Colour and its Impact on the Picture Completion Subtest in the WISC-IV

Brett Gregory

University of the Witwatersrand

2019

Author Note

The author is a registered student in the Department of Psychology, University of Witwatersrand, South Africa.

This research report was submitted in partial fulfilment of the requirements for the degree of the Masters in Social and Psychological Research.

Supplementary material has been made available for download from https://osf.io/mcuqy/?view_only=7adad91d1e304dc89e20f77133447d6f.

Correspondence concerning this article should be addressed to Brett Gregory.

E-mail: brettgregory95@gmail.com

Abstract

Limited research exists demonstrating whether the addition of colour to psychometric measures impact an individual's performance. Previous studies have found that colour neither has an incremental or decremental effect in performance, though both empirical and theoretical evidence exists that suggests colour should be affecting performance given the enhancement effects colour has on perception. Notably, previous attempts at analysing this question are outdated and sometimes suffer from methodological flaws, and thus their findings can arguably be questioned. The present study aimed at providing a more recent attempt at examining whether colour affects psychometric performance and whether this impact changes according to an individual's educational ability. To this end, a colourised and a greyscale version of the Picture Completion subtest in the WISC-IV were randomly administered to a sample of 87 children who differed in their educational ability level determined by the type of educational assistance they required. Children's overall score in the measure and their average time taken to provide a correct response to an item were recorded to provide insight into their overall performance in the measure. A series of robust two-way analysis of variances for both participants scores and times found a single significant main effect for the type of school the participant attended on their score, with the remaining main and interaction effects for both dependent variables being non-significant. These results suggest that colour did not significantly affect participants performance on the measure. However non-statistically significant but consistent differences found between the two versions of the measure that warrants for future investigation into this matter.

Keywords: WISC-IV, Picture Completion, South Africa, visual search, perception, attention, performance.

Plagiarism Declaration

Surname: Gregory

First Name/s: Brett

Student Number: 792376

Course: MASPR

Project Supervisor: Enid Schutte

Project Title: Colour and its Impact on the Picture Completion Subtest in the WISC-IV

I, *Brett Gregory*, know and accept that plagiarism (i.e., to use another's work and to pretend that it is one's own) is wrong. Consequently, I declare that

- the following assignment/proposal is my own work.
- I have correctly acknowledged all direct quotations and paraphrased ideas/ content. In addition, I have provided a complete, alphabetised reference list, as required by the APA method of referencing.
- I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work, or that I failed to acknowledge the source of the ideas or words in my writing.

Signed: Date: 15/12/2018

Acknowledgements

I want to acknowledge and express my sincere appreciation for those that helped make this research report possible:

- My research supervisor, Enid Schutte, for her outstanding guidance, time and motivation throughout the year.
- My family who has continuously encouraged me to pursue my dreams and have supported me through my entire academic journey.
- All my participants that gave their time to participate in my study and the schooling institutions that allowed me to conduct my study on their premises.

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

1.1 Brief Background to the Study

David Wechsler is well known for his contributions in psychometric testing, most notably for his development of a series of intelligence scales (Flanagan, Alfonso, Mascolo & Hale, 2010; Kaufman, Flanagan, Alfonso & Mascolo, 2006). The first scale in his series of Wechsler tests was the Wechsler-Bellevue Intelligence Scale (Wechsler, 1939). This was later revised and adapted in 1946 to produce Form II of the Wechsler-Bellevue Intelligence Scale (Flanagan et al., 2010; Kaufman et al., 2006). Wechsler's (1946) Form II served as the basis for the original children's battery, the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949). The original WISC has since undergone four revisions (1974, 1991, 2003a, 2014), with the WISC-V (Wechsler, 2014) being the latest edition of the scale (Canivez & Watkins, 2016; Kaufman et al., 2006).

The Picture Completion subtest has remained within the intelligence battery since the original WISC (Wechsler, 1949) and has been removed only recently in the WISC-V (Canivez & Watkins, 2016). The Picture Completion is a supplemental subtest that fell under the Perceptual Reasoning Index (PRI) in the WISC-IV (Flanagan et al., 2010). The subtest requires the examinee to view a picture presented to them and identify the essential part that is missing from it within a specified time limit (Flanagan et al., 2010). An important adaption that was made to the Picture Completion was during the revision of the WISC-R (Wechsler, 1974) to the WISC-III (Wechsler, 1991). Other than the number of items in the subtest being changed from 26 in the WISC-R to 30 items in the WISC-III, the artwork was enlarged and colourised to increase the attractiveness of the items (Mahone et al., 2003; Wechsler, 1991).

It remains unclear whether test adaptions similar to the one mentioned above, involving colour changes to items within a measure, has any effect on an individual's performance on it (Mahone et al., 2003). The addition of colour to items within a measure may unintentionally increase the demand placed on individuals executive control functions, which may inherently bias the testing procedure and could place children with learning difficulties and working memory impairments at a disadvantage (Mahone et al., 2003). Previous attempts (Husband & Hayden, 1996; Nelson, Arthur, Lautigar & Smith, 1994) have been made to address this issue. The findings tend to agree that no incremental or decremental effect on performance by the addition of colour is present, though, arguably these studies are either flawed, outdated or simply not generalisable to a South African context.

The present study aimed to contribute a more recent attempt at clarifying whether colour influences performance and to investigate whether this influence is universal across educational ability levels. To this end, an experimental study was conducted whereby either a coloured version or a greyscale version of the Picture Completion subtest of the WISC-IV (Wechsler, 2003a) was randomly administered to children who varied in their academic abilities. It was hypothesised that the children's performance on the task, measured by their overall score and their average time taken to give a correct response, would differ. Specifically, it was hypothesized that the groups of children who were presented with the coloured stimuli would perform better than those who received the greyscale stimuli. It was further hypothesised that these differences and the magnitude of the effects colour has on performance may differ according to children's educational ability.

Despite previous attempts by researchers to examine this issue, no results offer a clear and adequate account into whether colour affects psychometric performance. Thus, it was proposed that research attention is focused on providing a renewed attempt at addressing the problem of colour and performance, especially given that the results of such an attempt could have important implications for a wide variety of areas. Specifically, understanding the effects that the addition of colour to measures have on performance would be beneficial to our current understanding of psychometric testing, possibly highlighting the need to be conscious of how specific psychometric features relating to colour can impact individuals test performance. Moreover, the results could have implications for pedagogic strategies in educational settings which would be relevant in a South African context.

The current educational crisis being faced in South Africa has resulted in poor academic performance by the majority of the South African youth in various schooling institutions across the country (Spaull, 2013). A number of institutional and systematic factors exist that contribute to the educational crisis, though predominantly factors relating to poor management of the educational system, incompetence and weak capacity of school principals, lack of funding, poor resource allocation, lack of accountability and inadequate teacher performance seem to be at the forefront of the issue (Spaull, 2013). Given the limited budget allocated for education in South Africa, it is crucial that key decision makers in education involved in budget planning and resource allocation aim to maximise educational benefits by their decisions and cut costs where possible so that resources and funds may be directed elsewhere where it is needed. Results of the present study may facilitate those individuals in charge of making such decisions and thus could have broad implications for pedagogic strategies in educational settings. Decisions on, for example, whether funds should be allocated to print work and study books in colour in order to maximise educational performance or to cut costs by not printing them in colour so that the funds could be spent elsewhere could be facilitated by the findings of this study.

1.2 Structure of the Research Report

The research report begins with Chapter Two with an outline of the theoretical framework that was adopted in this study, followed by a comprehensive overview of the relevant theoretical and empirical literature. The study's methodology is then discussed in

Chapter Three, where the research design, sample, measures, procedures and analyses are described in detail. Next, Chapter Four is dedicated to reporting the results from the analyses, which are then later interpreted and discussed in Chapter Five. The research report ends in Chapter Six with a summary of the key findings.

Chapter Two: Theoretical Framework and Literature Review

The following chapter is devoted to outlining the study's theoretical framework and discussing relevant theoretical and empirical literature that exists regarding perception, colour and performance. The chapter begins by introducing the theoretical framework and arguing for this and the three models that were chosen. The introduction and argument concludes with a paragraph that details how all models were intended on complementing one another, with the primary intention being that they were able to help contextualise the study and offer explanations for the results found. Next, relevant literature is reviewed. The review starts with explaining the purpose of perception, later focusing on what roles colour vision has in the perceptual process. This is followed by a critical review of the existing research that attempted to examine the effects colour has on psychometric performance. Studies that demonstrated the effects cultural and subjective factors have on perception are then discussed. The literature review concludes with a brief discussion of perceptual decisions and visual search.

2.1 Theoretical Framework

The entire framework consisted of three theoretical models. The first model involves the processes of perception. This model was chosen since it details how one perceives and interprets external stimuli and scenes. Given that the intention of the present study is to examine the effects of colour on performance, having a clear understanding of the basics of the perceptual process is crucial before specific components of perception can be focused on. Without this foundational model, it would not be possible to know where colour, a feature of perception, fits into the entire process.

The second model that was adopted is the Feature Integration Theory (hereafter referred to as FIT) proposed by Treisman and Gelade (1980). This model focuses predominantly on the early processes of perception and provides a better understanding of how colour is broken down into a distinct feature, processed, and then later bound or grouped with other features to create a conscious perception. Additionally, a process known as visual search is also mentioned as it falls under the FIT. Finally, the third model discussed involves attention, specifically Kahneman's (1973) capacity model of attention which is helpful in explaining the idea of cognitive load.

The Perceptual Process. Since this study aimed to look at how colour adaptions may impact psychometric performance, it is only natural to briefly discuss the foundational perceptual processes that are involved when viewing an image, object or scene. It is much easier to navigate and pinpoint the stages in the perceptual process which is significantly impacted by colour if a foundational model of perception is first provided.

Perception is a somewhat complicated process in the brain, as many different brain regions are required to help make sense of what our eyes are seeing. To help navigate through this complicated process, Goldstein (2014) summarised seven primary steps that are typically involved in perception. These steps can be seen in Figure 1 below:

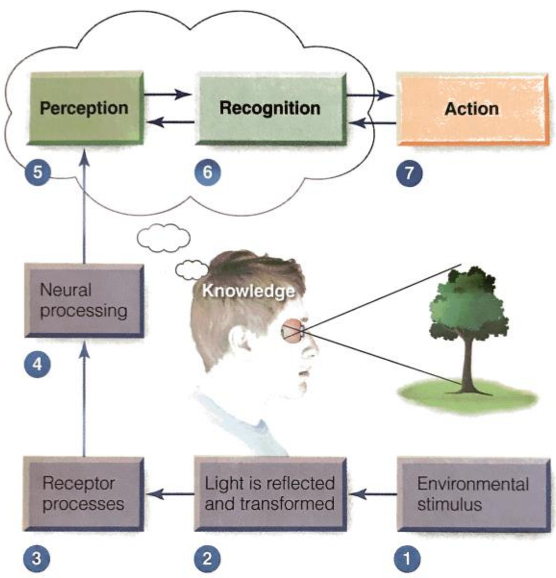


Figure 1. A visual representation of the steps involved during the perceptual process (Goldstein, 2014, p. 5)

Perception of external environmental stimuli requires that light be reflected off a particular object and be directed towards the eyes' optical system. This reflected light is of a particular wavelength, which needs to be roughly between 400 nm and 700 nm for it to be detected by the human visual system (Kolb & Whishaw, 2015). Notwithstanding the relative simplicity of the above statement, it is interesting to note that Winderickx et al. (1992) discovered that visual systems are not the same for all individuals, meaning that some may experience colours slightly differently to others. For example, males were found to experience the colour red differently, likely being due to "small variations in the absorption maxima of visual pigments" (Winderickx et al., 1992, p. 431). Despite these differences though, it remains that the human visual system requires reflected light to have a wavelength that falls within the narrow band of electromagnetic spectrum in order for it to be detected. The wavelength of light is partly determined by both the properties of the stimulus itself and the particular medium that the light travels through (e.g. air) before coming in contact with the eyes (Goldstein, 2014; Thompson, 1994).

It is important to briefly consider the finer aspects of light given that light is an essential carrier of visual sensory information and that slight changes to these aspects can affect the signal being transported to the visual system. Firstly, the wavelength of reflected light can change the way we experience external stimuli, specifically regarding the colour. To illustrate the importance of light wavelength, imagine a cube, if the light reflected from the cube has a wavelength of 480 nm, then the cube will be perceived as blue (Enns, 2004). However, if the same reflected light now travels through a filter that changes its wavelength from 480 nm to 565 nm before reaching the eye, the cube will no longer be experienced as blue but instead as yellow (Enns, 2004). Secondly, the intensity of the reflected light is important to consider since the light intensity affects the brightness of the stimuli (Enns, 2004). The higher the number of light photons that are reflected off the stimuli, the brighter it will be perceived to be, and vice versa.

The light that reaches the eye now passes through the various components within it. The anatomy of the eyes has been well understood for many years now, consisting of the cornea, the lens, the vitreous humor, and various fluids (Zusne, 1970). The eye is a very complex structure, to the point where the famous evolutionary theorist, Charles Darwin (1859), admitted that it would be difficult, though not impossible, for his theory to explain how it evolved to be the way it is today. Regardless of how this complex structure of the eye formed, it is known that it remains imperfect, especially when compared to human-made optical systems (Zusne, 1970). For example, the retina of the eye is not flat but concave, resulting in the light passing through it to become distorted (Zusne, 1970).

Nevertheless, the visual system seems efficient in providing a consistent representation of the world around us despite the many challenges it may face, specifically regarding small fluctuations in the spectrum of light that can distort the signal being transported towards our eyes (Radonjić & Brainard, 2016). Distortions in raw sensory data lead to the signals becoming impoverished, meaning that the visual system and brain regularly have to fill in details to our representations of external stimuli to ensure that they remain consistent (Chong, Familiar & Shim, 2015). Chong et al. (2015) examined how the visual system makes sense of sensory signals that are partial and degraded by using fMRI and a forward-encoding model of orientation. They demonstrated how “object-specific features (orientation) can be reconstructed from neural activity in early visual cortex (V1)” (Chong et al., 2015, p. 1453).

It may not be apparent yet from these beginning steps of the perceptual process that have been discussed, but our perceptions of the world are primarily based on *representations*. This idea is a lot easier to understand if we extend it to colour perception. From a Newtonian conception of colour, external stimuli are not coloured but instead are merely perceived to be (Thompson, 1994). In other words, “things do not look to be coloured because they are coloured; rather, things are coloured only because they look to be so” (Thompson, 1994, p. 12).

Returning to the steps presented in Figure 1, the light that enters the eye becomes slightly distorted as it passes through the cornea due to its concave shape (Kolb & Whishaw, 2015). The lens within the eye further bends the light to focus the image onto the back of the eye (Kolb & Whishaw, 2015). The light then travels through the retina, which is made up of photoreceptive cells known as cones and rods, where a process known as transduction occurs (Goldstein, 2014; Kolb & Whishaw, 2015). The process of transduction converts light enegery into chemical energy, which in turn, is converted to action potentials (Kolb & Whishaw, 2015). It is through the process of transduction that sensory signals are converted and directed to various regions of the brain to be processed further (Goldstein, 2014). These signals initially travel primarily along either the geniculostriate pathway or the tectopulvinar pathway (Kolb & Whishaw, 2015).

Up until this point, the various steps and processes mentioned occur unconsciously. It is only in step five when the processed electrical signals are transferred to one’s consciousness that the representation becomes *perceived*, and then, only after, becomes *recognised*. When an object is perceived, we merely become consciously aware of it, but it is only after the stage of recognition that the object that we have just become consciously aware of is given meaning through categorisation (Goldstein, 2014).

The last factor which we need to concern ourselves with to understand the perceptual process is knowledge. Goldstein (2014) describes knowledge as “any information that the perceiver brings to a situation” (p. 9). Knowledge is an essential factor since this internal knowledge can affect the steps within the perceptual process. The effect caused is due to both bottom-up processing and top-down processing which interact with each other, and it is this interaction that changes one’s perception (Goldstein, 2014). Perception usually begins with bottom-up processing, where a person receives incoming sensory data which activates the visual receptors within the eye (Goldstein, 2014). Top-down processing occurs when that person uses previous knowledge to help recognise the object through categorisation which in turn assigns meaning to what is being perceived (Goldstein, 2014).

In conclusion, the first model that was included provided a fundamental understanding of the perceptual process. This model acted as the foundation of the theoretical framework. The other two models in the framework built on from this core process by adding more complexity to the various steps. Ultimately, the perceptual model not only outlined the necessary steps involved in perception, but it also highlighted the impact that small fluctuations in the various components that help make perception possible can have on the final representation that we perceive.

Feature Integration Theory. The second model that was adopted, namely the Feature Integration Theory (FIT), builds on from the first model in that it aims to understand the early phases of perception further. More specifically, the FIT aims to provide an account for the role that focused attention has in the perceptual process. In short, the FIT proposed that features or dimensions of an object are automatically registered in the early phases of perception and are processed separately, where only later through focused attention does the brain effortfully bind together the various processed features to create a single conscious representation of the object (Treisman & Gelade, 1980). This process is presented in Figure 2 below:

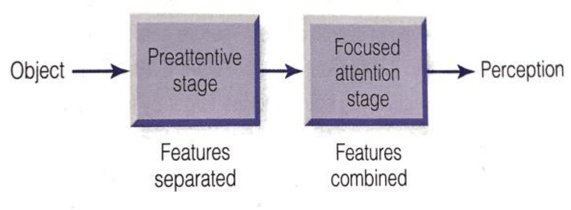


Figure 2. Flow diagram of the Feature Integration Theory (Goldstein, 2014, p. 143)

From the model presented in Figure 2, we can see that an object is first analysed and broken down into its separate features which exist independently from one another. Treisman believed that there existed a limited set of basic features (colour being one of them) and that these features could be processed in parallel preattentively (Treisman & Gelade, 1980; Wolfe, 2014). To expand on this idea more, first recall from the first model that once light energy is transformed into electrical energy through the process of transduction, the brain then sends the various electrical signals to different parts of the brain to be processed further. This is because different regions of the brain, specifically within the occipital lobe, are functional in that they are responsible for processing specific features or dimensions of the visual stimuli we are trying to perceive (Kolb & Whishaw, 2015). At least four different independent functions have been found, specifically: form, colour, movement, and location (Kellogg, 1997; Kolb & Whishaw, 2015). Only once the different features have been processed by their respective brain regions are they bound together, which requires focused attention (Treisman & Gelade, 1980).

