

Master of Science

Neurodiversity and Typeface: Do Specialized Typefaces Improve Reading Fluency and Accuracy in Dyslexic Individuals - A Systematic Review

Written by
Rumaisa Khusru

Supervised by
Dr. Elizabeth Wonnacott

Dissertation submitted in part-fulfilment of the requirements for the degree of Master of Science in Applied Linguistics and Second Language Acquisition

Trinity, 2023

Neurodiversity and Typeface: Do Specialized Typefaces Improve Reading Fluency and Accuracy in Dyslexic Individuals - A Systematic Review

Rumaisa Khusru
Department of Education, University of Oxford
St. Anne's College
Supervisor: Dr. Elizabeth Wonnacott

Dissertation submitted in part-fulfilment of the requirements for the degree of Master of Science in Applied Linguistics and Second Language Acquisition



August, 2023

Abstract

This systematic review aims to evaluate the extent of evidence on the effectiveness of specialized typefaces created for neurodiverse (dyslexic) users. Dyslexia is an extremely prevalent neurodevelopmental disorder, and it affects 1 in every 5 children. Living in an increasingly literate world where every field requires an advanced level of reading and writing, dyslexic individuals are at a serious disadvantage. Although there is no known cure for dyslexia, many tools have been designed to mitigate the struggles faced by dyslexics, specialized typefaces being one such tool.

There are several dyslexia-specialized typefaces available, and it is important to assess which one(s) are effective by evaluating the existing literature. In addition, the employment of specialized typefaces has become increasingly popular in the last decade and continues to gain popularity leading to a proliferation in the literature and research surrounding the effectiveness of these typefaces. As such, there is a critical need for a systematic evaluation of the outcomes of dyslexia-specialized typefaces.

A total of 12 studies met the criteria for this review and were included in the final list with the majority having participants as children and using the Dyslexie typeface for their intervention. OpenDyslexic, EasyReading and Lexie Readable were some of the other typefaces included in the studies. A risk of bias assessment revealed that most studies possess a weak-moderate strength of evidence as a whole. The findings from the body of included studies showed that the effect sizes were small and that specialized typefaces when compared to non-specialized typefaces did not significantly improve reading rate or accuracy.

Although the empirical research does not support claims that these specialized typefaces are beneficial, most studies had many limitations and small sample sizes that hindered calculating effect sizes. Multiple user reviews, anecdotes and qualitative responses suggest that specialized typefaces may have potential benefits for some dyslexic individuals to enhance readability and reduced visual strain.

Acknowledgements

There are many individuals without whom, the completion of this Master's degree would not have been possible.

Thank you to my beloved parents and siblings, who helped me during bouts of stress and uncertainty despite the physical distance separating us across continents. Your unwavering support, enthusiasm and sacrifices have been the cornerstone of my academic journey.

To the professors at the Department of Education and the University of Oxford whose dedication and commitment to academia have profoundly influenced my intellectual growth and have been instrumental in shaping my academic pursuits. I owe a special debt of gratitude to my supervisor, Professor Elizabeth Wonnacott who encouraged my interest in the field of neurolinguistics, neurodiversity and technological assistance for language learning among learners with developmental language disorders. I am grateful for your scholarly guidance and mentorship. Thank you also to Professor Hamish Chalmers for all his invaluable assistance and profound insights into the intricate realm of systematic reviews.

Thank you to the librarians at St. Anne's College and the Department of Education for helping with the arduous initial stage of this dissertation.

Finally, I would like to express my deepest appreciation to all the friends and colleagues I made during my time at Oxford who, in various ways, offered their assistance, encouragement, and understanding during this unique academic endeavour. Your support has been invaluable.

This dissertation and degree would not have been possible without the generosity and contributions of all those mentioned above, and for that, I am profoundly grateful.

Table of Contents

Abstract *3*

Acknowledgements *4*

List of Tables *6*

List of Figures *7*

List of Images *8*

List of Abbreviations *8*

Chapter One: Introduction *9*

1.1 Defining Dyslexia **9**

1.2 Defining Specialized Typeface **9**

1.3 Objective and Rationale for Review **10**

1.4 Dissertation Outline **11**

Chapter 2: Literature Review *12*

2.1 Background of Dyslexia **12**

2.2 Theoretical Ideas That Underpin Dyslexia **13**

2.2.1 Genetic and Neurobiological Factors 13

2.2.2 Phonological Deficit Theory 17

2.2.3 Visual Deficit Theories 17

2.3 What Makes a Typeface "Easy" or "Hard" **19**

2.3.1 Serif versus Sans-Serif Typefaces 20

2.4 Dyslexia Friendly Fonts **20**

2.4.1 Common Features of Dyslexia Specialized Typefaces 21

2.5 The Spread of Dyslexia Specialized Typefaces **26**

2.6 Why Do We Think These Typefaces Might Work? **28**

2.7 Measured Reading Outcomes **28**

2.8 Possible Benefits and Application of Dyslexia Specialized Typefaces **29**

2.9 Gaps in the Literature **30**

Chapter 3: Methodology *31*

3.1 Research Questions **31**

3.2 Inclusion / Exclusion Criteria for Studies **32**

3.3 Information Sources **34**

3.4 Search Strategy and Terms **35**

3.5 Data Management **37**

3.6 Selection Process **37**
3.6.1 Initial Screening 38
3.6.2 Full-text Screening 40
3.6.3 Ensuring Screening Reliability 40
3.7 Data Collection Process **41**
3.8 Data Items **41**
3.9 Risk of Bias of Individual Studies **41**
3.10 Data Synthesis **43**
Chapter Four: Results ***44***
4.1 General Characteristics **46**
4.1.1 Publication Details of Studies 48
4.1.2 Geographic Location of Studies 48
4.1.3 Participants 49
4.1.4 Language of Intervention 51
4.1.5 Typefaces in Studies 51
4.1.6 Measured Reading Skill(s) 54
4.2 Study findings **55**
4.2.1 Narrative Summaries of Individual Studies 55
4.2.2 Effect Size 61
4.3 Risk of Bias **64**
4.3.1 Category of Study Design 65
4.3.2 Global Risk of Bias Ranking 66
Chapter Five: Discussion ***67***
5.1 Research Question 1 ..... **67**
5.2 Research Question 2 ..... **68**
5.3 Other Findings **69**
Chapter 6: Conclusion ***70***
6.1 Implications for Dyslexic Individuals **70**
6.2 Limitations of the Review **71**
6.3 Directions for future research **72**
References[1] ***73***
Appendix A: IDESR Protocol ***80***
Appendix B: Data Extraction Form ***88***
Appendix C: Completed Data Extraction Form Example ***90***
Appendix D & E: Completed MMAT Form for Individual Studies ***100***
Appendix F: PRISMA Flow Diagram (Completed) ***112***

List of Tables

Table 1: Table 1: Eligibility Criteria – Page 32

Table 2: List of Databases – Page 35

Table 3: Boolean Search Terms – Page 37

Table 4: List of Full References of Included Studies – Page 44

Table 5: General Characteristics of Included Studies – Page 46

Table 6: Example of Dyslexia Specialized Typefaces Used in Studies – Page 53

Table 7: Example of Non-Specialized Typefaces Used in Studies – Page 54

Table 8: Effect Sizes by Study – Page 61

Table 9: Risk of Bias Rating for Individual Studies – Page 65

List of Figures

Figure 1: Prisma Flow Diagram – Page 39

Figure 2: Year of Publication of Studies – Page 48

Figure 3: Geographic Location of Included Studies – Page 49

Figure 4: Number of Participants in Studies – Page 50

Figure 5: Age of Participants – Page 50

Figure 6: Language of Intervention in Studies– Page 51

Figure 7: Dyslexia Specialized Typefaces in Studies– Page 52

Figure 8: Non-Specialized Typefaces in Studies– Page 53

Figure 9: Reading Skills Measured in Studies– Page 55

Figure 10: Study Designs of Included Studies– Page 66

Figure 11: Global Risk of Bias Ranking– Page 66

List of Images

Image 1: Nonimpaired reader, dyslexic reader, brain comparison – Page 15

Image 2: Typical Brain (top) compared to Dyslexic Brain (bottom) while reading – Page 16

Image 3: Enhanced Spacing – Page 22

Image 4: Unique Character Shapes – Page 22

Image 5: Inclined Letters – Page 23

Image 6: Letters with Lower Gravity Points – Page 24

Image 7: Letters with Lower Gravity Points – Page 24

Image 8: Letters with Lower Gravity Points – Page 25

Image 9: Comparison of Counterforms – Page 25

Image 10: Capitalization and Punctuation – Page 26

List of Abbreviations

APA: American Psychological Association

DF: Dyslexia Friendly

IDA: International Dyslexia Association

IDESR: International Database of Education Systematic Reviews

MMAT: Mixed Methods Appraisal Tool

OD: OpenDyslexic

ORF: Oral Reading Fluency

PICO: Participants, intervention, control, outcomes

PRISMA: Preferred Reporting Items of Systematic Reviews and Meta-Analyses

PROSPERO: International Prospective Register of Systematic Reviews

RD: Reading Disability

RoB: Risk of Bias

TNR: Times New Roman

WCAG: Web Content Accessibility Guidelines

Chapter One: Introduction

1.1 Defining Dyslexia

Dyslexia is one of the most prevalent learning disorders. It is a condition that is estimated to affect one in every 5 children, according to the National Center for Learning Disabilities (2017). Moreover, around 80% of people who are classified as having a learning disability are diagnosed with dyslexia, making it the most prevalent learning disability (Shaywitz, 1998). The learning disability usually impairs a person's reading, writing, and spelling capacity. In addition, the ability to recognise and manipulate language sounds, or phonological processing, can also pose an issue for those with dyslexia (Tunmer & Greaney, 2010). As a result, individuals may take longer to complete reading and writing chores than their classmates or avoid them entirely (McNulty, 2003). Although the disorder was previously associated with lower intelligence, research now suggests that individuals are diagnosed with dyslexia regardless of ability or educational experience (National Institutes of Health, 2015).

Reading is a prodigious mode of communication. We read for school, work, information, and pleasure and if we cannot decipher the written word quickly and fluidly, reading becomes an arduous process. Due to the commonality of this neurodevelopmental disorder and exposure to a world where literacy is given much importance, people create and sell typefaces as a solution to reading with dyslexia. Since there is no cure for dyslexia (Mayo Clinic, 2022), individuals with dyslexia often employ technological assistance (Gregor & Newell, 2000) to mitigate reading struggles and specialized typefaces have become one of the solutions individuals turn to for assistance.

1.2 Defining Specialized Typeface

A specialized typeface is essentially a font that serves or targets a specialised need (George, 2012). In the context of this paper, specialized typefaces will refer to fonts created specifically to aid dyslexic readers. Over the years, researchers and designers have created specialised dyslexia typefaces, often known as fonts, to assist in improving the symptoms of the disorder. To mitigate readability and prevent distortion, these typefaces employ distinct design elements such as increased letter spacing, stronger characters, and higher contrast between letters and backgrounds (Galliussi, Perondi, Chia, Gerbino, & Bernardis, 2020).

Recent research has raised the possibility of a connection between dyslexia and deficiencies in low-level visual processing, like excessive crowding. At the same time, "dyslexia-friendly" fonts that had been developed specifically to target the dyslexic audience entered the market. These fonts have since

created speculation regarding whether they have any impact on the reading speed and accuracy of dyslexic individuals (Galliussi et al., 2020).

1.3 Objective and Rationale for Review

There are several specialized typefaces available, and a plethora of reviews claim they work or deplore these fonts completely. This systematic review aims to answer the question of whether or not these fonts really impact the reading rate and accuracy of dyslexic readers by exploring the empirical research conducted on the topic. The rationale behind this systematic review is rooted in the proliferation of literature on the topic that has accumulated over the past few years.

Inextricably linked to this exploration of typefaces is the broader context of the digital age, where technology-driven language learning (like mobile applications) has become a prominent educational tool. As digital platforms grow and transform the learning landscape, an important question arises: Could the use of specialised typefaces contribute to an enhanced learning experience for dyslexic individuals who are learning a language? recognising the possible implications that typeface changes may have for the emerging realm of language learning technology. Essentially, the importance of this review stems not only from its ability to shed light on the empirical evidence on claims about specialised typefaces but also from its ability to inform practical applications and pedagogical strategies.

A systematic review of the chosen topic will seek to determine the efficacy of specialised dyslexia typefaces in enhancing reading rate and accuracy in people with dyslexia. Moreover, there are numerous specialised dyslexia typefaces available, each claiming to improve readability and alleviate the difficulties faced by people with dyslexia. A systematic review can evaluate and contrast various fonts, design concepts, and their impact on reading outcomes. This will provide useful information on which typefaces are more effective and why. This report can further offer evidence-based recommendations for the use of specialised dyslexia typefaces by synthesising the data of various studies. These recommendations can assist educators, parents, and individuals with dyslexia make educated decisions about these fonts.

According to Mallett, Hagen-Zanker, Slater, & Duvendack, (2012), a systematic review can be beneficial in identifying gaps in the existing literature and this report will aim to identify the gaps in research concerning specialised dyslexia typefaces.

1.4 Dissertation Outline

In this systematic review, I first review the existing literature on the effects of specialized typefaces on dyslexic readers. The literature review chapter will include background information on dyslexia as a neurodevelopmental disorder, its characteristics and underpinning theories explaining the causes of the disorder. In addition, a section on the various existing specialized typefaces will be included, explaining the features and rationale behind their creation. Next, I define and outline my research methodology which will include the research questions, the inclusion/exclusion criteria for the papers included in this report and the screening process. The methods chapter will be followed by the results section which provides an analysis and report on the findings of the included studies. This review will then conclude with a discussion of the found results, a conclusion and finally, recommendations for the trajectory of future research in the area.

Chapter 2: Literature Review

This chapter will focus on the existing literature surrounding the theories that underpin the causes of dyslexia and the use of a dyslexia-specialized font with the intent to mitigate reading struggles and enhance speed and accuracy. The first part of this chapter will give an overviewing background of dyslexia as a neurodevelopmental disorder and then discuss the various theories behind the reading disability. The next section of the literature review will focus on the specialized typefaces created so far their features and why these features may potentially be preferred over non-specialized typefaces among dyslexic individuals. The chapter will conclude by assessing the gaps in the literature and how this review will add to our current knowledge regarding the effects of specialized typefaces on neurodiverse, dyslexic individuals.

2.1 Background of Dyslexia

One of the earliest observances of Dyslexia was made by a German professor, Adolf Kassamul, who described the conditions of the disorder by terming it “word blindness” (Kassamul, 1877) Dyslexia, as a neurodevelopmental disorder, primarily affects reading and language-related skills, but more widely, it can inadvertently effect learning in other subjects and areas that require a fair amount of reading. Regardless of an individual's intelligence or educational background, they can develop this neurological disorder. Accurate and fluent word identification, spelling, and decoding skills are frequently challenging for dyslexics (Eden, Olulade, Krafnick, & Alkire, 2016).

German ophthalmologist Rudolf Berlin initially used the term "dyslexia" in 1883. Berlin used the phrase to designate a particular reading condition marked by challenges with reading and language processing that were not related to general intellectual decline, illiteracy, or visual acuity problems. The Greek words (δυσ) "dys," which means difficulty, and (λέξις) "lexis," which means word or language, are combined to form the English word dyslexia. As a result of Berlin's work, dyslexia was recognised as a genuine learning difference and is now frequently used to describe the unique reading difficulties that certain people suffer (Wagner, 1973).

The impact of dyslexia varies depending on the severity of the condition and the efficiency of training or rehabilitation. The primary challenge is with word recognition, reading fluency, spelling, and writing. Some people with dyslexia are able to acquire early reading and spelling activities, especially with exceptional teaching, but they eventually struggle with more complicated language abilities such as syntax, understanding textbook content, and writing essays (International Dyslexia Association, n.d.).

Dyslexia can also have an impact on a person's self-esteem. Students with dyslexia frequently feel less capable than they are. A student may grow discouraged about continuing in school after enduring a considerable amount of stress as a result of academic challenges (International Dyslexia Association, n.d.). However, it has been well established that dyslexia has no influence on an individual's intelligence. There are many revolutionary and intellectual people who have had dyslexia, suggesting that this learning disorder does not hinder intelligence or success. Some notable examples include Albert Einstein, John F. Kennedy and Pablo Picasso to name a few (Weitzman, 2022) These individuals, despite their struggles with dyslexia, exemplify that dyslexia does not affect intelligence and creativity.

2.2 Theoretical Ideas That Underpin Dyslexia

In order to understand why dyslexia-specialized fonts may have a beneficial impact on reading performance, it is crucial to first understand the underpinning theories that explain the causes of dyslexia. In this section of the literature overview, some of the predominant theories that underpin dyslexia will be highlighted to better understand the concept and reason behind the creation and design of specialized dyslexia typefaces.

As stated previously, the diagnosis of dyslexia has been around for several decades, and because it affects many people in a world that relies so heavily on literacy, numerous theories and models have been proposed to explain the causes of dyslexia. However, dyslexia is a complex disorder, and its precise causes and mechanisms are still being investigated. In addition to genetic and neurobiological factors, there are other main etiological causes for dyslexia that have been suggested, namely, language processing deficits (Marsh, Freidman and Welsh, 1981) and visual processing deficits (Lovegrove, Martin and Slaghuis, 1982). These theories are discussed at length below.

2.2.1 Genetic and Neurobiological Factors

There is substantial evidence to support the hypothesis that dyslexia has a hereditary component. For instance, research on twin studies (Bishop, 2015) provides insight into the heritability of dyslexia and has facilitated the identification of prospective candidate genes linked to this condition. DeFries, et. al (1987) conducted a study examining the development of reading disabilities in twins over time. The findings from a sample of 64 pairs of identical twins and 55 pairs of fraternal twins suggested that genetic factors influence the persistence of reading disabilities across different age groups.

There are multiple genes that have been pointed to as being responsible for the development of dyslexia, and it is considered that numerous variations or mutations in genes may be a contributor to the disorder. The particular genetic pathways that are involved in dyslexia are still being explored, despite the fact that it is likely that dyslexia is a complex genetic trait influenced by several genes (Fisher & DeFries, 2002).

According to research, genetic factors play a key role in dyslexia (Gilger, Hanebuth, Smith, & Pennington, 1996), which commonly runs in families. Several genes have been linked to dyslexia, although the exact genetic pathways and their interconnections are currently being researched. Scarborough (1990) reported that 23-65% of children who have a parent with dyslexia have similarly been identified as sharing the reading disorder.

Neurobiological research has revealed differences in brain structure and function between those with dyslexia and those who do not have the disorder. Functional brain imaging techniques, such as functional magnetic resonance imaging (fMRI), have revealed differences in the activation and connectivity of brain regions involved in language processing, such as those engaged in phonological processing and reading comprehension (Krishnan, Saloni, Watkins, Kate, Bishop, Dorothy, 2016). These distinctions are frequently observed in language-processing regions, such as the posterior regions of the left hemisphere.

The left hemisphere of the brain, particularly the posterior areas in the parietotemporal and occipitotemporal regions, is a crucial area of study. These areas are linked to phonological and visual word recognition, respectively. Differences in the development and functioning of certain brain regions may contribute to dyslexics' difficulties decoding and interpreting written language (Shaywitz, Shaywitz, Pugh, Mencl, Fulbright, Skudlarski, Constable, Marchione, Fletcher, Lyon, Gore, 2002).

Shaywitz et al. (2002) observed substantial changes in brain activity patterns between non-impaired and dyslexic readers. These variations emerged when phonological awareness was required in a non-word rhyming challenge. In dyslexic readers, there is a disruption in the posterior region involving the superior temporal gyrus and angular gyrus, with an increase in activation in the inferior frontal gyrus anteriorly. Children without dyslexia revealed much more activation in the left hemisphere than children with dyslexia. The researchers studied 144 right-handed children comprising 70 dyslexic readers, and 74 nonimpaired readers using functional magnetic resonance imaging (fMRI). The findings revealed that dyslexic children have a disruption in reading neural circuits including posterior brain regions such as the parietotemporal and occipitotemporal areas. The degree of activation in the

left occipitotemporal area was found to be positively linked with reading ability. Activation in the left and right inferior frontal gyri was larger in older dyslexic individuals compared to younger children.

Image 1: (Shaywitz et al., 2002) The left column represents a nonimpaired reader, the middle a dyslexic reader, and the right a comparison.

