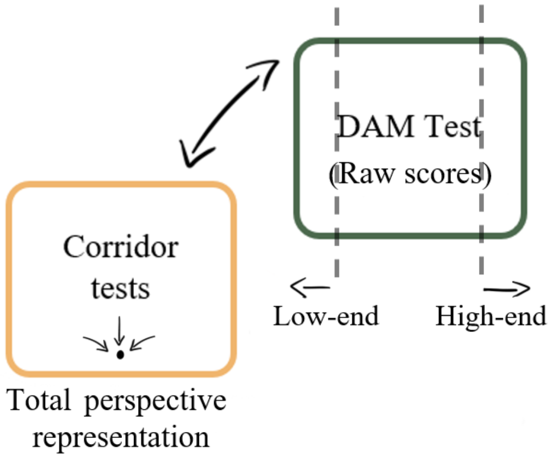


Figure 22: The negative relationship between the Corridor tests and DAM Test scores set me to investigate if there was a relationship between the extremes (full sample, 5:9 to 8:11 years).

Table 13: Crosstabulation between the Total Perspective Representation from the Corridor tests and High- and Low-end scores in the DAM Test (full sample, 5:9 to 8:11 years).

		DAM Test raw scores		
		Low-end raw score	High-end raw score	Total
Corridor tests	No TPR	12	2	14
	Yes TPR	10	0	10
Total		22	2	24

I also looked at the highest and lowest raw scores in the DAM (Figure 23), cutting higher-end scores at 29 and up, and lower-end scores at 15 and lower. I chose this because the standard deviation of the raw scores was 6,7, rounded up that means that the cutting off points were one standard deviation away from the mean (22) in both directions. Comparing the Low- and High-end scores with the Total Perspective score (Table 13), with the Fisher's exact test gave a significance of $p > .05$ (two-sided $p = .50$). The two variables were not related to each other in a statistically significant way.

Table 14: Individual data on the 5 participants who used linear perspective by all three measures (full sample, 5:9 to 8:11 years).

Participant	Gender	Age (y:m)	Balls test score	First Corridor test	DAM Test score
1	girl	7:5	Occlusion	Selection	26
2	boy	7:4	Occlusion	Selection	25
3	boy	7:7	Occlusion	Selection	19
4	boy	6:4	Occlusion	Selection	15
5	boy	7:4	Occlusion	Selection	14

Finally looking individually at those 5 cases where the participants drew and selected perspective representations in the Corridor tests there is the following pattern (Table 14). All five participants drew the balls in the Balls test as occluding. That means perspective, in the form of linear perspective (Corridor test) and occlusion (Balls test) was present in the results of both tests. The results of the DAM are intriguing, as the scores are not particularly high (26, 25, 19, 15 and 14). Participants were not particularly old either, since all were between 6 years and 4 months old and 7 years and 5 months old. With the exception of one girl, all were boys.

Although not all the children who did the Selection Test first chose and drew representations with perspective, out of those that did draw and select the perspective representation all had done the Selection Test first. This means that only those who had done the Selection Test first and chosen there a representation in one point perspective also drew the same representation later. Showing the children different options on how to draw the corridor might have influenced children to draw the representation in one-point-perspective. However, that is not a conclusion I can make within this master's thesis. For comparison 3 participants did the Selection Test first, chose a representation in perspective, but did not draw it on paper. 2 participants did the Selection Test first, selected the converging representation, and drew the convergence so that it was not visible to the naked eye, but only when measured with a ruler (Actual Measurement

System). Of those who did the drawing task first only 4 chose the representation in perspective in the Selection Test.

Overall, below a quarter of all participants selected the perspective representation (*n* 14, 21%), and selecting it seemed a precursor to drawing the corridor in perspective. Although the distribution in the Actual Measurement System scores was fairly even, 43% (*n* 29 in linear perspective) and 57% (*n* 38 no linear perspective) I trust this measure the least, as a convergence of a couple millimeters is easy to make on accident.

5.1.2 Interpretation of results

Put shortly, beyond a small but statistically significant relationship between the DAM Test scores and the Balls test, there was no other association between the tests. Relationships and the lack of relation tell us both interesting things to consider. There are some patterns that showed up without statistical significance as well.

It might be, that the results for the Corridor tests and any other variable are not as relevant because there were so few that drew and selected in perspective. This could indicate a floor effect, where a measure produces too many "low" scores by being too "difficult" (Coolican, 2019, 434), although it is questionable to consider a perspective representation as "pass" and one without a "fail". Both forms after all, do express to another person that there is a corridor. As representations they are both valid in our everyday world. This floor effect would support the idea that occlusion as a feature of perspective and linear perspective might show up in drawings at very different times, with linear perspective showing much later after occlusion shows. This is in line with what I concluded in the review of stage theories. In Lark-Horovitz and Norton's (1959, 435, 441) study, where they separated the 4th stage of *True-To-Appearance* and the 5th stage of *Perspective*. In the chart they present, the 5th stage only starts appearing in a very small portion of drawings whereas the 4th stage has a sharper rise and begins to be seen earlier (Lark-Horovitz & Norton, 1959, 442). In Littleton's (1991, 151–154) Experiment nine, all adolescents and adults used perspective, so it can be assumed that perspective drawing would emerge at some point. All those who did use linear perspective in all three Corridor tests also did use occlusion in the Balls test. If this test were done on adolescents, there might be a relationship between the two perspective tests. However, those who did use perspective were not particularly old, considering that the oldest child in the sample was almost nine.

It could also be that the test for occlusion as a measure of perspective and the test for linear perspective were measuring two completely different things and the construct of *perspective* is flawed. This was a view evident in Kindler's (2004, 234) writing as she told from experience that there is not a clear progression from one form of perspective to another. Other research, such as a study of 7–10-year-olds perspective drawing with different models that matched the retinal image more or less by Lange-Küttner (2014), have also called into question that perspective might not develop in the way stage theories suggest.

In the context of communication, occlusion and linear perspective could be considered simply different symbols unrelated to each other and used in contexts where they clarify meaning. In this sense, occlusion would have been more socially relevant than linear perspective. The concept of *behind* was in the wording of the Balls test, and showing it clearly is emphasized with occluding the ball that is behind. On the other hand, the Corridor tests did not include such adjectives that would demand linear perspective. I echo here Salminen's (Salminen & Koskinen, 2005, 112) views of the unmoving eye and the optic array in art, since drawing something *as it looks* does not necessarily mean viewing it from one eye to these participants. A corridor in perspective still "looks" rectangular because we move around in it. Participants also depicted a lot of the other features of the corridor floor, such as patterns, doorways and texture. Presumably they did this to communicate that the rectangle is not just a rectangle, but specifically a corridor. They also depicted a lot of the details from the balls, such as seams and shadows.

I'm not sure how to interpret the relationship between the DAM Test and the Corridor tests. It would seem that one had no bearing whatsoever on the other. The DAM Test does not contain measures of linear perspective, although it does include foreshortening. A drawing of a human simply is not shaped like a drawing of a long space going off into the distance. On the other hand, there is kind of a similarity between the two balls and the human figure. The DAM has several items that are either scored or not depending on whether the child has depicted things as opaque. This would explain the significant yet weak relationship between the DAM Test and the Balls test. The DAM has its own occlusion test built inside of it, but it is only a small part of the larger score. They are measuring the same thing, occlusion, one more precisely than the other.

5.2 The relationship between the DAM Test 1963 standardization and results in 2021/2022

5.2.1 Statistics

The second research question I chose to address was:

2. Do children in 2021/2022 get higher scores on average in the Harris-Goodenough Draw-A-Man Test than they did before?

This research question merited I use all the scores I have at hand, so I used the larger sample of participants aged between 5:9–8:11 years. I also excluded Littleton's (1991) results as there was no scores for the same participant in both perspective tests and the DAM Test in her results, since she did not investigate the DAM Test.

Figure 23: To answer my second research question I compared my (2021/2022) results with the standardization Harris had published in 1963

For this question I chose to look at the standardized scores, since they told how far each individual was from an average score for their age and gender (Harris, 1963, 294–296). Harris' (1963, 293) standardized score table has a normal distribution for each age group and for girls and boys separately. This means that his mean was 100, and the standard deviation is 15 (Harris, 1963, 293). His sample size was 2,975 US-children which represented the occupational distribution of the country in 1950 by their father's occupation (Harris, 1963, 100). Beyond this, I had no individual cases.

I was surprised to find that after passing all raw scores through Harris's (1963, 294 – 297) table to convert them into standard scores my mean was 100,14. This is very close to Harris' (1963, 293) mean of 100. The standard deviation for the standard scores I obtained was 12,99, which is a bit more uniform than Harris' 15. However, there are no dramatic differences. In other words, the scores I obtained in my study would be considered fairly average even in the 1950's US. The lowest score I obtained was 72, and the highest score went off the standardization scale provided by Harris and was not included. This last one going off the scale is certainly a score differing from Harris', but it is only one case. The highest score within the scale was 136.

A more precise test I performed, by the suggestion of Silén, was the one-sample *t*-test. After all, Harris (1963) aimed to get the average score of a population, not a small sample, for his standardization. The one-sample *t*-test determines whether your sample could have been picked from a population in the sense that their means are close enough (Lund Research Ltd., 2018a) There are four assumptions the data must meet for this test (Lund Research Ltd., 2018a). First, the data I compared to the population mean must be interval or continuous. Second, the data are independent. Third, outliers have been removed. Fourth, the variable compared to the population should be approximately normally distributed. The fourth criteria was met earlier in this master's thesis with the raw scores and should translate well enough into the standard scores. These were all obtained from a website outlining the process (Lund Research Ltd., 2018a).

Comparing the Standardized scores, I obtained in 2021/2022 to a population mean of 100 yielded a significance of $p > .05$ (one-sided, $p = .46$). This means that there was not a statistically significant difference between mine and Harris' scores.

Since both mine and Harris' (1963) samples were higher than 30 and I knew the variance in both I did a *Z*-test (Kumar, 2022). I calculated this by hand, and it was checked by Professor Pekka Vasari.

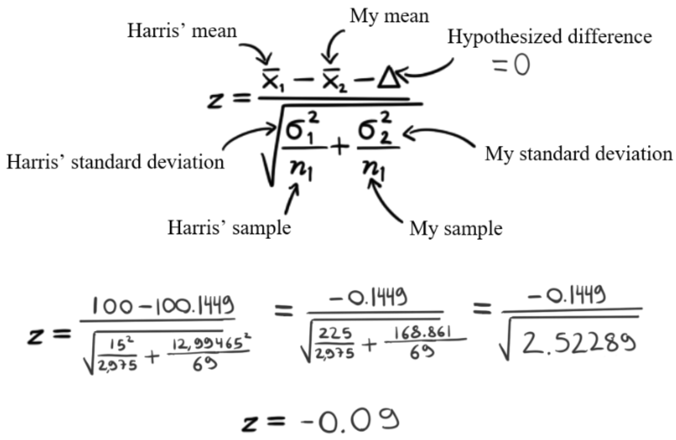


Figure 24: Calculation of the z score

My *z* score was -.09 (Figure 25) which matches a value of .5359 on a *z* score table. This gives us a one-sided probability of $p = 1 - 0.5359 = 0.4641$. Because $p > .05$ there was not a statistically significant difference between mine and Harris' (1963) scores. This should be obvious as well if we round up the mean in my results to 100, giving us a difference of means of 0, which divided by any number stays 0 and gives us a nonexistent chance that the two scores differed in any meaningful way.