Attention is an essential component in the focused attention stage in the model. Attention has a direct role and influence in combining the different features into a coherent representation, where "serial attention" is "needed to 'bind' features to object representations" (Wolfe, 2014, p. 14). This means that different levels of attention could affect the process of binding, resulting in the object being perceived to be misrepresented in one's consciousness. This means that the act of binding different features into one coherent conscious representation is affected by one's focused attention, where a lack of attention would result in features being incorrectly bound, creating more of an illusion than an accurate perception. This effect has been tested by Treisman and Schmidt (1982) who demonstrated that features could be incorrectly bound, often caused when attention is either diverted or overloaded, resulting in illusory conjunctions (Goldstein, 2008). Though, this is not the sole way that illusory conjunctions can occur. Memory decay or interference can also produce illusory conjunctions since the binding that once joined features of an object together disintegrate, allowing the features to float freely because the glue that held them together is no longer present or strong (Treisman & Gelade, 1980). This can result in the free-floating features to recombine to construct illusory conjunctions (Treisman & Gelade, 1980).

An important process that is involved in FIT is called *visual search*, a process that is often compared to a type of mental spotlight (Goldstein, 2014; Kellogg, 1997). We use visual search when we are trying to focus our attention on a specific object or an aspect of a particular scene or object. For example, we would use the process of visual search to find Wally in the famous Where's Wally books (Handford, 1987). Visual search, more specifically conjunction search, is essential when combining different features into a representation. (Goldstein, 2014). This type of visual search involves an individual actively searching or scanning a particular object/scene to focus their attention on a specific aspect of it. This means that the eye is rapidly moving, occurring on average between 2 to 5 times per second (Zusne, 1970). These sudden movements in the eye, known as saccades, can be either voluntary or involuntary (Zusne, 1970). In regards to the saccades that occur during visual search, these sudden movements involved when one actively scans an object or scene would be considered to be voluntary.

Interestingly, some features within a scene can guide one's attention to a specific aspect to it (Wolfe, 2014). When a visual search is guided, it means that specific basic features allow certain aspects of the visual scene to be prioritised "as more worthy of attention" (Wolfe, 2014, p. 21). This idea was demonstrated by Egeth, Virzi, and Garbart (1984) who showed that if you were to look at an image of different letters, some printed in black and some printed in red, and were asked to search for a red 'V', attention will ultimately be guided to items that only have the feature red (Wolfe, 2014). This effect of guiding one's visual search may take effect in coloured stimuli instead of in non-colourised stimuli in psychometric measures, in that colour may prioritise specific features over others, thus changing the way in which the picture is actively searched for the missing component. Wolfe (2014) highlights a number of different rules that govern the guidance of attention. However, one important rule to note is that "it is easier to find the presence of a feature than to find its absence" (p. 32).

FIT and visual search were vital to understand for the present study as it allowed for the exploration of how individuals actively scan a scene, break it down into different features to process and then combine the features into a representation which is later recognised and interpreted. A task such as the Picture Completion subtest of the WISC-IV involves visual search as participants have to narrow down their focus and attend to certain aspects of the picture to find what is missing from it. Furthermore, the FIT facilitated the interpretation of the study's results by enabling the researcher to explore whether the feature of colour is impacting their performance in any way. Studies relating to visual search are discussed later in the literature review.

Capacity Model of Attention. Finally, the third model adopted is the capacity model of attention, a model proposed by Daniel Kahneman (1973). Kahneman's (1973) model puts forward the idea that mental activity is limited, in the sense that it has a specific capacity to perform mental work. This limited capacity can be divided among different concurrent tasks, allowing for multiple tasks to be completed simultaneously (Kahneman, 1973; Kolb & Whishaw, 2015). Kahneman's (1973) model is represented in Figure 3 below:

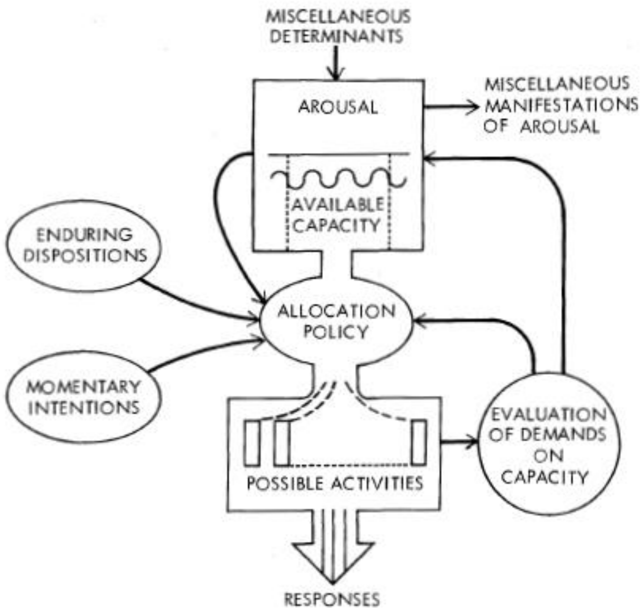


Figure 3. The capacity model of attention (Kahneman, 1973, p. 10).

Since the brain's capacity is said to be limited, this would mean that once one's capacity has been reached or the concurrent tasks require more mental resources than are currently available, then a decrease in performance on one or more of the tasks is expected. To help explain which of the tasks would most likely see a decrease in performance, Kahneman (1973) proposed that attention has a unique aspect to it which is effort (Kolb & Whishaw, 2015). The amount of effort one directs towards a specific task impacts how much resources are given to complete that task (Kahneman, 1973). Lower levels of effort mean that a lesser amount of mental resources are directed towards a task compared to if much higher levels of effort are applied (Kahneman, 1973).

Though, it should be mentioned that not all tasks demand the same amount of mental resources (Kahneman, 1973). Tasks can be classified as either an automatic or conscious process (Kolb & Whishaw, 2015). However, classifying tasks as being either one or the other is not so simple, as tasks often may shift between these two categories depending on specific contextual factors. Automatic processes require little focused attention to be performed, often taking place involuntarily in one's unconscious and rarely interferes with ongoing activities (Kahneman, 1973). Conscious processes, on the other hand, are effortful in that they require focused attention to be performed (Kahneman, 1973).

Automatic and conscious processes differ primarily in the number of resources they require to perform (Goldstein, 2008). Take, for example, the task of driving a car on an empty road. Such an activity may be an automatic process, especially if one has been driving for some years and, through consistent practice, the task has become familiar and makes minimal demands on one's attentional resources (Sarter, Gehring & Kozak, 2006). Thus, less focused attention is needed to be directed toward the act of driving. Despite this advanced level of driving proficiency, at times driving may occur on roads and in areas that are unfamiliar to the driver. In such situations more effort and an increase in focused attention is required to ensure that performance does not deteriorate (Sarter et al., 2006).

However, it is insufficient to focus soley on task demands as a determinant of attentional effort (Sarter et al., 2006). The fluctuations in how much attentional effort is directed towards difficult tasks at times seem to be explained less by the demands of the tasks and more by the performer's motivation (Sarter et al., 2006). Though the demands of a task are necessary to consider when talking about attentional effort, the inclusion of motivational forces that take into account goal-directed behaviours and cognitive processes are required as it provides a more coherent definition of attentional effort and provides a more precise insight into task performance (Berridge & Robinson, 2003; Sarter et al., 2006). Returning to the car example provided previously that demonstrated the impact that task demands have on attentional effort, one could imagine a situation in which the driver sees a police car driving nearby and is motivated to drive more carefully to ensure that no road rules are being broken. Here, motivational factors better explain the increase in attentional effort than task demands do.

Tomporowski and Tinsley (1996) demonstrated the impact that motivation has on attentional effort and in turn, the impact on task performance. The researchers aimed to examine the effects of monetary rewards on performance and whether this effect was consistent across different age groups. Their experiment involved young (20-24 years of age) and older (51-89 years of age) adult participants who were required to perform a cognitive vigilance task for 60 minutes (Tomporowski & Tinsley, 1996). The researchers selected subgroups from the young and older participants and offered them money in exchange for their participation. Their findings suggested that sustained attention for both the young and older participants who were paid did not differ, however the researchers did identify a decremental effect in performance for the young participants who were not paid for their participation (Tomporowski & Tinsley, 1996).

The relationship between motivation and attentional effort has also been demonstrated by Massar, Lim, Sasmita and Chee (2018) who found that performers are less willing to exert attentional effort when sleep deprived. The performance of 26 student participants was measured on a vigilance task, combined with an effort-based decision-making task and pupillometry (Massar et al., 2018). Participants were tested after they rested for nine hours and after sleep deprivation while under different incentive conditions. Their findings suggested that the cost of attentional effort can be perceived to be higher than it is in a state of sleep deprivation (Massar et al., 2018). The researchers also noted how impaired performance in the vigilance task could be explained by both a reduction in participants attentional capacity and by their diminished motivation (Massar et al., 2018).

Kahneman's (1973) capacity model of attention is not without its limitations and faults. Kahneman's (1973) model is situated in a limited resources paradigm, explaining that mental fatigue and a decrease in performance is to be expected as cognitive workload increases because the demand of cognitive effort temporary depletes the available resources (Kamza, Molińska, Skrzypska & Długiewicz, 2019). However, prior cognitive effort may not always result in mental fatigue and deterioration in task performance (Kamza et al., 2019). Instead, a cognitive warming-up effect may exist where the availability of cognitive resources actually increases after prior cognitive effort (Kamza et al., 2019).

This effect has been explored by Śpiewak, Ziaja and Doliński (2003), who initially hypothesised that a greater depletion of cognitive resources would occur after completing a difficult task. Their results, though, did not satisfy their initial hypothesis. Instead, their findings suggested that cognitive resources were under depleted and that after being 'warmed-up' their performance in subsequent tasks increased (Śpiewak et al., 2003; Kamza et al., 2019). These findings suggest that an underload account exists, contrary to the overload account provided by Kahneman's (1973) capacity model of attention, that is also able to provide explanations for why decrements in performance occur. This underload account explains that cognitive resources remain constant, and instead the allocation of these resources is what changes (Kamza et al., 2019). The modifications in resource allocation is primarily due to changes in arousal and motivation (Kamza et al., 2019). Therefore, even though Kahneman's (1973) model predominantly remains one of the primary accounts of attention that formed part of the present studies theoretical framework, the underload account was taken into consideration when the results were contextualised and discussed.

Combining the models into a single framework. The three models discussed were used throughout the study to guide the literature review and the adopted methodology. It is important to note that the models discussed were meant to complement each other. All models were used to jointly interpret the results obtained by the study in an integrative way, thus providing a more complex and coherent understanding of the impact colour is having on psychometric measures.

2.2 Literature Review

Purpose of colour vision. To understand the purpose of colour vision, it is first necessary to understand that colour is merely a feature of vision. Thus the general purpose of primary vision first needs to be considered before additional features of it can be focused on. At its core, vision informs us about the world (Enns, 2004). Many organisms possess the ability to detect light through photoreceptive organs (e.g. eyes), though, wavelength-discrimination, or colour vision, is only possible for vertebrates and arthropods (Koyanagi, Nagata, Katoh, Yamashita & Tokunaga, 2008).

The ability to see colour is not merely a luxury that a select few of species possess. Rather, colour vision played a fundamental role in ensuring the survival of many different species (Pinker, 1997). Take, for example, the brain of the rhesus macaque. This monkey has half of its brain dedicated to vision (Pinker, 1997), a development that did not occur by chance but instead evolved to provide an evolutionary advantage over other species. Colour vision too provided an evolutionary advantage to those that possessed it. The development of colour vision in our ancestral primates most likely served the primary purpose of making objects stand out from their backgrounds and aided them in their search for ripe fruit (Pinker, 1997). Colour vision also aided our ancestors to distinguishing between different berries and fruits, since ingesting the wrong food would have been fatal (Regan et al., 2001).

Taking the above into account, a more technical explanation of the purpose of vision could be provided that defines vision as an active process, in that it involves effortful and active guiding of our eyes to details of the environment in combination with active mental processes that help represent and interpret what we are trying to perceive (Enns, 2004). Furthermore, colour vision could be said to be merely a feature of vision, in the sense that colour supplements and enhances vision. One such enhancement colour has on vision is that it helps with the recognition of individual objects and facilitates the perception of complex scenes (Enns, 2004; Pinker, 1997).

Existing evidence suggests that colour plays an important and helpful role in both low-level and high-level vision (Tanaka, Weiskopf & Williams, 2001). Regarding low-level vision (edge and motion detection), colour helps segment complex visual scenes, thus allowing for objects to be distinguished from their backgrounds more easily (Herzog, Thunell & Ogmen, 2016; Tanaka et al., 2001). Regarding high-level vision (shapes and objects), objects and scenes that contain characteristic colours, such as a green lime or yellow lemon, are recognised more easily than if they did not have colour (Enns, 2004; Herzog et al., 2016; Tanaka et al., 2001). Taken together, colour may enhance the accuracy and speed at which we process and interpret complex scenes (Tanaka et al., 2001).

However, to what extent could these enhancement effects of accuracy and processing speed offered by colour vision be extended to psychometric testing? Asking questions similar to this is vital as such enhancements could have profound effects on performance in psychometric tasks. Research (Husband & Hayden, 1996; Nelson et al., 1994) suggests that colour in psychometric tests does not affect performance, though, as will be discussed shortly, these results are arguably flawed and outdated. Given the limited literature pertaining to this area, the possibility that colour affects psychometric performance cannot be ruled out entirely. Assuming that the enhancement effects of accuracy and processing speed offered by colour vision can be extended to psychometric testing, then greyscale measures could negatively impact performance by decreasing the accuracy of how they are processed and interpreted while increasing the time it takes the brain to process all the features and dimensions of the scene.

Should the absence of colour be found to make objects and scenes more difficult to process and interpret, then the performance by specific population groups may be impacted differently. For example, children diagnosed with Attention Deficit/Hyperactivity Disorder (ADHD) have been found to become more frustrated in tasks compared to their peers without ADHD, often resulting in the children being more likely to quit a task before completion (Scime & Norvilitis, 2006). Feelings of frustration have also been shown to negatively impact ones working memory, often by decreasing the available mental storage space within it which is now filled with intrusive and negative thoughts, directly resulting in a decrease in performance (Beilock & Carr, 2003). Given that it is known that working memory impairments often affect individuals with learning difficulties (Alloway, Gathercole, Willis, & Adams, 2005), feelings of frustration and anxiety would lead to their working memory capacity being further restricted in how much information it can store, directly resulting in a decrease in performance during tasks.

Colour and psychological measures. Adaptions made from the WISC-R to the WISC-III included “the addition of color to previously colorless test stimuli, addition/deletion of test items, and more stringent use of bonus points— rewarding speed of response” (Mahone et al., 2003, p. 332). Though the test manual mentions that the adaptions intended on increasing the attractiveness of the testing materials since it was believed that the addition of colour to the test would heighten children's attentiveness (Mahone et al., 2003; Nelson et al., 1994; Wechsler, 1991), the changes may have unintentionally placed higher demands on children's executive control functions (Mahone et al., 2003). A study by Mahone et al. (2003) compared children's performance on the WISC-R to the WISC-III. The researchers were mainly concerned with examining "the comparability of ability test revisions among clinical and non-clinical populations", finding that children with ADHD had lower performance IQ on the WISC-III compared to the WISC-R (Mahone et al., 2003, p.331). Their findings serve as a reminder to the importance of examining the consequences test adaptions may have on performance and how clinical and non-clinical populations may be impacted differently.

Research explicitly addressing the effect of the inclusion of colour in Psychometric instruments on performance is limited. Of that available in the literature, Husband and Hayden (1996) aimed to look at what effect the addition of colour to assessment instruments had on psychometric performance by administering colour and non-coloured versions of the WISC-III Picture Completion and Stanford-Binet (Fourth edition) Absurdities subtest to children. A sample of 80 children (16 of which were identified as remedially assisted students) from three different schools in Washington took part in their study. At the time, thirty-one of the children were in kindergarten/first grade, and forty-seven were in fifth/sixth grade. The sample contained equal numbers of boys and girls, and was mainly comprised of White students (83.4%).

It should be noted that no Black students were included in the sample, meaning that generalisability of their findings is limited given that their sample was not racially similar to a sample of children that one would expect to find in South Africa. Generalisability of their findings is further limited by the fact that their sample was solely comprised of children in Washington, who were most probably brought up in a different environment and society than that of South African children. This is especially true when we take into account the history of Apartheid in South Africa since its effects still impact and influence society to a large extent. Furthermore, there would be a vast difference in exposure to coloured reading material and media for children in under resourced areas around South Africa compared to the sample of children in Husband and Hayden's (1996) study, and given that ethnic factors are correlated to poverty in South Africa taking into account sample characteristics such as race is essential.