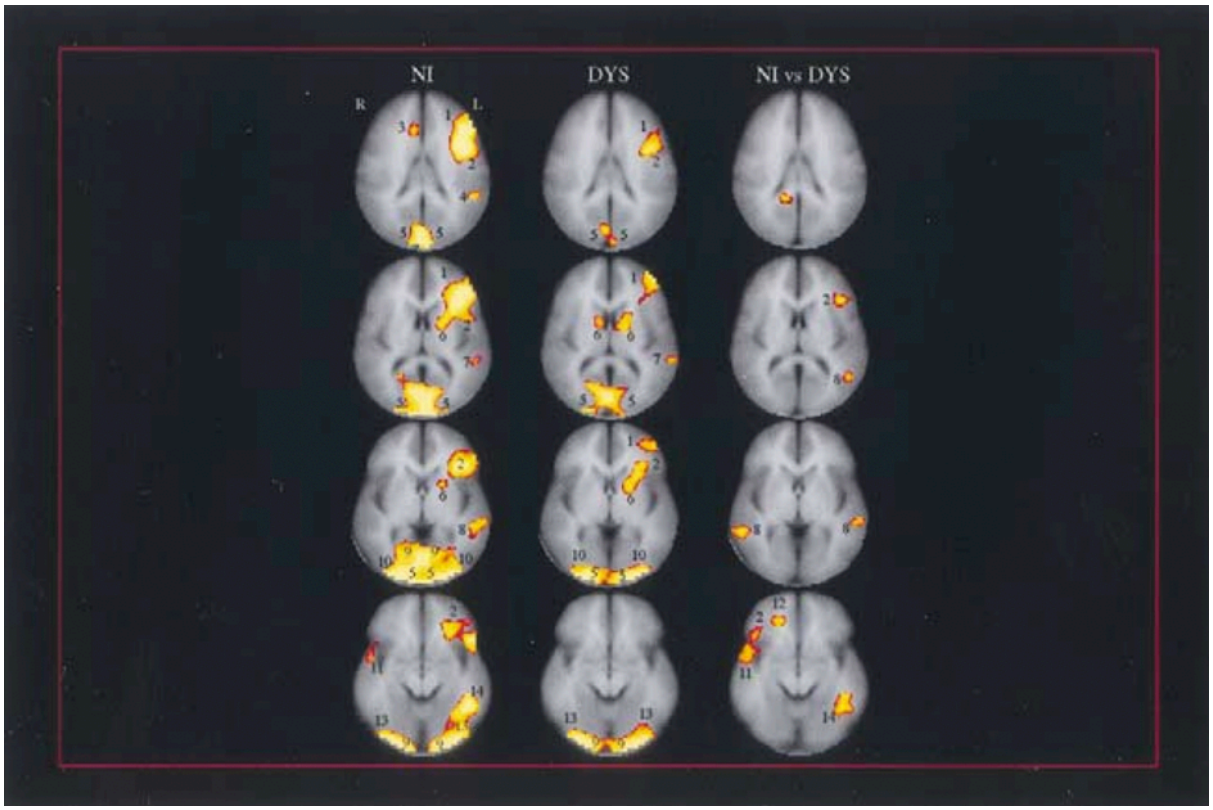
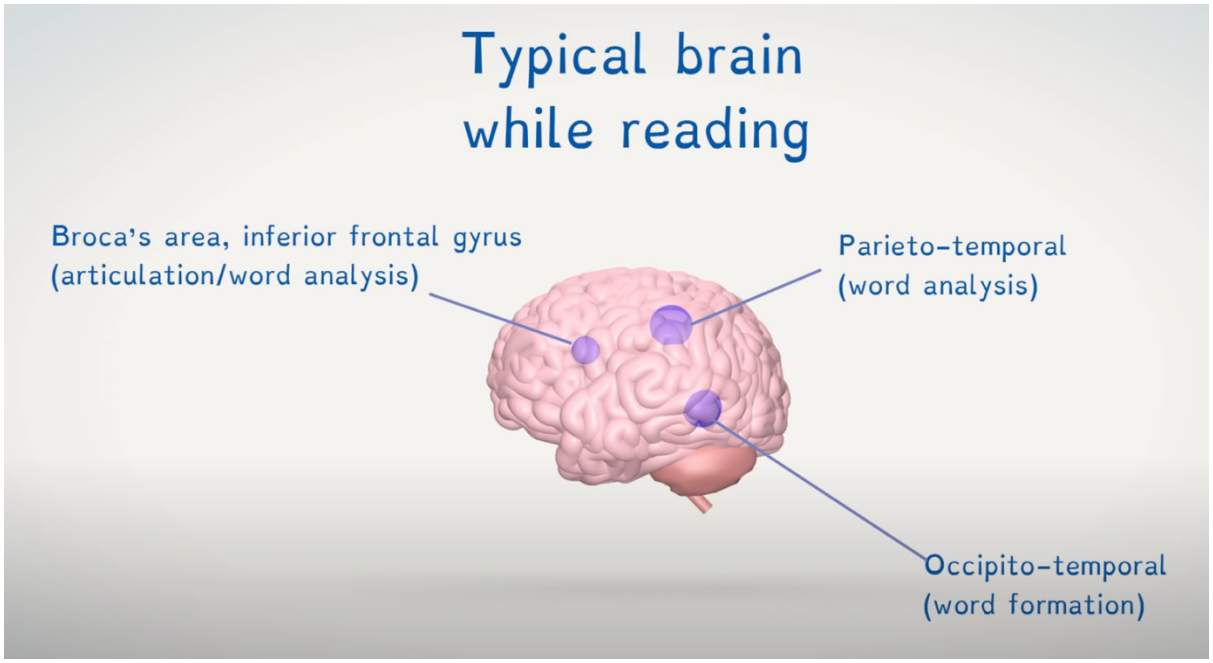
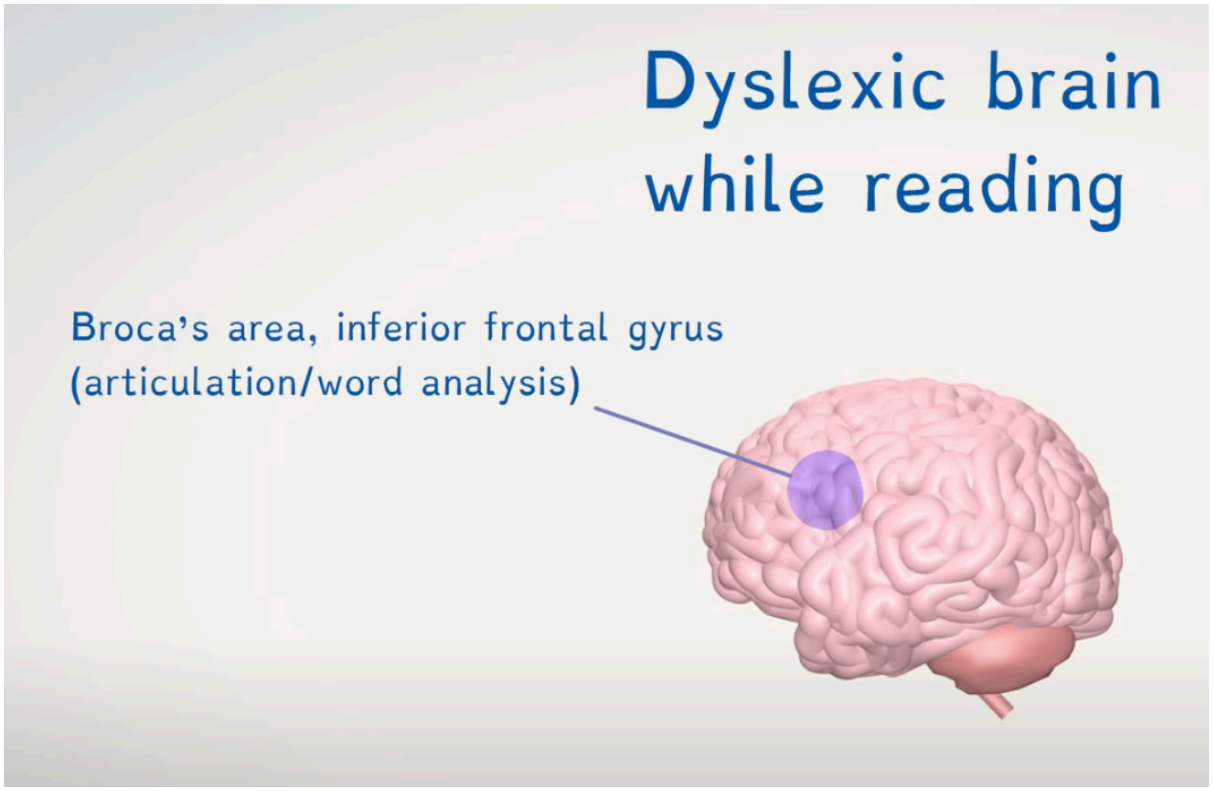


Image 2: Typical Brain (top) compared to Dyslexic Brain (bottom) while reading (Boer, 2015)





2.2.2 Phonological Deficit Theory

One of the most prevalent theories regarding the underlying cause of dyslexia is the Phonological Deficit Theory. This idea contends that a specific deficiency in language's phonological processing is the primary cause of dyslexia. Phonological awareness refers to the ability to identify and manipulate individual sounds, known as phonemes in spoken language (Blachman, 2000). Phonological awareness has a significant impact on literacy for several reasons including word recognition, vocabulary development, reading and writing. Almost all explanations about the causes of dyslexia include the assumption that phonological deficits play a role. The phonological deficit theory, in its most extreme form, claims that a phonological deficit is the only cause of dyslexia. Children with strong phonological skills, often known as "phonological awareness," excel as readers and spellers. Children with low phonological awareness, on the other hand, will struggle to memorise, recall, and manipulate the sounds of their language (Plessis, 2023).

According to the theory, people with dyslexia have trouble accurately and quickly processing phonological information. They might have trouble analysing and manipulating language's sounds, which makes it challenging for them to match sounds to letters and decode words. Phonological processing is the ability to distinguish, classify, and modify the various sounds (phonemes) that comprise spoken language. This includes knowledge of how words are made up of smaller sound components (phonemic awareness) (Ramus, Marshall, Rosen, & Van Der Lely, 2013).

Essentially, the phonological deficit hypothesis of dyslexia (Snowling, 2001) suggests that poor readers struggle to learn the relationships between letters and sounds. As a result, this theory possibly predicts that if children with dyslexia receive phonics training (learn grapheme-phoneme correspondence rules) in the setting of a typeface with visually distinct characters, they will learn faster.

2.2.3 Visual Deficit Theories

According to the visual deficit theories, dyslexia is an array of visual deficits that include difficulties learning to read, spelling, and processing symbolic information slowly. The disability has been linked to difficulty in processing sound forms, but additional issues experienced by dyslexics imply deficits in perceptual processes, motor coordination, attention, and memory (Everatt, Bradshaw, Hibbard, 1999). Reading requires both visual and phonological processing, therefore it stands to reason that reading difficulties could be caused by issues in either or both systems. The issue of visual crowding is relevant to the usage of specialised typefaces for dyslexic readers. Visual crowding is defined as having

trouble detecting objects (e.g., letters) when they are surrounded by comparable items (e.g., other letters) when presented in isolation (Bouma, 1970).

The Visual Spatial Attention Theory focuses on dyslexic spatial attention deficiencies. It suggests that dyslexic people may struggle with appropriately recognising and understanding spatial relationships between letters, words, and text. These challenges may have an impact on letter and word recognition as well as general reading fluency (Ackerman, Dykman, Gardner, 1990). Dyslexic people may struggle to perceive and understand the spatial arrangement of visual elements such as letters and words on a page. This covers letter spacing, word spacing, and alignment issues. The phenomenon of visual crowding occurs when the presence of surrounding visual stimuli interferes with the perception and identification of a target stimulus. Individuals who are dyslexic may be more susceptible to visual crowding, making it more difficult to detect individual letters or words within a cluttered visual field (Martelli, Di Filippo, Spinelli, & Zoccolotti, 2009). According to the Visual Spatial Attention Theory, dyslexic people may struggle to organise visual information in an organised and cohesive manner. This can have an effect on reading fluency and comprehension. The Visual Spatial Attention Theory emphasises the role of spatial perception and attentional processes in dyslexic reading difficulties. These spatial processing difficulties can have an impact on many areas of reading, such as letter and word recognition, tracking, and text comprehension. Understanding and correcting visual spatial attention deficiencies can inform interventions and instructional strategies to help dyslexic individuals build effective reading skills.

The Magnocellular Theory provides a potential explanation of what could underpin visual deficits in terms of neurobiology. The theory suggests that the magnocellular pathway of the visual system is impaired in people with dyslexia. One of the two primary pathways in the visual system, the magnocellular pathway, is in charge of processing rapidly changing visual information, including transient stimuli, contrast, and rapid motion. Magnocellular neurons, specialised cells that make up the magnocellular pathway, are in charge of detecting and processing low-frequency visual information with high temporal resolution. The magnocellular visual system is essential for controlling eye movements, processing motion, and vergence which is when both eyes point in the same direction. Aspects of visual processing, such as motion perception, temporal processing, and the capacity to recognise quick changes in the visual environment, may be impacted by these deficiencies. According to the theory, reading abilities can be impacted by magnocellular deficiencies. It may be more difficult to track and follow text due to difficulties in processing and detecting rapidly changing visual information, which makes it harder to read accurately and fluently (Stein and Walsh, 1997).

Another visual-based theory is the Impaired Visual Attention Span (VAS) Theory. According to the theory, dyslexic people may struggle to maintain their visual focus and precisely track the text as they read. As a result, they may skip words or lines, get disoriented, or have trouble reading smoothly and fluently. In most languages, children scan small symbols arranged in rows from left to right while reading. According to the notion of impaired visual attention span theory, reading development is limited in dyslexia because there are fewer individual elements accessible for simultaneous processing in the "attentional window," making such multi-letter parallel processing unusual (Valdois, Bosse, Tainturier, 2004).

Of all the various etiological causes for dyslexia and theories and hypotheses that underpin the disorder, visual-based deficits might be the main driving force behind the creation of specialized typefaces. While the phonological deficit theory remains a cornerstone in explaining dyslexia, the version that suggests it is the sole reason for the reading disorder seems to be extreme as the integration of visual-based deficits and their impact on reading cannot be overlooked. Designers like Christian Boer (Dyslexie) have spoken about the idea behind specialized typefaces being rooted in visual-based challenges. Visual issues have been addressed by adjusting letter forms, spacing, and other typographic elements. These changes attempt to improve the distinctness of individual letters, lessen crowding effects, and promote smoother visual tracking, making it easier for individuals with dyslexia to decode and comprehend text.

2.3 What Makes a Typeface “Easy” or “Hard”

Several elements influence a typeface's readability, making it "easy" or "hard" to read for readers in general, whether dyslexic or non-dyslexic. On average, readers prefer fonts with clear and distinct letterforms, as well as well-defined shapes and proportions. Typeface designs with minimal ornamentation or elaborate elements are considered easy to read. Readers have also shown an inclination towards more spacing to recognise individual words. Inadequate word space can make writing appear cluttered and make it difficult to read (British Dyslexia Association, n.d.).

A study carried out by Dressler (2019) investigated the effect of typefaces on reading comprehension. They compared a “difficult-to-read font” – Haettenschweiler (Haettenschweiler) and an “easy-to-read font” – Times New Roman (Times New Roman). The study involved 24 undergraduate or master's students and was carried out under time restriction where the participants read two passages and answered a set of multiple-choice questions. The hypothesis was that participants would perform poorly with the Haettenschweiler font when compared to Times New Roman due to more cognitive

resources being employed to decode the “harder” typeface. The results reflected the hypothesis positively, with the students scoring better reading in the Times New Roman typeface. A comparison of the two fonts used in the study shows that Haettenschweiler is more crowded with very little space between the letters and the letterforms don’t vary in shape, unlike Times New Roman.

In addition, the concept of word shape is a basic typographic element linked to legibility (Hiller, 2012). The word shape model was initially presented by psychologist James Cattell (1886). According to this paradigm, words are recognized as full units based on the outlines created by their shape. As a result, lowercase letters are more legible than uppercase letters, according to the word shape model, because ascenders and descenders (letter strokes that protrude above the x-height2 and below the baseline3 in lowercase letters) provide a more distinctive word shape for the reader: compare, for example, 'TYPEFACE' and 'typeface'. This is the first recognition model described in psychological literature, and it is still widely used in the field of typographic legibility currently.

While these factors contribute to readability, individual preferences and needs vary, especially for those with reading disabilities, special consideration is required which is why there are various specialized fonts to cater for their needs.

2.3.1 Serif versus Sans-Serif Typefaces

Another feature to consider is Serif vs. Sans Serif typefaces. Serif typefaces, which have little strokes or flourishes at the ends of letters, are commonly regarded as more readable for print, although sans-serif typefaces (which do not have those strokes) are frequently favoured for digital displays. When looking at examples of dyslexia-specialized typefaces (see Images 3-10) it can be seen that these typefaces are usually sans-serif. The British Dyslexia Association recommends using sans-serif typefaces (for example: Arial or Comic Sans) for dyslexic readers as letters can appear less crowded.

2.4 Dyslexia Friendly Fonts

Although dyslexia is a lifelong condition, people with dyslexia can develop coping strategies and abilities to manage their difficulties. Assistive technologies, such as text-to-speech software, speech recognition tools, and typefaces designed for those with dyslexia, can mitigate the effects of the neurodevelopmental disorder and improve reading and writing skills. In a predominantly literate world, dyslexia is a serious problem that must be addressed; specific dyslexia typefaces are one solution.

There are many fonts that are considered “dyslexia friendly” people claim Helvetica, Arial or Comic Sans aid dyslexic readers (Carmichael, Dyslexia Scotland, 2022) (British Dyslexia Association, n.d.) but in more recent times, the emergence of typefaces specifically designed targeting the dyslexic audience have entered the market. Typefaces such as Dyslexie, Opendyslexia, Sylexiad or Lexie Readable all have similar characteristics that are intended to promote easier readability among dyslexics.

OpenDyslexic, a specialised dyslexia font, has been designed to increase readability for people with dyslexia. These typefaces include elements that help dyslexic readers distinguish between related letters, such as b and d or p and q. The distinctive letterforms offer a more constant baseline, thicker bottoms, and distinctive shapes that improve letter recognition. If truly effective, these typefaces have the potential to help dyslexic individuals access information more effectively and lessen the frustration associated with reading by adopting these fonts in educational materials, publications, and digital platforms.

2.4.1 Common Features of Dyslexia Specialized Typefaces

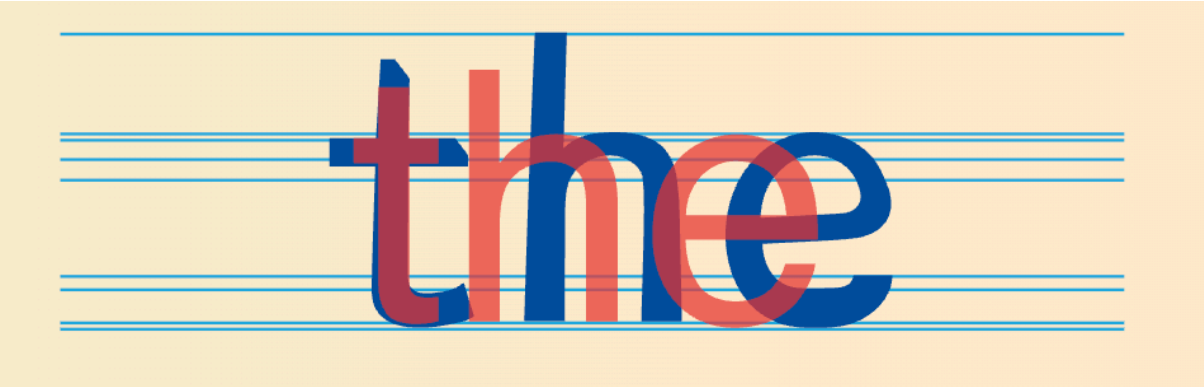
Dyslexia typefaces or fonts are specifically designed to address the difficulties that people with dyslexia face when reading text. These fonts include specific features designed to improve readability and reduce confusion caused by similar-looking letters. To make them easier to recognise and differentiate, these fonts are created by modifying traditional letterform shapes, spacing, size, and weight. Typefaces such as OpenDyslexic, Dyslexie, Sylexiad and EasyReading are some among the many dyslexia specialized typefaces available. These fonts usually share similar characteristics in the letter designs and these are explained in detail below with image examples taken from various typefaces’ webpages.

Enhanced Spacing

Increased spacing between letters and words may be used in dyslexia-specific typefaces. The increased spacing reduces visual crowding and improves letter recognition and word parsing.

There is also empirical research that highlights the effectiveness of larger letter spacing in improving reading performance in dyslexic individuals. The spacing between letters avoids the crowding effect for dyslexic readers.

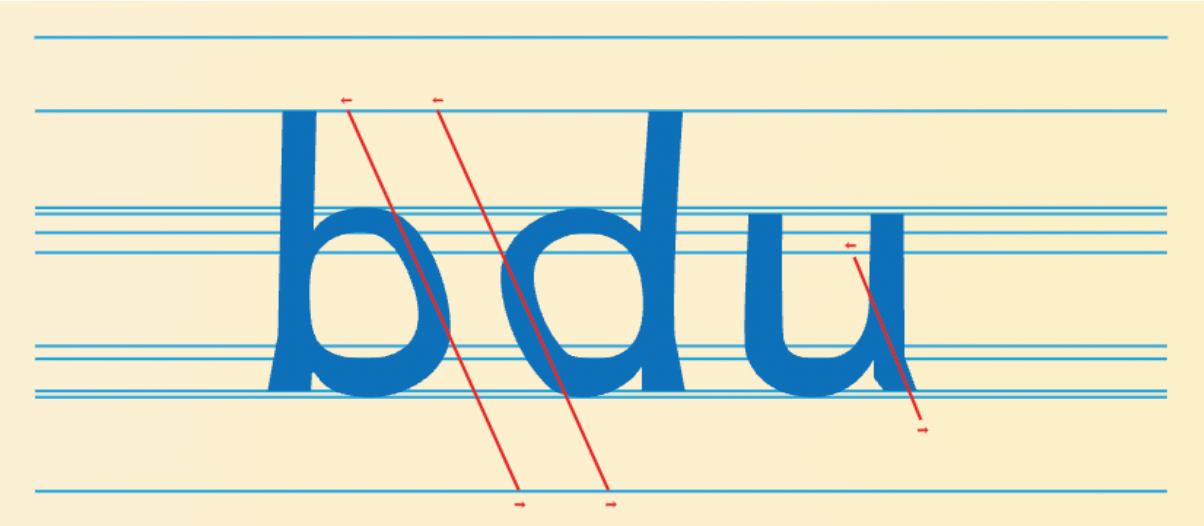
Image 3: Enhanced Spacing (from the Dyslexie website)



Unique Character Shapes

Some dyslexia-specific typefaces employ unusual shapes for specific characters. For example, the lowercase "a" could be written in a more circular shape, the uppercase "I" could have serifs, and the lowercase "g" could have a single-story design. These changes are intended to improve legibility and decrease letter reversals or confusions. The shapes and sizes of the letters are adjusted in a subtle and intentional manner that differs from other non-specialized fonts to prevent letters from mirroring, turning or swapping as they would normally.

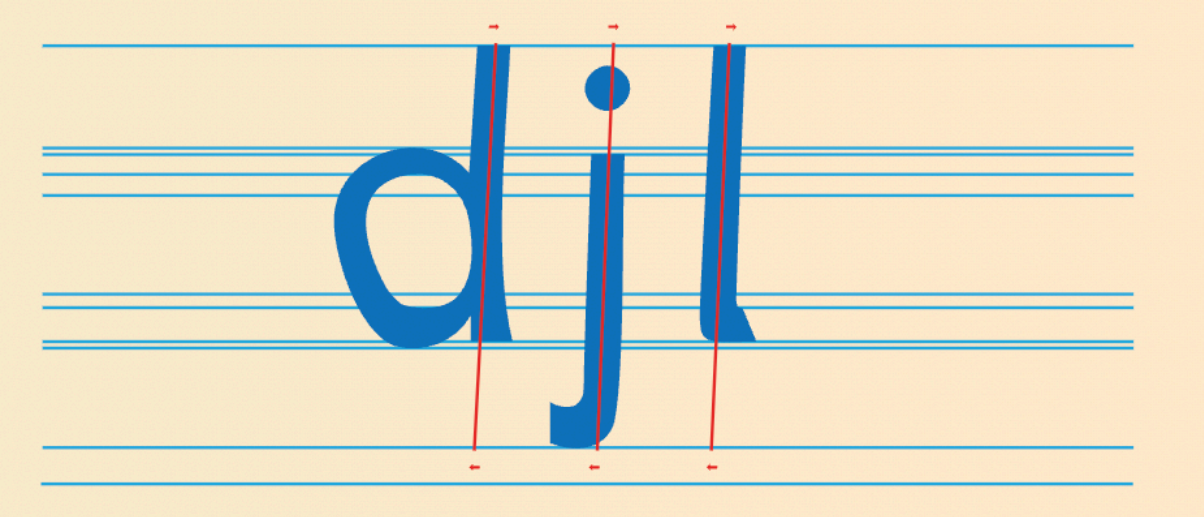
Image 4: Unique Character Shapes (from the Dyslexie website)



Clear Letterforms

Dyslexia-specific typefaces frequently highlight small distinctions between related letters. They may, for example, change the shapes or proportions of letters such as "b" and "d" or "p" and "q" to make them more distinct and easier to distinguish. Letters in the Dyslexie font, for example, are slightly slanted. Different letters are easier to discern this way.