There were some differences between boys' and girls' scores, which I felt the need to observe since Harris (1963, 294 – 297) had divided his own data into genders. The boys' mean was 93.40 (standard deviation 12.80) and the girls' 104.76 (standard deviation 11.07). Conducting an independent samples *t*-test in SPSS on these two groups gave us some numbers to estimate the significance of this difference. First, standard deviations had a difference with a significance of $p = .274$ ($F = 1.217$) according to Lavene's Test for Equality of Variances. This means that the variance in the two groups was similar enough for our purposes, as $p > .05$. The two groups are clearly sampled from the same population and are comparable with the assumption that there is about the same amount of variance in scores in both the boys' and girls' groups. An example of a different variance would be if all the boys got a score of 94 but the girl's scores differed wildly. This was not the case. As for the actual Independent Samples *t*-test the *t* value was 3.93 with the significance of this *t* value falling below .001 ($p < .001$). This means that differences between these two groups were highly statistically significant. Cohen's *d*, which estimates how large the effect of something applied to two different groups is, was .96. By Cohen's own scale, being larger than .8 would mean that the effect is large. However, caution must be exercised as gender is not a controlled element that can be applied to different groups at random inside one study (Coolican, 2019, 137) and therefore the wording of "*the effect of being a boy or a girl*" can be misleading. Nevertheless, there is a difference between groups, although we cannot know from my study's setting what is affecting what. To compare with Harris' (1963, 294–297), an average six-year-old boy is expected to get a DAM-test raw score between 18 and 19 to get a standardized score of 100, whereas an average girl is expected to get a raw score of 20 to get a 100. For a seven-year-old boy 22 raw points scores a standardized score of 99 and for seven-year-old girls this score of 99 comes from 23 points. This would mean that the detailedness of the DAM-drawings has moved into opposite directions for boys and girls and differentiated even more.

Table 15: Crosstabulation between the Balls test and gender (full sample, 5:9 to 8:11 years).

		Gender		
		Girl	Boy	Total
Balls test	No	15	12	27
	Yes	22	17	39
Total		37	29	66

I was curious whether there was a similar pattern of "higher" scores for girls in the Balls test. Therefore, I did a Chi-squared test for boys and girls and the Balls test scores (Table 15). This gave a significance of $p > .05$ (two-sided, $p = .95$). This means that there was no association between scores in the Balls test and gender.

There are some differences in children's drawing of a man I can note in a more anecdotal manner that came up when I was standardizing scores. For starters, the scoring manual did not have any examples of "sparkly" eyes or eyes where the pupil is dilated so that it takes up the entire eye as a black dot with a white highlight. My data did not include any drawings with two noses. A lot of the men depicted had long hair, which was not expected in the standardization manual, and probably reflects changes in fashion and gender roles. Even a few dress-like garments could be seen, which were probably also easier to draw. Changes in fashion were also apparent in the number of T-shirts and the lack of buttoned shirts, belts, and pants with a seam in front. Finnish culture also reflected in the hockey pants of one drawing. You can see some of the features depicted in Figure 26.

Figure 25: Long hair and T-shirts in drawings by children who gave permission for their drawings to be shown anonymously

One of the participants in the pretests asked if they could draw a *Minecraft man*. This means that they wanted to draw a character in the style of Minecraft, a popular game that has block-shaped characters and environments. Although this is not a drawing that I ended up scoring for this master's thesis, having everything be square shaped and lacking details would certainly have influenced the score. It is curious however, that despite changes in popular culture, fashion and gender roles, the drawings scored approximately the same as did drawings sixty years ago from children of the same age. At least, with some lenience in judgement, like considering the highlight in an eye on par with drawing the pupil separately.

Although my study had children of many ages from 5 to 8, I did not study whether the raw scores actually rose from one group to another. This is because the samples in each age group were not equal. I was originally trying to get only children aged 6–7 years in my study, and so my sample is exaggerated around this age. The 5- and 8-year-olds in my study were more the result of doing my studies in different grades and one joint grade with older and younger children.

5.2.2 Interpretation of results

It is almost a given that popular culture will reflect in human-centered drawings such as in the DAM Test. There are certainly different features present than in the 1950's such as sparkly eyes, graphic T-shirts and Minecraft men. This is in line with other feature differences people have found in children's drawings. Salminen (Salminen & Koskinen, 2005, 155) noticed that children no longer draw two noses on people as they did in the drawings collected by Corrado Ricci in 1887. The feature of a profile nose on the side of the face combined with a nose that's in the center of the face has not been present for decades and Salminen speculates it is because of the commonness of photographs which show the human face in only one perspective (Salminen & Koskinen, 2005, 68–69). Harris (1963, 139) noted some differences between his and Goodenough's participants from a previous generation. His children tended to draw the face more poorly in terms of the test, with less emphasis on the three-dimensional shapes of the nose, mouth and chin. The proportions of the legs are also portrayed "less suitably". He also notes that: "*Drawings in the present standardization tend consistently to excel on the presence of arms and trunk, attachment of limbs, correct number of fingers, depiction of hand, head and two-dimensional arms and legs, ears, eye details, chin and forehead. Children in the present group achieve notably higher performance on hair, finger detail, thumb opposition, proportion of the trunk, and coordination in drawing arms and legs*" (Harris, 1963, 139).

In times where was less of what we might call modern media Wilson (2004, 303) proposes something I will call *historical styles* have guided children's drawings. A Russian child named Onfim drew hands that looked like rakes between 1224 – 1238 (Wilson, 2004, 303). The same style of drawing hands was later found in the drawings collected by Sully, a child in Britain born 750 years later, and can be seen in late 19th-century European children's drawings (Wilson, 2004, 303). However, this style is not often found in the drawings of children outside Europe, and because of that, Wilson (2004, 303) proposes children have learned this way of drawing hands from other children in Europe for more than half a millennium. From other comparisons between drawings and paintings produced around the same area and time, Wilson (2004, 305) concludes that "*the self-initiated artlike things children made were products of their time*". Studied further in the future the drawings I have collected here may very well be seen as reflections of the historical style in fashion right now.

Despite the difference in content, the overall detailedness of the drawings seems to be much the same as seventy years ago. Harris (1963, 133) did not expect his test to be applicable across cultures and recommended it to be re-standardized for each culture studied. However, in this master's thesis study in 2021/2022 Finland, children's scores, when taken as a whole, did not differ markedly from those that were obtained in the 1950's in the US. This is in line with the results of Hentilä (2011) who in her master's thesis compared human drawings of children in her study with those of Golomb in 1974. Hentilä (2011, 72) concludes that much of what Golomb noticed was also present in her results. This is in contrast to the findings of Pataky (2020, 87), who also compared drawings of the human figure done in 2016 in her study to those published by Paál in 1974. Pataky (2020, 87) found that drawing skills, as measured by things such as detailedness, proportion and accuracy, were less developed than 40 years prior.

A drastic difference in my study was that of boys and girls. It would seem that girls' drawings have become more complex, whereas boys' drawings have become less so. I would like to distance myself again from the notion that detailedness and accuracy are necessarily good or better. Unless the DAM Test scores truly reflect some sort of lowering in intelligence in boys, a connection I have questioned earlier in this master's thesis, there is no need to consider the simplicity of boys' drawings a negative thing. In their day-to-day life, a stick figure is probably just good enough. I noticed I was making very simple pictures myself in the illustrations of this master's thesis, just so that I could communicate in a fast and clear way. There was also no gendered relationship in the results of the Balls test, but 4/5 of those who drew and selected the Corridor tests in perspective were boys.

In Harris' (1963, 294 – 301) standardization table girls get higher scores on average for the Draw-A-Man Test and the Draw-a-woman test. Goodenough (1926, 56, as cited in Harris, 1963, 103) also found this difference in the previous version of the test and hypothesized it to be because "*girls tend to make more rapid progress through the grades*". Harris (1963, 106) observed some of the progression through grades of different genders and decided to abandon this hypothesis Goodenough presented. Instead, he attributed the difference to another hypothesis Goodenough (1926, 58, as cited in Harris 1963, 106) presented which was that girls have generally more perseverance, "docility" and are more careful with details. I find this hypothesis to reflect gender roles of the time. Harris (1963, 106) also considers that cultural factors may influence girls to practice fine detailed work more than boys, and to be more attentive to people and clothing. Girls' attention to clothing was also something Hentilä (2011, 68) noticed when studying the human drawings of 5- to 6-year-olds.

Harris (1963, 106) also attributes the difference to a more rapid social and intellectual maturation seen in girls. This development speed difference can still be seen, as Lynn and Kanazawa (201, 321) found in their study that at the ages of 7 and 11 girls generally get a higher score in IQ tests, which changes in the same population to a reverse position at age 16, when boys' scores overtake that of girls on average. The means I obtained of 93.40 (boys) and 104.76 (girls) were close to results already obtained in the 80's by Rubin, Schachter and Ragins (1983). They found boy's mean to be 98.4 and that of girls 103.2 in the Draw-A-Person Test (Rubin et al., 1983, 656). From a more detailed perspective, a factor analysis done by Sinha (1970, as cited in Scott, 1981, 487) found that girls were more attentive to detail, whereas boys pay more attention to proportion. Gendered differences may also relate to the subject being a human figure. In a study by Bonoti and Metallidou (2010, 332) pre-primary school and primary school girls outperformed boys in drawing tests where the subject was a man and a house, but boys did a bit better than girls in more complex subjects such as a man in a boat and a tree in front of a house.

I feel it's important to note that all these tests are done in the context of schooling and education and what we culturally appreciate in it. They are not objective markers of good and bad. Intelligence tests, at least in the form of the Binet-Simon test, were created in an education context (Gibbons & Warne, 2003, 9). The DAM is usually done in a school context as well in the studies I have read for this master's thesis. When discussing the results of several studies on the supposedly intrinsic development of drawing Harris (1963, 177) points out that a teacher's influence on their pupils is historically considerable. The gendered difference can imply that we need to look further at how social circumstances may guide art making differently for boys and for girls in a school context.

In contrast to Harris' (1963, 106) evaluations of gendered culture influencing drawing Norema, Pietilä and Purtonen (2010) have published a report on how different genders are treated differently in Finnish schools closer to current times. There are some interactions relating to social control and attention to appearance and detailed patient work, that can be picked from this publication. For example, teachers allowed girls to regulate their own appearance far more, allowing makeup and such, whereas boys' dressing was controlled more by, for example, asking them regularly to remove their beanie caps (Norema et al., 2010, 24). There was also a general tendency in adjectives that labeled girls *proper* and *kind*, whereas boys were usually labelled *bad* and *big* (Norema et al.2010, 21). However more potential for development was seen in boys, whereas girls were thought to be more static (Norema et al., 2010, 21). An example of this is picked from an arts class, where boys got overwhelmingly more feedback than girls (Norema et al., 2010, 20). As noted in chapter 4.4.4, the teacher's supportiveness and permissiveness were related to higher Goodenough (not Harris-Goodenough) scores (Reichenberg-Hackett, 1953, as cited by Rubin et al., 1982, 663). Teachers also seemed to have a predisposition to what areas should interest boys and girls and guided them towards these topics (Norema et al, 2010, 17). Although one might think that heavily gendered problematic interactions have faded in some way in current times, Norema, Pietilä and Purtonen (2010, 36) report that many of the problems they found in how different genders are treated are similar to those already found in 1988 in Finland. I cannot help but reflect them to the remarks Harris made as well and see, that the cultural idea of girl- and boyhood remains much the same in a sense. Although a direction of relationships cannot be found in these tests or reports, I think it is important to further observe how gender and art interact in schools.