The measures were manually administered to the children in the study. The items of the subtests were split in an odd and even fashion. Each split was administered to every child, meaning that after the child had received the odd items first, they then would receive the even items. This allowed the children to act as their own control (Husband & Hayden, 1996). Half of the sample was randomly assigned the odd-numbered items to be colourised whereas the other half of the sample received the even-numbered items in colour. The authors decided this on the basis that this randomisation would help in reducing a possible order bias from occurring.

The results of their study indicated that the different versions of the subtests did not impact performance in any way, meaning that no enhancement effect or decrement was found (Husband & Hayden, 1996). A mixed design analysis for the Absurdities subtest found no significant differences between the colourised and non-colourised version. It was also discovered that children preferred measures that were colourised and concluded that this is sufficient reason to justify the addition of colour in assessments. Importantly, the researchers did note that the "Picture Completion subtest results were complicated by several interactions", specifically by factors relating to age differences between the groups of participants (Husband & Hayden, 1996, p. 147).

There are a few remarks that can be made about Husband & Hayden's (1996) attempt to examine the effects of colour on psychometric performance. Firstly, the results from the study are outdated given that the study was conducted 23 years ago and are merely not generalisable to a South African context. Secondly, the researchers failed to adequately control for a number of extraneous variables, which they did admit to. Two critical issues that required better control were the age of participants and the sample size of the different groups in the study. The remedially assisted children in the sample were on average older than the rest of their sample, which the researchers noted and pointed out that it would be natural that their raw scores would be higher than the younger participants (Husband & Hayden, 1996). Regarding the sample size, a small number of remedially assisted children were sampled into the study compared to the number of 'regular children' sampled, which would have resulted in the different groups in the study being unequal. This would mean that their factorial design in the study would have been unbalanced, and it is possible that the overall validity of their comparisons could be questioned if no correctional methods were used during their analysis (none were mentioned). This is because unbalanced factorial designs are known for having issues regarding homogeneity of variances (Landsheer & van den Wittenboer, 2015), and since the study used standard parametric statistical procedures it is possible that their overall analysis could have been affected.

A second study by Nelson et al. (1994) also examined the effects that colour has on performance by administering coloured and black and white versions of the Picture Completion and Picture Arrangement subtests to 90 elementary students. The study was similar to that of Husband and Hayden's (1996) study, except that they additionally examined whether gender differences had any effect on performance in both versions of the measures. In their study, Nelson et al. (1994) administered colourised and black and white versions of the Picture Completion and Picture Arrangement subtests in a counterbalanced order with the help of graduate students who were trained prior to testing. After the first administration of either the coloured or black and white Picture Completion and Picture Arrangement subtests, participants were required to complete the Kaufman Brief Intelligence Test so that a baseline of their ability levels could be established before the second administration of the Picture Completion and Picture Arrangement was given (Nelson et al., 1994). Their analyses consisted of a series of *t*-tests for related samples for both participants raw and scaled scores, which yield non-significant results (Nelson et al., 1994).

The study concluded that colour had minimal effects on performance in the measures, and that gender differences did not change the magnitude of this effect in any way (i.e. males and females performed at an equivalent level on both versions of the measures) (Nelson et al., 1994). No noticeable flaws in the study can be identified; however the findings are similar to the findings presented by Husband and Hayden (1996) in that they are also outdated and not generalisable to South African populations. Furthermore, though not guaranteed, it may be possible that the item order during testing had an influence on their final analyses. It may be that participants became fatigued after completing the Kaufman Brief Intelligence Test, which in turn may have affected their performance in the final phase of test administration. An important recommendation made by the researchers was that future studies be directed at examining if similar effects are found with samples that include “students with disabilities, such as Attention Deficit Hyperactivity Disorder” (p. 9).

The two studies presented above indicate that the addition of colour in the Picture Completion subtest had no noticeable effects on psychometric performance. However, it was still felt that renewed attempts to examine this possible effect were required, especially in a South African context where similar studies have never been performed. Not only would renewed attempts help add to the limited literature that exists regarding this effect, but it would help in terms of generalisability of previous findings to different populations that are culturally diverse and different to the populations used in previous studies.

Influence of cultural and subjective factors on perception. Opposing views exist in the literature on whether cultural differences affect visual processing (Ueda et al., 2018). Much of the literature that supports this idea suggests that any effects stemming from cultural differences are similar to the differences that have been found in higher level processes, namely thinking and reasoning (Ueda et al., 2018). For example, individuals from America tend to rely more on analytic or focused processing, whereby features or dimensions of important objects within complex scenes are analysed independently from the object’s context (Ueda et al., 2018). In contrast, individuals from East Asia tend to adopt more of a holistic processing procedure, in which the entire perceptual field as a whole is considered, allowing for links and relationships between multiple objects within scenes to be considered (Ueda et al., 2018).

A study by Ueda et al. (2018) attempted to explore this effect of cultural differences in more detail, specifically focusing on the allocation of visual attention, in an attempt to help bring clarity to the mixed views that seem to exist in the literature. Additionally, they were concerned with examining whether the analytic-holistic distinction mentioned earlier could help explain any of the differences found in their study. In a series of experiments, Ueda et al. (2018) demonstrated that there were significant differences in visual search between Western participants and East Asian participants, suggesting that cultural factors may impact underlying processes within visual search. In terms of applying the analytic-holistic distinction to help in their interpretation, they felt that these general differences in processing strategies were not sufficient in explaining the cultural differences observed in their study, and that instead that it’s possible that differences in much earlier levels of processing (e.g. coding of visual stimuli) could have some role (Ueda et al., 2018).

These results support the claim that it is possible for cultural differences to exist in the perceptual process, where not only the high levels of processing are impacted but the much lower and foundational levels too. Thus, findings by Husband and Hayden (1996) and by Nelson et al. (1994) that claim colour does not significantly impact psychometric performance in the Picture Completion cannot just be generalised to a South African population. Since South African populations are culturally diverse, its entirely plausible to argue that conducting a similar study to test colour effects on performance would yield slightly different results, especially if cultural differences in perception are present.

Concerning subjective factors that may influence perception, an article by Lebrecht, Bar, Barrett and Tarr (2012) suggest that micro-valences of objects affect the speed and quality of the perceptual process, specifically regarding object recognition and the effect this perception has on individual’s behaviour towards an object. It is typical for studies that look at the effect of valence on perception to focus on objects that have strong affective valences, for example, guns or roses (Lebrecht et al., 2012). Lebrecht et al. (2012) proposed that much more subtle valences are attached to everyday objects. These valences automatically form part of our mental representations of everyday objects when they are perceived (Lebrecht et al., 2012).

Such micro-valences possibly exist to optimise the visual perceptual process and facilitate higher levels of processing. Regarding the benefits micro-valency has on perception, it ensures that attention is orientated to objects with more positive valency and away from objects with negative valency (Lebrecht et al., 2012). This, in turn, facilitates higher levels of processing, especially when it comes to decisions, in that it helps reduce uncertainty (Lebrecht et al., 2012).

Valency in objects is an interesting idea. In general, this idea assumes that our perceptions of the world around us are not neutral, and instead, our perceptions are more similar to that of representations that can be coloured by our experiences and interactions that we have with particular objects (Lebrecht et al., 2012). It is necessary, then, to examine factors that may influence the valence that is attached to objects, especially factors that uniquely interact with micro-valences. Lebrecht et al. (2012) state that various visual features of objects contribute to the perceived valency we assign or attach to specific objects. Features such as shape, curvature, colour, material and symmetry are suggested as being influential in this process (Lebrecht et al., 2012).

Since the present study is concerned with analysing the feature of colour, it would be beneficial to look at how this specific dimension of an object impacts the valency we attach to objects and the impact it has on our behaviour. Something as simple as colour preference can affect object valency. For example, preferring the colour red over other colours may influence our future decisions with objects that contain red. If confronted by choice of which cup to use when presented with an array of coloured cups, it is more likely that a red cup would be chosen since its micro-valence is higher than the micro-valences of the rest of the non-red mugs.

Thus, the impact that micro-valence has on perception cannot be ignored. Concerning the present study, it is possible that coloured stimuli cards of the Picture Completion may have slightly higher micro-valences than the greyscale stimuli cards. This is, however, only an assumption. It is assumed that the colours, in general, would be more preferred over grey. This assumption is based mainly on the findings from Husband and Hayden (1996) who found that children stated that they had a preference for coloured measures over non-colourised measures. To this extent, the greyscale version of the Picture Completion may have a slight negative micro-valency attached to it, which may subtly change their behaviour by affecting the amount of effort they apply to the test. Based on the capacity model of attention, this decrease in effort would necessarily mean that lower cognitive resources would be allocated to the task, and thus it would be expected that task performance would decrease as a direct result.

Additionally, given that certain stimuli cards in the Picture Completion contains complex scenes which have multiple objects present within them, differing colours of the objects may mean that differing levels of micro-valances are attached to different objects within the scene. To recall, micro-valency also affects which objects are prioritised when being visually processed (Lebrecht et al., 2012). Thus, certain objects with specific colours and higher micro-valences may be prioritised over different objects within scenes, ultimately affecting the way the scene is being visually searched for clues of what aspect of the image is missing. Such an effect on one's visual search may be absent when a non-colourised version of the measure is used instead.

Perception and decisions. In the previous section, it was noted how our perceptions influence decisions. This section extends this idea further by explicitly focusing on the type of environments that are typically perceived when decisions are made. Decisions, as it is spoken about in this section, are defined as perceptual decisions which usually require an individual to make a choice between two binary alternatives, often being asked to judge whether a specific target stimulus was presented while distractor items are positioned within the same field of view. However, tasks such as this do not accurately represent the types of environments individuals are usually in when having to make some perceptual decisions, in addition to the fact that such decisions are usually more complicated than mere binary choices (Baldassi, Megna & Burr, 2006). Individuals usually have to make complex decisions in environments that contain clutter where multiple confounding distractors are present (Baldassi et al., 2006). Such a circumstance is very similar to the Picture Completion task, in that an object or complex scene is presented with multiple distractors existing within the artwork, where the individual is then required to make a decision on which aspect or part of the image is missing (Wechsler, 2003a). Thus, it is crucial to understand the impact that clutter within our perceptions has on the overall quality of our perceptual decisions.

This issue was examined by Baldassi et al. (2006). Their experiment involved two tasks. The first required participants to report the magnitude and direction of tilt of a grating patch presented to them (Baldassi et al., 2006). The second task required participants indicate the direction of tilt and rate their confidence in their decision (Baldassi et al., 2006). Both tasks intended on probing participants internal representations of orientation during visual search (Baldassi et al., 2006). Their findings suggested that the prevalence of errors and response threshold both increased when participants were presented with cluttered stimuli.

Additionally, they discovered that participants confidence in making decisions also increased when viewing cluttered visual stimuli. These results have significant practical implications. One implication is that it predicts that there would be “an increase in high-confidence errors when decisions are made in cluttered environments” (Baldassi et al., 2006, p. 393). Results such as these imply that decisions regarding what is missing in the Picture Completion subtest could be made with high degrees of confidence, regardless of whether they are correct or not, if the pictures presented to them are cluttered with distractors. Thus, more complex scenes could not only lead to an increase in errors made by the participants, but participants would be more confident about the answer they provided.

An interesting question could be posed in response to these results, that is, to what extent can colour as a feature act as a distracting factor that may increase the number of errors a participant makes during the test? It could be said that colour may act as a distractor, but more in terms of how it affects the overall mood and wellbeing of individuals and how this directly impacts one's decisions and performance in tasks. Studies have demonstrated that the children's performance in classrooms can be impacted by the classroom colours which are said to affect children's emotions and physiology (Barrett, Davies, Zhang & Barrett, 2015; Küller, Mikellides & Janssens, 2009). For example, colours such as red can affect children's emotional state by exciting it (Küller et al., 2009). Different studies have demonstrated the effects colour has on performance in specific tasks, showing that blue enhances performance on creative tasks whereas red enhances performance its tasks that are detail-orientated (Mehta & Zhu, 2009). Studies such as these in conjunction with the claims that colour influences the attachment of micro-valences to everyday objects point to the idea that colour in itself is not a neutral feature in perception.

Colour also has been said to be a distracting feature for children with specific learning difficulties (Sarason & Potter, 1947). According to a past study by Sarason and Potter (1947), colour may distract or even confuse children with specific learning difficulties during performance tasks. The authors concluded that "the presence of colored stimulation in a situation where abstraction regarding form relationships is involved prevents some maladjusted children from functioning up to the level of their intellectual capacity" (p. 206). This interference with intellectual functioning, they suggested, could be a result of emotional reactions that are stimulated by colour (Sarason & Potter, 1947). The authors' speculations that colours stimulated emotional responses within the children are supported by Küller et al. (2009). Furthermore, Sarason & Potter (1947) suggested that the interference is caused by an impact colour has on figure-ground relationships (important in perceptual grouping which is fundamental in how we perceptually organise parts into wholes and object recognition (Goldstein, 2008). In this sense, colour "makes visual comprehension of recognizable forms more difficult" (p. 206), thus increasing anxiety levels which have a direct impact on working memory and intellectual capacity (Beilock & Carr, 2003; Sarason & Potter, 1947).

Visual search. An interesting and unsolved debate involving the processes of visual search is whether attention can simultaneously be allocated to multiple objects that are located in different places within the same visual scene (Grubert & Eimer, 2015). Some discussion around this debate is required for the present study. In the Picture Completion subtest, it is more probable that participants are not searching for a single missing aspect in the artwork presented to them, but instead are searching for multiple aspects that could be missing simultaneously, constantly updating what aspects should be included in their search based on clues they are receiving while scanning the cards presented to them.

When individuals begin scanning an object or scene for a specified target feature or aspect, they rely on top-down attentional sets, also known as attentional templates, which help guide the searching process (Grubert & Eimer, 2015). Attentional or search templates play a vital role in the visual search process in that it helps direct (or guide) our mental spotlight to certain elements of an object or scene (Berggren & Eimer, 2016). These templates are activated prior to a search procedure. These templates are quite complex, containing multiple features from different dimensions (for example, a blue triangle) (Berggren & Eimer, 2016). These templates bias the visual search process by guiding one's attention to template-matching features of the object or scene (Berggren & Eimer, 2016).

An essential feature of this template-guided search is that it becomes relatively inefficient for objects/scenes that have features from the same dimension, for example, a red and blue object (Berggren & Eimer, 2016). This idea is supported by Grubert and Eimer (2015) who demonstrated that there exists "qualitative differences in the guidance of attentional target selection between single-color and multiple-color visual search" (p. 1433). They concluded that it is ultimately more effective to guide attention based on a single feature on one dimension. They mentioned that they are still unsure "why visual search for two different target colors is so much less efficient than search for a single color defined target", given that one's working memory has more than enough capacity to have two active colour templates (Grubert & Eimer, 2015, p. 1441).

Using this idea of template-guided visual search, it is possible to expect that someone viewing a greyscale Picture Completion picture, as opposed to a colourised one due to their being multiple features (in this case colour) on the same dimension, may incorporate this template-guided search. For example, if shown a picture of a belt, one may bring up a template of a belt which is used to guide one's visual spotlight to specific features of the picture to see if those features match their mental representation of what a belt looks like. Where a match cannot be made, maybe no holes are present on the belt in the picture but there are holes in their mental representation, then it leads one to believe that this is the aspect of the picture that is missing. In this way, colour may negatively impact one's template-guiding visual search process. However, this is not the only way people visually search, it is merely one way, but we can still see how a feature such as colour may affect an entire process that is crucial to the Picture Completion subtest.

Chapter Three: Methodology

The following section is dedicated to discussing the study's methodology. The chapter begins with outlining the research design that was implemented in this study. Next, a summary of the sampling procedure used is provided along with a description of the final sample that was obtained. The different measurement instruments and general data collection procedure is comprehensively outlined. Finally, the chapter ends by discussing the specific statistical analyses that were conducted.

3.1 Research Design

The present study adopted a 2 x 2 between-subjects factorial design. The two levels of the first factor had to do with the version of the measure that was administered to the participants, either being a coloured version or greyscale version of the Picture Completion subtest in the WISC-IV (2003a). The second factor was the type of school the participant belonged to, and this also had two levels, namely a mainstream school or a school that caters for children who have various learning difficulties and thus require specialised assistance. The factorial matrix below visually illustrates this design:

Table 1

A factorial matrix representing a 2x2 between-subjects design for the present study

	School Type	
Version of Measure	School for Learning Difficulties	Mainstream School
Coloured	*(Group A)*	*(Group B)*
Greyscale	*(Group C)*	*(Group D)*

As seen in in the factorial matrix above, four distinct groups were present in the study which have been labelled accordingly within each cell. Since the present study adopted a factorial design, it is necessary that each independent-dependent variable pair (hereafter referred to as IV-DV pair) is considered separately as this examines the extent to which conclusions drawn from either the main or interaction effects are justified. For the first factor, Measure, and the dependent variables, Score and Time: (a) the independent variable was manipulated; (b) there was an experimental and a control group; and (c) there was random assignment. Therefore, the design for the first IV-DV pair was a true experimental. For the second factor, School, and the dependent variables, Score and Time: (a) there was no manipulation of the independent variable since the type of school the participant belonged to cannot be directly altered due to the fact that participants already belonged to these pre-existing categories; (b) there were contrasts groups; and (c) there was no random assignment. Therefore, the design for the second IV-DV pair was a non-experimental, ex-post facto design.

Based on these individual designs, it is reasonable to conclude that causal conclusions made about the version of the measure administered in relation to the two dependent variables would be justified. Even though the design suggests that one should be cautious regarding causal conclusions about the type of school the participants belonged to in relation to the dependent variables, it is expected that the two schools would differ in both their average scores and times. Children with mild to moderate mental retardation or with learning difficulties tend to score on average lower than the general population (Wechsler, 2003a). Thus, causal conclusions regarding the second factor could be argued to be theoretically justified to a certain degree.