Image 5: Inclined Letters (from the Dyslexie website)



Increased Letter Weight

In addition, dyslexia fonts have significantly stronger bottoms or thicker strokes in the lower half of letters. This serves as a visual anchor and aids in determining the correct letter orientation. The lower gravity point in the letters is intended to avoid letters from turning upside down. The baselines of these fonts are also maintained in a consistent manner, ensuring they are aligned.

Image 6: Letters with Lower Gravity Points (from the Dyslexie website)



Image 7: Letters with Lower Gravity Points (from the OpenDyslexic website)

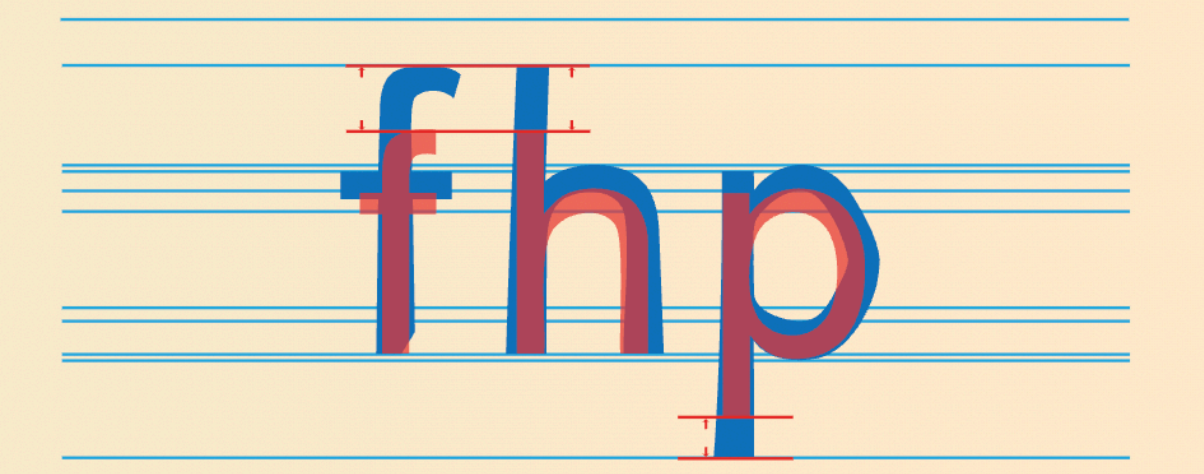


Elongated Letters

Some Dyslexia fonts have longer sticks on their letters, this helps to reduce letter switching and swapping during reading, which can be a problem for those with dyslexia. The longer sticks give visual signals and help dyslexic readers distinguish between similar letters.

Letters like "b" and "d" in Dyslexie font, for example, have longer sticks to make them more distinguishable and lessen the risk of mistake or reversal. This visual accent can help dyslexic readers identify and differentiate between these letters, reducing reading errors.

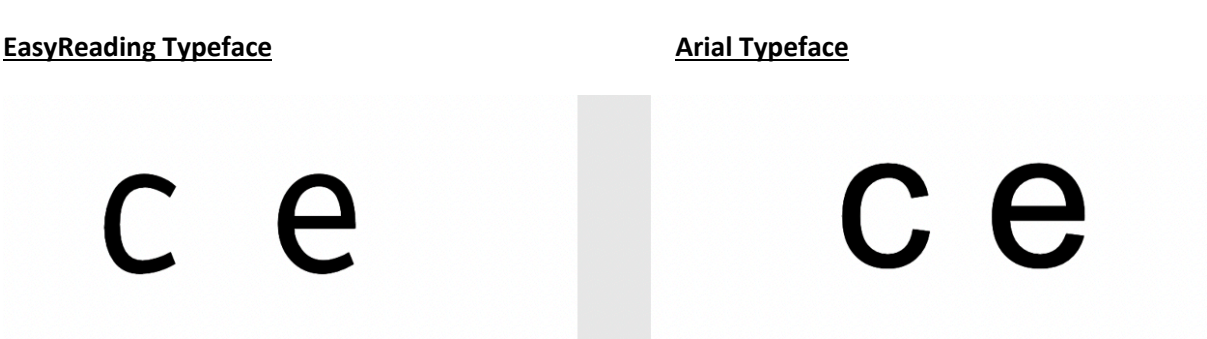
Image 8: Letters with Lower Gravity Points (from the Dyslexie website)



Open Counterforms

Dyslexia fonts frequently include larger, more open counterforms within letters. This refers to the enclosed or partially enclosed spaces within letters such as "c" or "e." The openness of these counterforms prevents them from seeming filled or deformed, which improves letter recognition.

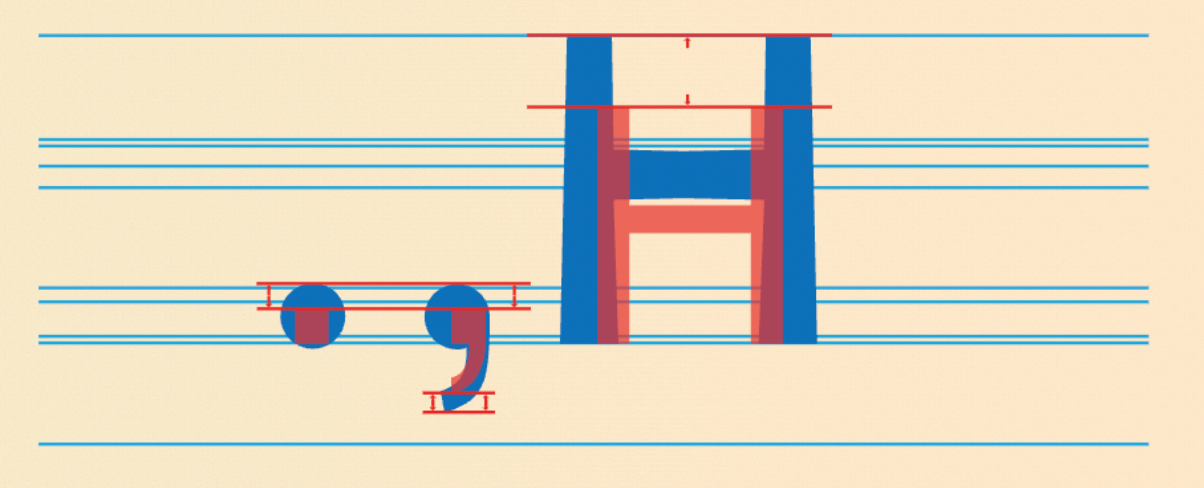
Image 9: Comparison of Counterforms (from the EasyReading website)



Punctuation

Punctuation marks and capital letters are highlighted, emphasising phrase breaks, ends, and beginnings. Dyslexia-specific typefaces often use separate and clearly discernible punctuation symbols. This involves adopting simple shapes and avoiding elaborate or decorative designs that may make recognition difficult.

Image 10: Capitalization and Punctuation (from the OpenDyslexic website)



2.5 The Spread of Dyslexia Specialized Typefaces

Over the past few years, and especially in the last decade, specialized dyslexia typefaces have grown in popularity, and graphic designers are increasingly creating new fonts with the same intention of mitigating reading struggles among dyslexic individuals. As a result, there are numerous such fonts on the market. But the question remains whether or not to make the effort to purchase these fonts and integrate them in a dyslexic learner's life. Or to implement them in learning apps, or to have printed textbooks and novels with these fonts.

The creation of specialized typefaces for dyslexic readers is possibly motivated by research findings and theories about the visual underpinnings of the disorder. Studies have suggested a possible link between dyslexia and deficits in low-level visual processing, such as excessive crowding (Whitne & Levi,2011); (Martelli, Di Filippo, Spinelli, & Zoccolotti, 2009). Dyslexia-friendly fonts present two main features that align with research findings: the particular graphic characteristics of the letterform designed to avoid confusion between similarly shaped letters, and wider inter-letter and inter-word spacing to reduce visual crowding. The open counterforms (see Image 9) in letters also reduces visual crowding.

As mentioned previously, there are a myriad of specialized typefaces on the market. Below are some of the popular typefaces.

Dyslexie Font

In 2008, a Dutch Graphic designer, Christian Boer, created the font to help with his own reading performance, being a dyslexic himself. The Dyslexie typeface is specifically designed to improve readability and reduce the challenges faced by individuals with dyslexia. Boer spoke on TEDx Talks about the many ways in which the font is designed to mitigate reading struggles (see section 2.5). The typeface was even a finalist for the Fast Company Innovate Through Design Award in 2012 (Dyslexie, n.d.). Dyslexie is one of the earliest specialized typefaces created and is available for commercial purchase. In addition, the typeface has received a lot of media attention worldwide (The Guardian, 2014); (USA Today, 2014) and is currently being used in children's publishing houses, mobile applications and electronic readers (Boer, 2016).

OpenDyslexic Font

OpenDyslexic font is another similar typeface specifically designed to assist individuals with dyslexia in reading and writing. The typeface was created in 2011 by Abelardo Gonzalez, it is an open-source font that can be freely used and shared. OpenDyslexic font incorporates several features intended to improve readability for dyslexic readers (OpenDyslexic, n.d.).

Sylexiad Font

The Sylexiad Font is an interesting dyslexia-specific font since it is a range of typefaces created specifically for adult dyslexic readers. It was designed by Dr. Robert Hillier as part of his PhD doctoral research to improve readability for adult individuals with dyslexia and has since been available to the public (Sylexiad, n.d.).

EasyReading Font

The EasyReading typeface was created in 2009 by Federico Alfonsetti. The website calls EasyReading a compensatory tool for dyslexic readers and it was designed to support various Latin-alphabet-based languages. (EasyReading, n.d). Much of the research using Italian as the language of intervention usually uses the EasyReading typeface in their experiments (Bachmann & Mengheri, 2018); (Damiano, Gena, & Venturini, 2019).

Differences Between the Fonts

Many specialized dyslexia typefaces entered the market around the same time, for instance, Dyslexie was created in 2008 and OpenDyslexic followed about four years later. Due to these close release dates, some typefaces share similar features with the same intention. Both Dyslexie and OpenDyslexic

have similar thick bases for the letters but the difference between these two fonts is the slight letter tilts in Dyslexie. The EasyReading and Sylexiad typefaces are similar as they both employ slimmer and longer letters without the thickness present in OpenDyslexic and Dyslexie.

Table 6 and Table 7 (see Chapter Four: Results) compare the specialized typefaces for a better visual representation of the differences and similarities between the fonts.

2.6 Why Do We Think These Typefaces Might Work?

All the characteristics outlined above provide reasons why these fonts may have a beneficial influence on the reading performance of dyslexics. Individuals with dyslexia as a neurodevelopmental disorder undergo many challenges with reading. Some of them are mentioned below.

Words on a page may appear twisted, jumbled, or difficult for dyslexic readers to understand. When letters are reversed or flipped, such as when "b" is seen as "d" or "p" is seen as "q," it can be confusing for them to recognise and distinguish between identical characters, among other difficulties that may damage their reading experience. A word may appear jumbled or scrambled depending on the order of its letters. It may be difficult to accurately digest and grasp the term as a result.

Letter substitutions, whereby people mistakenly use one letter for another when reading. For instance, kids might interpret "run" as "fun" or "cat" as "bat." Difficulty following where Word-to-word or line-to-line continuity issues may arise for dyslexic readers as a result of their inability to smoothly trace their gaze across the page. Reading comprehension and fluency may be impacted by this. When letters or words appear crowded or too close together, it might be challenging to tell which characters are whose. Readers with dyslexia who have trouble recognising letters may be particularly affected by this (Martelli, Di Filippo, Spinelli & Zoccolotti, 2009).

The main aim of dyslexia-specialized typefaces is to mitigate such struggles among readers. From the rationale provided for each characteristic choice, and how these choices align with the general preferences regarding “easy” and “hard” fonts, there could be hope for benefits.

2.7 Measured Reading Outcomes

This review focuses on the effect of specialized typefaces on reading accuracy and reading rate/speed as outlined by the research questions (see Chapter Three: Methodology). The two most widely used

parameters for diagnosing dyslexia are reading speed or the number of syllables read per second and accuracy or the number of errors committed during this reading (Damiano, Gena, & Venturini, 2019).

2.8 Possible Benefits and Application of Dyslexia Specialized Typefaces

If the specialized dyslexia typefaces do indeed have a beneficial effect on the reading performance of dyslexic individuals, it would mean an improved reading speed, accuracy and comprehension. In addition, if there is empirical evidence that a certain specialized font outperforms others there would be scope to adopt and apply the font(s) in various areas to assist dyslexic readers. A number of textbooks and novels could be printed in a specialized font to mitigate reading struggles and encourage reading habits. If they do not get the care they require, children with dyslexia run the danger of having worse academic success and life opportunities.

Research has examined anxiety levels in dyslexic students, often utilizing self-report questionnaires or structured interviews to assess anxiety symptoms. The findings indicate that dyslexic pupils may indeed experience higher anxiety levels than their non-dyslexic peers (International Dyslexia Association, n.d.). Carroll, Julia, and Jane E. Iles (2006) sought to investigate the anxiety levels of dyslexic students in their study. 16 students with dyslexia and 16 students without a history of cognitive difficulties were compared. Students were required to submit a verbal survey regarding trait anxiety levels. Students with dyslexia had slower reading rates than controls. They also had elevated levels of condition anxiety, academic and social anxiety, but not appearance anxiety. The anxiety levels of dyslexic students in higher education are significantly higher than those of students without cognitive difficulties. This anxiety extends beyond academic endeavours and into numerous social situations.

Another area could be adopting dyslexia-specific typefaces in learning apps can be a helpful step in assisting those who have the condition. There are a plethora of learning websites and language learning apps that could integrate a specialized font to cater to their dyslexic learners or to encourage dyslexic individuals to use these apps. There are already mobile applications such as “Dyslexia Baca” and “EasyLexia” but there could also be benefit in being able to switch to a dyslexia typeface in apps such as Duolingo or Rosetta Stone.

2.9 Gaps in the Literature

Given the fact that dyslexia is a common learning disorder, the importance to mitigate the struggles faced by dyslexic students and readers in general cannot be understated. According to the International Dyslexia Association (IDA), children who receive adequate interventions throughout kindergarten and first grade will have fewer issues learning to read at grade level with their peers than those who do not. The creation of dyslexia-specialized fonts may be an intervention for enhanced reading performance among dyslexic learners. While the potential benefits of the fonts have been outlined in this chapter, and claims via reviews and user anecdotes have been made about their benefits, it is essential to test this empirically. Since there are numerous typefaces and several papers studying their effects (more information on included studies in Chapter Three: Methodology), a systematic review investigating the empirical research would be a useful resource for understanding the current state of knowledge existing on these fonts and their reliability.

Finally, a systematic review has not been conducted to date on the effectiveness of specialized typefaces on dyslexic individuals' reading performance. This systematic review aims to contribute to the existing literature by evaluating and assessing the breadth of empirical research that has been carried out on the topic of typefaces and their effects on dyslexia reading.

Chapter 3: Methodology

The research conducted for this dissertation followed a systematic review design. A systematic review is a rigorous and meticulous synthesis of existing evidence pertaining to a particular research question or topic. It seeks to assemble evidence to answer a predetermined research query. This includes identifying all primary research pertinent to the defined review question, evaluating this research critically, and synthesising the findings. The purpose of a systematic review is to provide an objective and transparent summary of the best available evidence on a particular topic (Pollock and Berge, 2018). While it resembles a literature review, systematic reviews have the advantage of approaching a topic in an unbiased manner. In a literature review, the researcher has more extent of freedom in choosing the studies and interpreting the meaning of the results. Systematic reviews on the other hand, employ a specific research question(s) and are required to register a protocol to remain unbiased (Robinson and Lowe, 2015). Since the procedures of data extraction are transparent and methodical, systematic reviews also have the benefit of being repeated or expanded upon (Petticrew & Roberts, 2006).

In order to prevent unnecessary repetition of research, an initial search on SOLO, Google Scholar, the International Database of Education Systematic Reviews (IDESR) and the PROSPERO (The International Prospective Register of Systematic Reviews) database (The University of York, n.d.) was conducted to see if there were any existing reviews on the effects of typefaces on the reading performance of dyslexic individuals. It was found that although systematic reviews had been done on dyslexia as a disorder and the various underpinning causes, there were no results for reviews when the search terms “typeface” and “dyslexia” were put together.

To ensure transparency and accountability in the design of this systematic review, as well as to prevent potential research duplication, a protocol was registered and published with IDESR prior to the completion of the final search and data extraction.

3.1 Research Questions

The current state of our knowledge regarding dyslexia, specialised fonts, and the effects that these typefaces have on the reading ability of dyslexic individuals (see Chapter Two) has led to the identification of two primary issues that this systematic review intends to answer. These questions are based on the gaps that have been discovered in the existing body of research as well as an overview of our current understanding of dyslexia.

These questions are outlined below:

Research Question (1) - What is the extent and nature of empirical research investigating the impact of specialized typefaces on dyslexics' reading outcomes?

Research Question (2) - What is the extent of the evidence of the effectiveness of specialized typefaces on the reading performance of dyslexic readers?

(2. a) What impact do specialized typefaces have on reading rate?

(2. b) What impact do specialized typefaces have on reading accuracy?

3.2 Inclusion / Exclusion Criteria for Studies

The following table (Table 1) lists the inclusion/exclusion criteria established for the decision of either including a study as part of this systematic review or not. The table also includes a section for the rationale behind the decision of each inclusion/exclusion criterion.

There were ten items that were identified for inclusion and exclusion which were: intervention, participants, age of the participants, control groups, outcomes, language, language of publication, date of publication, bibliographic information and publication status.

Table 1: Eligibility Criteria

Item	Inclusion	Exclusion	Rationale
Intervention	To include studies that investigate the effects of specialised dyslexia fonts on the reading performance of people who have dyslexia	To exclude studies on general typography or fonts that have no direct relevance to dyslexia or reading difficulties.	This review focuses on specialized fonts created especially for dyslexic readers; thus, general fonts were not be considered.
Participants	To include studies involving people of any age who have	To exclude studies that do not involve individuals who have dyslexia or studies that are primarily	Since this systematic review focuses on the effects of specialized typefaces on the reading performance of individuals with dyslexia, the

	been diagnosed with dyslexia.	concerned with other learning difficulties or neurodevelopmental disorders and make no specific mention of dyslexia.	participants in the study need to be those diagnosed with the neurodevelopmental disorder.
Age	All age groups included.	No age restriction, age was considered as an influential variable in the analysis.	Although a majority of the research is conducted with children, there are some specialized typefaces (example: Sylexiad) that are specifically made for adults. In order to include as many studies on a variety of typefaces, no age restrictions were made.
Control groups	To include studies with randomized controlled trials (RCTs) or well-designed comparative studies.	To exclude studies that lack a control group or appropriate comparator condition.	Since this review focuses on
Outcomes	To include studies that test accuracy and/or reading rate.	To exclude studies testing other outcomes other than accuracy or reading rate.	Reading accuracy and fluency (rate and speed) are usually the most common testing outcomes for dyslexia reading research. Therefore, given the timeframe for this MSc dissertation, the outcomes have been limited to reading accuracy and rate.
Language	To include studies that conducted experiments using a language with Roman Alphabet (Examples: English, French, Italian).	To exclude studies using any other language.	Incorporating studies on languages with similar linguistic characteristics (e.g., Roman alphabet languages) enables this systematic review to examine dyslexia within a more homogenous group. This can help identify patterns and factors that are unique to these languages and shed light on literacy

			difficulties that are language-specific. Looking into specialized typefaces for languages that use logographic writing systems would require a different background.
Language of publication	To include studies published in English	To exclude studies published in different language.	Since the author's first language is English, it would be feasible to only include studies published in English.
Date of publication	To include papers from 1957 onwards	To exclude papers before 1957.	The oldest font considered dyslexia-friendly was created in 1957 (Slocombe, 2020).
Bibliographic information	To include studies with complete reference and sufficient bibliographic information	To exclude studies lacking sufficient bibliographic information.	The retrieval of works cannot be adequately performed without adequate bibliographic information.
Publication status	To include unpublished theses as well.	No studies are excluded because of their publication status.	To mitigate against the effects of publication bias and incorporate a larger range of literature in the study process, grey literature (unpublished theses) was included.

3.3 Information Sources

The table below lists the databases that were used for this systematic review. Well-established databases in the domains of linguistics, education and psychology were consulted as well as multidisciplinary sources that cover a larger overview of the social science fields, are included in this list. All databases were accessed electronically through the Bodleian Library (University of Oxford's) subscription.

The cell 'initial scoping search' in Table 2 refers to the initial search for systematic reviews in the field (to avoid duplication) as well as to gauge the extent and type of literature existing.

Table 2: List of Databases

DISCIPLINE	DATABASE
Education	ProQuest Education Collection (ERIC included) British Education Index (EBSCO)
Linguistics	ProQuest Linguistics Collection (LLBA included)
Psychology	PsycINFO (Ovid)
Multidisciplinary	Web of Science SCOPUS (Elsevier) PubMed
Grey literature	ProQuest Dissertations & Theses Global
For the initial scoping search	Google Scholar SOLO JSTOR

3.4 Search Strategy and Terms

For this review's initial literature search, a consultation was made with a University of Oxford college librarian (St. Anne's College) and an Education (Bodleian) librarian to create a list of search terms relevant to the current systematic review.

According to Brunton et al. (2017), an ideal systematic review finds a compromise between sensitivity and specificity, spreading a net far enough so that no qualifying studies are missed in the search, but also ensuring that its emphasis is restricted enough so that researchers do not have to filter through an extensive amount of research that is irrelevant to their question.