5.3 Relationship between Littleton's (1991) results and results in 2021/2022

5.3.1 Statistics

Lastly, I set out to answer the following research question

3. Do children in 2021/2022 draw with features of perspective at a younger age?

Figure 26: To answer my third research question I compared my (2021/2022) results with results Littleton collected in the 1980's and published in 1991

Reaching this part of the master's thesis has changed its meaning since I set out to answer this question. It would seem that the base assumptions set in research questions 1. and 2. are not met. There is not a strong association between all three tests so that scores in one test could reflect in the other tests. The strongest association, between the Balls and the DAM Test raw scores, was still a weak one. Both tests have a score for occlusion, it's just that in the DAM Test it counts as a fraction of the total score. Therefore, the connection is not that surprising. The other base assumption is that the Harris-Goodenough Draw-A-Man Test scores would be different in 2021/2022 and that in turn may or may not reflect in this last research question. Harris-Goodenough Draw-A-Man Test scores in 2021/2022 were essentially the same as in the 1950's, except for some differentiation between genders, which merits further studying.

In any case this last research question presents an interesting perspective on its own, without being tied to Flynn effects or abstract thinking. Children drawing or not drawing perspective tells us a story of what is put on paper, and that can already be related to generational differences or similarities, ideas of social structures and so on.

For this comparison I used the group of participants that was between 6:6–7:10 years old. Littleton (1991, 60, 68, 78, 82) also used this age range in Experiments One through Four and in the child participants of Experiment Nine. For the Balls test the oldest participant in Littleton's (1991, 121) Experiments was in Experiment Nine being seven years and eleven months old, and in all Experiments the youngest participant was six years and six months old. I decided to group my data just as strictly. I also cannot guarantee what the youngest or oldest participant was in the data I took from Littleton's (1991) thesis for the Balls test, as the participants were distributed into different conditions and for example the participant aged 6 years and 6 months might have been in a condition I did not compare to, such as a condition where the balls had faces drawn on them.

Table 16: Results in the perspective tests in 2021/2022 (sample with the narrow age range, 6:6 to 7:10 years)

	Balls test	Corridor test: selection	Corridor test: AM	Corridor test: Judgement
No perspective	19	35	24	40
Yes, perspective	26	9	20	4
Total	45	44	44	44

Figure 27: First, I investigated the relationship between my results in the Balls test, and Littleton's results in a similar test.

First, I investigated the relationship between my results in the Balls test, and Littleton's (1991) results. The condition I had selected from Littleton's test was "*Task 2: The Ball Behind a Ball Task (Occluding)*", which was used in in Experiments One through Four (Littleton (1991, 61). I only used it if it was presented first, so that other conditions presented before it did not skew the results. The total sample size obtained for comparison with the Balls test was 91, and for the Corridor tests the comparison sample size was 10. Below you can see the distribution in mine and Littleton's (1991) results (Table 17).

Table 17: Crosstabulation between the Balls test and similar tests in Littleton's (1991) thesis (sample with the narrow age range, 6:6 to 7:11 years).

		Study		
		Bencomo (2021/2022)	Littleton (1991)	Total
Balls test	No occlusion	19	75	94
	Yes, occlusion	26	16	42
Total		45	91	136

A Chi-squared test of Littleton's (1991) results and my results in the Balls test yielded a two-sided significance of $p < .001$. This means that there a highly statistically significant pattern to be seen in the results. φ was -.41 and Cramer's V .41. The direction of φ is irrelevant here, as there is no "rise" between me and Littleton as categories, it just tells that I labelled the data I collected with 0 and the data Littleton collected with 1. It also tells that the higher number of negative scores (0, no occlusion) was in Littleton's results. With a number higher than .3 and one degree of freedom Cramer's V can be considered medium (Coolican, 2019, 530).

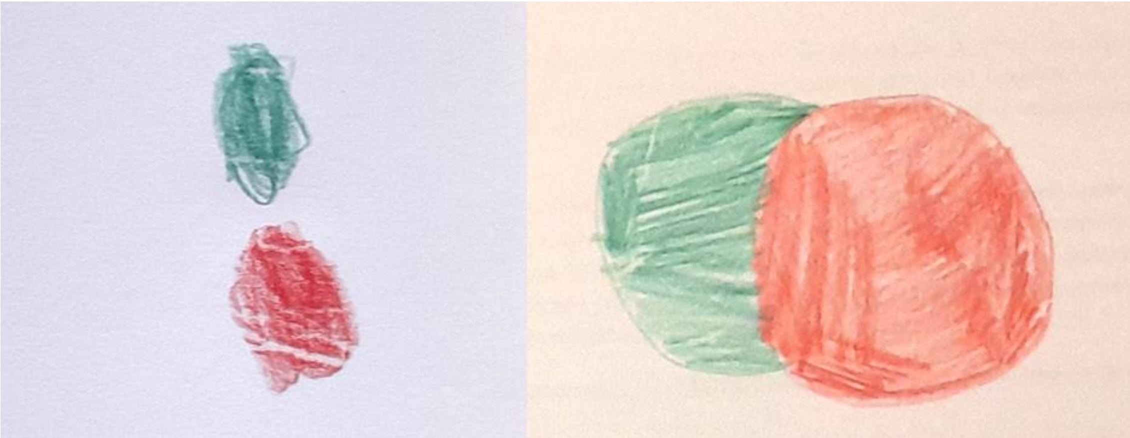


Figure 28: An example of a non-occluding drawing and an occluding drawing by participants who gave permission for their drawings to be shown anonymously.

At this point I wondered about a feature of the test design that would drastically affect how these results should be interpreted. In the Balls test that I was conducting 24% (*n* 11) of the participants in this narrower age range asked whether they should draw both balls from the model, or I had to tell them to draw both after they had finished the first one and said that they were done. This either implies a mistranslation of the directions on my part, or cultural differences in how these kinds of tasks are done in schools. When participants asked whether they should draw both balls, or when I instructed them to do so, it interrupted the task. Littleton (1991, 28) cites Ingram's (1983) and Stapeley and Cox's (1986a) experiments as showing that when the task of drawing two similar objects is interrupted by taking note of the further object (for example, by pointing at it), the participant is more likely to draw the closer object occluding the further one. In other words, more interruption leads to more occlusion. This was a serious consideration for the results in my master's thesis.

Table 18: Crosstabulation between my results in Balls test in 2021/2022 and similar tests in Littleton's (1991) thesis (sample with those who were asked to draw the second ball removed, age range, 6:6 to 7:11 years).

		Study		
		Bencomo (2021/2022)	Littleton (1991)	Total
Balls test	No occlusion	13	75	88
	Yes, occlusion	21	16	37
Total		34	91	125

Because of the issue of interruption, I decided to do a Chi-squared test just as before, but this time excluding all the cases in my results where the participant had to be told to draw the other ball (Table 18). I must confess that I was not as determined to note this variable when I was doing the test, since it came as a surprise, and therefore this variable might be more prone to human error. Deleting cases where there was a question related interruption lowered my sample size to 34 for the Balls test. Despite this the Chi-squared test yielded a two-sided *p* value below .001, which is still highly statistically significant. φ and Cramer's V actually increased a fractional amount with the former being –.43 and the later .43. It would seem that flaw in test design or not, there remains a strong relationship between when the data was collected and occlusion.

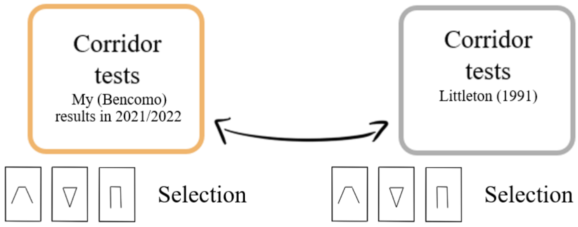


Figure 29: Second, I investigated the relationship between my results in the Corridor Selection test, and Littleton's results in a similar test.

Table 19: Crosstabulation between my results in the Corridor Selection test in 2021/2022 and a similar test in Littleton's (1991) thesis (sample with the narrow age range, 6:6 to 7:11 years).

		Study		
		Bencomo (2021/2022)	Littleton (1991)	Total
Corridor Selection test	No linear perspective	35	10	45
	Yes, linear pespective	9	0	9
Total		44	10	54

Second, I investigated the relationship between my results in the Corridor Selection test, and Littleton's (1991) results in a similar test (Table 19). As before, there were not enough cases for me to do a Chi-squared test with any of the Corridor results, so I used Fisher's exact instead. Comparison between my results in the Corridor Selection test and the Experiment Nine Selection Test had a significance of $p > .05$ (two-sided, $p = .18$) using Fisher's Exact. There was not a statistically significant association between when the tests were done and their results.

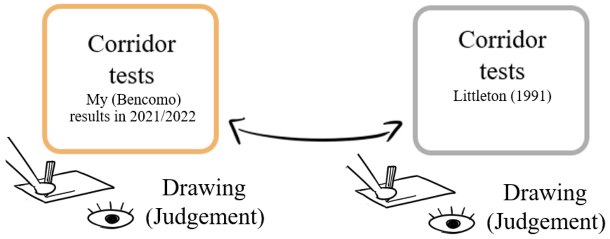


Figure 30: Third, I investigated the relationship between my results in the Corridor Drawing test, and Littleton's results in a similar test as measured by a judge's opinion.

Table 20: Crosstabulation between my results in the Corridor Drawing test as measured by judgement of intent in 2021/2022 and a similar test in Littleton's (1991) thesis (sample with the narrow age range, 6:6 to 7:11 years).

		Study		
		Bencomo (2021/2022)	Littleton (1991)	Total
Corridor Drawing test (Judgement)	No linear perspective	40	10	50
	Yes, linear pespective	4	0	4
Total		44	10	54

Third, I investigated the relationship between my results in the Corridor Drawing test, and Littleton's (1991) results in a similar test as measured by a judge's opinion (Table 20). Comparison between my results in the Corridor Drawing test with my interpretation of the child's intent (Judgement) and the Experiment Nine drawing task as rated by the judges had a significance of $p > .05$ (two-sided, $p = 1$), using Fisher's Exact. This being far above .05 tells that there was not a statistically significant association between the two categories.

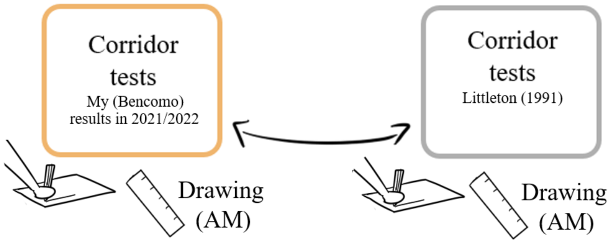


Figure 31: Fourth, I investigated the relationship between my results in the Corridor Drawing test, and Littleton's results in a similar test as measured by the Actual Measurement System score.

Table 21: Crosstabulation between my results in the Corridor Drawing test as measured by the AM System score in 2021/2022 and a similar test in Littleton's (1991) thesis (sample with the narrow age range, 6:6 to 7:11 years).

		Study		
		Bencomo (2021/2022)	Littleton (1991)	Total
Corridor Drawing test (AM)	No linear perspective	24	8	32
	Yes, linear pespective	20	2	22
Total		44	10	54

Fourth, I investigated the relationship between my results in the Corridor Drawing test, and Littleton's (1991) results in a similar test as measured by the Actual Measurement System score (Table 21). Comparison between my results in the Corridor Drawing test with the Actual Measurement System and Experiment Nine drawing task measured with the Actual Measurement System score had a significance of $p > .05$ (two-sided, $p = .173$) using Fisher's Exact. There was not a statistically significant association between the two categories.