3.2 Sample

Sampling procedure. The sample was obtained using a probability sampling method, specifically stratified random sampling. Children in each school were separated into two different strata. Children were first grouped according to their grade (Grade 1 to 7) and then according to their sex (male or female). This grouping strategy allowed for an equal number of males and females from each grade to be selected for the final sample, and thus it ensured some degree of control over age and sex which could act as extraneous variables in the study.

The parents of the children in multiple classes within each grade were provided with an information pack (see Appendix A-C) that detailed the study. Parents who were interested in allowing their child to participate in the study were required to sign the provided consent form and return it to the school before testing began. From the names of the students that returned the signed consent forms, a random selection of three males and females from each grade were chosen for the final sample.

Problems regarding response rate issues, especially in the lower grades, slightly impacted the implementation of this sampling method. It was noted that there was not always a large enough number of potential participants in each stratum to randomly select from, resulting in the random selection component for particular strata to be abandoned. This meant that it was not always possible to ensure an equal number of male and female participants were chosen for the final sample. Additionally, this issue meant that the different groups within the design (see Table 1) were unequal.

Sample description. The final sample size obtained was 87 children and was comprised of 46 males and 41 females ($M_{age} = 10.28, SD = 2.10$). Of the final sample, 52 participants attended a mainstream school and 35 attended a school that caters for children who require remedial assistance. In terms of ethnicity, most of the participants were Black (N = 65, 74.71%), followed by White (N = 16, 18.39%), Indian (N = 4, 4.60%) and Coloured (N = 2, 2.30%). Given that the present study adopted a factorial design, the table below provides an overview of the number of participants in each group in the design along with the groups average age:

Table 2

The number and the average age of participants belonging to each group within the study.

	School Type	
Version of Measure	School for Learning Difficulties	Mainstream School
Coloured		
N	18	27
M_{age} (SD)	11 (2.57)	9.89 (1.89)
Greyscale		
N	17	25
M_{age} (SD)	10.71 (2.09)	9.88 (1.88)

Supplementary bar graphs were constructed to provide a graphical representation of the sample characteristics. The figures presented in Appendix H depict the number of participants that were allocated to each condition and the number of males and females assigned to each version of the measure respectively.

An a priori computation of the required sample size for the design was performed using G*Power Analysis (Faul, Erdfelder, Lang, & Buchner, 2007) in order to determine the appropriate number of participants that were needed for the study. To compute the required sample size, the following input parameters were entered into G*Power analysis:

Effect size. $f = .40$, corresponding to a large effect size.

Alpha level. $\alpha = .025$ (see page 45 for discussion on the critical *p*-value adjusted using the Bonferroni method).

*Power***.** $1 - \beta = .975.$

*Numerator df***.** For the two main effects, one would take the number of levels of the factor and subtract one. Therefore, since the version of the measure had two levels and the school type had three levels, the numerator *df* would be equal to one and two respectively. For the interaction effects, the results for the main effects are multiplied together, and thus the numerator *df* would be equal to two.

*Number of groups***.** Four.

It was calculated that a minimum total sample size of 133 participants was needed ($N_{per\ cell} = 33$) in order to obtain a power level of .975 with an estimated effect size (*d*) of .40 (Howell, 2013; Richardson, 2011). However, given that a primary goal of sampling was aimed at selecting an equal number of males and females within each cell, the total sample size required was increased by three to allow the $N_{per\ cell}$ to be equal to an even number, 34.

By looking at Table 2, however, one would see that the cell sizes do not meet this minimum sample size of 24 and that the cell sizes are not equal. This is mainly due to the previously discussed sampling issues that were encountered. The number of participants willing to participate in the school for learning difficulties was lower than that from participants in the mainstream school. To compensate for this, a higher number of participants were sampled from the mainstream school to help increase the overall total sample size to the desired amount. By doing so, a total sample size of 87 participants was obtained, which is notably lower than the minimum sample size that was decided upon a priori. Regardless, given this sample size, a post hoc power analysis performed in G*Power estimated that a power level of .86 was achieved, which is well below the initial .975 power level that was aimed for but can still be regarded as acceptable.

3.3 Instruments

Three measuring instruments were used in the study, these being: (1) the Dvorine (1953) Colour Blindness Screening Test (2) a demographic questionnaire; and (3) the WISC-IV (2003a) Picture Completion subtest.

The Dvorine (1953) colour blindness screening test. Colour blindness was screened for using the Dvorine (1953). The Dvorine (1953) is a commonly used test of perception that can identify colour blindness and defective colour vision (Sattler & Dumont, 2004). Testing participants ability to see colour is vital since defective colour vision/colour blindness would affect the data in undesirable ways. The test consists of cards printed with different coloured dots on them (Lezak, Hannay, Howieson & Loring, 2004). The coloured dots form recognisable numbers on the cards, numbers that colour-blind individuals would be unable to recognise since all the dots would look the same to them (Lezak et al., 2004). The participants were shown the series of cards and asked to specify what number they were being shown.

This screening measure is satisfactory in testing for the two most common types of colour blindness (Lezak et al., 2004). A weakness in this screening measure, however, is that it fails to detect two rare forms of colour blindness (Lezak et al., 2004). This issue was considered to be unimportant since it was argued that it was unlikely that a large number of participants who suffer from such rare forms of colour blindness would be encountered, if any at all.

Demographic questionnaire. A demographic questionnaire was verbally administered to participants (see Appendix D). The purpose of the questionnaire was to obtain essential demographic information required to provide a descriptive overview of the overall sample. In addition to this, the questionnaire contained questions that asked participants about the type of reading material they are exposed to, the frequency at which they read, whether the material they read is in colour, and questions relating to television exposure. These questions intended on gathering information that would provide an idea of how much coloured media and reading material the participants are exposed to in their lives. It was believed that such information was essential to know since it was thought that children who are more exposed to colour throughout their lives might expect the measures administered to them to be colourised. These questions, however, were abandoned because the researcher noted that participants often battled to provide an answer to some of the questions they were asked. Often, participants would specify that they were unsure on how to answer the question, leading them to either guess an answer or simply not provide a response at all.

A shorter demographic questionnaire (see Appendix A) was also constructed that was sent home with the consent form for the parents to complete. Specifically, the parents were asked whether their child is prescribed any medication, wears glasses or has any medical conditions that the researcher should be made aware of. The questions were designed to provide further information that could be used to help determine whether the child should be excluded from the study. Children were excluded from the study if: (a) they suffered from any severe vision deficits; (b) suffered from any severe medical conditions (e.g. a mild traumatic brain injury); (c) were diagnosed with specific psychological disorders (e.g. autism); (d) if they wore glasses and failed to bring them along with them for testing; and (e) were on medication and did not take their prescribed dosage as per their medical practitioners guidelines.

Picture Completion. The Picture Completion is a supplemental subtest associated with the Perceptual Reasoning Index (PRI) in the WISC-IV (Kaufman et al., 2006). The Wechsler scales are said to be the most widely used intelligence batteries within psychology, so it is not surprising that the WISC is commonly used in both clinical and research settings to measures children's intellectual functioning in specified cognitive domains (Kaufman et al., 2006; Shuttleworth-Edwards et al., 2013; Wechsler, 2003a). Even though the WISC-V (2015) is currently available, the present study used the WISC-IV (2003a) because the Picture Completion was dropped from the WISC-IV when it was revised (Kaufman, Raiford & Coalson, 2015).

The WISC-IV is valid for children between 6 years and 16 years and 11 months, and the entire test usually takes just over an hour to administer (Aylward & Stancin, 2008). The "WISC-IV contains a Verbal Comprehension Index (Similarities, Vocabulary, Comprehension, *Information*, and *Word Reasoning*), a Perceptual Reasoning Index (Block Design, Picture Concepts, Matrix Reasoning, *Picture Completion*), a Working Memory Index (Digit Span, Letter-Number Sequencing, *Arithmetic*), and a Processing Speed Index (Coding, Symbol Search, *Cancellation*)" (Aylward & Stancin, 2008, p. 154, [emphasis added]; Wechsler, 2003a). The subtests that are italicised are supplementary tests that can be administered in place of or in addition to the core subtests in each index. The WISC-IV (2003a) also contains a measure of general intellectual function in addition to the four indexes previously mentioned which enables one to get a general composite score for the entire measure (Aylward & Stancin, 2008; Shuttleworth-Edwards et al., 2013).

Internal consistency coefficients for the WISC-IV (2003a) on a US population of children were found to be: .94 for VCI, .92 for PRI, .92 for WMI, .88 for PSI, and .97 for FSIQ (Flanagan & Kaufman, 2004; Sattler & Dumont, 2004). The WISC-IV (2003a) is also considered to be stable over time, with the majority of the indexes having stability coefficients above .80, with the exception of Processing Speed and Working Memory being just below (Sattler & Dumont, 2004). Also, the WISC-IV (2003a) has adequate criterion validity and is “supported by correlations with other global measures” (Flanagan & Kaufman, 2004, p. 37).

The present study did not utilise the entire WISC-IV (2003a), instead only the Picture Completion subtest was used. This subtest was chosen for the study since the test materials once lacked colour in the past, and were only colourised during the revision of the WISC-R (1974) to the WISC-III (1991) (Flanagan & Kaufman, 2004; Green, 1992). The Picture Completion contains a series of coloured artwork, or pictures, printed onto a set of stimuli cards. Each picture has a single piece or aspect of it that is missing. The subtest requires that the child locate and specify what they think is the single most important aspect missing from the picture within a specified time limit (Flanagan & Kaufman, 2004; Sattler & Dumont, 2004). If the response provided by the child is unclear, the researcher is allowed to ask the child to clarify what they mean. The average time it takes to administer the Picture Completion has been found to be 10 minutes and 26 seconds (Ryan, Glass & Brown, 2007). The researcher noted that administration time was typically around 10 minutes, but did fluctuate depending on the age and educational ability level of the participants.

The researcher manually administered two versions of the Picture Completion to the participants. The first version was the standard colourised Picture Completion that is found in the WISC-IV (2003a). The second version contained the same pictures, only that the artwork was edited so that they were no longer in colour. Both versions of the subtest were printed onto the same cardboard using the same high-quality printer. This was done to ensure that no differences in the material of the cards or the quality of the printing between the version could interfere with the results. Due to copyright laws, examples of the Picture Completion cards cannot be provided in this report.

Tracking program. A Python (version 3.6.5; Van Rossum, 1995) program was coded to facilitate the data collection process by helping with data management and was executed at the beginning of each new test administration session. The program was relatively simple in its design. The program was designed to ask the researcher to input the necessary data at each stage of test administration, these being: (a) the generation of the participant code; (b) consent and assent check; (c) Dvorine (1953) test; (d) the demographic questionnaire; and (e) the Picture Completion subtest.

3.4 Procedure

A standard email was initially sent to various schools located in the Johannesburg area and acted as an invitation that invited the schools to participate in the study while also serving the purpose of informing them about the study and its intentions (see Appendix E). In response to the initial emails, a mainstream school and a school that offers remedial assistance to learners that require it accepted the invitation to participate. Before collecting data, the researcher made contact with the two schools to schedule days within the week when children could be taken out of class individually in order to complete the measures without adversely impacting on teaching and learning activities. Meetings at the school were then arranged in order to discuss details about the study's procedure. Specifically, the meetings allowed the researcher to discuss the allocation of private areas on the school premises (such as empty classrooms or conference rooms) where testing could take place. This was discussed in order to prevent other students and staff from interrupting the testing procedure. The meeting also provided the schooling institutions with the opportunity to ask any questions and raise any concerns they might have regarding the study.

Each school allocated a private room which contained a table and two chairs. The chairs were arranged around the table such that the researcher sat at roughly a 90˚ angle to the participant. All testing materials were kept off the main table and were only placed in front of the participant when needed. A laptop was also positioned in front of the researcher which was used to execute the program intended on capturing participants data and to facilitate time management. Testing was done at each school individually over the course of a few days. To make the testing procedure more systematic, each grade was tested separately. Children from a particular grade were called out of class and tested individually. Once a grade had been completed, children from the next grade were called out of class to be tested, and so forth. Children were only asked to leave class and come to testing once it was their turn to be tested, which made sure that children were not kept out of class for too long.

The testing procedure began with collecting the participant from class and guiding them to the testing location. Once the participants took their seats, initial introductions were made whereby the researcher would introduce himself and outline the study to the participant in clear and simple English, which sometimes involved using an information sheet (see Appendix B) if required. The information sheet was typically referred to when the participant was very young as it allowed the researcher to outline the study in a more graphical way which helped better explain the study to the participant. The initial greetings and discussion about the study also served the purpose of rapport building to help ease any test anxieties the participant might have had, and thus they were kept brief and informal. During this introduction period, the tracking program was executed which asked the researcher to input the name of the participant, their grade and the version of the measure the researcher was planning on administering during the session. From this information, a participant code was generated (e.g. P1C1) and used throughout the administration to label the participant's information to ensure confidentiality and anonymity in the final report.

Once the purpose of the study and what the study entails was described to the participant, verbal assent was evaluated. Though the parents of the participant had already provided informed consent, where it was expected that the participant would have also provided verbal assent to their parents, participants were asked if they were still interested in participating in the study. The researcher required that verbal assent is given again before test administration to allow the participants the opportunity to change their minds about participating. The assent assessment was performed through the coded program. The program displayed a series of questions on the screen which the researcher asked the participant. Participants were required to provide yes or no answers to the questions. If this check failed (i.e. the participants responded no to any of the questions), the program would not continue.

Next, the participant's documents were briefly scanned, specifically the questionnaire that the parents answered, to check for any serious medical or psychological conditions that may have warranted that the participant be excluded from the study. For example, if the parents indicated that their child was diagnosed with autism, then the testing procedure would be terminated and the child would be sent back to class. Participants were also screened for colour blindness using the Dvorine (1953). Participants who performed poorly in this screening tool were excluded, and their names were documented by the tracking program so that their parents could be provided with information of nearby optometrists for them to take their child for a professional check-up.

The demographic questionnaire (see Appendix D) was verbally administered to the participants who passed the screening procedures. Testing was done verbally as it was believed that younger children and children with specific learning difficulties might struggle to answer the questions without assistance. Furthermore, by verbally administering the demographic questionnaire, questions that the participant was unsure about could be asked in a different, but equivalent, way. The demographic questions were shown on the screen by the tracking program which the researcher would ask the participant and then recorded their responses in the program so that they could be transferred to a sperate document.

Finally, the Picture Completion subtest was administered to the participant. The version of the measure (coloured or greyscale) that was administered was randomly assigned to each participant prior to them being called out of class. The tracking program did remind the researcher of this before it allowed participants data to be entered. The test and what was required of the participant was first explained in full, before allowing the participant the opportunity to ask any questions they may have regarding what was expected of them. The test was then administered and participants responses and time taken for each of the Picture Completion cards were recorded in the program. When a card was shown, the researcher would press a button and the program would begin a timer, which would be stopped when the researcher re-pressed the button indicating that the participant had responded. The program would then request the actual response the participant gave, whether it was correct or not, and finally whether the participant pointed. If six cards were consequently answered incorrectly, the program would terminate the testing session, as per the guidelines stated in the WISC-IV testing manual (Wechsler, 2003a). Once all items in the subtest had been completed or the test adminstration was terminated, the participant was given a final opportunity to ask any questions or raise any concerns they might have had before being escorted back to class.

3.5 Ethical Considerations

Ethical approval (protocol no.: H18/07/06) was obtained from the University of the Witwatersrand’s Human Research Ethics Committee (HREC Non-Medical) before the commencement of the study (see Appendix E). Since the sample comprised of children, both signed informed consent from the participant's parents and verbal assent from the participants was obtained before they were included in the study. Parents were informed about the study’s purpose and what it entails through an information sheet that was sent to them along with the consent form. Parents were made aware that they should discuss the study with their child and ensure that they verbally agree to participate in the study before signing and returning the consent form to the researcher. A child-friendly information sheet was also sent to the parents to help them explain the study to their child in a more simplistic way that they would easily understand. Children who returned the signed consent forms to the school were sampled into the study. Prior to test administration, the purpose of the study was re-explained to the participants, and verbal assent was obtained again to provide an opportunity for the participant to change their mind about whether they wanted to participate.

Parents were assured that their child's personal information and data would remain confidential. Participants all received a unique participant code which acted as a label for their data. This ensured that the participant's personal information remained separate from their data gathered by the instruments. All data was stored on a personal computer that was password protected in an encrypted folder.

3.6 Analysis

Software. This study incorporated quantitative techniques to analyse the participant's data. Before data analyses could begin, the data had to be cleaned to ensure that the data sets were constructed accurately and that all data were correctly entered into the various data matrices. Python (version 3.6.5; Van Rossum, 1995) was used to facilitate the data management and cleaning process. Once the data sets were processed, descriptive and inferential data analyses were performed using RStudio (version 1.1.456; RStudio Team, 2018), a free IDE (integrated development environment) for R (version 3.5.1).

Coding of variables. Before discussing the different data analyses that were used, it is important to briefly mention the two independent and two dependent variables in this study and describe how they were coded into the data sets. The two independent variables that acted as factors during the analysis were relatively simple to code. The version of the measure a participant received was coded as either 0 or 1, where the 0 represented the greyscale measure and 1 represented the coloured measure. The type of school the participant belonged to was coded as either 1 or 2, representing children who attended a mainstream school and children who attended a school which provided remedial assistance respectively.