For this reason, the Boolean search string (described below) underwent a few revisions and changes before it was finalized.

A pilot search containing the term “neurodiverse” was first used with “OR” “dyslexia”, but this yielded too many results with diverse and unrelated neurodevelopmental disorders such as ADHD and autism which were not the focus of this review, so the term was dropped and only the specific “dyslexia” term remained to ensure the results were relevant. The term “reading disability” is also often used in relation to dyslexia, so this term was added to avoid any eligible papers being left out.

The pilot search did not include specific fonts’ names, but it was seen that some important papers were left out so “dyslexie”, “lexie readable”, “sylexiad” and “opendyslexia” were then added to the search string. A well-known paper which met all the criteria for this systematic review was used as a reference check for the search string, it was anticipated that the results of the search string would include the particular paper.

The initial search looked at abstracts only and the Boolean string was slightly adapted relative to the specifications of the respective databases. The Boolean search string used in the SCOPUS database is presented as an exemplification below.

TITLE-ABS-KEY (dyslexia) OR TITLE-ABS-KEY ("reading disability") AND TITLE-ABS-KEY (typeface*) OR TITLE-ABS-KEY (font*) OR TITLE-ABS-KEY ("specialized font") OR TITLE-ABS-KEY ("specialized typeface") OR TITLE-ABS-KEY ("dyslexia font") OR TITLE-ABS-KEY (dyslexie) OR TITLE-ABS-KEY (opendyslexic) OR TITLE-ABS-KEY (“lexie readable”) OR TITLE-ABS-KEY (sylexiad) AND TITLE-ABS-KEY ("reading outcome*") OR TITLE-ABS-KEY ("reading rate") OR TITLE-ABS-KEY (“reading fluency") OR TITLE-ABS-KEY ("reading speed") OR TITLE-ABS-KEY (reading) OR TITLE-ABS-KEY ("reading accuracy")

Table 3 outlines the different terms used within each category (separated by the “OR” operator) and the categories are separated by the operator “AND”. These terms were individually inserted in each database as the different databases vary slightly in categorisation and the formatting of the Boolean search string.

Table 3: Boolean Search Terms

Neurodevelopmental Disorder	AND	Typefaces	AND	Outcomes
Dyslexia OR "reading disability"		typeface* OR font OR "specialized font" OR "specialized typeface" OR "dyslexia font" OR dyslexie OR opendyslexic OR "lexie readable" OR sylexiad		"reading outcome*" OR "reading rate" OR "reading fluency" OR "reading speed" OR "reading accuracy" OR reading

3.5 Data Management

After conducting the search, the lists of studies were uploaded into Rayyan, an internet-based, AI-powered programme designed to organise systematic reviews (Johnson & Phillips, 2018). During the screening process, Rayyan was utilised as the main platform of organisation, where duplicates of studies were eliminated, and initial screening of titles and abstracts was conducted. In addition, a reference manager (Zotero) was used to keep track of the research studies' bibliographic information (title, author, publication date, journal).

A Microsoft Excel table was used to document the inclusion/exclusion procedure, including references, the choice made at each screening step (which are the initial title and abstract stage, followed by the full-text screening), and the justification for exclusion (if relevant).

A MacBook Pro (version 12.5) was used throughout this research study.

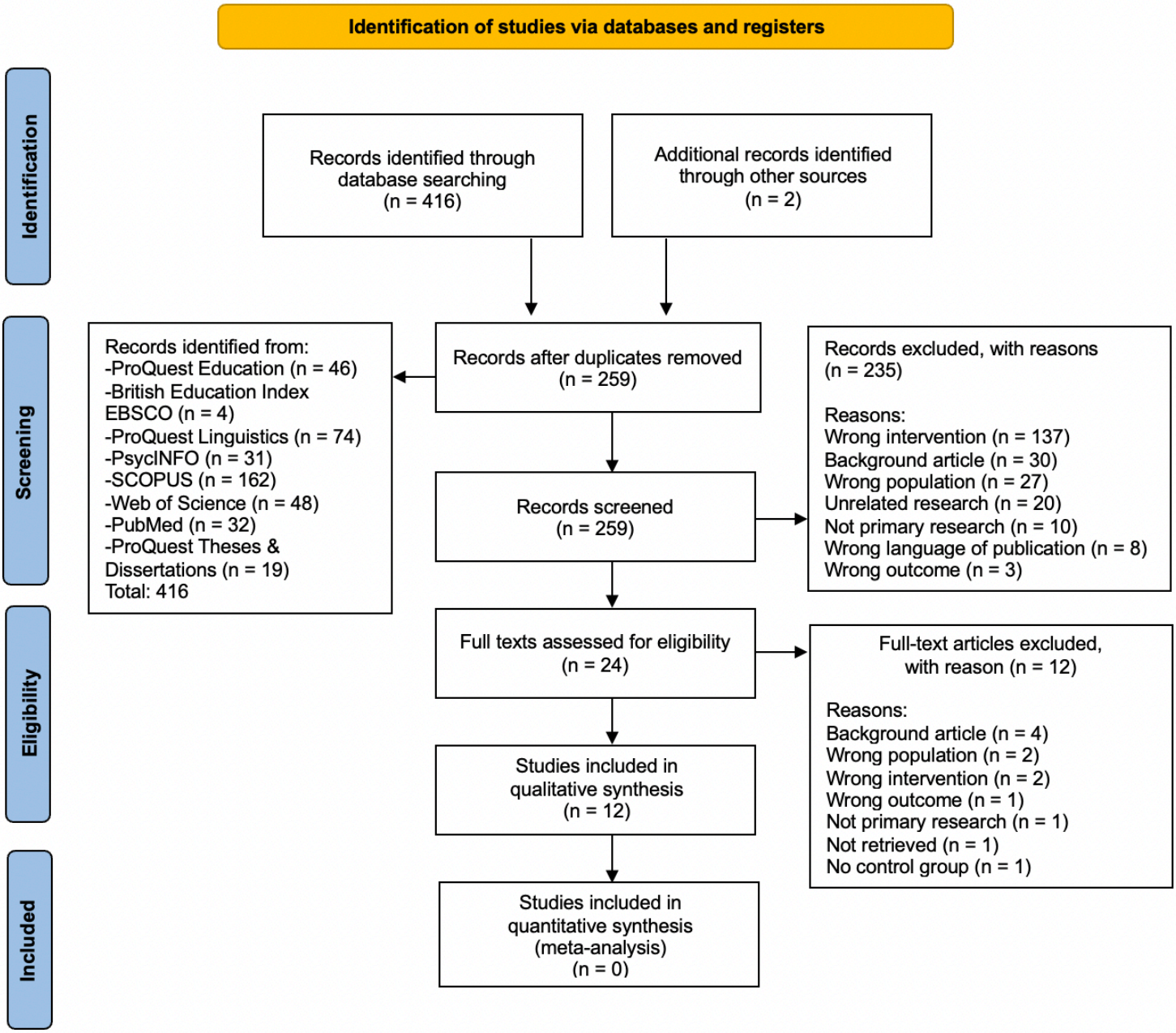
3.6 Selection Process

Figure 1, below, depicts a flow chart overview of the screening process. This section also gives a detailed account of the screening stages and process.

3.6.1 Initial Screening

The initial search of databases resulted in a total of 416 records which was subsequently reduced to 259 records after the duplicates were eliminated on Rayyan. The titles and abstracts of these 259 records were then screened against the eligibility criteria (see Table 1) and 24 studies were included for the next stage – full-text screening. The abstract of a report must not expressly contravene any of the exclusion criteria in order for it to be eligible for full-text screening. If the title or abstract did not match a criterion, the study was removed, and the reason was noted. A total of 235 records were excluded during the initial screening stage for the following reasons: wrong intervention (137), background article (30), wrong population (27), unrelated research (20), not primary research (10), wrong language of publication (8) and wrong outcome (3).

Figure 1. Flow Diagram of the Screening Process



3.6.2 Full-text Screening

Full reports were collected from the relevant databases after the initial screening. In the instance where complete texts were unavailable, they were excluded. Although ideally, we would reach out to the author(s) when full texts are unavailable (as is common practice), in the context of this MSc and due to time constraints, this was not possible. Both screening results were documented in line with the PRISMA flow diagram (Preferred Reporting Items of Systematic Reviews and Meta-Analyses) (Page et al., 2021), with a rationale for all decisions.

During the full-text screening stage, only the studies that clearly met the inclusion criteria were included in the final list of studies for this systematic review. This stage resulted in the elimination of 12 studies for the following reasons: background article (4), wrong population (2), wrong intervention (2), wrong outcome (1), not primary research (1), no control group (1) and not retrieved (1). This finally resulted in 12 studies being included in this review.

3.6.3 Ensuring Screening Reliability

To reduce bias and subjectivity during the study selection process, a second reviewer was recruited. The second reviewer had a strong background in Applied Linguistics and appropriate knowledge of the systematic reviewing approach. The second reviewer was thoroughly briefed on the inclusion and exclusion criteria for this stage of screening before the process began and was given the chance to ask any questions about the review's purpose and methods, before performing a screening on a fraction of the articles found in the search. The number of items undergoing a second evaluation was roughly 5% of abstracts (13 studies) of the total number of studies; more than 5% was not practicable due to time constraints. The second reviewer was "blind" to the decision made by the first reviewer until the final discussion. In the field of education, a Kappa value of 0.7 is typically deemed satisfactory (Frey, 2018). If there was less of an agreement between the researchers, it was intended that the original researcher and the second reviewer would discuss their disagreements and reach a consensus.

There was a perfect mutual agreement (1.0) on the decision of inclusion/exclusion of papers during this stage. It was determined that the inclusion and exclusion criteria were applied appropriately and that the initial selection procedure posed a low risk of systematic bias.

3.7 Data Collection Process

The studies that made it through both the primary and secondary screenings served as the review's data sources. The Cochrane (2017) good practice data extraction form served as the basis for the creation of a data extraction form (see Appendix B). The form was modified to meet the requirements of the works that were included in the final list, taking into account factors such as the variety of research designs. The form employed the PICO (participants, intervention, comparison, outcome) system when extracting data to include all essential categories (Petticrew and Roberts, 2006)

3.8 Data Items

The following data items were considered important for this review:

General information: author(s), title, reference, year of publication, source.

Study characteristics and methods: study design, research question(s), methods, funding (if applicable).

Participants' characteristics: sample size, age, language background (if applicable).

Task/intervention: type of task, typeface used, methods, comparators, control group(s).

Results: Outcome type, outcomes as reported by the authors, conclusions as reported by the authors, limitations.

Additional notes were also extracted if important and applicable.

3.9 Risk of Bias of Individual Studies

This review adheres to Slavin's (1986) best-evidence synthesis method, which is not biased against any research design in study selection but rather evaluates the methodological rigour and potential bias of the overall work. Bias is defined as a systematic inaccuracy or deviation from the truth that can affect a study's outcomes positively or negatively and varies in extent (Boutron et al., 2019). Even if bias cannot be completely eliminated, there are reasonable measures that researchers should employ to both decrease general biases in their work and to clearly identify any lingering bias through transparent reporting. The overall bias generated by each individual study must therefore be taken into account when conducting a systematic review, and the results must be weighted accordingly.

One of the most important criteria in identifying bias is having tools that are adaptable to the structure of each research design (Dixon-Woods et al., 2005). The Mixed Methods Appraisal Tool (MMAT) was used to assess the quality of the included studies (Hong et al., 2018). The MMAT was the only critical evaluation tool built specifically for a systematic mixed methods review (Crowe and Sheppard, 2011). The MMAT quality appraisal tool has also been used in many systematic reviews in the field of education (Scott, Rotter, Flynn, Brooks, Plesuk, Bannar-Martin, Chambers & Hartling, 2019) (Ritcher, 2021) The instrument considers a variety of study designs, categorising them into five distinct categories: qualitative, randomised controlled trials, non-randomised, quantitative descriptive, and mixed methods studies. After two general screening queries (used to exclude non-empirical papers), the instrument features quality evaluation questions for each category, to which the researcher must respond with 'yes', 'no', or 'can't tell,' as well as a remark box to explain/justify ratings. The MMAT converts each of the five criteria (in each category) into a quality score ranging from 0 to 5. A study employing a mixed-methods design was graded in all relevant categories. The MMAT user guide and template can be found in Appendix D. The final list of papers all reported quantitative data out of which 11 studies were quantitative non-randomized and 1 study was a quantitative randomized control trial. Thus, only two core sections of the MMAT were used for quality appraisal in this review. The MMAT's developers advise against providing a global ranking of strength to each study, since this may obscure the individual qualities and limitations of each study (Hong et al., 2019). However, a global assessment of the risk of bias for each study makes it more feasible to create a comparison between studies. Moreover, an overall strength rating is a typical component of many risk of bias tools (Li et al., 2017). Therefore, in addition to including the individual risk of bias assessment for individual studies, this systematic review also allocated a global weight of evidence score of “weak”, “moderate”, or “strong” depending on the possibility of bias. Guidelines for assigning global ratings were based on other tools such as the Effective Public Health Practice Project Quality Assessment Tool (Armijo-Olivo et al., 2012). A study was given a "strong" global rating when none of the criteria were evaluated as 'no,' and at least four as ‘yes,’ a study was assigned a “moderate” rating when the answers were no more than one ‘no’ and at least three ‘yes,’ and “weak” ratings were assigned to studies with two or more ‘no’ answers or fewer than three ‘yes.’ The quality assessments of the studies were taken into account when discussing the findings.

It is recommended that there be more than one reviewer involved in this process since determining the risk of bias and the overall weight of the evidence is subjective (Petticrew & Roberts, 2006) (Hong et al., 2018). As per the protocol, a second reviewer conducted 10% (1 study) of the risk of bias evaluations, and similar to the initial abstract screening the expected threshold for interrater reliability

was a Kappa value of 0.7. There was a perfect agreement (1.0) between the two reviewers during this process as well.

3.10 Data Synthesis

The main method for presenting data is a narrative synthesis. This review aimed to follow Popay et al.' s (2006) guidelines for narrative synthesis: (1) Developing a theory of how the intervention works, why and for whom (2) Developing a preliminary synthesis of findings of included studies (3) Exploring relationships in the data (4) Assessing the robustness of the synthesis.

It was specified in the protocol that if several included studies were adequately homogeneous in their designs, reported measures for outcomes and the studies report enough comparable data, a meta-analysis would be carried out. However, the studies were extremely heterogeneous in study characteristics. Thus, no meta-analyses were included in this review.

Chapter Four: Results

This chapter reports the overall characteristics and results of the final set of research studies found during the search process. After the full-text screening, twelve studies met the criteria for this systematic review.

In this section, the characteristics extracted using the data extraction form (see Appendix B) of the included studies are presented in the form of tables, charts and graphs. The first section of this chapter focuses on the studies' general characteristics. The subsequent section focuses on reporting the individual characteristics of studies, including a narrative summary of each study, the effect sizes of the measured outcomes and the risk of bias / quality assessment for each included study.

Table 4: List of Full References of Included Studies

No.	Full Reference
1	Joseph, H., & Powell, D. (2022). Does a specialist typeface affect how fluently children with and without dyslexia process letters, words, and passages?. Dyslexia, 28(4), 448-470.
2	Malherbe, H. M. (2017). Comparing dyslexie and Times New Roman fonts for reading accuracy and rate in grade 4 learners. University of Johannesburg (South Africa).
3	Duranovic, M., Senka, S., & Babic-Gavric, B. (2018). Influence of increased letter spacing and font type on the reading ability of dyslexic children. Annals of Dyslexia, 68, 218-228.
4	Bachmann, C., & Mengheri, L. (2018). Dyslexia and fonts: Is a specific font useful?. Brain sciences, 8(5), 89.
5	Powell, S. L., & Trice, A. D. (2020). The impact of a specialized font on the reading performance of elementary children with reading disability. Contemporary School Psychology, 24, 34-40.
6	Wery, J. J., & Diliberto, J. A. (2017). The effect of a specialized dyslexia font, OpenDyslexic, on reading rate and accuracy. Annals of dyslexia, 67, 114-127.
7	Ramsey, J. R. (2014). The effects of the font dyslexie on oral reading fluency skills in students grades 8 through 12 with and without reading disabilities. Western Carolina University.
8	Galliussi, J., Perondi, L., Chia, G., Gerbino, W., & Bernardis, P. (2020). Inter-letter spacing, inter-word spacing, and font with dyslexia-friendly features: testing text readability in people with and without dyslexia. Annals of dyslexia, 70(1), 141-152.

9	Damiano, R., Gena, C., & Venturini, G. (2019). Testing web-based solutions for improving reading tasks in dyslexic and neuro-typical users. Multimedia Tools and Applications, 78, 13489-13515.
10	Heiner, B., & Tanja, M. (2017). Contrast and font affect reading speeds of adolescents with and without a need for language-based learning support. Training, Language and Culture, 1(4), 39-55.
11	Kuster, Sanne M., Van Weerdenburg, Marjolijn, Gompel, Marjolein, & Bosman, Anna M. T. (2018). Dyslexie font does not benefit reading in children with or without dyslexia. Annals of Dyslexia, 68(1), 25-42.
12	Md Mahtar, I. F. (2018). Digital text presentation and navigation to support people with dyslexia (Doctoral dissertation, University of York).

4.1 General Characteristics

Table 5 summarizes the general characteristics of the twelve studies included. The subsections that follow provide a deeper review and individual condensation of the data for each characteristic such as participant's age group, geographic location of study and the measured reading skill(s).

Table 5: General Characteristics of Included Studies

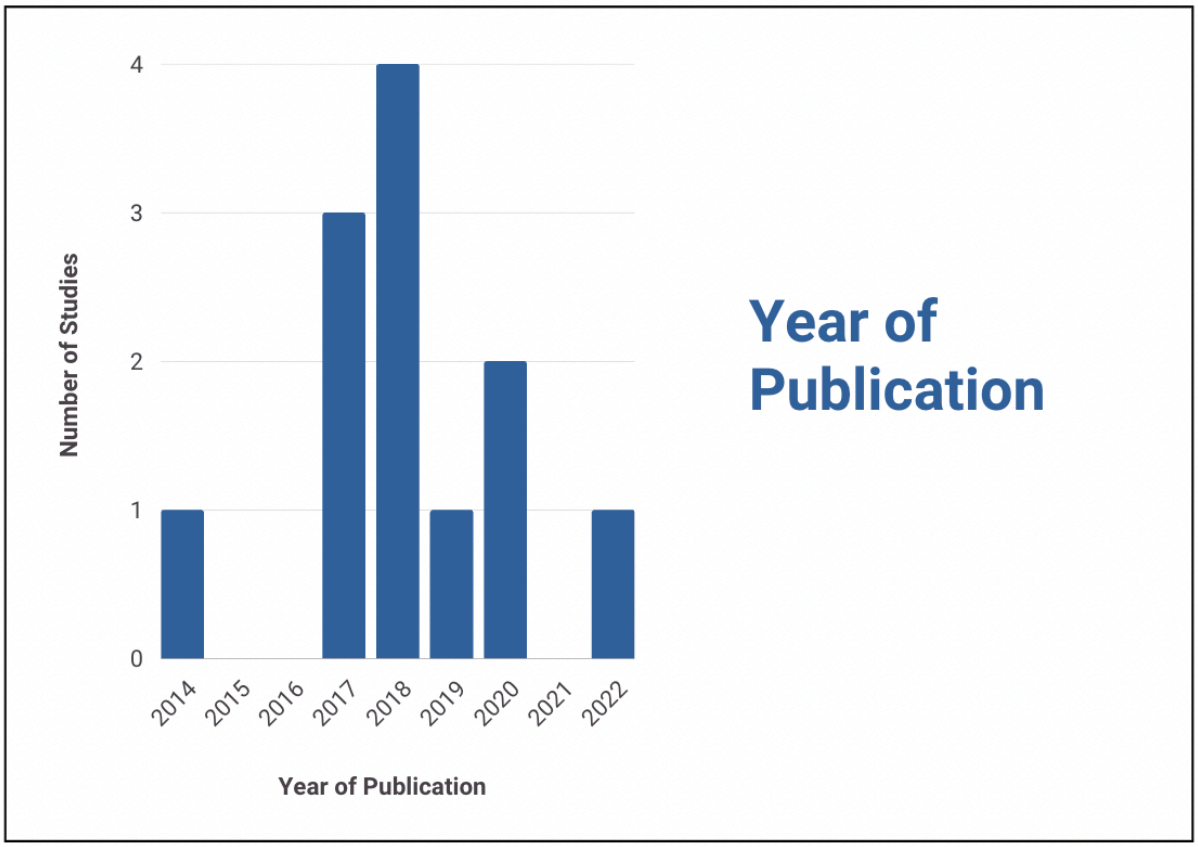
No.	Reference	Database	Participants' Age Group	Sample Size	Location	Intervention Language	Specialized Typeface	Non-specialized Typeface	Measured Reading Skills
1.	Joseph & Powell (2022)	British Education Index	Children (18<)	71	United Kingdom	English	Dyslexie	Calibri	Reading Fluency
2.	Malherbe (2017)	ProQuest Dissertations & Theses Global	Children (18<)	7	South Africa	English	Dyslexie	Times New Roman	Reading rate and accuracy
3.	Duranovic, Senka, & Babic-Gavric (2018)	PsycINFO	Children (18<)	69	Bosnia and Herzegovina	Bosnian (in the Latin script)	Dyslexie	Times New Roman	Reading rate and accuracy
4.	Bachmann & Mengheri (2018)	PubMed	Children (18<)	533	Italy	Italian	EasyReading	Times New Roman	Reading fluency and accuracy
5.	Powell & Trice (2020)	ProQuest Education Collection	Children (18<)	36	United States of America	English	Dyslexie	Arial and Times New Roman	Reading rate and accuracy

6.	Wery & Diliberto (2017)	Web of Science	Children (18<)	12	United States of America	English	OpenDyslexic	Arial and Times New Roman	Reading rate and accuracy
7.	Ramsey (2014)	ProQuest Dissertations & Theses Global	Children (18<)	34	United States of America	English	Dyslexie	Times New Roman	Reading Fluency
8.	Galliussi, Perondi, Chia, Gerbino, & Bernardis (2020)	ProQuest Education Collection	Children (18<)	128	Italy	Italian	EasyReading (Similar)	Verdana Pro Condensed	Reading rate and accuracy
9.	Damiano, Gena, & Venturini (2019)	Scopus	Children (18<)	26	Italy	Italian	EasyReading	Arial	Reading rate
10.	Heiner & Tanja (2017)	Scopus	Children (18<)	47	Germany	English	OpenDyslexic	Times New Roman, Arial and Courier	Reading rate
11.	Kuster, Sanne, Van Weerdenburg, Marjolijn, Gompel, Marjolein, & Bosman, Anna. (2018)	Web of Science	Children (18<)	170	Netherlands	English	Dyslexie	Arial and Times New Roman	Reading rate and accuracy
12.	Md Mahtar (2018)	ProQuest Dissertations & Theses Global	Adults (18+)	36	United Kingdom	English	OpenDyslexic and Lexie Readable	Times New Roman, Georgia, Verdana and Arial	Reading rate

4.1.1 Publication Details of Studies

Figure 1 illustrates the number of studies published over the years. Although the search criteria included studies from much earlier (1957) since certain fonts created then had dyslexia-friendly features, specialized dyslexia-friendly fonts became prominent only recently (Boer's Dyslexie font was created in 2010). The included studies in this review ranged from the earliest being published in 2014[7] to the latest being published in 2022[1] and the majority of studies being published in 2017[2,6,10] and 2018[3,4,11,12].