5.3.2 Interpretation

There was a statistically significant difference between the tests done in the 1980's and the ones I did in 2021/2022. However, this was not the case for the Corridor test, as results were similar to those Littleton (1991) obtained in the 1980's. This leads me to interpret that there has been a change in how children draw occlusion, but not linear perspective.

This difference can be taken to mean that a developmental milestone (occlusion as a feature of visual realism) is reached by more children earlier than before. This is not without precedent. Harris (1963, 139) also noticed that his participants drew less transparencies than those of Goodenough. Salminen notes how the first stage of drawing, called *scribbling*, has emerged faster and faster in new generations (Salminen & Koskinen, 2005, 69). Salminen speculates that the reason why children pass through the stages of drawing even faster is partially because physical and psychological development overall has also gotten faster (Salminen & Koskinen, 2005, 69).

Effects of arts education on tasks that mimic the view seen from one unmoving eye seem to be more pronounced at earlier ages. Haanstra (1996, 203) compiled from different studies the observation that competence in visual-spatial tasks (such as perceiving, remembering, and manipulating visual forms on a two-dimensional surface) were trainable for children aged 4–6 years using arts education that focused on perception as training. Beyond that age range arts education focused on perceptual training did not seem to make a difference in the visual-spatial skills of the students (Haanstra, 1996, 203). I questioned whether it was possible that perceptual training in art classes has influenced this change in drawing occlusion. A document that gives us a good view of Art Education the 80's UK, where Littleton (1991) did her Experiments, is *A View of the Curriculum* (1980) by the Department of Education and Science. This document recommends that "*Skills of observation, listening and touching need to be developed so that children possess information on which their imaginations can work and be expressed through painting, modelling, music-making, dancing and storytelling*" (Department of Education and Science, 1980, 11). This is still far from the idea of rigorous training of perceptual skills. Even documents from the 1940's UK tell that children will "*draw out of their heads*" and that that is acceptable and should only be gently criticized while letting the children experiment and explore (Ministry of Education, 1946, 11).

The goal of training perception is not central the current Finnish pre-primary school and primary school curriculum either (Opetushallitus 2014a; Opetushallitus 2014b) Rather, the Finnish core curriculum for basic education (Opetushallitus, 2014b, 144) emphasizes Art Education as a way to learn about the worlds of art through participation. The social and cultural side are more important than "*drawing what you see*", though I cannot rule out that the previous sentiment is not present in a lot of the art making ways the students explore. However, it is not central to "*draw what you see*" in every artmaking tradition. In the national core curriculum for pre-primary schools (Opetushallitus, 2014a, 31) multisensory artmaking through many different mediums is a way for children to learn to express themselves, participate and learn to interact with the world and its cultures.

Salminen considers that of far vaster importance than intellectual development is the fact that children are exposed to more pictures and have far easier access to drawing materials, and that *that* is likely the reason why children's drawings look different than before (Salminen & Koskinen, 2005, 69). Children certainly draw as part of their education at this age. Those who have compared three-dimensional and two-dimensional skills express concerns how children are given far more chances to draw than sculpt in Finland and Hungary (Hentilä, 201, 5; Pataky, 2020, 87). According to Golomb (2002, 40), who studied children's drawings, having time to experiment with different graphic solutions plays a role in how drawings look. The more a child gets chances to draw the more diverse their drawing vocabulary will be. Hentilä (2011, 73) also expresses that this experimentation is not individual, as the sculptures in her study were far more heterogenous than drawings. She attributes this to children not sculpting together as often and therefore not using the same ways of expression (Hentilä, 2011, 73). It could be, that using occlusion has become more fashionable in the artmaking sessions the participants in this study have together.

It might be fruitful to study whether children draw together more than before and have more opportunities to spread different fashions. This again, relates to the idea of historical styles and pictures as a language that is learned through copying. Another thing that supports this is that the age at which children start drawing linear perspective has not shown a remarkable change in my results. Perspective itself does not show earlier, but occlusion does. However, this could only be tested by studying the drawings of older children, as it might be that occlusion and linear perspective are linked but show at ages that are very far apart.

Comparing with other studies, it could also be that this change has occurred far earlier near the shift of the new millennia. Morra (2002, 425–426) also tested children’s drawings for occlusion using not only balls but things such as cars and cones as well with a total of 119 children aged 5:11 to 7:10. In some conditions even 87% of the participants used occlusion, and in the balls condition the total percentage of participants using occlusion was 57%, (Morra, 2002, 428). This is particularly close to my result, as 59% of the participants used occlusion in my full sample. Morra had also studied this setting of drawing a ball in front of a ball with children of the same age range earlier in his 1997 publication with Angi and Tomat. In the results the frequency of occlusion was again near the 50% mark in different conditions (Morra et al., 1996, 283–287).

This brings us to the importance of the frequency of occlusion. If half of the children draw with occlusion, be it either because it’s socially acceptable or because it reflects their cognitive skills, it does not change much in how art in art lessons looks. It is not a complete overtaking of all expression by this way of drawing, but rather it’s a marker of the diversity in drawings that children in pre-primary school and primary school can produce. It reflects individual variability. It emphasizes the goal that is already central in Finnish core curriculums. That is, individual students should be met as individuals, with their own needs and goals (Opetushallitus, 2014a, 9, 15, 24; Opetushallitus, 2014b, 10, 15). Where previously children’s use of occlusion might have been more homogenous, now there is a more evenly distributed variety of expression.

However, as I expressed before, I also deem the overview of children's drawings as important, and it is also part of what I have studied as part of my Art Education degree. If we do view occlusion as a marker of visual realism and consider this stage to be reached slightly earlier in this regard, then we need to consider what social factors are related to visual realism. As Salminen (Salminen, & Koskinen, 2005, 47) puts it: "*Drawing at this stage means more disappointments than satisfaction."* Salminen cites Read (1956) and Arnheim (1970) and expresses that at this stage children usually stop drawing because there are other ways to express competency among peers, such as by succeeding in other school subjects that are deemed more important by the school system (Salminen, & Koskinen, 2005, 49). Salminen also cites Nordström (1975) as saying that children stop making art when they notice their pictures are not as good as commercial ones. This is in contrast to the younger mindset where children know their images do not look exactly like the model, or might not be culturally appreciated, but do not care as much and do whatever they want to (Salminen, & Koskinen, 2005, 45; Cox, 2005, 121; Golomb, 2002, 91).

In general, children begin to consider the opinions of others on their behavior at a fairly young age. Psychological research in other actions besides drawing puts this milestone of *"caring what other's think"* far earlier. Children's self-presentation, reputation management and understanding of an audience has generally been thought to develop at age 3–5 (Engelmann et al., 2012; Fu et al., 2016, as cited in Botto & Rochat, 2018, 1732). However, a study by Botto and Rochat (2018, 1732) found by investigating even younger children that this development was present in two-year-olds as well. Perhaps in the future children's confidence in their own drawings could be studied. It would be useful to determine at what age caring about other's opinions on whether the drawing is *good* might start to slow the process of trying and experimenting.

6. Discussion

6.1 Implications for Art Education and Art Education research

It was important for me to study whether children's drawings had changed for two reasons. First, I cannot fully unlearn artmaking, and therefore I must partially rely on theories that outline a probable process of artistic maturation. In my opinion it is easier for me to consider the viewpoints of children who don't have as much experience as me when I know what their drawing process and learning might look like. I cannot fit the world inside my head. Therefore, I must rely somewhat on averages and probabilities. These average paths of artistic development form what is called stage theories, and their most relevant stages here are logical and visual realism. These, in short, are respectively the stages of drawing what you know through ordinary living perception, and drawing what you would see from one unmoving eye like a camera (Salminen & Koskinen, 2005, 43, 121; Cox, 2005, 116).

Second, the shift from logical realism into visual realism brings with it new hardships in drawing, since the child also starts to consider what the world thinks of their drawings (Salminen, & Koskinen, 2005, 49). If this occurred earlier than what had been established, we would need to consider not just how it affected what was shown on paper, but also how it would reflect the child's ability to think differently than previous generations. Genovese (2018) had already studied earlier changes in children's drawings of the human figure. In his findings he associated the change he found, among other things, to a shift in cognitive skills and namely abstract thinking (Genovese, 2018, 181). He suggested that if children master a more abstract form of thinking earlier, then they would also draw in perspective earlier, since he assumed that perspective drawing would involve these more abstract thinking skills (Genovese, 2018, 181).

First, I aimed to answer whether there was any relationship between results in Human Figure Drawing Tests and results in perspective tests. My research question was: "*Do children get similar results in perspective tests as they do in a Human Figure Drawing Test?"* I had chosen the Draw-A-Man (DAM) test by Harris (1963) and Goodenough as my Human Figure Drawing Test, and two setups from Littleton's (1991) thesis as my perspective tests. In my master's thesis the perspective tests are called the Balls test and the Corridor test, and they measure occlusion and linear perspective use respectively. I chose participants aged 6 to 7 years for this thesis so that I could compare my results to those with Littleton (1991). In the process I ended up getting 5-year-old and 8-year-old participants as well.

The answer to the first research question is broadly speaking *"No"* within this study. There was not a relationship between all three tests. When comparing the tests with each other, I found a statistically significant but only moderate connection between the raw scores in the DAM Test and the Balls test. The DAM Test and the Corridor test, or the Corridor test and the Balls test did not seem to have a connection to each other. In fact, some of the only people who drew and selected linear perspective in the Corridor test had fairly low scores in the DAM Test. This implied that the concept of artistic maturation as a unified whole might be flawed. It would seem that getting "better" in one test did not reflect in the others, with the exception of the DAM Test and Balls test. However, these two tests measured the same thing almost explicitly since the DAM Test had some items that considered occlusion. In that sense, the DAM Test is just a less specific occlusion test like the Balls test. This finding is in line with Matthews (2004, 270) suggestion that categorizing features into these larger groups of visual and logical realism might not be the best way to study drawings. The almost nonexistent connection between the two perspective tests also calls into question the entire concept of perspective as a unified whole. This is in line with the criticism presented by Kindler (2004, 234) and Lange-Küttner (2014), who have questioned the linear and stage-like development of perspective drawing.

Perhaps, it would be more fruitful to study individual features. Although, this is already the case from what I have seen from the research. Researchers are already studying occlusion, linear perspective, theme et cetera separately, though they are usually associated with stage theories as a base. We might also be able to find new features if we look beyond those described in stage theories. It might also be interesting to start relating these features in new ways, to find new wholes, without relying as much on the already established stage theories. Stage theories have already been heavily criticized for their implicit biases towards realistic depiction and modern western art (Hamblen, 1993, 45, as cited in Kindler, 2004, 241; Matthews, 2004, 271). If we try to acknowledge these biases and look at drawings from new perspectives, there might be new paths to be found.

It's also possible that the age range and distribution I selected for my study might not be the best to relate occlusion and linear perspective, because very few were drawing in linear perspective this young. However, all of those who selected and drew in linear perspective also used occlusion in the Balls test. Studying older children might reveal a stronger connection between occlusion and linear perspective if more children draw in perspective when they get older. For example, the adolescents and adults in Littleton's (1991, 151–154) Experiment Nine did draw mostly in perspective.

I wanted to see if results in these tests were different in newer generations. All tests presumably also gave a higher number of points or a higher amount of perspective representations in older participants. Therefore, all tests measured a representation that would also change as the child got older.