In regards to the dependent variables, the first dependent variable, the participant's score, was scaled using the WISC-IV scoring manual (2003b) to produce a standardised score. The second dependent variable, the average time taken to give a correct response, had to be calculated since multiple time scores existed for each participant. For example, if one participant answered nine questions in the Picture Completion correctly, a measure of central tendency had to be calculated to represent those nine time scores. It was decided that the traditional arithmetic mean was an inappropriate measure of central tendency given that each participant’s set of time contained outliers that distorted the mean. For example, say a participant set of time for correct responses contained the following values:

2.63 3.57 3.68 4.25 4.42 17.69 18.01

The arithmetic mean for the above data set can be calculated using the following equation (Wilcox, 2012):

$$\bar{X} = \frac{1}{n}\sum X_i$$

Using this equation on the data set above, the arithmetic is calculated to be 7.75. It is clear that the mean is being distorted by the two outliers (17.69 and 18.01). Differences in item difficulty in the Picture Completion subtest resulted in multiple cases similar to the example represented above, and thus an arithmetic mean was considered to be inappropriate as it gave an inaccurate representation of the average time it took participants to give a correct response.

Therefore, a robust measure of central tendency was used instead, specifically Huber's M-estimator. Briefly, Wilcox (2017) describes how the M-estimator of location is represented by the value $\hat{\mu}_m$, which is calculated by solving

$$\sum \psi \left(\frac{X_i - \hat{\mu}_m}{MADN} \right) = 0$$

using iterative estimation procedures (see Wilcox (2017) for a detailed discussion on how to compute an M-estimator). The M-estimator for the above data set is 4.21, which is a much better measure of central tendency for the data set compared to the arithmetic mean. The *WRS* package (Wilcox & Schönbrodt, 2018) which contains numerous R functions was used to calculate an M-estimator for each participant using their time taken for each correct response.

Exploration of data. The data was checked to see if it violated any assumptions that standard parametric analyses rely on, specifically focusing on testing the assumption of normality, the assumption that no significant outliers are present in the data and the assumption of homogeneity of variance. The dependent variables were separately checked for normality for each of the four groups using the Shapiro-Wilks test which was performed using the function *shapiro.test()*. Outliers were detected using graphical boxplots in combination with a modification of the common boxplot rule using the *adjboxout()* function. Since the common boxplot rule has been criticised for declaring too many points outliers where there is skewness in the data (Wilcox, 2017), a modified version of boxplot rule that is based on a robust measure of skewness was also used to see if both detection methods agree with one another. The assumption of homogeneity of variance of the two dependent variables was tested separately for each of the four groups using Levene's test which was performed using the function *levene.Test()*. After the initial assumptions were checked, descriptive statistics and histograms were constructed to help explore the data. All graphs were constructed using the GGPlot2 library (Wickham, 2016).

Factorial ANOVA. The assumption checks that were performed found that there were issues with normality and outliers were present within the groups. It is typical in psychology for researchers to correct these problems by removing any outliers that were detected (which also helps with normality issues), by transforming the data using a variety of different techniques or by performing non-parametric tests (Erceg-Hurn, Wilcox, & Keselman, 2013). However, such approaches are not advised. Transformations often fail to correctly restore normality and complicate the interpretation of results (Erceg-Hurn et al., 2013). Additionally, non-parametric techniques have been known to also suffer from the same problems that parametric tests encounter (Erceg-Hurn et al., 2013). A more effective way of dealing with assumption violations is to conduct a robust statistical method.

Researchers in psychology commonly apply classical parametric statistics, even when the assumptions that those test procedures are built on have been violated (Erceg-Hurn et al., 2013). Literature has shown (see Micceri (1989)) that psychological data is often not normal, and thus the use of traditional parametric measures has been increasingly questioned over the years (Erceg-Hurn et al., 2013). A much more attractive way of dealing with non-normal data is to apply robust methods since they are not drastically impacted by deviations in assumptions (Erceg-Hurn et al., 2013). Therefore, a robust version of a two-way ANOVA was performed for both dependent variables separately given that the data did violate parametric test assumptions.

The function *t2way()* was used to perform a robust two-way ANOVA based on trimmed means. It is important to note that the function *t2way()* does not report the degrees of freedom since an adjusted critical value is used (Mair & Wilcox, 2018), and thus they will not be mentioned when the results are presented in the following chapter. A Bonferroni correction was applied to the critical *p*-value since multiple ANOVAs were being performed. Thus, a critical *p*-value of .025 was used to determine whether the results of the ANOVAs were significant. Effect sizes were calculated for all significant effects founds using the function *yuenv2()*, which computes a robust explanatory measure of effect size ξ. It was decided to avoid using conventional measures of effect sizes, such as Cohen's d or η^2, since they tend to be somewhat biased, especially when there is non-normality (Lakens, 2013; Wilcox, 2017). Thus, the measure of effect size ξ was used instead, where $\xi = .15, .35$, and .50 correspond to small, medium, and large effect sizes respectively (Wilcox, 2012; Wilcox, 2017). The interaction graphs were constructed using robust measures of central tendency, specifically 20% trimmed means in conjunction with their standard errors.

At this point, it is possible that the choice of conducting two factorial ANOVA's instead of a single multivariate ANOVA (often referred to as MANOVA) is being questioned, especially since this approach does increase the familywise error rate. A MANOVA assumes that there exists a linear relationship between the two dependent variables and that they correlate with each other by some degree. This assumption is why one would decide to analyse the two dependent variables together in the first place. The two dependent variables in this study lacked this linear relationship and correlation that a MANOVA requires, and thus it was argued that the two dependent variables be analysed separately using two factorial ANOVA's.

Chapter Four: Results

The following chapter is dedicated to discussing all results obtained from the different analyses performed. The chapter begins with reporting the results from the assumption checks that were conducted, followed by general descriptive statistics that were used to explore the data. Finally, the results for both the factorial ANOVA's are presented and interpreted. Supplementary material has been made available to demonstrate how the analyses were executed in R by providing the R code that was used along with the outputs that were produced.[1]

4.1 Assumption Checks

Normality. The Shapiro-Wilks test of normality was separately performed for each group that was presented in the factorial matrix (see Table 1). The results of the tests for each dependent variable are shown below:

Table 3

A table representing the results for the Shapiro-Wilks tests of normality.

	Values	
Group	W	*p*-value
Coloured Measure		
Mainstream School		
Score	.97	.72
Time	.94	.33
School for Learning Difficulties		
Score	.93	.08
Time	.93	.08
Greyscale Measure		
Mainstream School		
Score	.96	.63
Time	.98	.90
School for Learning Difficulties		
Score	.96	.40
Time	.86	.00*

*Significance at $p < .05$

The results from Table 2 indicate that the distribution of the Time variable for the group who received the greyscale measure in the school which caters for learning difficulties is not normal (p = .00). Furthermore, the normality of the distributions for both the Score and Time variables for the group who received the coloured measure in the school which caters for learning difficulties could also be questioned, given that the p-values are close to significance (p = .08). Based on these results, the assumption of normality has been violated and it was expected that significant outliers within the groups exist that were distorting the distributions.

Outlier detection. It is essential to identify any significant outliers that exist in the groups since the assumption of normality was violated. The first method used to identify probable outliers was graphical boxplots. These can be found in Appendix G. As noted earlier, the common boxplot rule has been criticised for declaring too many points outliers when there is skewness, and thus a modification of the boxplot rule was used as an additional outlier detection procedure. Based on results for both detection methods presented in the table below, it can be seen that the common boxplot rule detected a total of seven outliers within the groups whereas the modified boxplot rule detected four. Regardless of what detection method was used, it is clear that significant outliers exist within the groups. The presence of these outliers is most likely the reason why some groups distributions are not normal. The participants that were declared as outliers were not removed for the remainder of the analyses since robust statistical methods were utilised.

Table 4

A table showing the number of outliers detected within the groups based on the detection method used.

	Outlier Detection Method	
Group	Boxplot Rule	Modified Boxplot Rule
Coloured Measure		
Mainstream School		
Score	1	0
Time	0	0
School for Learning Difficulties		
Score	1	1
Time	2	0
Greyscale Measure		
Mainstream School		
Score	0	0
Time	0	0
School for Learning Difficulties		
Score	0	0
Time	3	3

Homogeneity of variances. The Levene's test was used to test the assumption of homogeneity of variances for each group. Levene's tests were performed across the different Measure and School groups separately for each dependent variable. For the Measure, non-significant results were found for both the Score and Time, $F(1,85) = .91$, $p = .34$ and $F(1,85) = .04$, $p = .85$ respectively. Non-significant results were also found for the School, resulting in $F(1,85) = 3.03$, $p = .09$ for the Score and $F(1,85) = .11$, $p = .74$ for the Time. However, given that the interaction between the variables is the primary concern, Levene's tests were performed to test whether the variance between the four groups differed. This resulted in a non-significant result for the Score, $F(3,83) = 1.38$, $p = .25$, and for the Time, $F(3,83) = .04$, $p = .99$. Overall, all tests were found to be non-significant (i.e. $p > .05$), meaning that the assumption of homogeneity of variances was not violated.

4.2 Descriptive Statistics

Since robust statistical methods were utilised in this study, it makes sense to report descriptive statistics that are based on robust measures of location and dispersion. Therefore, the arithmetic mean and standard deviation are replaced by the 20 % trimmed mean ($\bar{X}_t, \gamma = 0.2$) and the 20% trimmed standard deviation (SD_t) respectively. These descriptive statistics for each cell of the design are shown in the Table 5 and 6 below. Please note, the calculations of the arithmetic means and standard deviations have been provided in the supplementary material should one wish to view them (see Footnote 1).

Table 5

Descriptive statistics for the score variable.

Measure	School	$\bar{X}_t$	SD_t	N
Colour	Mainstream	10.65	2.60	27
	School for LD	8.92	3.18	18
	Total	10.07	2.75	45
Greyscale	Mainstream	9.87	2.25	25
	School for LD	7.64	2.47	17
	Total	9.04	2.33	42
Total	Mainstream	10.28	2.50	52
	School for LD	8.24	3.60	35

Note. $\bar{X}_t$: 20% trimmed mean; SD_t = 20% trimmed standard deviation

Table 6

Descriptive statistics for the time variable.

Measure	School	$\bar{X}_t$	SD_t	N
Colour	Mainstream	4.90	1.36	27
	School for LD	4.51	1.67	18
	Total	4.75	1.57	45
Greyscale	Mainstream	4.74	1.20	25
	School for LD	4.87	1.51	17
	Total	4.79	1.34	42
Total	Mainstream	4.80	1.27	52
	School for LD	4.68	1.71	35

Note. $\bar{X}_t$: 20% trimmed mean; SD_t = 20% trimmed standard deviation

4.3 Factorial ANOVA

To examine if any differences between the groups existed, a series of robust two-way ANOVA's based on trimmed means were performed for both the Score and Time as dependent variables, with the Measure (Colour vs Greyscale) and School (Mainstream vs School for LD) as between factors. To reiterate, robust ANOVAs were performed due to the assumption of normality being violated and since significant outliers existed within the different groups, thus ensuring that these assumption violations do not affect the analysis.

The first robust ANOVA, with the Score as the dependent variable, revealed a non-significant main effect of Measure, Q = 2.90, p = .10. This indicates that the scores obtained by the participants who received the coloured measure ($\bar{X}_t = 10.07, SD_t = 2.75$) were not significantly different than the scores obtained by participants who received the greyscale measure ($\bar{X}_t = 9.04, SD_t = 2.33$). However, the scores for the coloured measure were on average numerically higher than the scores for the greyscale measure. A significant main effect of the School was found, Q = 10.72, p = .00, $\hat{\xi} = .51$, revealing that the scores obtained by participants in the mainstream school ($\bar{X}_t = 10.28, SD_t = 2.50$) were significantly higher than the scores obtained by participants in the school for learning difficulties ($\bar{X}_t = 8.24, SD_t = 3.60$). Lastly, there was a non-significant interaction effect between the version of the measure administered and the different schools, on the score obtained in the Picture Completion, Q = .17, *p* = .68. This indicates that participants that belonged to different schools were not affected differently by the version of the measure that was given to them in terms of their score.

The second ANOVA that was performed, with Time as the dependent variable, found non-significant main effects of both the Measure (Q = .08, *p* = .78) and School (Q = .13, *p* = .72) on the average time taken by participants to provide a correct response. Thus, it can be said that the average time taken to provide a correct response in the coloured measure ($\bar{X}_t = 4.75, SD_t = 1.57$) was not significantly different to the average time taken in the greyscale measure ($\bar{X}_t = 4.79, SD_t = 1.34$). Furthermore, no significant differences in the average time taken to provide a correct response was present between the mainstream school ($\bar{X}_t = 4.80, SD_t = 1.27$) and the school for learning difficulties ($\bar{X}_t = 4.68, SD_t = 1.71$). The ANOVA also revealed a non-significant interaction effect between the version of the measure administered and the different schools, on Time, Q = .55, *p* = .47. This suggests that the average time taken to provide a correct response by participants in different schools was not affected by the version of the measure that was administered.

The graphs depicting the interaction effects for both the dependent variables are shown below:

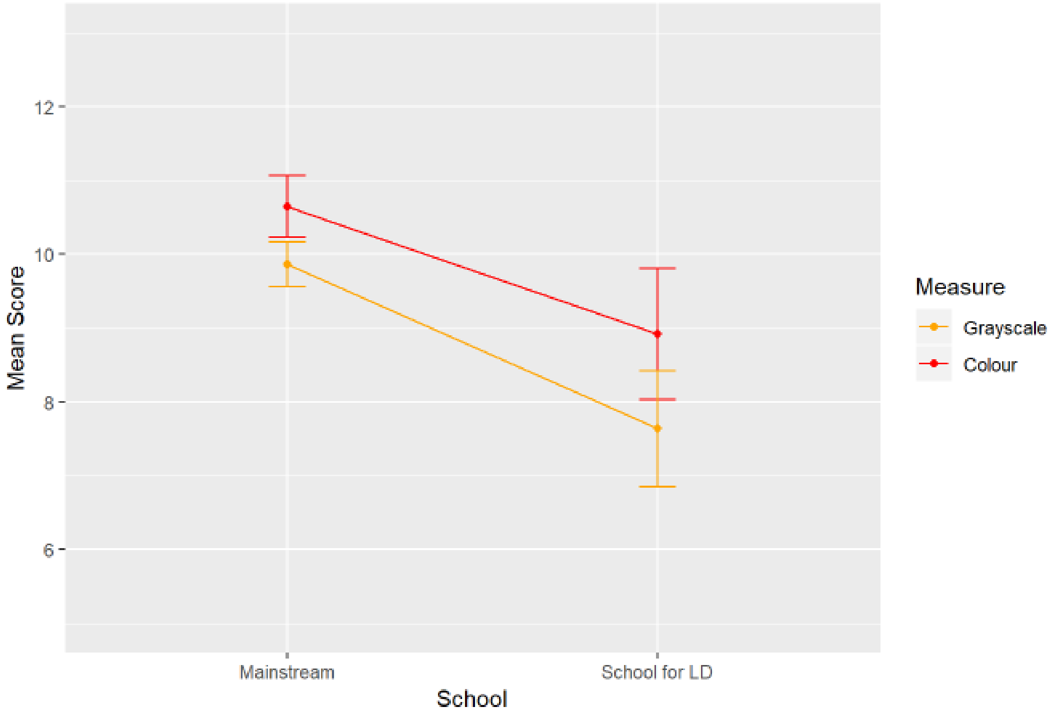


Figure 4 Interaction graph, dependent variable = Score.

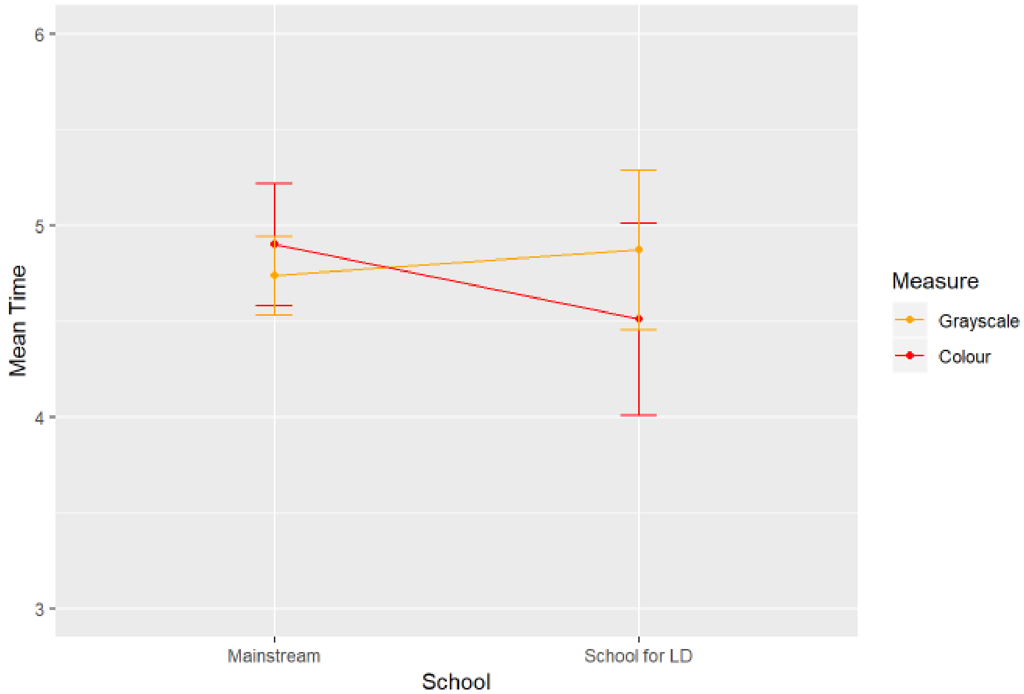


Figure 5 Interaction graph, dependent variable = Time.