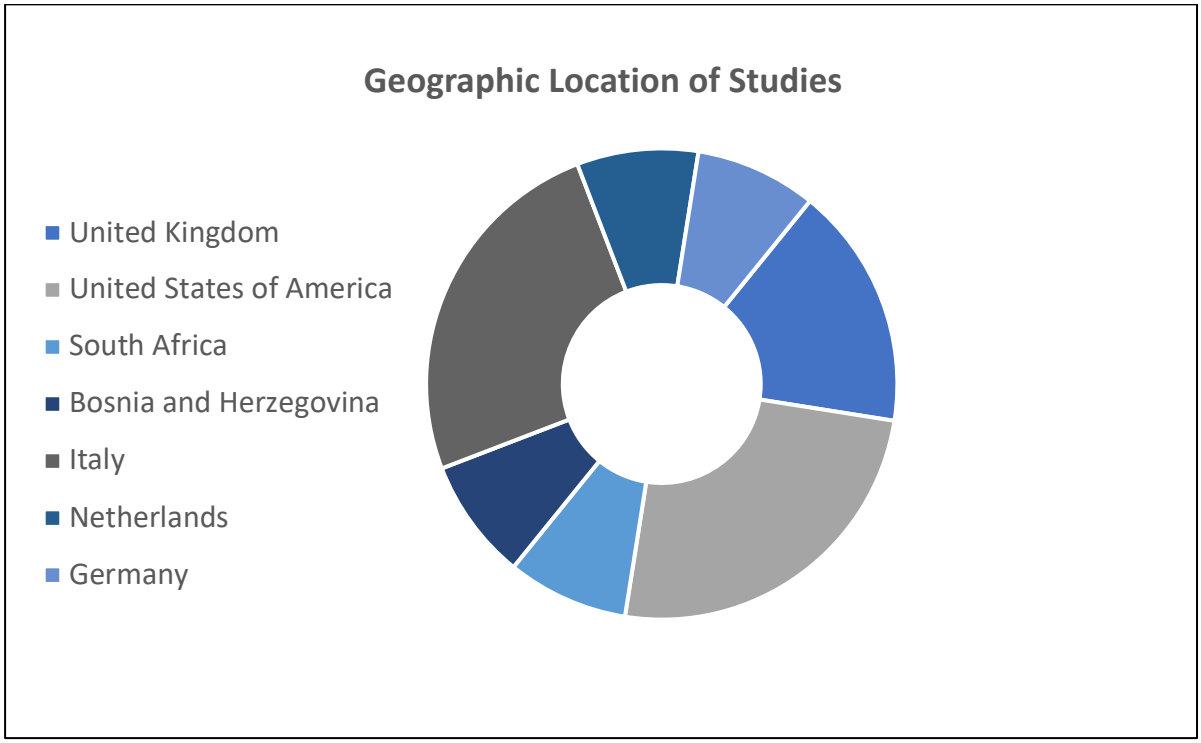
Figure 2: Year of Publication of Studies



4.1.2 Geographic Location of Studies

The geographic locations of the included studies are summarized in Figure 2. The studies were conducted in a number of countries including The United Kingdom, The United States of America, South Africa, Bosnia and Herzegovina, Italy, the Netherlands and Germany. The United States[5,6,7] (25%) and Italy[4,8,9] (25%) had the most number of studies followed by the United Kingdom[1,12] (16.7%) and the rest of the countries had one study each – South Africa[2] (8.3%), Bosnia and Herzegovina[3] (8.3%), the Netherlands[11] (8.3%) and Germany[10] (8.3%).

Figure 3: Geographic Location of Included Studies



4.1.3 Participants

The number of participants varied across the twelve included studies. Figure 3 illustrates the sample size across the studies in a box plot. The outlier was the study with 533 participants[4], by far the study with the most participants. The smallest sample size had 7 participants[2] and the median of the values was 41.5. Figure 4 summarizes the age groups of the participants, a majority of the studies – 91.7% (11 studies) included participants that were children (below 18) and one study[12] – 8.3% had adult participants (above 18).

Figure 4: Number of Participants in Studies

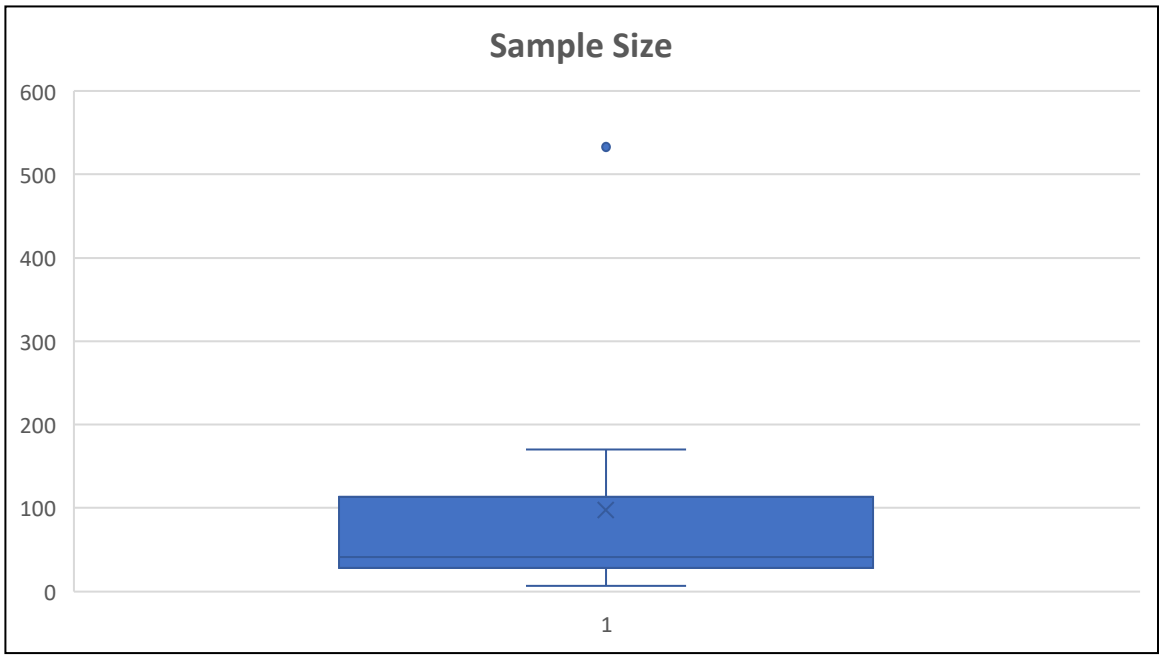
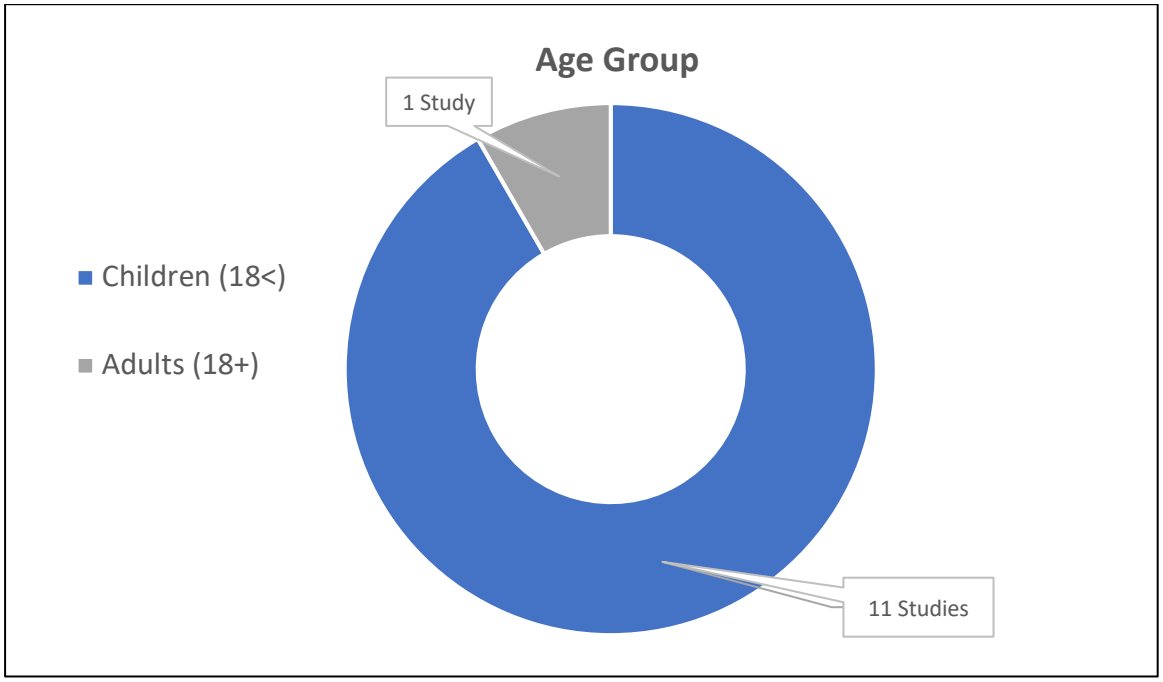


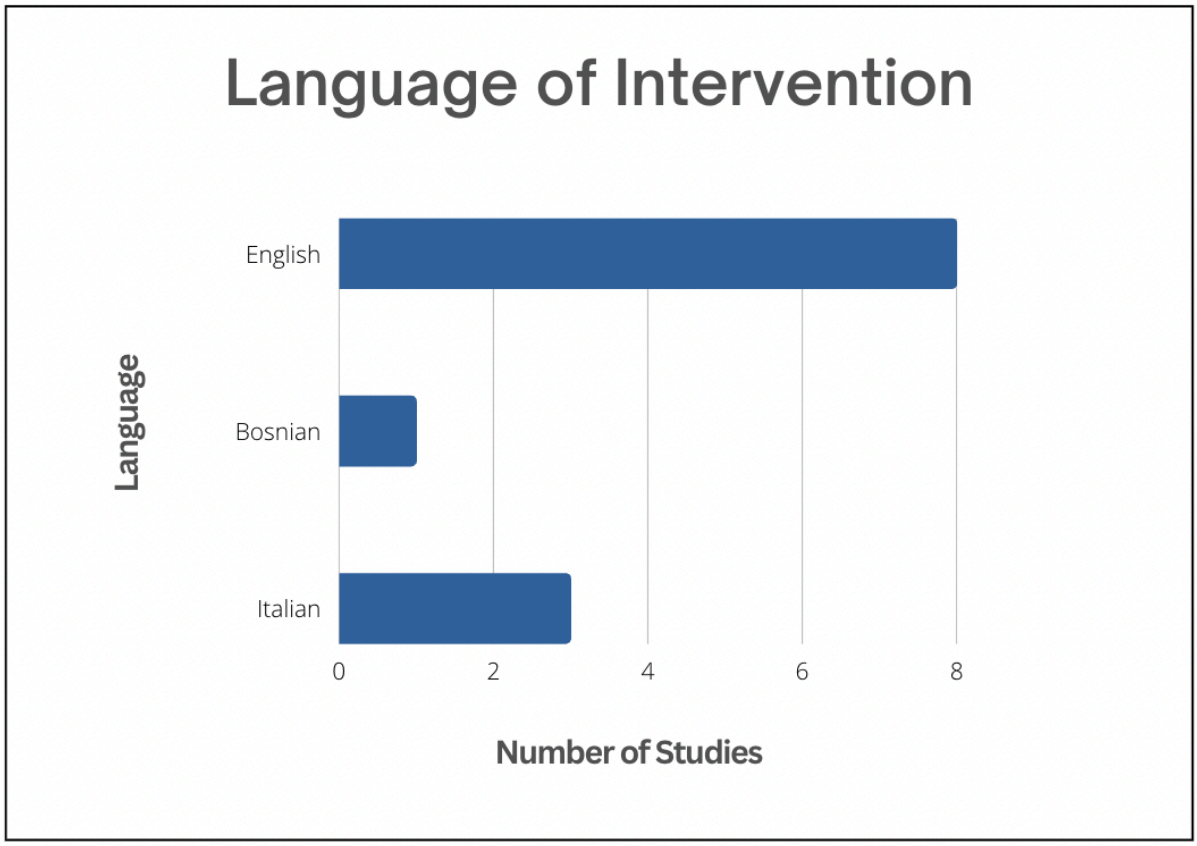
Figure 5: Age of Participants



4.1.4 Language of Intervention

Figure 5 depicts the languages used for the experiments in the included studies. The most recurrent language was English, used in the interventions of 8 studies[1,2,5,6,7,10,11,12] or 66.7% of the papers. 3 studies[4,8,9] or 25% used Italian as the experimental language and 1 study[3] or 8.3% used Bosnian (in the Latin script).

Figure 6: Language of Intervention in Studies



4.1.5 Typefaces in Studies

This subsection elucidates the typefaces used in the experiments of the included studies. Figure 6 illustrates the dyslexia-specialized typefaces used in the studies. The Dyslexie typeface was used most frequently, with 50% of studies[1,2,3,5,7,11] (6 studies) using the font in their experiments, 25% of the studies[4,8,9] (3 studies) using the EasyReading font, 17% of the studies[6,10] using OpenDyslexic (2 studies) and 8% or 1 study[12] used two specialized typefaces in the experiment – both OpenDyslexic and the Lexie Readable fonts.

Figure 7 summarizes the non-specialized typefaces used as part of the control variable in the studies. The most commonly used typeface was Times New Roman which was part of 42% of the included studies[2,3,4,7], 3 studies[5,6,11] – 25% – used both Arial and Times New Roman fonts in their intervention, while only the use of Arial was done by 1 study[9] – 8%. The rest of the fonts were also used in only one study – Calibri[1], Verdana Pro Condensed[8] and a study[10] using Times New Roman, Arial and Courier and another study[12] using a combination of Times New Roman, Georgia, Verdana and Arial.

Table 6 exemplifies the four specialized typefaces used in the experiments of included studies – Dyslexie, EasyReading, OpenDyslexic and Lexie Readable. Similarly, Table 7 includes examples of the non-specialized typefaces used in the studies – Arial, Times New Roman, Calibri, Courier, Verdana Pro Condensed, and Georgia.

Figure 7: Dyslexia Specialized Typefaces in Studies

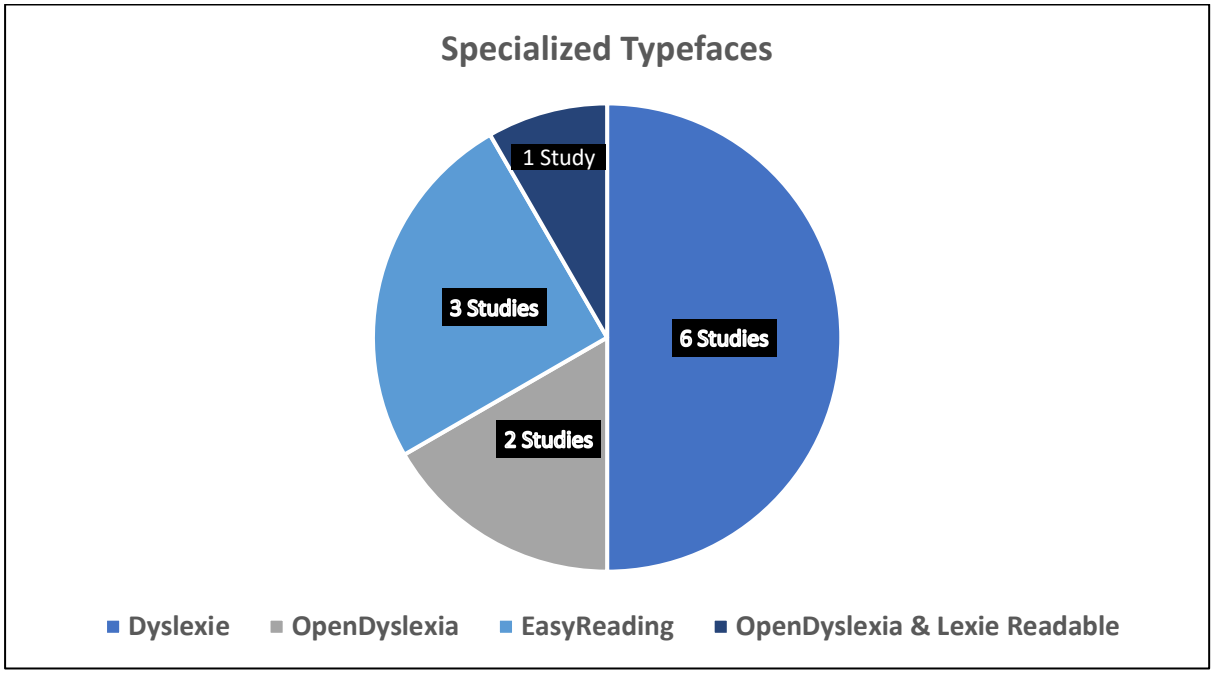


Figure 8: Non-Specialized Typefaces in Studies

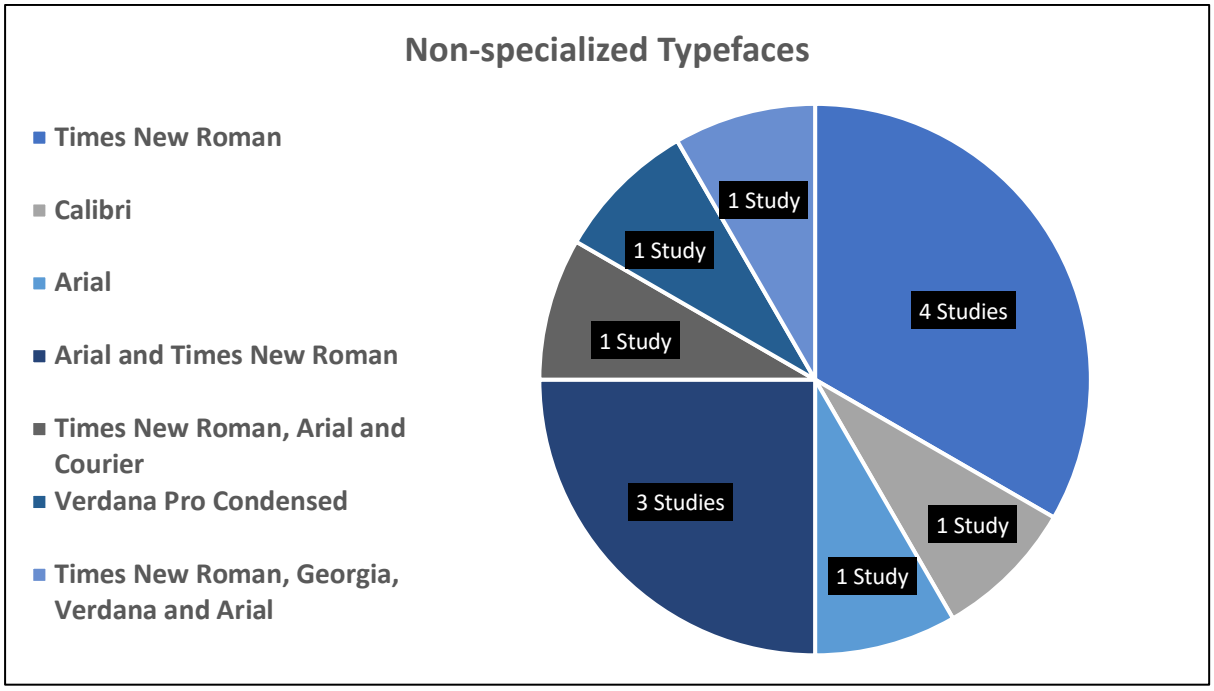


Table 6: Example of Dyslexia Specialized Typefaces Used in Studies

Typeface Name	Typeface Example
Dyslexie	dyslexie
Lexie Readable	Lexie Readable
EasyReading	EasyReading
OpenDyslexic	OpenDyslexic
Sylexiad	Sylexiad

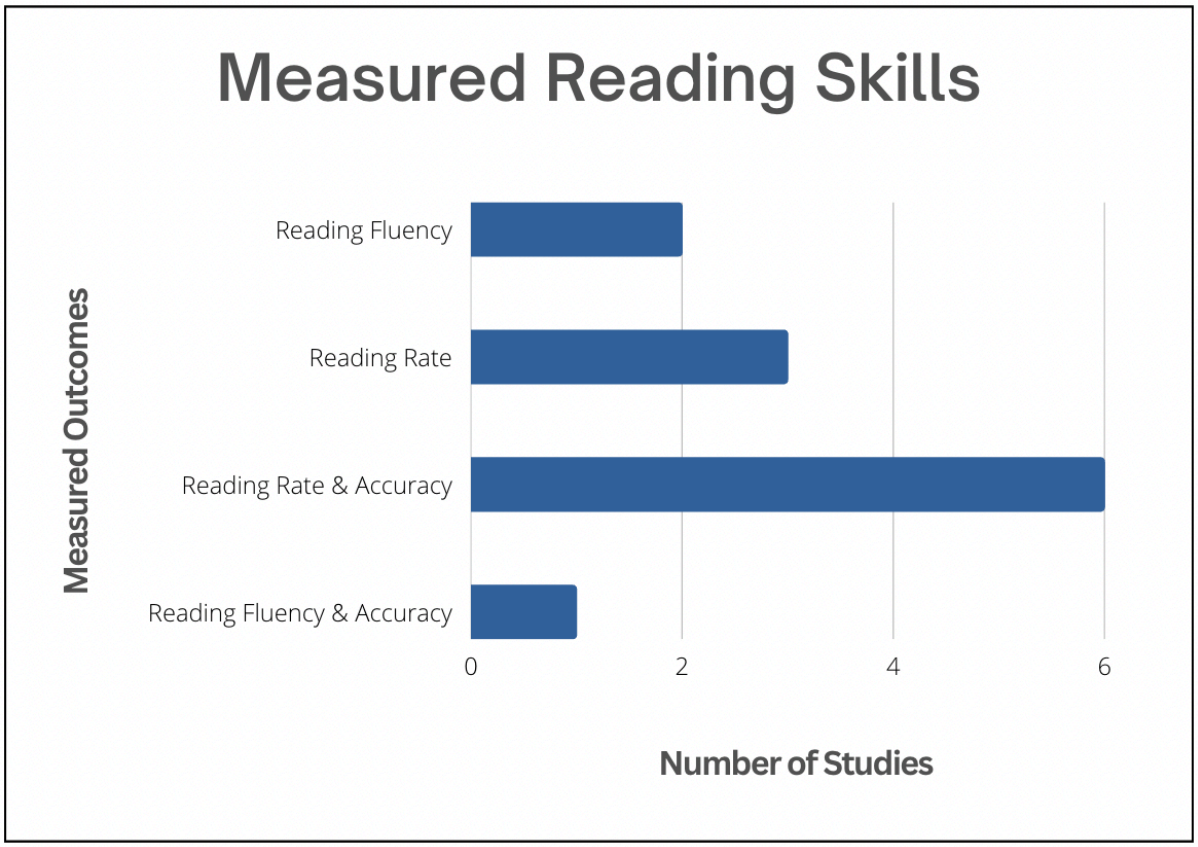
Table 7: Example of Non-Specialized Typefaces Used in Studies

Font Name	Font Example
Arial	Arial Typeface
Times New Roman	Times New Roman Typeface
Calibri	Calibri Typeface
Courier	Courier Typeface
Verdana Pro Condensed	Verdana Pro Condensed Typeface
Georgia	Georgia Typeface

4.1.6 Measured Reading Skill(s)

Figure 7 shows the reading skills that were measured in the studies. A combination of reading rate and accuracy was measured the most, with 50% (6) studies[2,3,5,6,8,11]. A measure of just reading rate was done by 25% (3) of studies[9,10,12], a measure of only reading fluency was done by 17% (2) of studies[1,7] and a measure of both reading fluency and accuracy was done by 8% (1) of studies[4].