To really connect my findings with those of Genovese (2018) I would have to see if the rise in scores of Human Figure Drawing Tests he found in earlier generations had continued to date. My second research question was: *"Do children in 2021/2022 get higher scores on average in the Harris-Goodenough Draw-A-Man Test than they did before?"* Genovese (2018, 177) had also reported some other statistically significant and statistically insignificant rises in Human Figure Drawing Test scores by other researchers closer to current times.

The answer to this second research question was a firm *"No"*. I found that my results were remarkably similar to Harris' (1963) standardization, despite containing features reflecting current culture in Finland. This is in line with a similar comparative master's thesis study in Finland by Hentilä (2011, 72), who concluded that the human figure drawings she obtained were similar to those of Golomb in 1974. However, in Hungary the results of Pataky (2020, 87) spoke of less detailedness, proportion and accuracy in human figure drawings by children in current times compared to children 40 years ago.

When I separated my results by gender, boys and girls scores had separated further away from each other than they were in the 1950's. This result was in line with findings by Rubin, Schachter and Ragins (1983, 656), who also found a separation between boys' and girl's scores in the Draw-A-Person Test. This gendered difference did not show in the Balls test, but there were more boys in the Corridor test depicting linear perspective than there were girls. Looking at the results of my study, different subjects might incite differences between boys and girls, which is in line with the findings of Bonoti and Metallidou (2010, 332).

It could be fruitful to research how different treatment of different genders reflects in their artmaking. However, at the same time it's important not to think of one way of making art as better than the other. Difference in interactions, and the possible problematic features of it, should be in focus, not how to push any gender to make "good art" as a narrow concept.

Lastly, I wanted to investigate Genovese's (2018) suggestion that children might draw in perspective earlier. My third research question was: *"Do children in 2021/2022 draw with features of perspective at a younger age?"* For this, I compared my results to those in similar tests by Littleton (1991), who told me via e-mail that she had done the tests in the 1980's. By this point the concept of *features of perspective* had changed its meaning, as there didn't seem to be a unified concept of perspective in this study.

The answer to this third research question is: "*Only partially".* There was a medium sized difference between mine and Littleton's (1991) results in the Balls test, but there was no statistically significant difference between the results of the Corridor test. Morra (2002, 428; Morra et al., 1996, 283) obtained similar proportions of drawing occlusion in children of approximately the same age as my participants. The timing of his studies was not particularly far from that of Littleton's publication (1991). This would either indicate a rapid shift, simple chance difference or that the tests were set up too differently so that for example the language of instruction influenced results since his studies were in another language as well (Italian).

These results do not have major implications for the every-day art class. Overall, they highlight differences in drawings produced by different individuals. There has not been a major overhaul by one way of drawing or the other. If we take occlusion as a feature of visual realism, the results are close to a cut-off age as old as that presented by Piaget and Inhelder (1967, 51), who determined that the first forms of visual realism start emerging at age 7–8. A difference of one year is not major, though it might be important. It would be most fruitful to study more closely the implied changes in the classroom that the different stages may have. Confidence in one's drawings and social competence by making drawings are the factors that ultimately influence how Art Education should be structured to raise confident children who know how to express themselves and make friends.

The answer to my main research question: "*Have children's drawings changed?*" is "*Only a little*". My study noticed only small differences between studies done before. Children's drawings have changed in some ways, but these changes are only small as seen within the scope of this master's thesis. The changes are located in the differentiation between how boys and girls draw the human figure, and the frequency of using partial occlusion as a drawing technique to indicate perspective. Larger concepts and larger datasets are needed to truly gauge whether children's drawings have changed. My results are only a small part of what I hope can be studied someday as a large tapestry. When it comes to the concept of historical styles, we might still be too close to our own time to notice changes and patterns that would be more obvious as seen from the future. We could also study whether children have more opportunities to copy from each other.

6.2 Limitations of the study

I have detailed the differences of my tests to the original ones in their respective sections in chapter 4. There are also many biases which I have presented and partially tried to remove in this master's thesis. Other issues that arise are the generalization of the results and my own inexperience with statistical methods.

My contextualization of drawings as part of human life can go only so far, since the context in which these children made the drawings explored in this master's thesis is not spontaneous or entirely natural. I can tell anecdotally, from my own limited experience working with children of this age in their homes and in their schools, that most drawings I have witnessed have been prompted by adults in some way. This is in line with Pataky's (2020, 89) findings that half of the children in her study prefer create drawings among their family members at home. During this master's thesis, I've come to appreciate the insight qualitative research can bring in a new way. Qualitative material made its way into my material and writing in the form of anecdotes and experiences that might have warranted closer analysis. Unfortunately, closer analysis was not possible within the scope I had for my master's thesis. I believe that in the future employing both qualitative and quantitative research side by side will be most fruitful. Genovese has guided me via e-mail to find more about Bayesian approaches to both. I will be studying this in the future.

I am inexperienced when it comes to statistics. This master's thesis has been a learning experience and I hope that my understanding of different methods to study Art Education deepens in the future as well. I believe that statistics does have a place in Art Education too, as it provided me with exact information, and was a conversation partner of sorts. The numerical results here should be observed with skepticism. Though I have checked and rechecked my calculations and data categorization many times, there are still many underlying mathematical concepts I do not understand. I can provide the fully anonymized dataset upon request via the e-mail sandra.m.bencomo@gmail.com. The dataset also includes a lot of variables I did not end up using, such as whether the drawing had details. Unfortunately, I do not have high quality photos of the DAM Test scoring and standardization table to attach, as the Harris (1963) book is in another city as I am finishing this master's thesis.

The Harris-Goodenough version of the Draw-A-Man test was the only one I had access to, as the manual was at a library I could loan from. If I had to do this master's thesis again, I would choose Naglieri's (1988, as cited in Ganesh, 2012, 217) Draw-A-Person test, since it is the most up-to-date version of a Human Figure Drawing Test that correlates with cognitive measures. This would be a better measure to test the link between cognitive skills and drawings.

7. Conclusions

In this master's thesis, I set out to study whether children's drawings had changed. Since this question was broad, I narrowed it to Genovese's (2018, 181) suggestion that children might draw perspective at an earlier age. I also tested two of the concepts implicit as factors behind a shift in perspective drawing that I interpreted from Genovese's (2018) paper. The first, was that there was a relationship between Human Figure Drawing Tests and perspective drawing. The second of these was that children's drawing of a human figure would also have changed, as measured by a Human Figure Drawing Test. All the tests I used had also been used in previous generations in the 1950's and 1980's. All of the tests are assumed to measure features of drawings that would show differently in younger children than in older children, adolescents and adults. They all yielded numerical results, and therefore I compared results quantitatively. For dichotomous variables I compared results using the Chi-squared test, and when this was not possible, I used the Fisher's exact test. For two continuous variables I used the *Z*-test, and *t*-test. For one continuous and one dichotomous I used the Point-Biseral Correlation. For effect sizes I observed the Kappa statistic when I had two measurements measuring the same concept in the same participant, and Cramer's V and φ for dichotomous variables. For continuous variables, I estimated effects with Cohen's *d*.

Overall, the results in this master's thesis did not tell of any major differences in the way children draw that would warrant immediate changes in classroom practice. Only a medium difference was found between results in tests measuring use of occlusion in 2021/2022 and in the 1980's. In the Harris-Goodenough Draw-A-Man (DAM) Test, results between boys and girls had become more polarized. However, there is no one way of drawing that would have taken over the children's drawings. Rather, results highlight that each child draws differently, and that it's important for an art teacher to know their students.

However, results hint at many possibly fruitful courses for research to follow. Results also bring into question the conceptualizations present in stage theories of drawing, such as *perspective drawing* as a unified whole. The results also did not support that there would be a central variable that would guide results in both the Harris-Goodenough DAM Test and the two perspective tests I chose. However, there was a small relationship between the DAM Test and a perspective test measuring the use of occlusion. This can be explained by the DAM Test measuring occlusion in a few of its items.

References

Aikman, K. G., Belter, R. W., & Finch, A. J. (1992). Human figure drawings: Validity in assessing intellectual level and academic achievement. *Journal of Clinical Psychology, 48*(1), 114–120. https://doi.org/10.1002/1097-4679(199201)48:1<114::AID-JCLP2270480116>3.0.CO;2-Y

Bhandari, P. (2021, June 12). *What Is Quantitative Research? | Definition, Uses and Methods.* Scribbr. https://www.scribbr.com/methodology/quantitative-research/ [Accessed: 14.5.2021].

Bonoti, F. & Metallidou, P. (2010). Children's Judgments and Feelings about Their Own Drawings. *Psychology*, *1*(5), 329–336. https://doi.org/10.4236/psych.2010.15042

Botto, S.V., & Rochat, P. (2018). Sensitivity to the Evaluation of Others Emerges by 24 Months. *Developmental Psychology, 54*(9), 1723–1734. https://doi.org/10.1037/dev0000548

Cohn, N. (2021). Explaining 'I Can't Draw': Parallels between the Structure and Development of Language and Drawing. *Human Development 2012*(55), 167–192. https://doi.org/10.1159/000341842

Coolican, H. (2019) *Research Methods and Statistics in Psychology*. (7th edition). Routledge.

Cox, S. (2005). Intention and Meaning in Young Children's Drawing. *International Journal of Art & Design Education, 24*(2), 115–125. https://doi.org/10.1111/j.1476-8070.2005.00432.x

Department of Education and Science. (1980). *A View of the Curriculum.* Her Majesty's Stationery Office. http://www.educationengland.org.uk/documents/hmidiscussion/viewofthecurric.htm [Accessed: 19.5.2022].

Freeman, N. H. & Janikoun, R. (1972). Intellectual Realism in Children's Drawings of a Familiar Object with Distinctive Features. *Child Development, 43*(3), 1116–1121. https://doi.org/10.2307/1127668

Ganesh, T. G. (2012). Children-Produced Drawings: An Interpretive and Analytic Tool for Researchers. In E. Margolis & L. Pauwels (Eds.), *The SAGE Handbook of Visual Research Methods* (pp. 214–240). SAGE Publications. https://doi.org/10.4135/9781446268278

Genovese, J. E. C. (2018). Evidence of a Flynn Effect in Children's Human Figure Drawings (1902–1968). *The Journal of Genetic Psychology, 179*(4), 176–182. https://doi.org/10.1080/00221325.2018.1469113

Gibbons, A. & Warne, R. T. (2019). First publication of subtests in the Stanford-Binet 5, WAIS-IV, WISC-V, and WPPSI-IV. *Intelligence, 75*, pages 9–18. https://doi.org/10.1016/j.intell.2019.02.005

Gilbert, D., King, G., Pettigrew, S. & Wilson, T. (2016). Comment on “Estimating the reproducibility of psychological science”. *Science, 351*(6277), 1037. https://doi.org/10.1126/science.aad7243

Glen, S. (2016, June 20). *Phi Coefficient (Mean Square Contingency Coefficient).* https://www.statisticshowto.com/phi-coefficient-mean-square-contingency-coefficient/ [Accessed: 13.5.2022].

Glen, S. (2022). *Z Test: Definition & Two Proportion Z-test.* https://www.statisticshowto.com/probability-and-statistics/hypothesis-testing/Z-test/ [Accessed: 13.5.2022].

Golomb, C. (2002). *Child Art in Context. A Cultural and Comparative Perspective.* American Psychological association.

Harris, D. B. (1963). *Children's Drawings as Measures of Intellectual Maturity.* Harcourt, Brace & World.