In Figure 4, the lines are very close to being parallel, which suggests that no interaction effect would be present. The results from the analysis confirm this finding. The graph suggests that the scores obtained by the participants who attend a mainstream school were higher than the scores obtained by the participants who attend the school for learning difficulties, a finding that would be expected and has been demonstrated to exist (Wechsler, 2003a). The graph also shows that participants who received the coloured measure scored numerically higher than participates who received the greyscale measure, regardless of what school the participant belonged to. It should be noted that this difference was not found to be statistically significant during the analysis, though this consistent pattern of difference between the two versions of the measures will be discussed in more detail in the chapter that follows.

In Figure 5, the lines are not parallel as in the first graph but instead cross, implying that a significant interaction effect may exist. However, results from the analysis found a non-significant interaction effect. Part of the reason for this finding is that non-parallel lines do not automatically mean that a significant interaction exists, instead, the interaction effect is dependent on the degree to which the lines are not parallel (Field, Miles, & Field, 2012). Another interesting detail to note about Figure 5 is the large error bars for participants who attend a school for learning difficulties. Reasons for why such large standard errors exist are focused on more during the discussion. For now, it is more beneficial to note that the large error bars seem to indicate that a large amount of variability within the data may exist.

Chapter Five: Discussion

5.1 General Discussion

The primary objective of the present study was to investigate whether the addition of colour to psychometric measures impacts an individual's performance. This link between test adaptions involving colour changes to measures and psychometric performance was explored by examining whether the addition of colour to the Picture Completion had an incremental or decremental effect on individuals overall score for the subtest and average time taken to provide a correct response to an item. It was initially believed that the coloured version of the Picture Completion subtest would have a significant positive effect on psychometric performance in contrast to the greyscale version, and that this effect would differ depending on participants educational needs. This hypothesis seemed well supported by much of the literature previously reviewed which provided theoretical and empirical evidence which justified the importance of colour in perception, especially regarding the benefits colour has in the perceptual process as well as in the various complex perceptual structures that are within it.

However, the results of the present study yielded some interesting findings which seem to suggest that this initial hypothesis was incorrect. The findings indicated that participants performance on the coloured and greyscale Picture Completion did not significantly differ. Additionally, the non-significant interaction effects between the version of the measure and the type of school on both participants score and time suggest that test adaptions concerning colour changes to test stimuli do not impact the performance of children who differ in their educational ability level. Both findings are consistent with those of Husband and Hayden (1996) and of Nelson et al. (1994) who found that colour adaptions did not impact performance in psychometric measures. Given that the present study examined the same area of interest and shared certain procedural elements from both studies, finding similar results contribute in corroborating these past findings.

The single main effect found to be significant suggested that the scores obtained by children who differed in educational ability level were statistically different. Husband and Hayden (1996) detected a similar result, and thus the presence of this significant main effected contributed in corroborating their findings. Moreover, independent examiners and researchers were tasked in examining the performance of special group samples in the WISC-IV to evaluate the scales clinical utility given that the scale is commonly used as a diagnostic assessment tool (Wechsler, 2003a). They found that the group with Attention Deficit/Hyperactivity Disorder (ADHD) and Learning Disorders, which often have high comorbidity, had substantially lower average composite scores to the matched control group (Wechsler, 2003a). Of particular importance, moderate effect sizes were found for the Perceptual Reasoning Index (PRI), under which the Picture Completion subtest is located (Wechsler, 2003a).

Given these findings presented in the test manual, it is not surprising that participants with lower educational ability levels obtained significantly poorer scores than children with higher educational ability levels. Individuals diagnosed with Learning Disorders/ADHD often perform substantially lower than what is expected of them given their chronological age and education (Barlow & Durand, 2015). The learning difficulties experienced by these individuals often have multiple origins, one example being from reading disorders, though subtle brain impairments have also been thought to be a contributing factor (Barlow & Durand, 2015). Brain impairments seem to affect primarily the regions within the left hemisphere, specifically the Broca's area, the intraparietal sulcus and regions within the left occipitotemporal and left parietotemporal area (Ashkenazi, Black, Abrams, Hoeft, & Menon, 2013; Barlow & Durand, 2015; Shaywitz et al., 2006). Moreover, learning difficulties and deficits in working memory often co-exist. It has been found that children with greater learning difficulties had more severe working memory impairments compared to children with milder learning difficulties (Alloway et al., 2005). Taken together, the significant main effect found could most likely be explained by the ability and cognitive impairments that often accompany Learning Disorders and neurodevelopmental disorders such as ADHD.

The finding of non-significant interaction effects between the version of the measure and school type was interesting. It is known that visual deficits are highly prevalent among children diagnosed with Learning Disorders (Stegemann, 2016) and neurodevelopmental disorders such as ADHD (DeCarlo, Swanson, McGwin, Visscher, & Owsley, 2016). It is common that deficits in performance of individuals with Learning Disorders are primarily due to vision and reading problems (Interagency Committee on Learning Disabilities, 1987; Shaywitz, Fletcher & Shaywitz, 1994), and thus given the beneficial effects colour has on low-level and high-level vision (Tanaka et al., 2001) it would be expected that coloured test stimuli would aid performance to some degree. As previously mentioned, colour is known to help segment complex visual scenes by enabling objects to be more easily distinguished from one another and their backgrounds (Tanaka et al., 2001). Colour also helps with the recognition of objects, where sometimes colour may be the defining feature that helps one recognise a particular object (Enns, 2004; Tanaka et al., 2001).

Given the above, it was surprising that no incremental effect was found to be statistically significant. Though, it should be pointed out participants from the school who received remedial assistance did on average score lower on the greyscale version of the measure and higher on the coloured version. Also, the same participants had numerically higher time averages on the greyscale version compared to the coloured version. Thus, participants from the remedially assisted school would score higher on the coloured measure but took longer to respond, and would score lower on the greyscale measure but would respond faster. Similarly, participants from the mainstream school were also found to have numerically higher score averages for the coloured measure than for the greyscale measure. A tentative explanation for this pattern of results may be that the coloured measure had a small incremental effect on performance, and it is this small impact that may help explain why the effect was found to be non-significant. Concerning the time, the greyscale stimuli may have aided the processing speed of participants with specific learning difficulties, but failed to aid in the accuracy of how it was processed.

An alternative explanation could be that this pattern of results occurred by chance and that it is just as likely that slightly different patterns could be found should the experiment be reattempted. This explanation is largely based upon the large error bars that can be seen in Figures 4 and 5. Not only do the error bars indicate that a large amount of variability exists (Cumming, Fidler & Vaux, 2007), though this alone does not invalidate the mean represented on the interaction graphs, the error bars can be seen to overlap, more so for the interaction graph for the time dependent variable. The larger the gap between the two error bars (i.e. the less they overlap with one another), the smaller the *p*-value gets and the evidence for a true effect become stronger (Cumming et al., 2007). Given that the error bars indicate high levels of overlap, it would be expected that the analysis would have yielded non-significant results and that a large amount of variability would exist. The possible causes for this large amount of variability are reflected on later in the limitations section that follows shortly.

It is felt that the second explanation offered to explain the pattern of results is more likely than the first explanation provided which is somewhat tentative. With that being said, one should not be so quick to completely rule out the possibility that colour adaptions affect psychometric performance. Both the present study and the study conducted by Nelson et al. (1994) observed that scaled scores in coloured versions of the Picture Completion subtest was numerically, though not significantly, higher than the greyscale versions. Furthermore, it should be noted that the present study noted a larger difference (mainstream school: $\Delta\bar{x}_t =$ .78; school for LD: $\Delta\bar{x}_t = 1.28$) between the scaled scores of the coloured and greyscale Picture Completion than the difference that was observed by Nelson et al. (1994) ($\Delta\bar{x} = .2$).

This increased difference between the two measures may be due to the robust statistical procedures utilised in the present study that is more capable of handling bias within data, thus ensuring that more accurate and meaningful results were obtained. Given Nelson et al. (1994) utilised standard arithmetic means in their calculations, it is possible that inaccurate representations of the average scaled score were obtained. On the other hand, the present study utilised robust 20% trimmed means in calculating the average scaled score for the coloured and greyscale Picture Completion which ensured that statistical bias was not an influencing factor in the calculations, thereby increasing the overall confidence in knowing that an accurate measure of central tendency was obtained. Given this, it may be that the use of trimmed means as opposed to standard arithmetic means provided a more accurate representation of the average scores for each version of the measure which, in turn, highlighted a greater difference. Though, it is equally possible that this increase in the magnitude of score difference between the two measures is due to other factors such as medication status, the severity of symptoms or other comorbid disorders (Mahone et al., 2003).

This increased difference between coloured and greyscale stimuli is an important and consistent trend to note as its implications would affect diagnostic procedures in clinical settings. As previously mentioned, the WISC-IV is widely used as a diagnostic tool in clinical practices (Wechsler, 2003a), often providing useful information that health care practitioners rely on when making a diagnosis (Mahone et al., 2003). Noting how colour adaptions affect children's performance in measures is important since fluctuations in performance that is primarily due to colour changes of stimuli may unintentionally result in larger or fewer learning disorders and disabilities being detected by practitioners (Mahone et al., 2003). Therefore, though the difference between the two measures was found to be non-significant, it is felt that this consistent pattern of results that suggest an increased difference between coloured and grayscale measures warrants further investigation.

5.3 Limitations

The present study is not without its limitations. It could be argued that the this study suffered from limitations in control over extraneous variables as Husband and Hayden's (1996) study had. Even with the aim of attempting to exert greater control over sample characteristics (sample size and age of participants) to ensure no extraneous variables would impact the result, not enough attention was directed towards controlling the types of learning difficulties that should be included in the final sample. Grouping multiple learning difficulties under one category may have been a primary reason for a large amount of variability and statistical noise that was observed in participants performance. Furthermore, the overall sample size could be said to be too small to obtain an optimal power level to detect a true effect. Future studies would benefit from greater control over what types of learning difficulties are focused on by ensuring stricter inclusion criteria are established prior to sampling. Additionally, future studies should focus on increasing the overall sample size to ensure that an adequate power level is achieved, especially for a study such as this one where statistical power is a crucial aspect to the extent to which claims about the findings can be made.

A second limitation of the present study has to with the expected effect size the researcher chose. The chosen effect size of .40 is considered to be large, though, literature does suggest that the effect that may exist would be much smaller given that it was argued in the literature review that previous research attempts to study the same phenomenon failed to detect any significant effects. Although the researcher chose a large effect size for practical reasons, most notably due to the limited resources available and time constraints that would have made collecting a much larger sample size nearly impossible, future attempts should assume the effect size to be much smaller and should sample accordingly to achieve adequate statistical power to detect a real effect.

A second limitation is the extent to which these findings can be generalised to different psychometric measures in a variety of different contexts and settings. The present study utilised only a single subtest from the WISC-IV (2003a), and thus one should be cautious in making claims about different psychometric measures regarding colour adaptions. Psychometric measures and subtests are not identical, often having different instructions that testees are required to follow and often attempt to measure different psychological constructs. Given this, it is possible that colour adaptions may impact psychometric performance differently in one test compared to another.

Finally, failure to pilot the demographic questionnaire meant that items were unable to be tested and adapted to ensure that participants were able to respond to them. As previously mentioned, items within the demographic questionnaire were intended on exploring and gathering information regarding children’s exposure to coloured reading material and media. However, many of the participants, most notably the younger participants, struggled to respond to specific items which impacted the response rate to these items. This poor response rate is what lead the researcher in deciding to drop these questions in the final analysis, meaning that certain relationships involving expectations were unable to be examined. The failure to pilot the demographic questionnaire was primarily due to lack of resources and time restraints, though similar studies in the future would benefit in exploring the impact expectations may have on the results.

5.4 Implications and Contributions

The support provided by the present findings is strengthened further by the fundamental differences in sample characteristics and data analysis techniques utilised that helped distinguish the study from past research attempts. The inclusion of a racially and culturally diverse sample helped increase the generalisability of past findings to different population groups. A primary limitation of past findings was in their poor generalisability, mostly due to the failures of ensuring that a broad range of participants who varied in race, culture and location were included in their samples. For example, 83.4% of the sample collected by Husband and Hayden (1996) were White children from the United States and no Black children were included. Sampling issues such as this limited the extent to which their findings could be generalised and extended to other population groups that were very different. Thus, the present study contributed to improving on the extent to which previous findings can be generalised to different population groups and settings. Though the generalisability of findings were improved upon, a greater number of future studies involving samples with different characteristics from various socioeconomic backgrounds would be required to increase the generalisability further.

The utilisation of robust statistical procedures provided a renewed and alternative approach to identifying whether colour adaptions to measures have an incremental or decremental effect in psychometric performance. Finding similar results to past research by investigating the area of interest from a differing angle and method provides greater support for the conclusions that are drawn from the results with the possibility that greater insight into an issue can be gained. Knowledge production is not accomplished through a single approach or worldview (Goduka, 2012). Instead, multiple possible methods to inquiry exist, and only by attempting to approach topics from differing angles are we able to produce new knowledge that can add more meaningful insights and depth to our current understanding of specific issues (Goduka, 2012).

It is believed the statistical analyses utilised in the study contributed to demonstrating the appropriateness and effectiveness of applying robust statistical procedures to psychological data collected in South Africa. Researchers within the psychological domain in South Africa, for the most part, still heavily rely upon standard parametric statistical methods, which are primarily taught in statistical courses in universities across the country. Often, little importance is given to reflecting upon methodological issues that may severely impact the usefulness of such methods, and more crucially towards alternative ways of overcoming methodological issues along with their implications. Thus, the present paper hopes to highlight to researchers within South Africa the importance of not solely relying on traditional analyses in their investigations. Instead, that further reflection is given to adopting robust methods that are arguably superior (see Erceg-Hurn et al., 2013), often enabling the researcher to: (a) detect more genuine relationships between variables, while increasing statistical power and decreasing the chances of making a type II error; (b) to report on results that are less spurious; (c) to better estimate more relevant effect sizes and other parameters; and (d) to compute more accurate parameter confidence intervals with greater probability coverage (Erceg-Hurn et al., 2013). It is felt that an increase in the utilisation of robust techniques would benefit psychological research within South Africa, especially given the high degrees of cultural variability within South African populations which no doubt affects the normality and variability of the data that is often collected. Relying more on techniques that are better equipped at handling data that violates statistical assumptions would be highly beneficial in offering ways to overcome bias during data analysis, thus ensuring that more genuine and meaningful relationships are uncovered that provide more insightful explanations regarding specific issues.

5.5 Recommendations for Future Research

Future studies would benefit from incorporating higher degrees of control over population factors when sampling participants and by adjusting the angle from which they approach the issue of colour adaptions to psychometric measures and performance. By outlining more specific selection criteria before sampling, researchers may sample a more homogenous group of participants that could result in data that has less variability and statistical noise, thus ensuring that data analyses may more accurately detect any relationships that may exist.

Regarding the different approaches available to researchers interested in investigating this area in the future, it would be both beneficial and informative for future studies to focus on examining whether subtle differences in performance exists as a result of colour adaptions. As noted from the findings, a small difference between the measures was discovered which may require further investigation to provide insight into whether this small difference is significant and could be considered to be an influencing factor in psychometric performance. A possible approach to explore this subtle difference is to investigate the effect colour may have on visual search, specifically whether colour changes the way test stimuli are actively scanned and searched. Incorporating eye tracking software would assist in investigating this effect, but may also allow for the possibility of measuring the effect colour has on cognitive load by measuring individuals' pupil sizes. Kahneman and Beatty (1996) proposed that pupil diameter is an effective index is providing insight into an individual's cognitive load as they perform a specific task. Though careful procedures would have to be followed since multiple confounding factors may impact the measurement of a pupil's diameter, such as ambient light, the use of eye tracking software and technology may prove to be useful as a non-invasive technique to measure cognitive load in tasks that differed in colour (Krejtz, Duchowski, Niedzielska, Biele & Krejtz, 2018).

Future studies may also benefit in reconsidering the units of measurement used when tracking performance. The current study, for example, used seconds as a unit of measurement for the average time taken to provide a correct response. Though, it may be that seconds may be too large of a measurement to detect whether a difference in time exists, especially if this difference in time is particularly small.

Chapter Six: Conclusion

When item and test adaptions are made to psychometric measures, the implications those adaptions may have on performance needs to be considered. One such test adaption that has not been explored in enough detail was to do with colour adaptions of test stimuli. The present study mainly focused on examining the impact that colour adaptions to the Picture Completion subtest had on individual's performance in it and whether this impact was different for individuals who differed in educational ability level. This issue was explored through a factorial design whereby a greyscale and coloured version of the Picture Completion were given to different groups of participants. Robust statistical analyses indicated that colour adaptions did not have a significant incremental or decremental effect on psychometric performance, though, an expected main effect of the type of school the participant attended was found to be significant. These findings in conjunction with past studies provide support that colour adaptions made to test stimuli are not an important influencing factor when it comes to performance in psychometric measures. However, with that being said it was pointed out that a consistent pattern of results was observed that might suggest that though colour adaptions do not yield significant differences in performance, subtle differences in performance between coloured and greyscale measures may exist. Coloured measures have consistently been observed to be numerically higher than greyscale measures, and this result warrants that future investigation is required.

References

Alloway, T. P., Gathercole, S. E., , Willis, C., & Adams, A. M. (2005). Working memory abilities in children with special educational needs. *Educational and Child Psychology*, *22*(4), 56-57.