Figure 9: Reading Skills Measured in Studies



4.2 Study findings

The section above focused on the general characteristics of the included studies dealing with the sample size, participants, outcomes, typefaces used and geographic location of studies. Moving beyond the general characteristics of the studies, the following section focuses on each study individually. This section includes the narrative summaries of individual studies, the size of effects and significance of outcomes reported in each study (see table 8) and the risk of bias/quality assessment of individual studies (see table 9 for a summary of rankings and Appendix E for the completed form for each study).

4.2.1 Narrative Summaries of Individual Studies

According to Popay et al. (2006), it is considered best practice to include a narrative description of the studies included in a systematic review to provide contextual understanding. For each of the twelve included studies in this systematic review, a brief narrative summary of the results, study design and conclusion are provided below.

1. Joseph & Powell (2022) looked at the possible advantages of Dyslexie font over standard typefaces in identifying letters, reading words, and reading passages. A rapid letter naming task, a word reading efficiency task, and a passage reading task were completed by 71 children aged 8 to 12, 37 of whom had a dyslexia diagnosis. The results revealed no changes in word or passage reading between the two typefaces in a within-participants study design, although letter naming appeared to be more fluent for all participants when letters were presented in the Dyslexie typeface rather than Calibri font. The findings indicate that a font with different letters may be useful for letter identification, and that an early childhood intervention in which children are taught letters in a specialised typeface is worth considering. According to the study's findings, fonts with more distinct letters, such as Dyslexie, may be effective for letter recognition.

2. Malherbe (2017) examined if Dyslexie typeface affects reading rate and accuracy in different groups. This quasi-experimental study compares the reading rates and accuracy of grade 4 students using two text fonts. The study uses the Neale Analysis of Reading Ability Second Edition (NARA-II) test in Times New Roman and Dyslexie fonts. Seven students in grade 4 who were either dyslexic readers, struggling readers, or on-par readers participated in this study to test these assertions in a South African environment and participants were not randomly assigned to groups. The study compared the reading pace and accuracy of the Times New Roman typeface with the Dyslexie font using three statistical approaches to evaluate any potential effects - the Wilcoxon matched-pairs signed rank test, between groups analysis and individual analysis. The NARA-II test results were also utilised to establish the participants' reading proficiency. Results for reading rate and accuracy were not significant at $p \leq 0.05$.

3. Duranovic, Senka, & Babic-Gavric (2018) looked at how spacing affects the reading of different typefaces for children with and without dyslexia. The results revealed that only spacing plays a role in improving dyslexic persons' reading abilities, as there was no difference between Dyslexie and the Times New Roman interspaced typeface. Increased interletter space mitigated the negative effects of unfriendly typefaces - Times New Roman (TNR) Italic and Curlz MT. The research included 69 Bosnian and Herzegovina children (mean age = 10.77 years), all native Bosnian speakers, who were evaluated individually for reading ability and speed. The study used recording and timed reading to examine the reading speed and mistakes of dyslexic individuals. The results indicated that among dyslexics,

Dyslexie had the lowest mean number of reading errors and reading time, whereas TNR default interspaced typeface had the largest number of errors. The results of a repeated measures ANOVA for dyslexic and typically developing readers revealed significant differences in reading accuracy and speed. Dyslexics read less accurately and slower in properly spaced sentences in TNR than in substantially spaced sentences in italic and Curlz MT fonts, which are not recommended for dyslexic readers. Dyslexics fared better in the Dyslexie typeface than in the TNR interspaced font, but not in the TNR Italic or Curlz MT fonts, according to the research. The findings suggest that increasing letter space in text may aid dyslexics, as crowding reduces reading speed and accuracy, particularly in dyslexics.

4. Bachmann & Mengheri (2018) evaluated the impact of the EasyReading typeface on children in a fourth-grade primary school. 533 fourth-grade children participated in the study. The Raven's Coloured Progressive Matrices CPM were employed in the study to evaluate intellectual developmental problems and nonverbal IQ. The outcomes demonstrated that the EasyReading version greatly increased reading fluency and accuracy. Normal readers read at a rate of 3.73 syllables per second on average, compared to 2.67 syllables per second for dyslexics. Children with dyslexia did so with an average fluency of 2.39 syllables per second and a performance that was inside the acceptable performance range. The study emphasises EasyReading's potential as a compensatory tool for dyslexic readers and a font that makes reading easier for all readers.

The EasyReading font had a beneficial effect on reading fluency across all reading tests and the results showed that reading fluency significantly improved within all groups when the text was presented in the EasyReading version. Both dyslexic and non-dyslexic participants read more accurately and fluently while using the EasyReading version.

5. Powell & Trice (2020) investigated the reading rate and accuracy in 36 fourth and fifth-grade children with specific learning difficulties and compared three different typefaces: Dyslexie, Times New Roman, and Arial. The participants were asked to read three short stories out loud using each font. The study recorded the reading time and the number of mistakes made, and the findings indicated no significant differences in terms of reading errors among the three typefaces.

All three fonts produced the same number of accurate comprehension questions, suggesting that Dyslexie did not have a notable impact on reading accuracy compared to Arial and Times New Roman. The study also found no differences in reading rates or accuracy across the different fonts. Dyslexie

did not enhance readers' ability to read accurately, comprehend the text, or increase reading speed when compared to Arial and Times New Roman. The study's results showed consistent reading fluency across all three typefaces, hinting at the possibility that using a more readable font could potentially improve readers' retention of information. However, the advantages of Dyslexie seemed to diminish when larger and more widely spaced-out typefaces were employed, suggesting that the benefits are not solely attributed to the font's size and spacing. The study found no evidence to justify the usage of Dyslexie above other common fonts that were made larger and with more inter-word and inter-line spacing.

6. Wery & Diliberto (2017) employed a single-subject alternating treatment design to examine the impact of OpenDyslexic (OD), on the reading rate and accuracy of elementary students with dyslexia. They compared Arial and Times New Roman in three reading tasks: letter naming, word reading, and nonsense word reading. The results did not reveal any noticeable improvement in reading rate or accuracy for individual students with dyslexia or the group as a whole. On all three reading tasks, the OD font showed significantly worse reading accuracy when contrasted with both Arial and TNR.

The effect sizes, denoted by negative values, and the overlapping confidence intervals demonstrated substantial differences, favouring Arial and TNR. Consequently, these fonts yielded better reading accuracy outcomes in comparison to the OD font for the tasks under examination.

7. Ramsey (2014) investigated the effectiveness of the "Dyslexie" font in improving oral reading fluency (ORF) for individuals with reading disabilities. The research involved two groups: Group 1 (No RD Group) with individuals without reading disabilities and Group 2 (RD Group) with individuals diagnosed with reading disabilities. The study presented both groups with six passages, three in the commonly used Times New Roman (TNR) font and three in the Dyslexie font, and assessed their ORF scores. The study had a sample size of 34 students in grades 8 through 12.

However, the study faced challenges due to a lower-than-expected number of participants in the RD Group, making statistical analysis difficult. Nevertheless, the results indicate that both groups generally achieved higher median ORF scores when reading in the TNR font than in the Dyslexie font. Some participants in the RD Group did perform better with the Dyslexie font, highlighting the need for individualized font selection. Moreover, participants' preferences for a specific font did not always correlate with their actual reading abilities. The study emphasizes the lack of conclusive evidence regarding the effectiveness of specialized fonts for reading disabilities.

8. Galliussi, Perondi, Chia, Gerbino, & Bernardis (2020) aimed to investigate the impact of dyslexia-friendly fonts on reading accuracy and speed in children with and without dyslexia. Two groups, each consisting of 64 children each, with and without dyslexia, participated in the study. The study used 8 different texts with various typographical conditions. The results showed that the letterform (standard vs. dyslexia-friendly) did not significantly affect reading performance for both groups. However, the spacing between letters and words had some impact. An increase in inter-letter spacing without adequate inter-word spacing led to decreased reading speed in both groups. Additionally, for children with dyslexia, the dyslexia-friendly letterform combined with increased inter-word spacing resulted in faster reading, while the same combination did not significantly affect typical readers. Reading accuracy was not influenced by the letterform or spacing for either group. Overall, the study suggests that dyslexia-friendly fonts may not consistently improve reading accuracy and speed and that spacing plays a role in legibility.

9. Damiano, Gena, & Venturini (2019) aimed to improve web accessibility for users with dyslexia by assessing the readability of web pages and proposing new criteria in addition to the Web Content Accessibility Guidelines (WCAG). The researchers designed a test for students diagnosed with dyslexia and compared two websites, Site A (complying with WCAG) and Site B (adding specific solutions) using a single factor within-subjects design. The study identified nine features that could enhance accessibility: font type and size, column width, line spacing, paragraph spacing, title and menus, text and background colour, glossary, and layout. The focus for this systematic review remained on the area concerned with the typeface effect.

A repeated measures ANOVA with a Greenhouse-Geisser correction determined that the differences between the two conditions were not significant. The findings showed that users with dyslexia preferred the modifications in Site B, especially the use of the EasyReading typeface and larger font size. However, the changes in Site B did not significantly affect the reading time or text comprehension.

10. Heiner & Tanja (2017) investigated the effects of different fonts on the reading speed of individuals. They found that the OpenDyslexic font was read significantly faster than other fonts, with participants taking the least time to read text displayed in this font. The sample consisted of 47 participants with different learning support statuses (with and without dyslexia) and age groups

ranging from students in the 6th grade to students in the 12th grade. The study used eye-tracking technology to collect data and the OpenDyslexic font was found to be particularly effective for individuals requiring learning support, while participants without learning disabilities also read it faster. The study suggests that using OpenDyslexic in inclusive settings may improve the performance of both supported and non-supported students.

11. Kuster, Sanne, Van Weerdenburg, Marjolijn, Gompel, Marjolein, & Bosman (2018) conducted two experiments testing the claim that the dyslexia-specific font Dyslexie improves reading proficiency in children with and without dyslexia in a randomized controlled trial. In Experiment 1, 170 dyslexic children read Arial type material faster and more accurately than Dyslexie font and the main effect of Time (rate of reading) was significant. In Experiment 2, dyslexic (n = 102) and non-dyslexic (n = 45) children read word lists in Dyslexie, Arial, and Times New Roman fonts. Dyslexie typeface did not improve reading speed for both groups. Moreover, Arial and Times New Roman were preferred over Dyslexie, but reading performance was unrelated for both groups. The conclusion was that these experiments demonstrate that the Dyslexie typeface neither helps nor impedes dyslexic children's reading.

12. Md Mahtar (2018) investigated the aspects that could make online materials accessible for people with dyslexia. One of these elements was font type and therefore, study 1 focused on the effects of typefaces and font size. The study employed several typefaces - for the serif level, Times New Roman and Georgia were used; for the sans serif level, Verdana and Arial were used; and for the dyslexia-optimised level, OpenDyslexic and Lexie Readable. The study compared English-speaking adults with and without dyslexia on eye gazing, performance, preferences, and views. The study used a mixed-participants design, with participant groups as the between-participants variable and typeface, typeface examples, and font size as the within-participant variable. A total of 36 participants were divided into groups of 12 - non-dyslexic, mild dyslexic, and moderate dyslexic groups based on a diagnostic checklist. The study used eye gaze tracking to measure participants' reading behaviour and performance. The dependent variables were categorized into three types: eye gaze behaviour (number of fixations and average fixation duration), reading performance (reading time and comprehension scores), and participants' preferences and opinions (rating of ease of reading, preferred typeface and font size combination). The study analysed the data using parametric tests (ANOVA) for normally distributed and heterogeneous data. Non-normal data were analysed using

non-parametric tests. The conclusion was that sans serif fonts and greater line spacing were favoured while Dyslexic and non-dyslexic participants felt differently about dyslexia-optimized typefaces like OpenDyslexic.

4.2.2 Effect Size

Table 8 shows the reported results of the outcomes in individual studies.

Table 8: Effect Sizes by Study

Study	Outcome	Reported Effect
Joseph & Powell (2022)	Reading Fluency	nonword reading: no effect of typeface, $F(1, 68) = .844$, $p = .32$, $\eta p2 = 0.012$, no effect of group, $F(1, 68) = 1.83$, $p = .18$, $\eta p2 = 0.0003$, and no interaction, $F(1, 68) = .018$, $p = .89$, $\eta p2 = 0.0003$.
Malherbe (2017)	Reading rate and accuracy	Sample size very small (7 participants). Effect size not mentioned.
Duranovic, Senka, & Babic-Gavric (2018)	Reading rate and accuracy	Rate: the main effect of font condition was significant for dyslexics, $F(4.88) = 6.26$, $p < 0.001$. Accuracy: the main effect of font condition was significant, $F(4.88)=5.28$, $p<0.001$.
Bachmann & Mengheri (2018)	Reading fluency and accuracy	Fluency: in the word task, the average reading fluency was 3.03 in the original version and went up to 3.33 ($t(532) = -18.14$, $p < 0.001$) in the EasyReading one, while in the non-word task, it increased from 1.86 to 2.04 ($t(532) = -10.37$, $p < 0.001$)

		Accuracy: In the word task, students' mistakes decreased from 5.49 on average (using the original format) to 4.14 ($t(532) = 9.56$, $p < 0.001$) in the EasyReading format, while in the non-word task, mistakes were reduced from 7.72 to 6.49 ($t(532) = 8.41$, $p < 0.001$)
Powell & Trice (2020)	Reading rate and accuracy	Stories (3) differed significantly in the number of errors participants made while reading, $F(2, 70) = 5.623$, $p = .005$, $\eta 2 = .138$. Rate: no significant differences in the amount of time in seconds that participants took to read stories in Arial ($M = 202.11$, $SD = 76.48$), Times New Roman ($M = 213.44$, $SD = 96.0$), or Dyslexie ($M = 211.61$, $SD = 116.22$), $F(1.307, 45.729) = 1.031$, $p = .336$ Accuracy: no significant differences in errors for Arial ($M = 22.91$, $SD = 13.22$), Times New Roman ($M = 22.13$, $SD = 11.64$), or Dyslexie ($M = 21.22$, $SD = 8.6$), $F(2, 70) = .522$, $p = .298$ (one-tailed)
Wery & Diliberto (2017)	Reading rate and accuracy	Rate: OD produced negative results, or decreased students' outcomes compared to both Arial and TNR, on all three reading tasks (i.e., letter naming, word decoding, nonsense word decoding). Accuracy: OD produced negative results compared to both Arial and TNR, on all three reading tasks.

Ramsey (2014)	Reading Fluency	Due to the low number of participants and the wide range of characteristics within the Reading Disability (RD) Group, statistical analysis of the data was not reported and could not be conducted.
Galliussi, Perondi, Chia, Gerbino, & Bernardis (2020)	Reading rate and accuracy	Rate: Dyslexics' reading speed (M = 2.91 syll/s, SD = 0.82) was approximately 35% lower than that of controls (M = 4.49 syll/s, SD = 0.69) ($F(1,122) = 139$, $p < 0.001$, $\eta p2 = 0.533$). Accuracy: dyslexics' accuracy level (M = 1.50, SD = 0.33) was approximately 27.5% lower than that of controls (M = 1.82, SD = 0.28) ($F(1,122) = 34.9$, $p < 0.001$, $\eta p2 = 0.222$). The model did not show a main effect of letterform ($F(1,875) = 0.496$, $p = 0.482$).
Damiano, Gena, & Venturini(2019)	Reading rate	The reading task performed in Site A (without dyslexia specialized typeface) obtained an average time to be completed of 11.5 min (SD = 6.36), while the average time in Site B (with dyslexia specialized typeface) was 10.92 min (SD = 5.09). A repeated measures ANOVA with a Greenhouse-Geisser correction determined that differences between the two conditions were not significant ($F(1,25) = 5,15$, $p = .48$). The observed and expected distributions differ significantly for Arial/Easy Reading and no preferences.

Heiner & Tanja (2017)	Reading rate	Not enough data provided to calculate effect size.
Kuster, Sanne, Van Weerdenburg, Marjolijn, Gompel, Marjolein, & Bosman, Anna. (2018)	Reading rate and accuracy	Rate: the main effect of Time (i.e., repetition) on reading speed was highly significant, $F(1, 167) = 33.86$, $p < .0001$, $\eta 2 = 17$. Accuracy: number of errors did not decrease significantly after repeated reading $F(1, 167) = 1.08$, $p < .30$, $\eta 2 = .006$.
Md Mahtar (2018)	Reading rate	There was a significant two-way interaction between Typeface and Font Size with moderate to large effect size, $F(2, 66) = 4.15$, $p = .02$, $hp2 = .11$

4.3 Risk of Bias

The risk of bias was measured using the MMAT assessment tool and a global score was given for each study (see Chapter Three: Methodology for more details on how the global score was calculated). The filled-out MMAT form for each study that was used to calculate the global rating can be found in Appendix E.

This section summarizes the studies' design and risk of bias rating. Table 9 shows the category of design and the risk of bias for each study. Out of the 12 included studies, 5 studies were ranked as moderate, 3 as weak and 4 as strong. The MMAT has five subsections of questions per type of study design. This means that, in the MMAT form, the studies received a moderate rating when the answers to the subsection questions were no more than one 'no' and at least three 'yes'. For the weak studies, it was two or more 'no' answers or fewer than three 'yes.' And for studies ranked as strong, none of the criteria was evaluated as 'no,' and at least four as 'yes'.

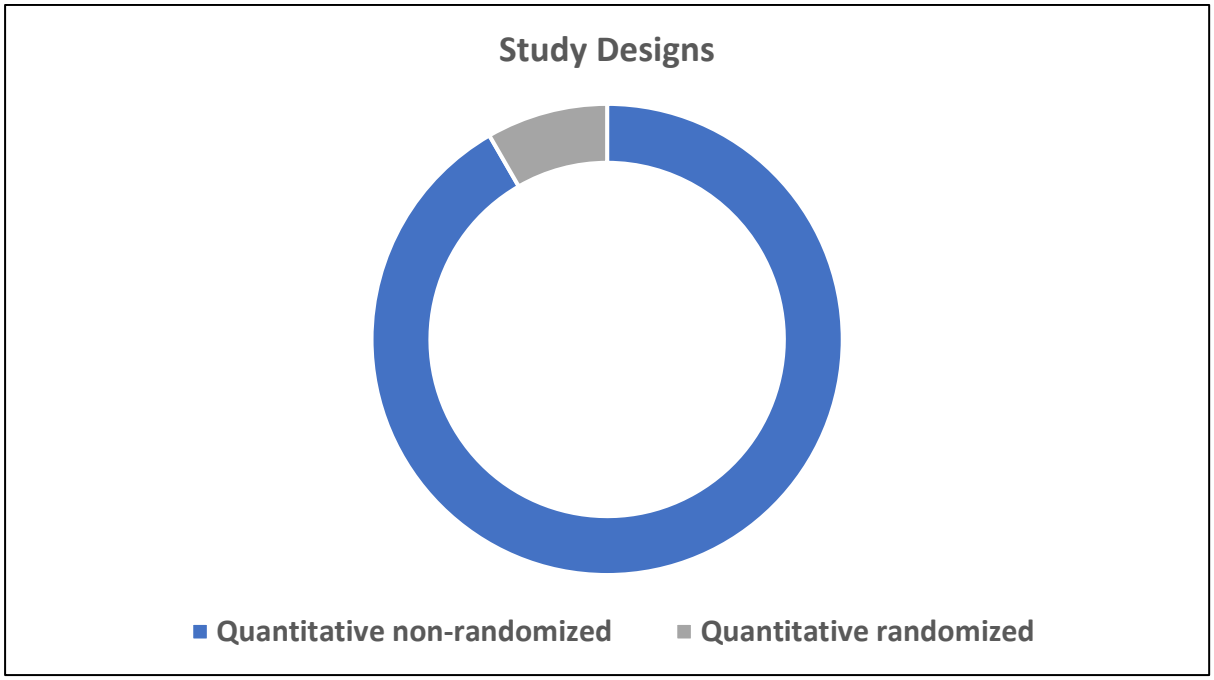
Table 9: Risk of Bias Rating for Individual Studies

Study	Category of Study Designs	Global RoB Rating
Joseph & Powell (2022)	Quantitative non-randomized	Strong
Malherbe (2017)	Quantitative non-randomized	Weak
Duranovic, et al. (2018)	Quantitative non-randomized	Strong
Bachmann & Mengheri (2018)	Quantitative non-randomized	Moderate
Powell & Trice (2020)	Quantitative non-randomized	Moderate
Wery & Diliberto (2017)	Quantitative non-randomized	Moderate
Ramsey (2014)	Quantitative non-randomized	Weak
Galliussi, et al.(2020)	Quantitative non-randomized	Strong
Damiano, et al. (2019)	Quantitative non-randomized	Weak
Heiner & Tanja (2017)	Quantitative non-randomized	Moderate
Kuster, et al. (2018)	Quantitative randomized controlled trials	Moderate
Md Mahtar (2018)	Quantitative non-randomized	Strong

4.3.1 Category of Study Design

Based on Table 9, Figure 8 shows the distribution of the study design categories. The MMAT can be used for Qualitative, Quantitative randomized controlled trials, Quantitative non-randomized, Quantitative descriptive and Mixed methods studies. In this review, 92% or 11 studies were Quantitative non-randomized and 8% or 1 study was Quantitative randomized controlled trials.