Hentilä, L. (2011). *5-6 -vuotiaiden lasten ihmisen kuvaaminen; muovailu- ja piirustusilmaisun vertaileva tutkimus [5-6-years old children's human figure representation. A comparative study of the pictorial expression of modelling and drawing].* [Master's thesis, University of Lapland]. https://urn.fi/URN:NBN:fi:ula-201110031169

Kalakoski, V. (2014, October 10). *Onko tilastollisesti merkitsevillä tutkimustuloksilla käytännön merkitystä? [Do statistically significant results matter in actuality?].* Psyologia. http://www.psykologia.fi/maanantaiblogi/313-onko-tilastollisesti-merkitsevillae-tutkimustuloksilla-kaeytaennoen-merkitystae [Accessed: 14.5.2022].

Kindler, A. M. (2004). Researching Impossible? Models of Artistic Development Reconsidered. In E. W. Eisner & M. D. Day (Eds.), *Handbook of Research and Policy in Art Education,* (pp. 233–252). Taylor & Francis Group.

Kozbelt, A. (2017). Learning to see by learning to draw: Probing the perceptual bases and consequences of highly skilled artistic drawing. *High Ability Studies, 28*(1), 93–105. https://doi.org/10.1080/13598139.2017.1298995

Kumar, A. (2022, February 1). *When to Use Z-test vs T-test: Differences, Examples. Data Analytics.* https://vitalflux.com/when-to-use-Z-test-vs-t-test-differences-examples/ [Accessed: 14.5.2022].

Lange-Küttner, C. (2014). Do Drawing Stages Really Exist? Children’s Early Mapping of Perspective. *Psychology of Aesthetics, Creativity, and the Arts*, 8(2), 168–182. : https://doi.org/10.1037/a0036199

Lark-Horovitz, B., & Norton, J. (1959). Children's Art Abilities: Developmental Trends of Art Characteristics. *Child Development, 30*(4), 433–452. https://doi.org/10.2307/1126111

Laukka, M. (2005). Kuvailmaisu ja tulkinta [Image expression and interpretation]. In I. Koskinen (Ed.), *Pääjalkainen [Tadpole drawing],* (pp. 13–15). Taideteollinen korkeakoulu.

Littleton, K. S. (1991). *The Representation of Depth in Children’s Drawings.* [Doctoral dissertation, University of York]. https://etheses.whiterose.ac.uk/4251/1/DX094375.pdf

Lund Research Ltd. (2018a). *One-Sample T-Test using SPSS Statistics.* https://statistics.laerd.com/spss-tutorials/one-sample-t-test-using-spss-statistics.php [Accessed: 13.5.2022].

Lund Research Ltd. (2018b). *Point-Biserial Correlation using SPSS Statistics.* https://statistics.laerd.com/spss-tutorials/point-biserial-correlation-using-spss-statistics.php [Accessed: 13.5.2022].

Lynn, R. & Kanazawa, S. (2011). A longitudinal study of sex differences in intelligence at ages 7, 11 and 16 years. *Personality and Individual Differences, 51*(3), 321–324. https://doi.org/10.1016/j.paid.2011.02.028

Läärä, E. & Seppo, A. (1988). Mittausmenetelmien hyvyys ja sen tilastollinen arviointi [The virtues of measuring methods and their statistical evaluation]. *Duodecim, 104*, 40–52.

Matthews, J. (2004). The Art of Infancy. In E. W. Eisner & M. D. Day (Eds.), *Handbook of Research and Policy in Art Education*, (pp. 253–298). Taylor & Francis Group.

Metsämuuronen, J. (2011) *Tutkimuksen tekemisen perusteet ihmistieteissä [The basics of research in human sciences].* International Methelp.

Ministry of Education. (1946). *Art Education.* His Majesty's Stationery Office. http://www.educationengland.org.uk/documents/minofed/pamphlet-06.html [Accessed: 19.5.2022]

Opetushallitus [Ministry of Education]. (2014a). *Esiopetuksen opetussuunnitelman perusteet [National Core Curriculum for Pre-primary Education].* Ministry of Education. https://www.oph.fi/fi/koulutus-ja-tutkinnot/esiopetuksen-opetussuunnitelmien-perusteet [Accessed: 15.5.2022]

Opetushallitus [Ministry of Education]. (2014b). *Perusopetuksen opetussuunnitelman perusteet [National core curriculum for basic education].* Ministry of Education. https://www.oph.fi/fi/koulutus-ja-tutkinnot/perusopetuksen-opetussuunnitelman-perusteet [Accessed: 15.5.2022]

Morra, S., Angi, A. & Tomat, L. (1996). Planning, Encoding, and Overcoming Conflict in Partial Occlusion Drawing: A Neo-Piagetian Model and an Experimental Analysis. *Journal of experimental child psychology. 61*, 276–301. https://doi.org/10.1006/jecp.1996.0017

Morra, S. (2002). On the relationship between partial occlusion drawing, M capacity and field independence. *British Journal of Developmental Psychology, 20*(3), 421–438. https://doi.org/10.1348/026151002320620244

Norema, A., Pietilä, P. & Purtonen, T. (2010). *Pipot tiedon tiellä. Sukupuolihuomioita opettajankoulutuksessa [Beanie caps in the way of knowledge. Notes about gender in teacher education].* University of Helsinki.

Pataky, G. (2020). »Do not touch it!« Today's Children's Visual Competencies in Comparison with the Pre-Digital Era in Light of their Art Educational Environment. *Art, for Children's Sake! At the Crossroads of Making, Understanding and Teaching Visual Art, 10*(4), 75–96. https://doi.org/10.26529/cepsj.925

Piaget, J. & Inhelder, B. (1967). *The child's conception of space.* (F.J. Langdon & J.L. Lunzer Trans.). Norton & Company. (Original work published in 1948).

PsychED. (2016, January 27). *Tutorial: Effect sizes – Part 3 (Phi, Cramer's V) [Video file].* YouTube. https://www.youtube.com/watch?v=YxLIPoJQ-Os [Accessed: 14.5.2022].

Psychology.emory.edu, (n.d.). *Effect Size.* http://www.psychology.emory.edu/clinical/bliwise/Tutorials/SCATTER/scatterplots/effect.htm [Accessed: 22.5.2022].

Rautopuro, J. & Malin, A. (2008). Miksei asioista puhuta niiden oikeilla nimillä? Tilastollisten menetelmien tarpeellisuudesta kasvatustieteissä [Why are things not called by their real names? The necessity of Statistical Methods in Educational Sciences]. *Kasvatus, 39 (2),* 108–118. https://urn.fi/URN:NBN:fi:ELE-1480167

Riley, H. (2008). Drawing: Towards an Intelligence of Seeing. In S. Garner (Ed.), *Writing on Drawing. Essays on Drawing Practice and Research* (pp. 153 – 167). Intellect books.

Rubin, J. A., Schachter, J., & Ragins, N. (1983). Intraindividual variability in human figure drawings: *A developmental study. American Journal of Orthopsychiatry, 53*(4), 654–667. https://doi.org/10.1111/j.1939-0025.1983.tb03409.x

Salminen, A. & Koskinen, I. (Ed.). (2005). *Pääjalkainen [Tadpole drawing].* Taideteollinen korkeakoulu.

Scott, L. H. (1981). Measuring intelligence with the Goodenough-Harris Drawing Test. *Psychological Bulletin, 89*(3), 483–505. https://doi.org/10.1037/0033-2909.89.3.483

Statistics Solutions. (2022a). *Effect Size.* https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/effect-size/ [Accessed: 13.5.2022]

Statistics Solutions. (2022b). *Fisher's exact test* https://www.statisticssolutions.com/fisher-exact-test/ [Accessed: 13.5.2022]

van den Berg, R. G. (2022a). *Cramér's V – What and Why?* https://www.spss-tutorials.com/cramers-v-what-and-why/ [Accessed: 22.5.2022].

van den Berg, R. G. (2022b). *How to Run Levene's Test in SPSS?* https://www.spss-tutorials.com/levenes-test-in-spss/ [Accessed: 22.5.2022].

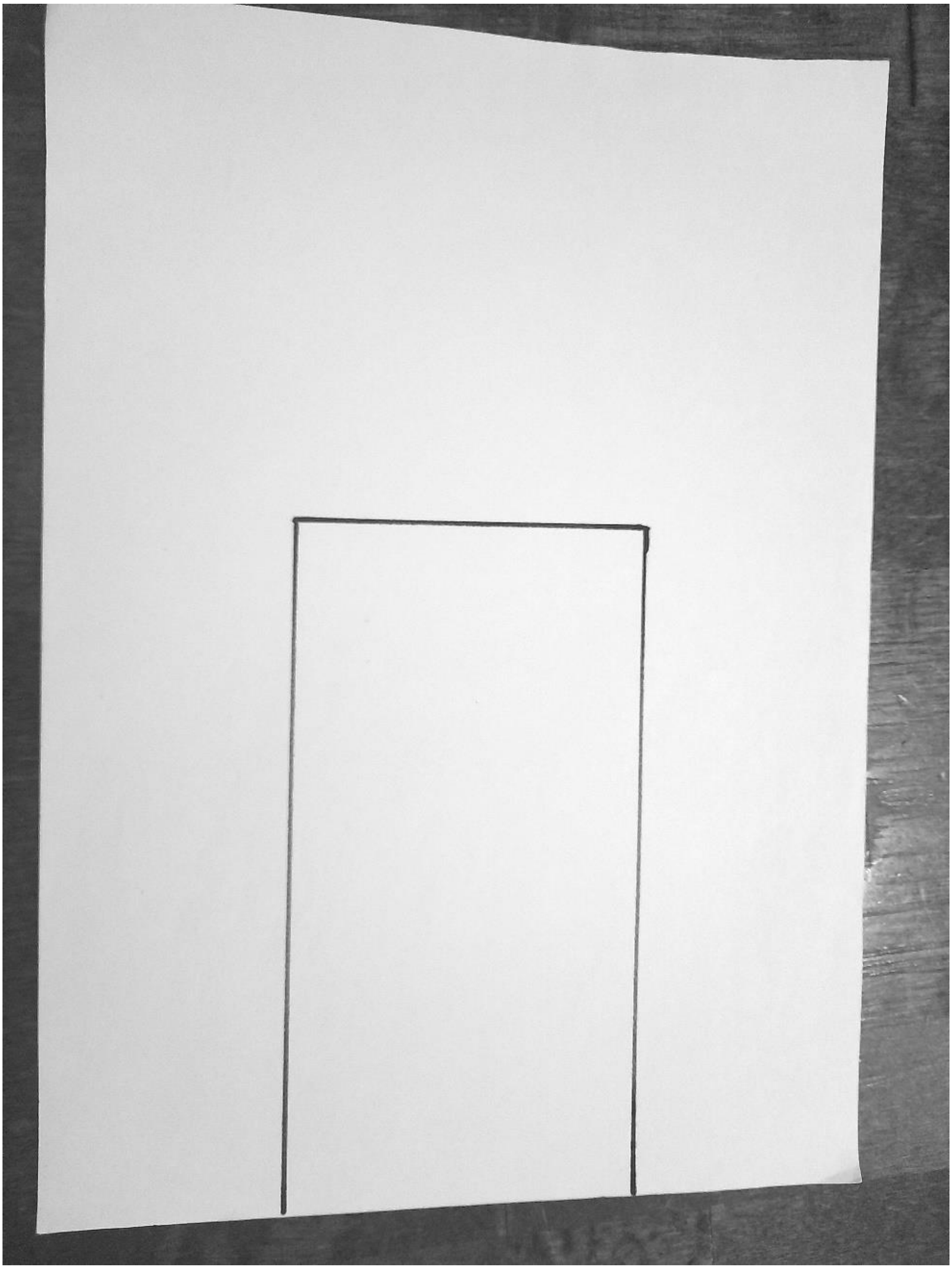
Wilson, B. (2004). Child Art After Mordernism: Visual Culture and New Narratives. In E. W. Eisner & M. D. Day (Eds.), *Handbook of Research and Policy in Art Education,* (pp. 299 – 328). Taylor & Francis Group.