Ashkenazi, S., Black, J. M., Abrams, D. A., Hoeft, F., & Menon, V. (2013). Neurobiological underpinnings of math and reading learning disabilities. *Jounral of Learning Disabilities.* doi: 10.1177/0022219413483174

Aylward, G. P., & Stancin, T. (2008). Assessment of Development and Behavior. In Wolraich, M. (Ed.) *Developmental-Behavioral Pediatrics: Evidence and Practice* (pp. 154 – 155), Philadelphia: Mosby/Elsevier.

Baldassi, S., Megna, N., & Burr, D. C. (2006). Visual Clutter Causes High-Magnitude Errors. *PLoS Biology*, *4*(3), 387-394. doi: 10.1371/journal.pbio.0040056.

Barlow, D. H., & Durand, V. M. (2015). *Abnormal Psychology: An Intergrative Approach* (7th Edition). Canada: Cengage Learning.

Barrett, P., Davies, F., Zhang, Y., & Barrett, L. (2015). The impact of classroom design on pupils' learning: Final results ofaholistic, multi-level analysis. *Building and Environment*, *89*, 118–133. doi:10.1016/j.buildenv.2015.02.013.

Beilock, S. L. & Carr, T. H. (2003). When High-Powered People Fail: Working memory and "Choking Under Pressure" in Math. *American Psychological Society, 16*(2), 101-105. doi: 10.1111/j.0956-7976.2005.00789.x.

Berggren, N., & Eimer, M. (2016). The Guidance of Spatial Attention During Visual Search for Color Combinations and Color Configurations. *Journal of Experimental Psychology*, *42*(9), 1282-1296. doi: 10.1037/xhp0000225.

Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *TRENDS in Neurosciences, 26*(9), 507-513.

Canivez, G. L., & Watkins, M. W. (2016). Review of the Wechsler Intelligence Scale for Children-Fifth Edition: Critique, commentary, and independent analyses. In A. S. Kaufman, S. E. Raiford, & D. L. Coalson (Eds.), *Intelligent testing with the WISC–V* (pp. 683–702). Hoboken, NJ: Wiley.

Chong, E., Familiar, A., & Shim, W. M. (2015). Reconstructing Representations of Dynamic Visual Objects in Early Visual Cortex. *PNAS, 113*(5), 1453-1548. doi: 10.1073/pnas.1512144113.

Cumming, G., Fidler, F., Vaux, D. L. (2007). Error bars in experimental biology. *The Journal of Cell Biology*, *177* (1), 7–11. doi:10.1083/jcb.200611141.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life.* London: John Murray.

DeCarlo, D. K., Swanson, M., McGwin, G., Visscher, K., & Owsley, C. (2016). ADHD and Vision Problems in the National Survey of Children's Health. *Optometry and vision science : official publication of the American Academy of Optometry, 93*(5), 459–465. doi:10.1097/OPX.0000000000000823

Dvorine, I. (1953). *Dvorine Pseudo- Isochromatic Plates (2nd ed.)*. Baltimore: Waverly Press.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance, 10*(1), 32-39. doi: 10.1037/0096-1523.10.1.32.

Enns, J. T. (2004). *The thinking eye, the seeing brain: Explorations in visual cognition*. New York: Norton.

Erceg-Hurn, D. M., Wilcox, R. R., & Keselman, H. J. (2013). Robust Statistical Estimation. In T. D. Little (Ed.), *The Oxford Handbook of Quantitative Methods* (Vol. 1, pp. 388-406). Oxford University Press.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. doi: 10.3758/BF03193146.

Field, A., Miles, J., & Field, Z. (2012). *Discovering Statistics Using R*. London: SAGE.

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment*. Hoboken, NJ: John Wiley & Sons.

Flanagan, D. P., Alfonso, V. C., Mascolo, J. T., & Hale, J. B. (2010). *The Wechsler Intelligence Scale for Children–Fourth Edition in Neuropsychological Practice*.

Goduka, N. (2012). From Positivism to Indigenous Science: A Reflection on World Views, Paradigms and Philosophical Assumptions. *Africa Insight*, *41*(4), 123-138.

Goldstein, E. B. (2008). *Cognitive Psychology: connecting mind, research and everyday experience*. Australia: Wadsworth Cengage Learning.

Goldstein, E. B. (2014). *Sensation and Perception*. Austrailia: Wadsworth Cengage Learning.

Green, H. (1992). A comparison study of the WISC-III and WISC-R with a special education population. *Student Work*. Paper 143.

Grubert, A., & Eimer, M. (2015). Rapid parallel attentional target selection in single-color and multiple-color visual search. Journal of Experimental Psychology. *Human Perception and Performance, 41*(1), 86–101. doi: 10.1037/xhp0000019.

Handford, M. (1987). *Where's Wally?* United Kingdom: Walker Books.

Herzog, M. H., Thunell, E., & Ogmen, H. (2016). Putting low-level vision into global context: Why vision cannot be reduced to basic circuits. *Vision Research, 126*, 9-18. doi: 10.1016/j.visres.2015.09.009.

Howell, D. C. (2013). *Statistical methods for psychology (8th ed*.). Belmont, CA: Wadsworth Cengage Learning.

Husband, T. H., & Hayden, D. C. (1996). Effects of the addition of color to assessment instruments. *Journal of Psychoeducational Assessment, 14*, 147-151.

Interagency Committee on Learning Disabilities. (1987). *Learning disabilities: a report to the U.S. Congress*. Washington, DC: Government Printing Office.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, N.J.: Prentice-Hall.

Kahneman, D., & Beatty, J. (1996). Pupil Diameter and Load on Memory. *Science, 154*(3756), 1583-1585. doi: 10.1126/science.154.3756.1583.

Kamza, A., Molińska, M., Skrzypska, N., & Długiewicz, P. (2019). Can sustained attention adapt to prior cognitive effort? An evidence from experimental study. *Acta Psychologica, 192*, 181-193. doi: 10.1016/j.actpsy.2018.11.007.

Kaufman, A. S., Flanagan, D. P., Alfonso, V. C., & Mascolo, J. T. (2006). Test review: Wechsler intelligence scale for children, (WISC-IV). *Journal of Psychoeducational Assessment, 24*(3), 278-295. doi: 10.1177/0734282906288389.

Kaufman, A., Raiford, S. and Coalson, D. (2015*). Intelligent testing with the WISC-V*. John Wiley & Sons, p.653.

Kellogg, R. T. (1997). *Cognitive Psychology*. London: SAGE.

Kolb, B., & Whishaw, I. Q. (2015). *Fundamentals of human neuropsychology*. New York: Macmillan.

Koyanagi, M., Nagata, T., Katoh, K., Yamashita, S., & Tokunaga, F. (2008). Molecular Evolution of Arthropod Color Vision Deduced from Multiple Opsin Genes of Jumping Spiders. *Journal of Molecular Evolution, 66*(2), 130-137. doi: 10.1007/s00239-008-9065-9.

Krejtz, K., Duchowski, A. T., Niedzielska, A., Biele, C., & Krejtz, I. (2018). Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. *PLOS ONE 13*(9), e0203629. doi: 10.1371/journal.pone.0203629.

Küller, R., Mikellides, B., & Janssens, J. (2009). Color, arousal, and performance - A comparison of three experiments. *Color Research and Application*, *34*(2), 141–152. doi: 10.1002/col.20476.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*, 1-12. doi: 10.3389/fpsyg.2013.00863.

Landsheer, J. A., & van den Wittenboer, G. (2015) Unbalanced 2 x 2 Factorial Designs and the Interaction Effect: A Troublesome Combination. *PLOS ONE*, *10*(3), 1-18.doi: 10.1371/journal.pone.0121412.

Lebrecht, S., Bar, M., Barrett, L. F., & Tarr, M. J. (2012). Micro-valences: Perceiving affective valence in everyday objects. *Frontiers in Psychology, 3*, 1–5. doi: 10.3389/fpsyg.2012.00107.

Lezak, M. D., Hannay, H. J., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment (4th ed.).* Oxford: University Press.

Mahone, E. M., Miller, T. L., Koth, C. W., Mostofsky, S. H., Goldberg, M. C., & Denckla, M. B. (2003). Differences between WISC-R and WISC-III Performance Scale among Children with Adhd. *Psychology in the Schools, 40*(4), 331-340. doi: 10.1002/pits.10094.

Mair, P., & Wilcox, R. (2018). *Package 'WRS2'*. Retrieved from: https://cran.r-project.org/web/packages/WRS2/WRS2.pdf

Massar, S. A. A., Lim, J., Sasmita, K., & Chee, M. W. L. (2018). Sleep deprivation increases the costs of attentional effort: Performance, preference and pupil size. *Neuropsychologia, 123*, 169-177. doi: 10.1016/j.neuropsychologia.2018.03.032.

Mehta, R., & Zhu, R. J. (2009). Blue or Red Exploring the Effect of Color on Cognitive Performances. *Science, 323*(5918), 1226–1229. doi: 10.1126/science.1169144.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*, 156-166. doi: 10.1037/0033-2909.105.1.156.

Nelson, K. M., Arthur, P., Lautigar, J., & Smith, D. K. (1994). *Does the Use of Color on the WISC-III Affect Student Performance?* Paper presented at the annual meeting of the National Association of School Psychologists, Seattle, WA.

Pinker, S. (1997). *How the mind works*. London: Penguin Books.

R Core Team. (2018). *A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.r-project.org/

Radonjić, A., & Brainard, D. H. (2016). The Nature of Instructional Effects in Color Constancy. *Journal of Experimental Psychology, 42*(6), 847-865. doi: 10.1037/xhp0000184.

Regan, B. C., Julliot, C., Simmen, B., Viénot, F., Charles-Dominique, P., & Mollon, J. D. (2001). Fruits, foliage and the evolution of primate colour vision. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 356*, 229-83. doi: 10.1098/rstb.2000.0773.

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educational Research Review, 6*(2), 135-157. Doi: 10.1016/j.edurev.2010.12.001.

RStudio Team. (2015). *RStudio: Integrated Development for R*. RSudio, Inc. Boston, MA. Retrieved from http://www.rstudio.com/

Ryan, J. J., Glass, L. A., & Brown, C. N. (2007). Administration time estimates for Wechsler Intelligence Scale for Children-IV subtests, composites, and short forms. *Journal of Clinical Psychology, 63*(4), 309–318. doi: 10.1002/jclp.20343.

Sarason, S. B., & Potter, E. H. (1947). Color in the Rorschach and Kohs block designs. *Journal of Consulting Psychology, 11*(4), 202–206. doi: 10.1037/h0056816.

Sarter, M., Gehring, W. J., & Kozak, R. (2006). More attention must be paid: the neurobiology of attentional effort. *Brain Research Reviews, 51*(2), 145-160. doi: 10.1016/j.brainresrev.2005.11.002.

Sattler, J. M., & Dumont, R. (2004). *Assessment of children: WISC-IV and WPPSI-III Supplement*. San Diego: Sattler.

Scime, M., & Norvilitis, J. M. (2006). Task Performance and Response to Frustration in Children with Attention Defecit Hyperactivity Disorder. *Psychology in the School, 43*(3), 377-386. doi: 10.1002/pits.20151.

Shaywitz S. E., Fletcher J. M., & Shaywitz B. A. (1994). Issues in the definition and classification of attention deficit disorder. *Topics in Language Disorders, 14*(4), 1-25. doi: 10.1097/00011363-199408000-00003.

Shaywitz, S. E., Mody, M., & Shaywitz, B. A. (2006). Neural mechanisms in dyslexia. *Current Directions in Psychological Science*, *15*, 278-281.

Shuttleworth-Edwards, A. B., van der Merwe, A. S., van Tonder, P., & Radloff, S. E. (2013). WISC-IV test performance in the South African context: a collation of cross-cultural norms. In Laher, S., & Cockcroft, K. (Eds.) *Psychological assessment in South Africa: research and applications* (pp. 33 – 45). Johannesburg, South Africa: Wits University Press.

Slawomir, J. (2012). *mvnormtest: Normality test for multivariate variables* (version 0.1-9). Retrieved from https://CRAN.R-project.org/package=mvnormtest

Spaull, N. (2013). *South Africa's Education Crisis: The quality of education in South Africa 1994-2011*. Retrieved from: https://pdfs.semanticscholar.org/8284/240c243cbd0172ee1451486af141f3fd2e09.pdf

Śpiewak, S., Ziaja, J., & Doliński, D. (2003). Wpływ przeciążenia poznawczego nadostępnośćzasobów: efekt rozgrzania poznawczego [The impact of cognitive load onthe availability of cognitive resources: The warm-up effect]. *Przegląd Psychologiczny,46*(3), 291–306.

Stegemann, K. C. (2016). Learning disabilities in Canada. *Learning Disabilities: A Contemporary Journal, 14*(1), 53-62.

Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences, 5*(5), pp. 211-215. doi: 10.1016/S1364-6613(00)01626-0.

Thompson, E. (1994). *Colour vision: A study in cognitive science and philosophy of science*. London: Routledge.

Tomporowski, P. & Tinsley, F. V. (1996). Effects of Memory Demand and Motivation on Sustained Attention in Young and Older Adults. *The American journal of psychology, 109*(2), 187-204. doi: 10.2307/1423272.

Treisman, A., & Gelade, G. (1980). A feature intergreation theory of attention. *Cognitive Psychology, 12*(1), 97-136. doi:10.1016/0010-0285(80)90005-5.

Treisman, A. M., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14*(1), 107-141. doi: 10.1016/0010-0285(82)90006-8.

Ueda, Y., Chen, L., Kopecky, J., Cramer, E. S., Rensink, R. A., Meyer, D. E., … Saiki, J. (2018). Cultural Differences in Visual Search for Geometric Figures. *Cognitive Science, 42*(1), 286–310. doi: 10.1111/cogs.12490.

Van Rossum, G. (1995). *Python tutorial. Technical Report CS-R9526*. Amsterdam: Centrum voor Wiskunde en Informatica (CWI).

Wechsler, D. (1939). *Wechsler-Bellevue Intelligence Scale*. New York: Psychological Corporation.

Wechsler, D. (1946). *Wechsler-Bellevue intelligence Scale, Form II*. San Antonio, TX, US: Psychological Corporation.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. New York: Psychological Corporation

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children - Revised*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2003a). *Wechsler Intelligence Scale for Children-Fourth Edition*: Technical and interpretive manual. San Antonio: The Psychological Corporation.

Wechsler, D. (2003b). *Wechsler Intelligence Scale for Children-Fourth Edition: Administration and scoring manual*. San Antonio: The Psychological Corporation.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children - fifth edition*. Bloomington, MN: Pearson.

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer.

Wilcox, R. (2012). *Modern Statistics for the Social and Behavioural Sciences: A Practical Introduction*. Elsevier.

Wilcox, R. (2017). *Introduction to Robust Estimation and Hypothesis Testing (4th ed.)*. Elsevier.

Wilcox, R., & Schönbrodt, F. D. (2018). *The WRS package for robust statistics in R (version 0.35)*. Retrieved from Retrieved from https://github.com/nicebread/WRS

Winderickx, J., Lindsay, D. T., Sanocki, E., Teller, D. Y., Motulsky, B. G., & Deeb, S. S. (1992). Polymorphism in red photopigment underlies variation in color matching. *Nature, 356*(6368), 431-433. doi: 10.1038/356431a0.

Wolfe, J. M. (2014). Approaches to Visual Search: Feature Integration Theory and Guided Search. In Nobre, A. C., & Kastner, S. (Eds.). *The Oxford handbook of attention* (pp. 11–55). Oxford (UK): Oxford University Press.

Zusne, L. (1970). *Visual Perception of Form*. New York: Academic Press.

Footnotes

[1] Supplementary material for this manuscript has been made available:

https://osf.io/mcuqy/?view_only=7adad91d1e304dc89e20f77133447d6f

Appendices

Appendix A: Parental Consent Form

Parental Permission for Participation of a Child

University of Witwatersrand

Dear Parents,

Title of Research Project	Colour and its Impact on the Picture Completion Subtest in the WISC-IV

The study has been described to me in language that I understand and I freely and voluntarily agree to allow my child, ___________________ ***(name)***, to participate. My questions about the study have been answered, either via the provided information sheet or via direct contact with the researchers. I understand that my child's identity will not be made known and that I may withdraw my child from the study without giving a reason at any time and this will not negatively affect my child in any way. **I have spoken to my child about the study and they have verbally indicated to me that they would like to participate in the study**. (Your child will also be asked if they want to participate in this study prior to test administration by the researcher, Brett Gregory, in case your child does change their mind).

Parent Name (optional)	
Parent Signature	
Date	

Should you have any questions regarding this study or wish to report any problems you have experienced related to the study, please contact the researcher or research supervisor:

Researcher's Name	Brett Gregory
Cell phone number	0749029962
Email	792376@students.wits.ac.za
Research Supervisor's Name	Enid Schutte
Email	Enid.Schutte@wits.ac.za

Appendix B: Information Sheet

Information Sheet

Dear Sir/Madam,

My name is Brett Gregory and I am currently enrolled at the University of the Witwatersrand in Johannesburg where I am studying a Masters in Social and Psychological Research. Part of my studies requires that I undertake a research project, and I have decided to investigate **Colour and its Impact on the Picture Completion Subtest in the WISC-IV**. The aim of my study is to see if children's performance in a test (the Picture Completion test) changes depending on whether that test is printed in colour or black and white.