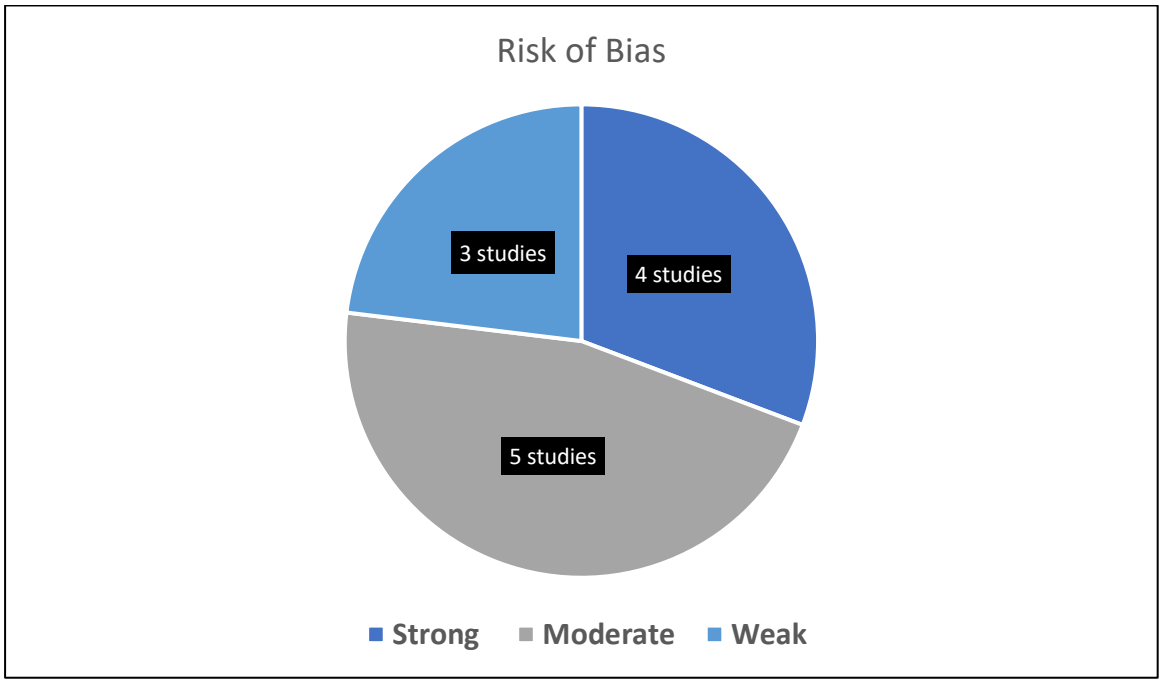
Figure 10: Study Designs of Included Studies



4.3.2 Global Risk of Bias Ranking

Figure 9 shows the Risk of Bias ranking after completing the MMAT quality assessment check for each of the twelve studies. It was found that 31% of studies or 4 studies were ranked as "strong", 46% or 5 studies were ranked as "moderate" and 23% or 3 studies were ranked as "weak" (for more information on how the studies were ranked, see section 3.9 in the Methodology chapter).

Figure 11: Global Risk of Bias Ranking



Chapter Five: Discussion

This systematic review was guided by two main research questions. These questions are answered below in reference to the data and results collected from the twelve included studies (Chapter Four: Results).

5.1 Research Question 1- What is the extent and nature of empirical research investigating the role of specialized typefaces on dyslexics' reading outcomes?

The data extraction process and the results section (Chapter Four) revealed the extent and nature of the empirical research investigating the role of specialized typefaces on dyslexics' reading outcomes. All the studies included in this review were published from the year 2014 up to 2022 and the majority of the studies used the English language for their interventions.

In terms of the age groups found in the literature, only 1 study (Md Mahtar, 2018) in this review experimented with the effects of specialized typefaces on adults while the rest of the studies focused on children in school. Although there is a specialized typeface – Sylexiad – made specifically for adults, there was no empirical research found on it. The lack of research on adults with dyslexia suggests that the main audience for these typefaces is children in school. It could also mean that most of these typefaces were developed with the intention of being integrated into educational systems.

A recurring problem with many of the included studies was the sample size. These studies were treated as pre-tests with smaller sample sizes and stated that further research with a larger sample size is needed to draw definitive conclusions. Malherbe (2017) had a sample size of only 7 participants, Wery & Diliberto (2017) had 12 participants and Damiano, Gena, & Venturini (2019) has 26 participants in their sample. The small sample size in many studies made it difficult to calculate the effect sizes and gauge significance levels.

As for the language of intervention, only 1 study, Duranovic, Senka, & Babic-Gavric (2018), experimented with Bosnian, three studies (Bachmann & Mengheri, 2018); (Galliussi, Perondi, Chia, Gerbino, & Bernardis, 2020) and (Damiano, Gena, & Venturini, 2019) experimented with Italian and the rest used English. Italian was used in some studies since the EasyReading font was created specifically for Italian. Most of the other existing specialized typefaces, however, were made with the English alphabet as intent.

Lastly, the typefaces used in the experiments were mostly the same or a combination of similar fonts. A majority of included studies (Joseph & Powell, 2022); (Malherbe, 2017); (Duranovic, Senka, & Babic-

Gavric, 2018); (Powell & Trice, 2020); (Ramsey, 2014); (Kuster, et al., 2018) experimented with the Dyslexie typeface since the font did receive much attention and praise (Wainwright, 2014). The specialized typefaces were usually compared with widely used typefaces like Times New Roman and Arial. As mentioned in the literature review, dyslexic individuals have shown a preference towards sans-serif fonts, so it was interesting to see that although Times New Roman, a serif font, was used in many studies, there was still no significant preference towards the dyslexia-specialized sans-serif typefaces.

5.2 Research Question 2- What is the extent of the evidence of the effectiveness of specialized typefaces on the reading performance of dyslexic readers?

The empirical evidence that was reported in this review raises important considerations regarding the efficiency of dyslexia-specialised typefaces on reading rate and accuracy. The findings show that the overall usefulness of these typefaces might be more complex than was previously anticipated, despite the fact that they have received attention and interest as potential tools that could assist those with dyslexia.

A significant aspect to consider is the lack of consistent and significant positive effect of the Dyslexie, OpenDyslexic or EasyReading typeface on reading rate or accuracy observed across the included studies.

RQ 2a) What impact do specialized typefaces have on reading rate?

Most studies looked at both reading rate and accuracy such as (Wery & Diliberto, 2017) and (Powell & Trice, 2020) some studies isolated reading rate only (Damiano, Gena, & Venturini, 2019) and (Heiner & Tanja, 2017). None of the studies found a significant positive outcome in reading speed using a dyslexia-specialized typeface. Even looking at the different specialized typefaces involved – Dyslexie, OpenDyslexia and EasyReading – none of them showed an increase in reading speed in dyslexic individuals.

RQ 2b) What impact do specialized typefaces have on reading accuracy?

The same results as reading rate were found for reading accuracy and fluency. Ramsey (2014) and Joseph & Powell (2022) were the two studies that looked only at accuracy/fluency. Both studies

concluded that there were no differences in errors made comparing the specialized typefaces and the non-specialized ones. Other studies mentioned above that reported on reading rate and accuracy also showed that the specialized typefaces had no significant results in improved accuracy. As for fluency, Bachmann & Mengheri (2018), found that the EasyReading typeface had a statistically significant benefit on dyslexic readers, concluding that the specialized typeface was effective. It is worth noting that Bachmann's study was the only one with a large sample size of 533 participants. Most of the other studies were treated as pilot studies which would potentially explain the lack of statistical significance in their findings.

5.3 Other Findings

Many studies had a main focus on other aspects – such as spacing (Duranovic, Senka, & Babic-Gavric, 2018); (Galliussi, Perondi, Chia, Gerbino, & Bernardis, 2020), web design (Damiano, Gena, & Venturini, 2019); (Md Mahtar, 2018) and colour contrast (Heiner & Tanja, 2017). This focus on other elements of the typeface effect overshadowed the effects of the dyslexia-specialized typeface and was not given enough attention.

Some studies suggested that the enhanced spacing between letters and lines turned out to be more beneficial than the use of a specialized typeface itself. When comparing Dyslexie to Times New Roman on reading rate and accuracy, Duranovic, Senka, & Babic-Gavric (2018), found that spacing the letters and lines in a text resulted in improved reading performance among dyslexic readers. Similarly, another paper (Marinus, Eva, Mostard, Segers, Schubert, Madelaine, and Wheldall, 2016) came to the same conclusion that the efficiency of the Dyslexie font might not be attributed to its specific shape but rather to its particular spacing setting. However, Maurinus et.al (2016) study was not included in this review since the participants were not specifically dyslexic individuals and this violated the inclusion/exclusion criteria.

An assessment of the risk of bias finds that the general quality of research is low due to poor study design and reporting practices. The small sample sizes played a major role in incomplete outcome reporting and the limited empirical research shows the novelty of specialized typefaces especially when noted that all studies were conducted after 2014.

Chapter 6: Conclusion

A total of 12 studies were included in this systematic review that were conducted in various geographical locations and used a number of different typefaces in the interventions with the Dyslexie typeface used most frequently. The studies' sample included mostly children and English was the most common experimental language with Italian being a close second. A conclusion can be drawn from the findings of these papers that although markets are flooding with various typefaces claiming to be specialized in mitigating the struggles of dyslexic individuals' reading performance, there is a clear lack of empirical evidence to back up these claims. While anecdotal evidence and user testimonies indicate the possible benefits of dyslexia-specialized typefaces, more thorough empirical research is needed to assess their usefulness.

From the initial screening process, it was noted that much of the content regarding specialized typefaces came from magazine articles, reviews and personal opinions without empirical research to back these claims. In addition, the risk of bias assessment revealed that most studies included in this review were of moderate to weak global ratings regarding the study design and reporting of results. However, it is worth noting that most studies conducted on these typefaces took place in the last decade and more research is being produced in the area as the popularity of these typefaces grows.

It is crucial to emphasise, however, that dyslexia support extends beyond fonts. Creating a dyslexia-friendly environment entails a variety of strategies, including multisensory learning approaches, the use of assistive technologies such as text-to-speech software, personalised instruction, and promoting a supportive and understanding environment in educational settings. Awareness and training for educators, parents, and peers are critical to ensure that individuals with dyslexia receive the modifications and support they require to excel academically and professionally. By adopting these inclusive practices, we may work towards a future in which dyslexia is viewed as a learning difference rather than a constraint, allowing individuals to attain their full potential.

6.1 Implications for Dyslexic Individuals

Although majority of the current empirical research does not support the claim that dyslexia-specialized typefaces have a significant effect on the improvement of reading skills in dyslexic individuals, many user testimonials and individual experiences with the typefaces have been positive. Some users of the Dyslexie font have stated that the font is more "comfortable" and made a "big difference" to their reading (Dyslexie website).

This apparent divergence between empirical findings and anecdotal evidence might be attributed to individual preferences and subjective experiences. While quantitative evidence from controlled studies may not be conclusive in proving the efficacy of dyslexia-specific typefaces, qualitative and experiential feedback from actual users adds an intriguing dimension to the debate. Thus, it is possible that particular individuals, influenced by distinct cognitive and perceptual variables, receive genuine benefits from adopting certain typefaces.

6.2 Limitations of the Review

During this review process, every effort was made to remain methodological, comprehensive, and concise while screening studies, extracting data and reporting the results. Every effort was also made to limit potential bias by including second reviewers for abstract screening and quality appraisal. However, there were some limitations that were inevitable. These limitations will be discussed in this section.

The first limitation of this review is that languages not using the Roman alphabet were not included in lieu of typeface characteristic differences and to make the research question myopic to fit the scale of the project. This could have potential impacts on the findings since during the screening projects, studies using specialized typefaces on Arabic, Chinese and Cyrillic alphabets were discovered. The complex interplay of various writing systems, language structures, and cognitive processes demands additional research to fully appreciate the impact of specialised typefaces outside the constraints of Latin scripts.

Another limitation of this review is that during the full-text screening, 1 study (Sparrow, Dei Cas, & Khelifi, 2016) was not retrievable from any database and due to time constraints contacting the authors was not possible. Another limitation was the absence of a second reviewer in the data extraction process. While this is unlikely to have significantly impacted the results, it would have further reduced the risk of bias in the review but again, due to time constraints, it was not feasible.

Since the field of specialized typefaces is still new and rapidly expanding, there are more studies being conducted on the matter, for instance, there was a study published on the effect of dyslexia typeface on reading fluency a few months ago but it was not added in this review due to becoming available later. It is impossible to incorporate new studies on a constant basis while forming this report. However, the introduction of these new studies indicates a rise in interest in this field. The current review provides a starting point for future work in the field.

6.3 Directions for future research

Future studies in specialized typefaces could employ a longitudinal study design to track the effects of the typefaces over a longer period of time. All the studies included were cross-sectional studies and a drawback to this could be that the participants did not have time to get used to the font. This point about the new exposure to typefaces was also discussed by Joseph and Powell (2022), who expressed an interest in future studies accounting for the novelty of the typeface when introduced to participants. Joseph and Powell (2022) also stated that the novelty of these fonts could have had an adverse effect on dyslexic participants since they faced the additional challenge of identifying the letters in a brand-new typeface they hadn't encountered before. Researchers like Ossen (2012) have proposed that if children are first familiarised with the font, the effects may be strengthened. Font familiarity has also been shown to benefit adult readers (Gauthier, Isabel, Wong, Hayward, and Cheung, 2006).

The majority of research so far has centred around Latin scripts and the development of specialized fonts tailored to the Latin alphabet. However, numerous languages, such as Arabic, Chinese, and Japanese, employ non-Latin alphabets and provide a compelling avenue for future research to explore the potential advantages of typefaces within these linguistic contexts.

Future research could also compare different dyslexia-specialized typefaces to see whether individual design components contribute the most to improved reading outcomes. This could help to guide the development of optimised typefaces that take advantage of the most effective characteristics for improving reading fluency.

Finally, there may be opportunities to expand research outside controlled settings and investigate the usefulness of specialised typefaces in real-world reading circumstances such as classroom settings and everyday reading activities. This can provide useful information regarding the fonts' practical utility and ecological validity.

References[1]

Armijo-Olivo, S., Stiles, C. R., Hagen, N. A., Biondo, P. D., & Cummings, G. G. (2012) Assessment of study quality for systematic reviews: a comparison of the Cochrane Collaboration Risk of Bias Tool and the Effective Public Health Practice Project Quality Assessment Tool: methodological research. Journal of evaluation in clinical practice, 18(1), 12-18.

Bachmann, C., & Mengheri, L. (2018). Dyslexia and fonts: Is a specific font useful?. Brain sciences, 8(5), 89.

Bishop, D. V. (2015). The interface between genetics and psychology: lessons from developmental dyslexia. Proceedings of the Royal Society B: Biological Sciences, 282(1806), 20143139.

Blachman, B. A. (2000). Phonological awareness. Handbook of reading research, 3, 483-502.

Boland, A., Cherry, M., & Dickson, R. (2017). Doing a systematic review: A student's guide (2nd ed.). Los Angeles, London: SAGE.

British Dyslexia Association. (n.d.-a). Dyslexia friendly style guide. Retrieved from https://www.bdadyslexia.org.uk/advice/employers/creating-a-dyslexia-friendly-workplace/dyslexia-friendly-style-guide

Brunton, G., Stansfield, C., Caird, J., & Thomas, J. (2017). Finding relevant studies. In D. Gough, S. Oliver, & J. Thomas (Eds.), An introduction to systematic reviews (2nd edition, pp. 93–122). SAGE Publications.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. Nature, 226(5241), 177-178.

Boutron, I., Page, M. J., Higgins, J. P., Altman, D. G., Lundh, A., Hróbjartsson, A., & Cochrane Bias Methods Group. (2019). Considering bias and conflicts of interest among the included studies. Cochrane handbook for systematic reviews of interventions, 177-204.

Carmichael, K. (2022). New International Dyslexia Campaign speaks directly to the design sector: Dyslexia Scotland. Retrieved from https://dyslexiascotland.org.uk/news/new-international-dyslexia-campaign-speaks-directly-to-the-design-sector/

[1]This dissertation adheres to the citation conventions of the APA style guide (American Psychological Association, 2020).

Carroll, J. M., & Iles, J. E. (2006). An assessment of anxiety levels in dyslexic students in higher education. British journal of educational psychology, 76(3), 651-662.

Cochrane Effective Practice and Organisation of Care (EPOC). Data collection form. EPOC Resources for review authors, 2017. epoc.cochrane.org/resources/epoc-specific-resources-review-authors [accessed 18 June, 2023]

Crowe, M., & Sheppard, L. (2011). A review of critical appraisal tools show they lack rigor: alternative tool structure is proposed. Journal of clinical epidemiology, 64(1), 79-89.

Damiano, R., Gena, C., & Venturini, G. (2019). Testing web-based solutions for improving reading tasks in dyslexic and neuro-typical users. Multimedia Tools and Applications, 78, 13489-13515.

DeFries, J. C., Fulker, D. W., & LaBuda, M. C. (1987). Evidence for a genetic aetiology in reading disability of twins. Nature, 329(6139), 537-539.

Deutsch, L. (2014). Man creates font designed to help dyslexics read. Retrieved from https://eu.usatoday.com/story/news/nation-now/2014/11/12/dyslexic-font/18915297/

Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., & Sutton, A. (2005). Synthesising qualitative and quantitative evidence: a review of possible methods. Journal of health services research & policy, 10(1), 45-53.

Dressler, E. (2019). Understanding the effect of font type on reading comprehension/memory under time-constraints.

Duranovic, M., Senka, S., & Babic-Gavric, B. (2018). Influence of increased letter spacing and font type on the reading ability of dyslexic children. Annals of Dyslexia, 68, 218-228.

Dyslexia basics. (2020). Retrieved from https://dyslexiaida.org/dyslexia-basics/

Dyslexia. (2022). Retrieved from https://www.mayoclinic.org/diseases-conditions/dyslexia/symptoms-causes/syc-20353552

Dyslexia in the classroom: What every teacher needs to know. (2017). Retrieved from https://dyslexiaida.org/dyslexia-in-the-classroom/

Eden, G. F., Olulade, O. A., Evans, T. M., Krafnick, A. J., & Alkire, D. R. (2016). Developmental dyslexia. In Neurobiology of language (pp. 815-826). Academic Press.

Everatt, J., Bradshaw, M. F., & Hibbard, P. B. (1999). Visual processing and dyslexia. Perception, 28(2), 243-254.

Fisher, S. E., & DeFries, J. C. (2002). Developmental dyslexia: genetic dissection of a complex

cognitive trait. *Nature Reviews Neuroscience*, *3*(10), 767-780.

Frey, B. B. (Ed.). (2018). The SAGE encyclopedia of educational research, measurement, and evaluation. Sage Publications.

Galliussi, J., Perondi, L., Chia, G., Gerbino, W., & Bernardis, P. (2020). Inter-letter spacing, inter-word spacing, and font with dyslexia-friendly features: testing text readability in people with and without dyslexia. *Annals of dyslexia*, *70*(1), 141-152.

Gauthier, Isabel, Alan CN Wong, William G. Hayward, and Olivia S. Cheung. "Font tuning associated with expertise in letter perception." Perception 35, no. 4 (2006): 541-559.

George, J. (2012). 7 Fonts With Very Specialized Purposes. Retrieved from https://www.sitepoint.com/7-fonts-with-very-specialized-purposes/

Gilger, J. W., Hanebuth, E., Smith, S. D., & Pennington, B. F. (1996). Differential risk for developmental reading disorders in the offspring of compensated versus noncompensated parents. Reading and Writing, 8, 407-417.

Goswami, U. (2022). Theories of Dyslexia. In M. Skeide (Ed.), *The Cambridge Handbook of Dyslexia and Dyscalculia* (Cambridge Handbooks in Psychology, pp. 5-24). Cambridge: Cambridge University Press. doi:10.1017/9781108973595.002

Gregor, P., & Newell, A. F. (2000, November). An empirical investigation of ways in which some of the problems encountered by some dyslexics may be alleviated using computer techniques. In Proceedings of the fourth international ACM conference on Assistive technologies (pp. 85-91).

Heiner, B., & Tanja, M. (2017). Contrast and font affect reading speeds of adolescents with and without a need for language-based learning support. Training, Language and Culture, 1(4), 39-55.

Hillier, R. (2012). The Design and Development of the Sylexiad Typeface. Supporting Dyslexic Adults in Higher Education and the Workplace, 185-196.

Hoffmeister, S. (2016). The impact of font type on reading.

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., ... & Vedel, I. (2018). Mixed methods appraisal tool (MMAT), version 2018. Registration of copyright, 1148552(10).

Johnson, N., & Phillips, M. (2018). Rayyan for systematic reviews. Journal of Electronic Resources Librarianship, 30(1), 46-48.

Joseph, H., & Powell, D. (2022). Does a specialist typeface affect how fluently children with and without dyslexia process letters, words, and passages?. Dyslexia, 28(4), 448-470.

Krishnan, Saloni, Watkins, Kate E, & Bishop, Dorothy V.M. (2016). Neurobiological Basis of Language Learning Difficulties. Trends in Cognitive Sciences, 20(9), 701-714.

Kussmaul, A. (1877). Disturbance of speech. Cyclopedia of the practice of medicine, 581-875.

Kuster, Sanne M., Van Weerdenburg, Marjolijn, Gompel, Marjolein, & Bosman, Anna M. T. (2018). Dyslexie font does not benefit reading in children with or without dyslexia. Annals of Dyslexia, 68(1), 25-42.

Li, Y., Chen, W., Liu, G., Wei, L., Liu, C., Si, L., Zhang, J., & Yang, K. (2017). Analysis of quality assessment tools in Campbell systematic reviews. Abstracts of the Global Evidence Summit, 9 Supp 1.