Yantis, S. & Abrams, R. (2017). *Sensation and Perception.* (2nd ed.). Worth Publishers Inc.

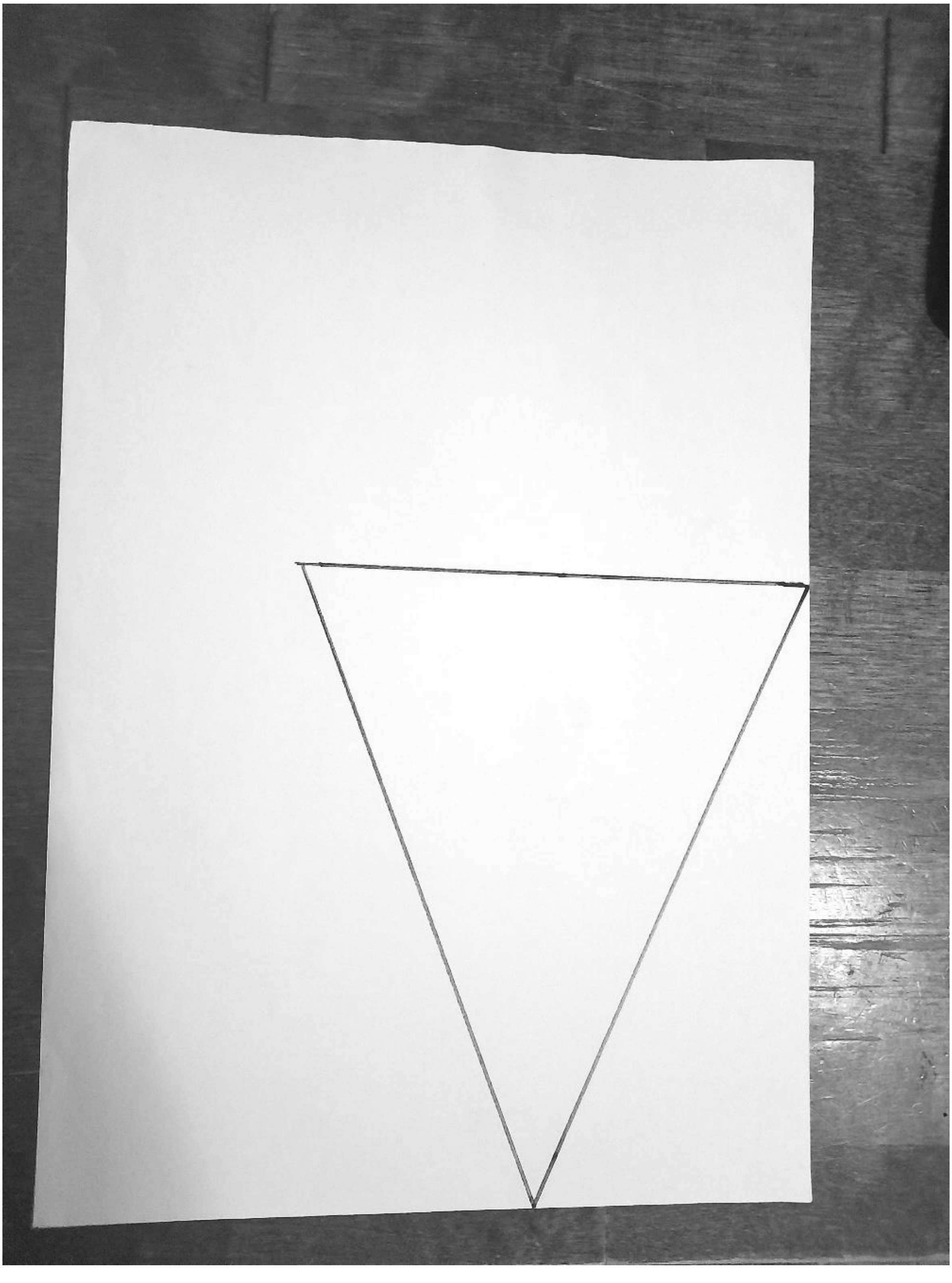
Zaiontz, C. (2022). *Effect Size for Chi-squared Test.* https://www.real-statistics.com/chi-square-and-f-distributions/effect-size-chi-square [Accessed: 22.5.2022].

Attachments

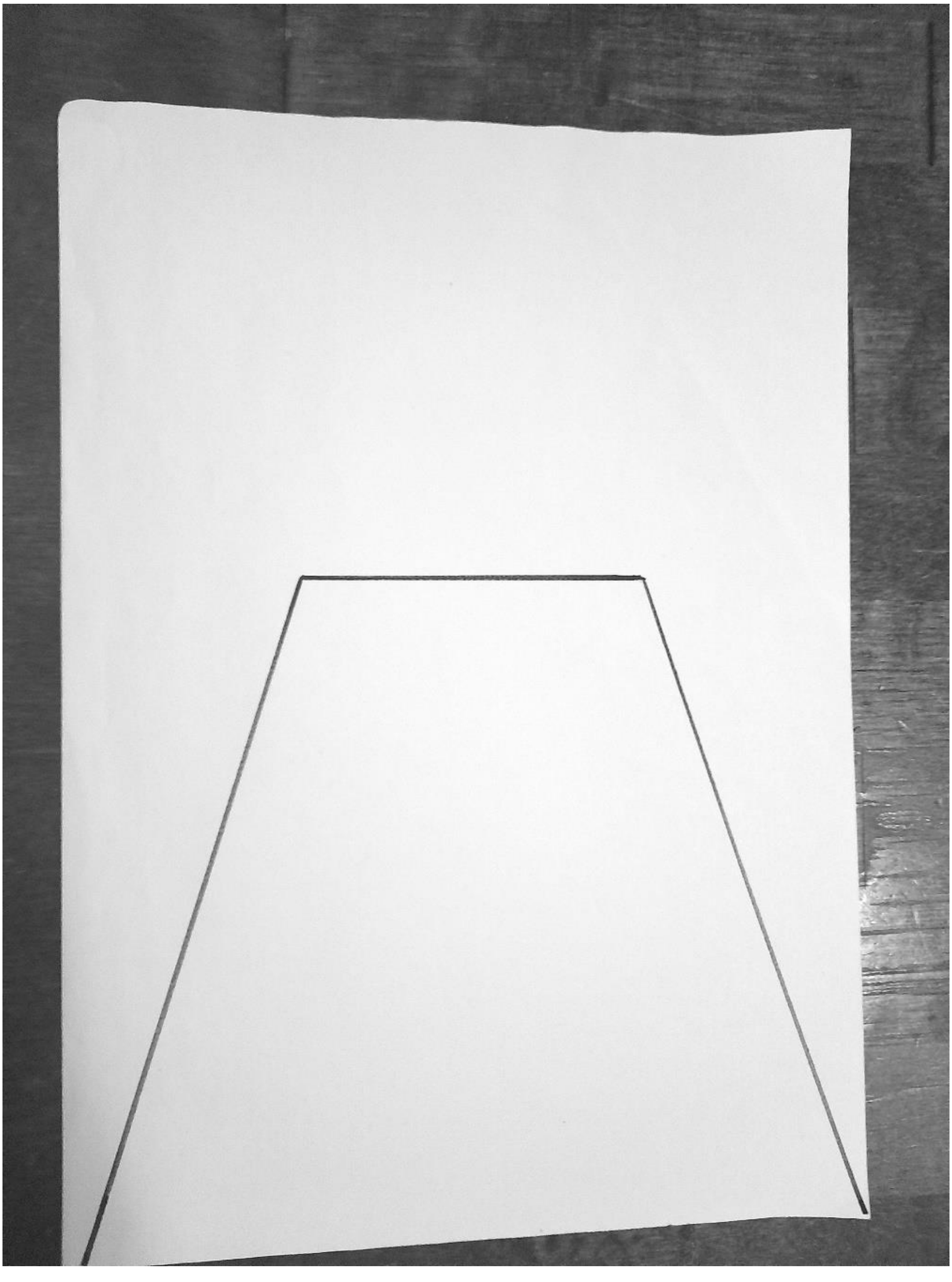
Line drawings for the Selection test



Attachment 1: Line-drawing of a corridor with no perspective



Attachment 2: Line-drawing of a corridor with no perspective



Attachment 3: Line-drawing of a corridor with perspective

Old and new versions of the consent form

TUTKIMUKSEEN SUOSTUMINEN

Koululla, tulee olemaan tutkimus, jossa Sandra Bencomo tutkii lasten piirtämistä. Sinua pyydetään osallistumaan tutkimukseen ja piirtämään ainakin neljä piirustusta erilaisiin tehtäviin, jotka tehdään koululla. Lisäksi sinun tulee eräässä tehtävässä valita kolmesta vaihtoehdosta yksi kuva. Ei ole oikeaa tapaa tehdä piirustuksia, vaan tutkija haluaa selvittää, miten juuri sinä piirrät ja minkä kuvan juuri sinä valitset.

Tutkija pitää piirustukset omassa kodissaan ja omalla tietokoneellaan sekä sellaisessa paikassa internetissä, joka on muilta piilossa.

Osallistumalla tutkimukseen autat aikuisia selvittämään miten lapsen ajattelu ja piirtäminen toimii. Piirustuksiasi ei näytetä muille, ellet niin halua.

☐ Kyllä, haluan näyttää piirustukseni kaikille

☐ Kyllä, haluan näyttää piirustukseni kaikille ja sen vieressä saa lukea minun nimeni.

Jos et halua osallistua tutkimukseen, voit jättää tämän paperin tyhjäksi. Sinun ei ole pakko osallistua tutkimukseen. Voit myös lopettaa osallistumisen kesken.

Jos haluat osallistua tutkimukseen, laita nimesi tähän. Voit myös laittaa allekirjoituksesi, jos sinulla on sellainen. Jos kirjoitat tähän nimesi, tutkija säilyttää tämän paperin.

Nimi: ____________________

Allekirjoitus_______________

Tutkijan, Sandra Bencomo, allekirjoitus: ____________________

Attachment 4: Old consent form for child participant. Sent to some schools and pre-primary schools that participated.

TUTKIMUKSEEN SUOSTUMINEN

Koululla, tulee olemaan tutkimus, jossa Sandra Bencomo tutkii lasten piirtämistä. Sinua pyydetään osallistumaan tutkimukseen ja piirtämään ainakin kaksi piirustusta erilaisiin tehtäviin, jotka tehdään koululla. Lisäksi sinun tulee eräässä tehtävässä valita kolmesta vaihtoehdosta yksi kuva. Ei ole oikeaa tapaa tehdä piirustuksia, vaan tutkija haluaa selvittää, miten juuri sinä piirrät ja minkä kuvan juuri sinä valitset.

Tutkija pitää piirustukset omassa kodissaan ja omalla tietokoneellaan sekä sellaisessa paikassa internetissä, joka on muilta piilossa.

Osallistumalla tutkimukseen autat aikuisia selvittämään miten lapsen ajattelu ja piirtäminen toimii. Piirustuksiasi ei näytetä muille, ellet niin halua.

☐ Kyllä, haluan näyttää piirustukseni kaikille

☐ Kyllä, haluan näyttää piirustukseni kaikille ja sen vieressä saa lukea minun nimeni.