This letter serves as an invitation inviting your child to take part in my research project. <u>Please ensure that you do discuss this study with your child and allow them the opportunity to verbally indicate that they would like to participate in the study.</u> The research project will involve three measures which your child would complete. The first is a screening test for colour blindness which would require your child to look at cards with coloured dots on them and asked to specify which number they are being shown. The second measure is a demographic questionnaire that asks general questions about themselves (age, grade, gender, race, etc.) and about their exposure to reading and television. There is also a smaller demographic questionnaire that should be filled out by you which should be given to your child to return to me along with the signed consent form. Finally, the third measure, the Picture Completion test, involves your child looking at pictures of objects or scenes and specifying any details that are missing from the picture (E.g. a door handle that is missing from a picture of a door). The entire study will take around 10 to 15 minutes to complete.

Both you and your child will not receive any direct benefits from participating in this study, and there are no disadvantages or penalties for not participating. Your child may choose to withdraw from the study at any time and does not have to answer any questions that they do not want to. Your child’s personal details and data from the tests will be kept confidential and will not be shared with anyone else. All sheets of paper containing personal data will be stored in a safe file that will be locked away when not being used, and all data that is transferred to my personal computer will be password protected. Should your child for any reason experience any distress or discomfort, the test/s will be stopped. Furthermore, procedures will be in place that stops the tests should your child consequentially fail to respond correctly to prevent them from feeling frustrated. Finally, if we do suspect that your child has a potential perceptual difficulty, such as colour blindness, we will bring this to your attention along with contact details of a nearby optometrist so that you can take your child for a professional check-up.

If you have any further questions about my research project, please feel free to contact either myself or my supervisor on the details listed below. The study will be written up as a research report that will be submitted as an assessment that will be graded since the completion of my degree requires it. If you wish to receive a summary of this report, I will be happy to send it to you upon request (optional). The final research report will also be put forward for publication in online academic journals. If you have any queries, concerns or complaints regarding the ethical procedures of this study, you are welcome to contact the University Human Research Ethics Committee (non-medical), telephone + 27(0)11 717 1408, email Shaun.Schoeman@wits.co.za.

Your sincerely,

Brett Gregory

Contact Details

Brett Gregory, 792376@students.wits.ac.za, 074 902 9962.

Enid Schutte, Enid.Schutte@wits.ac.za, 082 920 6731.

Appendix C: Assent Form

Children Assent

TO: ______________________________________

[Full Name of Volunteer]

My name is Brett Gregory. I am a student at the University of Witwatersrand and I am inviting you to help with my study. You can decide if you would want to help with my study or not. Before saying yes, make sure to talk to your parents about it first and listen to what the study is about. You then can choose if you want to be in it.

Why is this study being done?

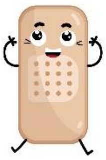
I am looking to see if a child's mark in a test would change or stay the same depending on if the test was printed in colour or is printed in black and white.

What will happen to me?

You will go through some questions with me and then be shown some pictures. Your job will be to tell me what is missing from every picture.

Let's practice to show you what I mean! What is missing from the cat below?

The ear is missing! Poor kitty!



Will it be safe?

Yes, it will. You will not be hurt in anyway.

What are the good things about being in the study?

You would be helping us learn more about children and colour.

How much time will it take?

You would only be with me for about 10-15 minutes.

Do I have other choices?

Of course! You can decide if you want to be in the study or not. You are also allowed to decide to pull-out of the study whenever you feel like it.

Will people know that I helping you?

I won't tell anyone. But, you parents and possibly your teacher would know you are helping me.

Who should I ask if I have questions?

You can ask me. You can also ask your parents or teacher who would tell me what your question is.

Is it ok if I say "No, I don't want to be in the study?"

Yes. You are not forced to be in the study and no one will be upset with you if you do not want to be in the study.

Okay, so:

Do you understand what you would be helping me in?	Yes / No
Do you have any questions for me?	Yes / No
Do you know that you don't have to say yes to help me?	Yes / No
Do you want to help me in my study?	Yes / No

Signature of Researcher: ______________________

Date: ______________________

Appendix D: Children Demographic Questionnaire

Basic Information:

Gender	Male	Female

Race	Black	White	Coloured	Indian	Asian	Other

Age	

Grade	1	2	3	4	5	6	7

Do you wear glasses?	Yes	No
Have you suffered any sort of head injury in your life?	Yes	No
Are you on any medication?	Yes	No

Questions:

Questions about reading				
Is reading one of your hobbies?			Yes	No
Do you enjoy reading?			Yes	No
Do your friends read a lot?			Yes	No
Do you own a lot of books at home?			Yes	No
How often do you read?				
Almost Never	Rarely	Sometimes	Often	Almost Always
How many hours a week do you think you spend time reading?				

Do you read any of the following?	**Tick if yes**
Magazines	
Books	
Poems	
Comics	

Cereal Boxes	

Do you read books with pictures in them?	Yes	No
If yes, are the pictures in colour?	Yes	No

Name some books you have read/enjoy:
Name some magazines you read/enjoy:

Do you own a TV?	Yes	No		
Do you watch TV?	Yes	No		
How often do you watch TV?				
Almost Never	Rarely	Sometimes	Often	Almost Always
How many hours a week do you think spend time watching TV?				

Name some TV shows you enjoy watching:
Name some of your favourite movies:

Appendix E: Letter for School Permission

Dear X,

I am writing to request permission to conduct a research study at your school, [Enter School Name]. I am currently enrolled in a psychology research masters at the University of Witwatersrand, Johannesburg, and am in the process of writing my Master's thesis (supervised by Enid Schutte). The study is titled: Colour and its Impact on the Picture Completion Subtest in the WISC-IV.

I hope that the school administration will allow me to recruit roughly 56 children (28 boys and 28 girls) from the school to anonymously take part in the study. Interested students, who volunteer to participate, will be given a consent form to be signed by their parent or guardian (consent form attached) and returned to the primary researcher at the beginning of the testing process. A short demographic questionnaire will also be sent home to the parents which they would need to complete should they choose to allow their child to participate in the study. Children will also be given an assent form (attached) which will indicate their own willingness to take part in the study. Information sheets will be provided to the parents to inform them of the study and what it entails (information sheet attached).

If approval is granted, student participants will complete a demographic questionnaire and the Picture Completion test in a classroom or other quiet setting on the school site. Testing can take place anytime during the school day, however, we will discuss this matter further with the teachers to ensure that disruption to regular class and school activities are kept to a minimum. The tests will be individually administered by myself. The testing process should take no longer than 15 minutes per child. The results will be pooled for the thesis project and individual results of this study will remain absolutely confidential and anonymous. Should this study be published, only pooled results will be documented. No costs will be incurred by either your school or the individual participants. More information regarding the study can be found in an information sheet attached to this email.

Your approval to conduct this study will be greatly appreciated. I will follow up with a call next week and would be happy to answer any questions or concerns that you may have at that time (a meeting may be scheduled if preferred). You may contact me at my email address: 792376@students.wits.ac.za. In addition, my supervisor, Enid Schutte, may also be contacted should you have any questions for her: Enid.Schutte@wits.ac.za.

If you agree, kindly submit a signed letter of permission on your institution's letterhead acknowledging your consent and permission for me to conduct this study at your institution.

Sincerely,

Brett Gregory

Appendix F: Parent Demographic Questionnaire

Please answer the questions below and attach it to the consent form so that your child may return it to the researcher prior to test administration.

Basic Information:

Please tick either yes or no regarding the following questions:		
Has your child ever been diagnosed with a *psychological disorder*?	Yes	No
Does your child have any *medical concerns* we should be made aware of?	Yes	No
Is your child on any *medication*?	Yes	No
Does your child wear *glasses*?	Yes	No
Has your child ever suffered from a *traumatic brain injury (TBI)*?	Yes	No

If you answered yes to any of the above questions please elaborate:

Appendix G: Ethical Approval

Research Office

HUMAN RESEARCH ETHICS COMMITTEE (NON-MEDICAL)
R14/49 Gregory

CLEARANCE CERTIFICATE	**PROTOCOL NUMBER: H18/07/06**
PROJECT TITLE	Colour and its impact on the picture completion subset in the WISC-IV
INVESTIGATOR(S)	Mr B Gregory
SCHOOL/DEPARTMENT	Social Sciences/
DATE CONSIDERED	20 July 2018
DECISION OF THE COMMITTEE	Approved
EXPIRY DATE	11 September 2021

DATE 12 September 2018 **CHAIRPERSON** ______________________
(Professor J Knight)

cc: Supervisor : Professor E Schutte

DECLARATION OF INVESTIGATOR(S)

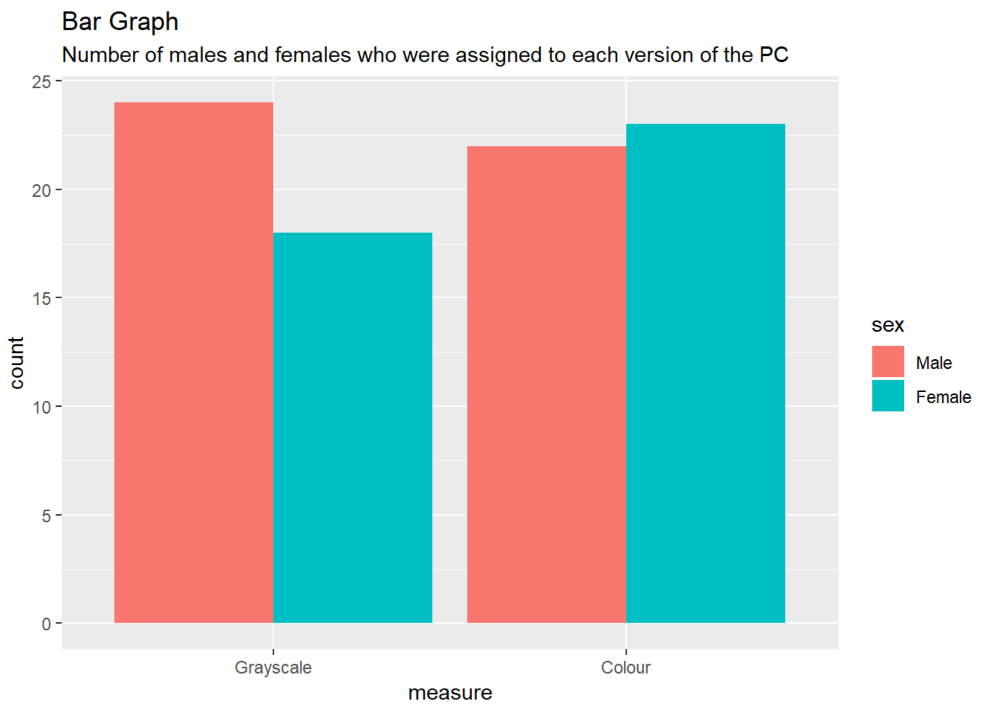
To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10004, 10th Floor, Senate House, University. Unreported changes to the application may invalidate the clearance given by the HREC (Non-Medical)

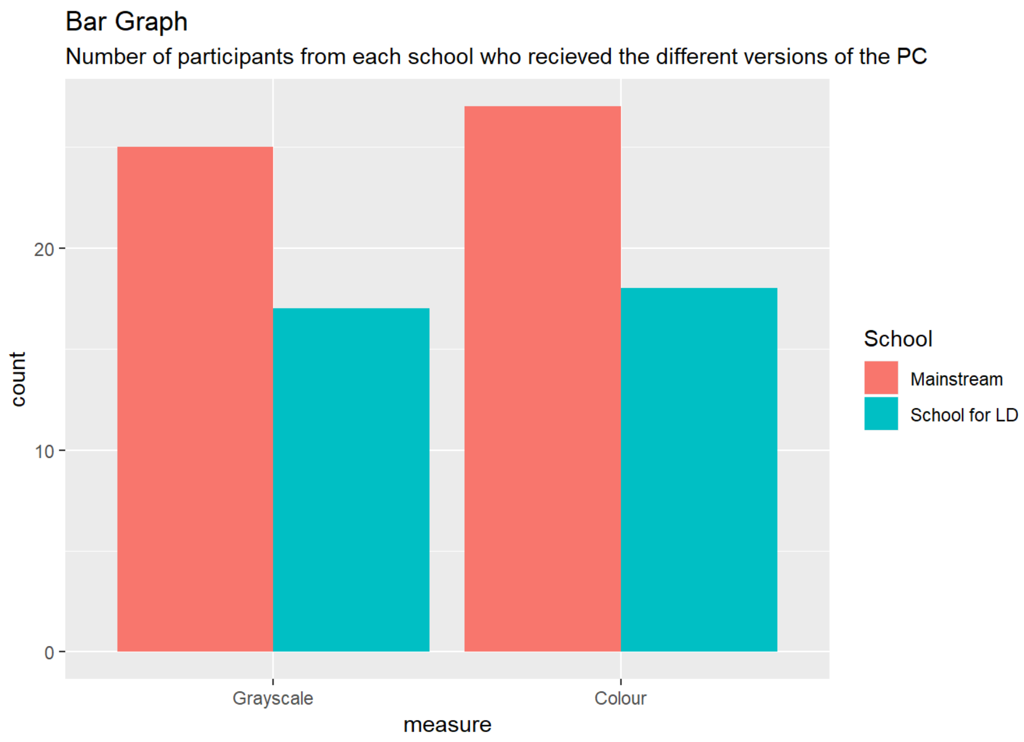
I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to completion of a yearly progress report.**

______________________ ____/____/____
Signature Date

PLEASE QUOTE THE PROTOCOL NUMBER ON ALL ENQUIRIES

Appendix H: Bar Graphs





Appendix I: Boxplots

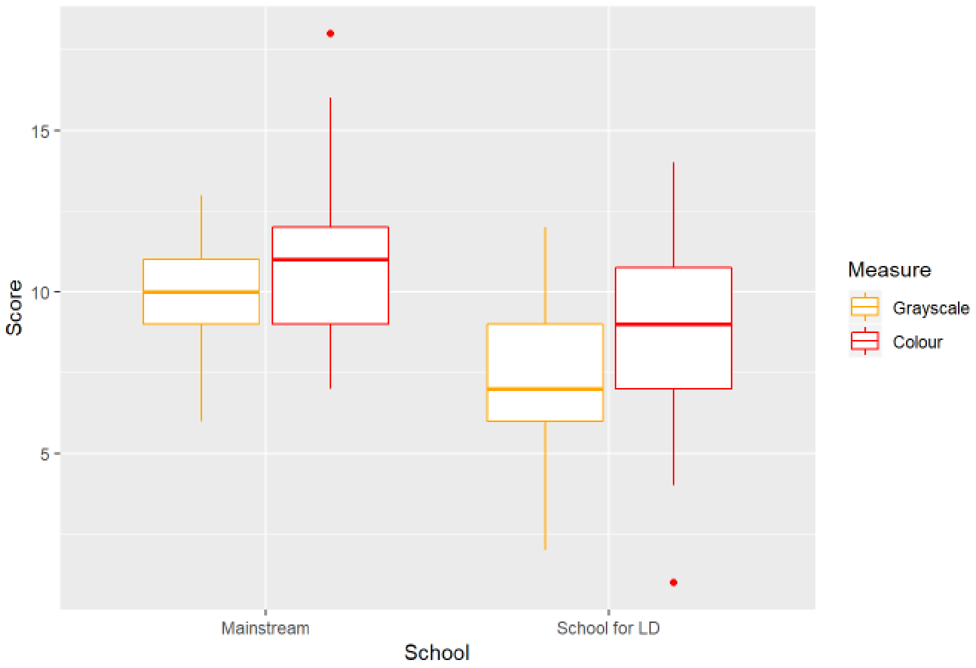


Figure 6. Boxplots indicating outliers within each group for the variable Score.

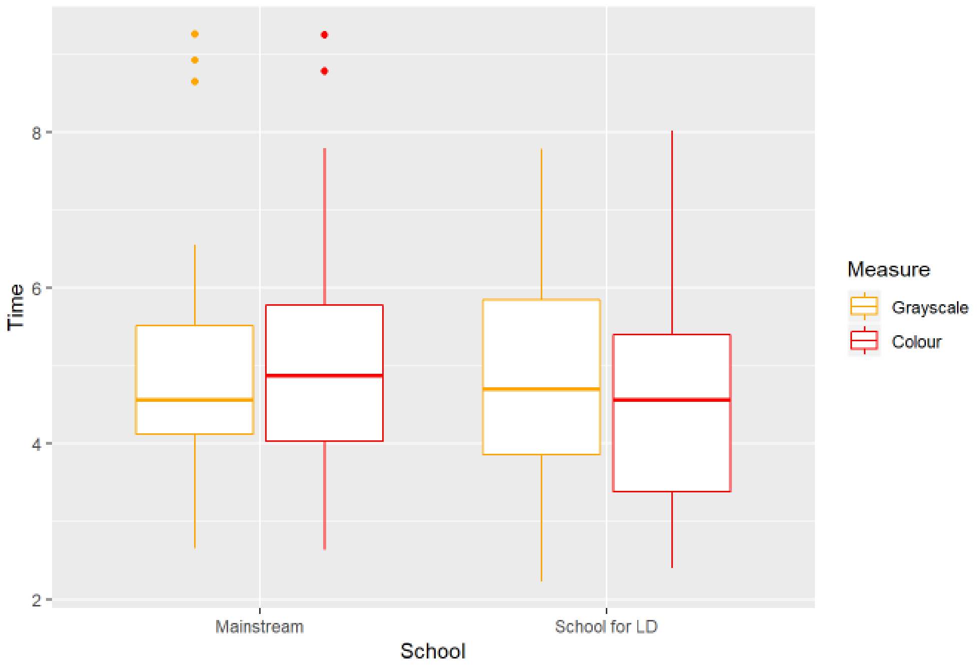


Figure 7. Boxplots indicating outliers within each group for the variable Time