☐ Ei, kuviani ei saa näyttää kenellekään

Jos et halua osallistua tutkimukseen, voit jättää tämän paperin tyhjäksi. Sinun ei ole pakko osallistua tutkimukseen. Voit myös lopettaa osallistumisen kesken.

Jos haluat osallistua tutkimukseen, kirjoita nimesi tähän. Voit myös laittaa rastin, jos et osaa kirjoittaa nimeäsi. Tutkija säilyttää tämän paperin, kun se on täytetty.

Lapsen itse kirjoittama nimi tai rasti: ______________________

Tutkijan, Sandra Bencomo, allekirjoitus: _____________________

Attachment 5: New consent form for child participant. Clarified after some questions from the parents.

TUTKIMUKSEEN OSALLISTUVAN LAPSEN HUOLTAJAN SUOSTUMUS

Muutokset lasten piirustuksissa

Lastasi on pyydetty osallistumaan Sandra Bencomon pro gradu -tutkimukseen, joka tähtää maisterintutkintoon Lapin yliopistossa. Luethan suostumustekstin huolellisesti ennen allekirjoittamista ja kysyt, mikäli kaipaat lisätietoja jostakin yksityiskohdasta.

Tutkimuksen tavoitteena on selvittää, onko lasten piirtämisessä tapahtunut muutoksia. Muutoksia selvitetään suhteessa aiempiin kohortteihin.

Mikäli päätät osallistua, lastasi pyydetään osallistumaan seuraaviin testeihin:

- Harris-Goodenough Draw-a-person testistö (kesto n. 15 min, ryhmässä suoritettava)

- Kahden värillisen pallon havaintopiirtäminen Littletonin (1991, s.60) suorittaman testin mukaisesti (kesto n. 15min per lapsi, yksittäin suoritettava)

- Viivapiirroksen valitseminen kolmesta annetusta vaihtoehdosta sopivaksi havaintoon loittonevasta tiestä tai käytävästä Littletonin (1991, s.142) suorittaman testin mukaisesti. (kesto n. 5min per lapsi, yksittäin suoritettava)

Tutkija toivoo, ettette harjoittele kyseisiä testejä ennakkoon. Osallistuminen ei aiheuta lapsellesi minkäänlaista haittaa, vaaraa tai uhkaa.

Tutkimuksessa kerätään seuraavat tiedot: Ikä, syntymävuosi, syntymäkuukausi, nimi (vain, jotta samat kuvat saadaan ryhmitettyä saman osallistujan tekemiksi, lapsenne nimeä ei julkaista pro gradu -tutkielmassa), kokeen tekemisen päivämäärä, lapsenne piirustukset ja tiedot siitä, minkä viivapiirroksen lapsenne on valinnut viimeisessä testissä.

Tutkimuksessa kerätyt tiedot säilytetään Sandra Bencomon henkilökohtaisella tietokoneella ja Google Drive -pilvipalvelussa. Molemmat ovat salasanasuojattuja säilytyspaikkoja. Toinen tutkija voi päästä käsiksi aineistoon, jos hän suorittaa tutkimusta yhteydessä johonkin yliopistoon tai vastaavaan instituutioon. Myös pro gradun ohjaajilla on pääsy aineistoon. Muille annettavassa aineistossa lapsenne nimeä ei julkaista, vaan hän on aineistossa koodinimellä (esim. oppilas A).

Aineistoa tai osia siitä voidaan raportoida, esittää tai julkaista pro gradu -tutkielmassa ja sen esittelyssä omaa pro graduaan aloittaville oppilaille.

Osallistumisesi ei tuota välitöntä hyötyä sinulle tai muille, mutta se voi auttaa selvittämään ajoittuuko lasten kritiikille herkkä kehityskausi (visuaalinen realismi) aiempaan ikävuoteen kuin ennen. Tämä edistää kuvataidekasvatuksen alaa ja tietoa lasten ajattelusta.

Tutkimuksesta ei makseta korvausta.

Lähtökohtaisesti aineistossa esiintyvien henkilöllisyys salataan ja luottamuksellisuus turvataan siten, etteivät osallistujat ole tunnistettavissa julkaistavista tai julkisesti esitettävistä aineisto-otteista. Sandra Bencomo ja Lapin yliopisto ovat sitoutuneet yksityisyytesi ja henkilötietojesi suojaamiseen. Aineistoa käsitellään lainmukaisesti ja tutkimuseettisiä periaatteita kunnioittaen.

Voit halutessasi antaa luvan myös sellaisen piirustusaineiston julkaisemiseen, jonka voi tunnistaa lapsenne tekemäksi, jos on nähnyt hänen tekevän sen. Tämä piirustus julkaistaan nimettömästi, ellette ehdottomasti halua lapsenne nimeä kuvan yhteyteen.

☐ Kyllä, annan erikseen luvan julkaista lapseni piirustus

☐ Kyllä, haluan ehdottomasti, että lapseni nimi on piirustuksen yhteydessä

Osallistuminen on täysin vapaaehtoista. Sinulla on oikeus keskeyttää osallistuminen tutkimukseen milloin tahansa ilman seurauksia. Suostumuksesta on laadittu kappale sekä tutkijalle että osallistujalle. Tutkijalle luovutettu allekirjoitettu suostumus säilytetään osana tutkimusaineistoa.

_________________________________ _________________________________

NIMENSELVENNYS / Osallistuja PÄIVÄYS (pp/kk/vvvv)

_________________________________ _________________________________

ALLEKIRJOITUS / Osallistuja ALLEKIRJOITUS / tutkija Sandra Bencomo

Jotta tutkimuksen aineiston kerääminen on helpompaa, täytäthän myös seuraavat tiedot, jos suostutte lapsenne osallistumiseen.

Lapsen nimi: _____________________________________

Lapsen ikä _______________________________________

Lapsen syntymävuosi ja syntymäkuukausi: ______________

Mikäli sinulla on kysyttävää tutkimukseen liittyen, voit ottaa yhteyttä:

sbencomo@ulapland.fi

sandra.m.bencomo@gmail.com

Puh. 0458461101

Attachment 6: Old consent form for parents of the participants. Sent to some schools and pre-primary schools that participated.

TUTKIMUKSEEN OSALLISTUVAN LAPSEN HUOLTAJAN SUOSTUMUS

Muutokset lasten piirustuksissa

Lastasi on pyydetty osallistumaan Sandra Bencomon pro gradu -tutkimukseen, joka tähtää maisterintutkintoon Lapin yliopistossa. Luethan suostumustekstin huolellisesti ennen allekirjoittamista ja kysyt, mikäli kaipaat lisätietoja jostakin yksityiskohdasta.

Tutkimuksen tavoitteena on selvittää, onko lasten piirtämisessä tapahtunut muutoksia. Muutoksia selvitetään suhteessa samanikäisiin lapsiin 1960 ja 1980-luvuilta.

Mikäli päätät osallistua, lastasi pyydetään osallistumaan seuraaviin testeihin:

- Harris-Goodenough Draw-a-person testistö (kesto n. 10min, voidaan suorittaa ryhmässä). Lapsi piirtää kuvan miehestä.

- Kahden värillisen pallon havaintopiirtäminen Littletonin (1991, s.60) suorittaman testin mukaisesti (kesto n. 10min per lapsi, yksittäin suoritettava)

- Käytävän/tien piirtäminen. + Viivapiirroksen valitseminen kolmesta annetusta vaihtoehdosta sopivaksi havaintoon loittonevasta tiestä tai käytävästä Littletonin (1991, s.142) suorittaman testin mukaisesti. (kesto n. 5min per lapsi, yksittäin suoritettava).

Tutkija toivoo, ettette harjoittele kyseisiä testejä ennakkoon. Osallistuminen ei aiheuta lapsellesi minkäänlaista haittaa, vaaraa tai uhkaa. Hän saa kieltäytyä tutkimuksesta missä vaiheessa tahansa. Osallistuminen on täysin vapaaehtoista.

Tutkimuksessa kerätään seuraavat tiedot: Ikä, syntymävuosi, syntymäkuukausi, nimi (vain, jotta samat kuvat saadaan ryhmitettyä saman osallistujan tekemiksi, lapsenne nimeä ei julkaista pro gradu -tutkielmassa, ellette erityisesti halua niin), kokeen tekemisen päivämäärä, lapsenne piirustukset ja tiedot siitä, minkä viivapiirroksen lapsenne on valinnut viimeisessä testissä.

Tutkimuksessa kerätyt tiedot säilytetään Sandra Bencomon henkilökohtaisella tietokoneella, älypuhelimessa ja Google Drive -pilvipalvelussa. Kaikki ovat salasanasuojattuja säilytyspaikkoja. Myös kuvat säilytetään Sandra Bencomon asunnossa. Toinen tutkija voi päästä käsiksi aineistoon, jos hän suorittaa tutkimusta yhteydessä johonkin yliopistoon tai vastaavaan instituutioon. Myös pro gradun ohjaajilla on pääsy aineistoon. Muille annettavassa aineistossa lapsenne nimeä ei julkaista, vaan hän on aineistossa koodinimellä (esim. oppilas A).

Aineistoa tai osia siitä voidaan raportoida, esittää tai julkaista pro gradu -tutkielmassa ja sen esittelyssä omaa pro graduaan aloittaville oppilaille.

Osallistumisesi ei tuota välitöntä hyötyä sinulle tai muille, mutta se voi auttaa selvittämään ajoittuuko lasten kritiikille herkkä kehityskausi (visuaalinen realismi) aiempaan ikävuoteen kuin ennen. Tämä edistää kuvataidekasvatuksen alaa ja tietoa lasten ajattelusta.

Tutkimuksesta ei ensisijaisesti makseta korvausta.

Lähtökohtaisesti aineistossa esiintyvien henkilöllisyys salataan ja luottamuksellisuus turvataan siten, etteivät osallistujat ole tunnistettavissa julkaistavista tai julkisesti esitettävistä aineisto-otteista. Sandra Bencomo ja Lapin yliopisto ovat sitoutuneet yksityisyytesi ja henkilötietojesi suojaamiseen. Aineistoa käsitellään lainmukaisesti ja tutkimuseettisiä periaatteita kunnioittaen.

Voit halutessasi antaa luvan myös sellaisen piirustusaineiston julkaisemiseen, jonka voi tunnistaa lapsenne tekemäksi. Tämä piirustus julkaistaan nimettömästi, ellette ehdottomasti halua lapsenne nimeä kuvan yhteyteen. Suostumus tähän on lapsen lomakkeessa.

Tutkijalle luovutettu allekirjoitettu suostumus säilytetään osana tutkimusaineistoa. **Muista allekirjoittaa myös toinen samanlainen suostumus,** joka jää itsellesi. Toinen suostumus palautetaan koulun kautta tutkijalle.

___________________________________ ___________________________________

NIMENSELVENNYS / Huoltaja PÄIVÄYS (pp/kk/vvvv)

___________________________________ ___________________________________

ALLEKIRJOITUS / Huoltaja ALLEKIRJOITUS / tutkija Sandra Bencomo

Jotta tutkimuksen aineiston kerääminen on helpompaa, täytäthän myös seuraavat tiedot, jos suostutte lapsenne osallistumiseen.

Lapsen nimi: _______________________________________

Lapsen syntymävuosi ja syntymäkuukausi: ______________

Mikäli sinulla on kysyttävää tutkimukseen liittyen, voit ottaa yhteyttä:

sbencomo@ulapland.fi

sandra.m.bencomo@gmail.com

Puh. 0458461101

Attachment 7: New consent form for parents of the participants. Clarified after some questions from the parents.