Lovegrove, W., Martin, F., & Slaghuis, W. (1986). A theoretical and experimental case for a visual deficit in specific reading disability. Cognitive neuropsychology, 3(2), 225-267.

Lowe, K., Tennent, C., Burgess, C., Moodie, N., Vass, G., & Guenther, J. (2023). Developing a systematic methodology to explore research in Indigenous education. In Assessing the Evidence in Indigenous Education Research (pp. 19-39). Palgrave Macmillan, Cham.

Malherbe, H. M. (2017). Comparing dyslexie and Times New Roman fonts for reading accuracy and rate in grade 4 learners. University of Johannesburg (South Africa). McNulty, M. A. (2003). Dyslexia and the life course. *Journal of learning disabilities*, *36*(4), 363-381.

Mallett, R., Hagen-Zanker, J., Slater, R., & Duvendack, M. (2012). The benefits and challenges of using systematic reviews in international development research. Journal of development effectiveness, 4(3), 445-455.

Mangal, C. (2019a). The state of LD: Understanding the 1 in 5. Retrieved from https://www.ncld.org/news/newsroom/the-state-of-ld-understanding-the-1-in-5/

Marinus, E., Mostard, M., Segers, E., Schubert, T. M., Madelaine, A., & Wheldall, K. (2016). A special font for people with dyslexia: Does it work and, if so, why?. Dyslexia, 22(3), 233-244.

Marsh, G., Friedman, M., Welch, V., & Desberg, P. (1981). A cognitive-development theory of reading acquisition Reading Research: Advances in Theory and Practice. 3.

Martelli, M., Di Filippo, G., Spinelli, D., & Zoccolotti, P. (2009). Crowding, reading, and developmental dyslexia. Journal of vision, 9(4), 14-14.

Md Mahtar, I. F. (2018). Digital text presentation and navigation to support people with dyslexia (Doctoral dissertation, University of York).

Moher, D., Liberati A., Tetzlaff J., Altman D. G., The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. DOI:10.1371/journal.pmed1000097

Ossen, A. (2012). Effecten van het Lettertype Dyslexie in het basisonderwijs-een praktijkonderzoek. Beter begeleiden digitaal.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *International journal of surgery*, *88*, 105906.

Plessis, S. du. (2023). Dyslexia theories and models - edublox online tutor. Retrieved from https://www.edubloxtutor.com/dyslexia-theories-and-models/#Phonological

Pollock, A., & Berge, E. (2018). How to do a systematic review. International Journal of Stroke, 13(2), 138-156.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version, 1(1), b92.

Powell, S. L., & Trice, A. D. (2020). The impact of a specialized font on the reading performance of elementary children with reading disability. Contemporary School Psychology, 24, 34-40.

Ramsey, J. R. (2014). The effects of the font dyslexie on oral reading fluency skills in students grades 8 through 12 with and without reading disabilities. Western Carolina University.

Ramus, F., Marshall, C. R., Rosen, S., & van der Lely, H. K. (2013). Phonological deficits in specific language impairment and developmental dyslexia: Towards a multidimensional model. *Brain*, *136*(2), 630-645.

Robinson, P., & Lowe, J. (2015). Literature reviews vs systematic reviews. *Australian and New Zealand journal of public health*, *39*(2), 103-103.

Scarborough, H. S. (1990). Very early language deficits in dyslexic children. Child development, 61(6), 1728-1743.

Scott, S. D., Rotter, T., Flynn, R., Brooks, H. M., Plesuk, T., Bannar-Martin, K. H., ... & Hartling, L. (2019). Systematic review of the use of process evaluations in knowledge translation research. Systematic reviews, 8(1), 1-10.

Shaywitz, S. E. (1998). Dyslexia. *New England Journal of Medicine*, *338*(5), 307-312.

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Mencl, W. E., Fulbright, R. K., Skudlarski, P., ... &

Gore, J. C. (2002). Disruption of posterior brain systems for reading in children with developmental dyslexia. Biological psychiatry, 52(2), 101-110.

Shaywitz, B. A., Shaywitz, S. E., Blachman, B. A., Pugh, K. R., Fulbright, R. K., Skudlarski, P., ... & Gore, J. C. (2004). Development of left occipitotemporal systems for skilled reading in children after a phonologically-based intervention. Biological psychiatry, 55(9), 926-933.

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational researcher*, *15*(9), 5-11.

Slocombe, A. (2020, November 10). Dyslexia friendly fonts: The top 10 dyslexia styles. Exceptional Individuals. https://exceptionalindividuals.com/about-us/blog/our-top-10-dyslexia-friendly-fonts/

Snowling, M. J. (2001). From language to reading and dyslexia 1. Dyslexia, 7(1), 37-46.

Snowling, M., Stackhouse, J., & Rack, J. (1986). Phonological dyslexia and dysgraphia—a developmental analysis. Cognitive Neuropsychology, 3(3), 309-339.

Stein, J., & Walsh, V. (1997). To see but not to read; the magnocellular theory of dyslexia. Trends in neurosciences, 20(4), 147-152.

TEDx Talks. (2015). How a font can help people with dyslexia to read | Christian Boer | TEDxFultonStreet. Retrieved from https://www.youtube.com/watch?v=qVaeGOflF7w

Trollbäck, L. (2004). READING WITHOUT TEARS. Print, 58(1), 16. Retrieved from https://www.proquest.com/trade-journals/reading-without-tears/docview/231181770/se-2

Tunmer, W., & Greaney, K. (2010). Defining dyslexia. Journal of learning Disabilities, 43(3), 229-243.

Valdois, S., Bosse, M. L., & Tainturier, M. J. (2004). The cognitive deficits responsible for developmental dyslexia: Review of evidence for a selective visual attentional disorder. Dyslexia, 10(4), 339-363.

Wagner, R. F. (1973). Rudolf Berlin: Originator of the term dyslexia. Bulletin of the Orton Society, 57-63.

Wainwright, O. (2014). Dyslexie: The Chubby-ankled font that makes reading easier for dyslexics. Retrieved from https://www.theguardian.com/artanddesign/architecture-design-blog/2014/nov/12/dyslexie-new-font-that-makes-reading-easier-with-dyslexia

Weitzman, C. (2022). Famous people with dyslexia. Retrieved from

https://speechify.com/blog/famous-people-dyslexia

Wery, J. J., & Diliberto, J. A. (2017). The effect of a specialized dyslexia font, OpenDyslexic, on reading rate and accuracy. Annals of dyslexia, 67, 114-127.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. Trends in cognitive sciences, 15(4), 160-168.

(2015). Retrieved from https://www.nih.gov/news-events/news-releases/nih-funded-study-finds-dyslexia-not-tied-iq.

Appendix A: IDESR Protocol Document

Protocol Registration Form

The content of this form is based on Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. (2015). Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1). doi: 10.1186/2046-4053-4-1. We strongly recommended that this form is completed in accordance with the recommendations in that paper.

Title

Neurodiversity and Typeface: Do Specialized Typefaces Improve Reading Fluency and Accuracy in Dyslexic Individuals – Protocol for a Systematic Review
The title should identify the report as a protocol of a systematic review. If this is an update of an existing review, indicate this in the title. Examples: The effectiveness of task-based language teaching for adult learners of foreign or second languages. Protocol for a systematic review. *Or* The effectiveness of written corrective feedback for the acquisition of L2 grammar among primary-aged language learners. A protocol for an update of the systematic review by Bloggs and Smith (2020).

Main Contact/Corresponding Author

Name: Rumaisa Khusru
Institutional affiliation: University of Oxford
e-mail address: rumaisa.khusru@education.ox.ac.uk
Physical mailing address: 15 Norham Gardens, Oxford OX2 6PY, Department of Education

Additional Authors

Name:
Institutional affiliation:
e-mail address:
On the IDESR website, you will be able to add as many additional authors as you need.

Review Question(s)

RQ 1- What is the extent and nature of empirical research investigating the role of specialized typefaces on dyslexics' reading outcomes?

RQ 2- What is the extent of the evidence of the effectiveness of specialized typefaces on the reading performance of dyslexic readers?

a) What impact do specialized typefaces have on reading rate?

B) What impact do specialized typefaces have on reading accuracy?

Provide the review questions. For reviews of interventions, include reference to PI(E)CO, as appropriate (Participants, Intervention (Exposure), Comparator, Outcomes). Examples: What are the effects of study abroad compared to classroom teaching on vocabulary acquisition among adolescent learners of a foreign language? *Or* What is the impact on academic attainment of attending a bilingual school compared to a target language only school among minority language users?

Rationale

According to Mallett, Hagen-Zanker, Slater, & Duvendack, (2012), a systematic review can be beneficial in identifying gaps in the existing literature and this report will aim to identify the gaps in research concerning specialised dyslexia typefaces. This systematic review aims to answer the question of whether or not these fonts really impact the reading fluency and accuracy of dyslexic readers by exploring the range of studies conducted on the effects of typefaces on the reading performance of dyslexic readers. Moreover, there are numerous specialised dyslexia typefaces available, each claiming to improve readability and alleviate the difficulties faced by people with dyslexia. A systematic review can evaluate and contrast various fonts, design concepts, and their impact on reading outcomes. This would provide useful information on which typefaces are more effective and why. This report can further offer evidence-based recommendations for the use of specialised dyslexia typefaces by synthesising the data of various studies. These recommendations can assist educators, parents, and individuals with dyslexia make educated decisions about these fonts.

In no more than 300 words, describe the rationale for the review in the context of what is already known.

Inclusion Criteria

1) Intervention: To include studies that investigate the effects of specialised dyslexia fonts on the reading performance of people who have dyslexia. To exclude studies on general typography or fonts that have no direct relevance to dyslexia or reading difficulties.
2) Participants: To include studies involving people of any age who have been diagnosed with dyslexia. To exclude studies that do not involve individuals who have dyslexia or studies that are primarily concerned with other learning difficulties or neurodevelopmental disorders and make no specific mention of dyslexia.
3) Age: No age restriction, age will be considered as an influential variable in the analysis.

4) Control groups: To include studies with randomized controlled trials (RCTs) or well-designed comparative studies. To exclude studies that lack a control group or appropriate comparator condition.
5) Outcomes: To include studies that test accuracy and/or reading rate. To exclude studies testing other outcomes other than accuracy or reading rate.
6) Language: To include studies that experiment with a language of Roman Alphabet (Examples: English, French, Italian). To exclude studies using any other language.
7) Language of publication: To include studies published in English. To exclude studies published in different language.
8) Date of publication: To include papers from 1957 onwards. To exclude papers before 1957. Rationale: the oldest font considered dyslexia-friendly was created in 1957 (Slocombe, 2020).
9) Bibliographic information: To include studies with complete reference and sufficient bibliographic information. To exclude studies lacking sufficient bibliographic information.
10) Publication status: No studies are excluded because of their publication status. In order to prevent publication bias and incorporate a larger range of literature in the study process, grey literature (unpublished theses) will be included.

List here the criteria to be used for inclusion in the review. As appropriate, include information about the population, interventions, comparators, primary outcomes, setting, study design(s), time frame, publication types, language(s) of publication, etc.

Information Sources

Below is a list of the databases that will be used for this systematic review. Popular databases in the domains of education, linguistics, and psychology, as well as multidisciplinary sources that cover a larger range of social science fields, are included in this list. All databases will be accessed electronically through the Bodleian Library (University of Oxford's) subscription.

Education: Education Collection; ERIC; British Education Index
Linguistics: ProQuest Linguistics Collection
Psychology: PsycINFO
Multidiscipline: Web of Science; SCOPUS; PubMed
Grey literature: ProQuest Dissertations & Theses Global
For the initial scoping search: Google Scholar; SOLO; JSTOR

Describe all intended information sources. Include the names of electronic databases, journals or websites that will be hand searched, contact with study authors, grey literature sources, etc.

Search Strategy

For this review's initial literature search, a consultation was made with a University of Oxford college librarian (St. Anne's) and an Education (Bodleian) librarian to create a list of search terms relevant to the current systematic review.

A pilot search containing the term "neurodiverse" was first used, but this yielded too many results with diverse and unrelated neurodevelopmental disorders, so the term was dropped and only the specific "dyslexia" term remained to ensure the results were relevant. The term "reading disability" is also often used in relation to dyslexia, so this term was added.
The pilot search did not include specific fonts' names, but it was seen that some important papers were left out so "dyslexie", "lexie readable", "sylexiad" and "opendyslexia" were then added to the search terms.

The initial search looked at abstracts only and the Boolean string was slightly adapted relative to the specifications of the respective data bases but the crux of it is mentioned below:

TITLE-ABS-KEY (dyslexia) OR TITLE-ABS-KEY ("reading disability") AND TITLE-ABS-KEY (typeface*) OR TITLE-ABS-KEY (font*) OR TITLE-ABS-KEY ("specialised font") OR TITLE-ABS-KEY ("specialised typeface") OR TITLE-ABS-KEY ("dyslexia font") OR TITLE-ABS-KEY (dyslexie) OR TITLE-ABS-KEY (opendyslexic) OR TITLE-ABS-KEY ("lexie readable") OR TITLE-ABS-KEY (sylexiad) AND TITLE-ABS-KEY ("reading outcome*") OR TITLE-ABS-KEY ("reading rate") OR TITLE-ABS-KEY ("reading fluency") OR TITLE-ABS-KEY ("reading speed") OR TITLE-ABS-KEY ("reading accuracy")

(The Boolean string mentioned in this form is from the SCOPUS database search)

Present the search strategy to be used for at least one electronic database such that it could be repeated by a third party. Include planned limiters, for example, date range, and location in the text (e.g. Title, Abstract, or Full Text). Present these as a Boolean phrase if possible. If Boolean phrasing is inappropriate for your review, present the search strategy in a way that can allow replication by a third party.

Data Management

After conducting the search, the list of studies will be uploaded into Rayyan, an internet-based, AI-powered programme designed to organise systematic reviews (Johnson & Phillips, 2018). During the screening process, Rayyan will be utilised as the main technique of organisation, where duplicates of studies will be eliminated, and initial screening of titles and abstracts will be conducted. In addition, a reference manager (Zotero) will be used to keep track of the research studies' bibliographic information (title, author, publication date, journal).

A Microsoft Excel table will be used to document the inclusion/exclusion procedure, including references, the choice made at each screening step (which are the initial title and abstract stage, followed by the full-text screening), and the justification for exclusion (if relevant). Additionally, the search procedure will be recorded and continually updated on a Word document.

A MacBook Pro (version 12.5) will be used throughout this research study.

Describe the mechanisms by which the data will be managed throughout the review. For example, say which data management software the review team use, e.g. Rayyan, EPPI Reviewer, Excel, Covidence, etc. Describe if these will change with different phases of the review (abstract screening, full text screening, data extraction, etc.).

Selection Process

Following the initial search, Rayyan, an AI tool, will remove any duplicate studies. The remainder of the titles and abstracts will be evaluated for eligibility before being assigned for full-text screening. The abstract of a report must not expressly contravene any of the exclusion criteria in order for it to be eligible for full-text screening. If the title or abstract do not match a criterion, the study will be removed, and the reason will be noted.

Full reports will be collected from the relevant databases after the initial screening. In the instance where complete texts are unavailable, they will be excluded. Although it is common practice to reach out to the author(s) when full texts are unavailable, in the context of this MSc and due to time constraints, this would not be possible. Both screening results will be documented in line with the PRISMA flow diagram (Page et al., 2021), with a rationale for all decisions.

To reduce bias and subjectivity during the study selection process, a second reviewer will be recruited. The second reviewer will have an applied linguistics background and appropriate knowledge of the systematic reviewing approach. The second reviewer will be given the opportunity to become acquainted with the present protocol and will be given the chance to ask any questions about the review's purpose and methods, before performing a screening on a fraction of the articles found from the search. The number of items undergoing second evaluation will be roughly 5% of the total number of studies; more than 5% is not practicable due to time constraints. The second reviewer will be "blind" to the decision by the first reviewer until the final discussion. To achieve inter-rater reliability, 90% agreement amongst researchers is typically regarded as sufficient. If there is less than 90% agreement between the researchers, the original researcher and the second reviewer will discuss their disagreements and reach a consensus.

Describe the method by which studies will be selected for inclusion at each stage of the review. For example, how many reviewers will screen abstracts/full texts? What quality assurance procedures will be in place in each of these phases (dual screening of all records, percentage dual screened then checked for consistency, etc.)?

Data Collection Process

The studies that make it through both the primary and secondary screenings will serve as the review's data sources. The Cochrane (2017) good practice data extraction form will serve as the basis for the creation of a data extraction form. The form will be modified to meet the requirements of the works that are included in the final list, taking into account factors such as the variety of research designs.

Cochrane Effective Practice and Organisation of Care (EPOC). Data collection form. EPOC Resources for review authors, 2017. epoc.cochrane.org/resources/epoc-specific-resources-review-authors [accessed 18 June, 2023]

Describe how data will be extracted from reports. Will a data extraction form be used? Will this be piloted? Will data be extracted independently by multiple reviewers? What is the process for obtaining data not contained in the reports (e.g. contacting authors directly)?

Data Items

The following data items are considered important for this review:
General information: author(s), title, reference, year of publication, source;
Study characteristics and methods: study design, research question(s), methods, funding (if applicable);
Participants characteristics: sample size, age, language background (if applicable);
Task/intervention: type of task, typeface used, methods, comparators, control group(s);
Results: Outcome type, outcomes as reported by the authors, conclusions as reported by the authors, limitations.

Additional notes will also be extracted if applicable.

List and define all data items that will be extracted (e.g. participant info, outcome measures, sources of funding, study design, etc.).

Risk of bias/trustworthiness of individual studies

This review adheres to Slavin's (1986) best-evidence synthesis method, which is not biased against any research design in study selection, but rather evaluates the methodological rigour and potential bias of the overall work.

The Mixed Methods Appraisal Tool (MMAT) will be used to assess the quality of the included studies (Hong et al., 2018). The instrument considers a variety of study designs, categorising them into five distinct categories: qualitative, randomised controlled trials, non-randomised, quantitative descriptive, and mixed methods studies. After two general screening queries (used to exclude non-empirical papers), the instrument features quality evaluation questions for each category, to which the researcher must respond with 'yes', 'no', or 'can't tell,' as well as a remark box to explain/justify ratings. The MMAT converts each of the five criteria (in each category) into a quality score ranging from 0 to 5. A study employing a mixed-methods design will be graded in all relevant categories.

These quality evaluations will be taken into account when discussing the findings.

A second reader will conduct 5-10% of the risk of bias evaluations, similar to the initial screening. It is considered best practice to include more than one reviewer in this process due to the subjective nature of assessing risk of bias (Petticrew & Roberts, 2006).

Slavin, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. Educational researcher, 15(9), 5-11.

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., ... & Vedel, I. (2018). Mixed methods appraisal tool (MMAT), version 2018. Registration of copyright, 1148552(10).

Describe how risk of bias, trustworthiness, or quality of individual studies will be assessed. Name any specific tools, e.g. Gorard's Sieve, Maryland Scientific Methods Scale, Cochrane Risk of Bias Tool, EPPI Weight of Evidence Tool, etc. State how this information will be used in the synthesis.

Data Synthesis

The main method for presenting data will be a narrative synthesis.

This review will aim to follow Popay et al.' s (2006) guidelines for narrative synthesis:

- Developing a theory of how the intervention works, why and for whom;
- Developing a preliminary synthesis of findings of included studies;
- Exploring relationships in the data;
- Assessing the robustness of the synthesis.

Meta-analyses will only be carried out if the methodologies of several studies (i.e. intervention characteristics and outcome measures) are adequately similar to one another and the studies report enough comparable data.

Boland, A., Cherry, M., & Dickson, R. (2017). Doing a systematic review: A student's guide (2nd ed.). Los Angeles, London: SAGE.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., ... & Duffy, S. (2006). Guidance on the conduct of narrative synthesis in systematic reviews. A product from the ESRC methods programme Version, 1(1), b92.

Describe criteria under which quantitative synthesis will be performed. If quantitative synthesis is appropriate, describe preferred summary measure (Cohen's D, Hedges' G, etc.) and how these will be combined. Describe any additional planned analyses (e.g. sub-group analysis). If quantitative synthesis is not appropriate, describe how data will be combined and summarised.

Meta-biases

In order to reduce publication bias, this review incorporates searches of "grey" literature, such as theses and other unpublished papers that meet the criteria for inclusion.

Describe how meta-biases (publication bias, selective outcome reporting, etc.) will be addressed.

Confidence in cumulative evidence

The strength of the cumulative evidence will be evaluated based on the results of the Mixed Methods Appraisal Tool's evaluation of the individual credibility of each study.

If time allows, the CASP checklist will be used as an additional way for determining the strength of the cumulative evidence.

Critical Appraisal Skills Programme (2022). CASP Systematic Review Checklist. [online] Retrieved 18th June, 2023, from: https://casp-uk.b-cdn.net/wp-content/uploads/2018/03/CASP-Systematic-Review-Checklist-2018_fillable-form.pdf

Describe how the strength of the body of evidence will be assessed.

Sources of Funding

N/A

Specify any financial or other support for the review. Provide name for the review funder and/or sponsor. Include funding reference number if available.

Role of Funders

N/A

Describe roles of funder(s), sponsor(s), and/or institution(s), if any, in developing the protocol.

Anticipated or actual start date: May, 2023

Anticipated completion date: August, 2023

Other language resources

N/A

If you have other language versions of your protocol, please include a link here.

Current Status

☑ Ongoing
☐ Completed but not published
☐ Completed and published

For initial submissions select 'ongoing'. When you have completed the review, you should return to update the record accordingly. Your review status will be displayed next to your record in IDESR.

Appendix B: Data Extraction Form

Data Extraction Form

	Item	Data	Description
General Information	Date form completed		Use format dd/mm/yyyy
	Reference citation		List the full reference
	Study Title		Full title of study
	Year of Publication		Use format dd/mm/yyyy
	Study author(s)		Full name(s)
	Location		Country of study
	Document source		Source database or website
Study Overview	Research Questions		List the study's research question(s)
	Study design		e.g. RCT, observational study
	Data type		Quantitative/qualitative
	Language of publication		Language study was published in
	Language of Study Intervention		Language used in study's intervention
Intervention	Type of task		Task description as outlined by author(s)
	Typeface(s) used		e.g. Dyslexie or OpenDyslexic
	Control		Control groups / comparators
	Attrition		Did any participants leave the study? How many, and for what reasons?

Participants	Sample size		Number of participants
	Age		Age of participants
	Language background		If applicable
Outcomes	Outcome name		Reading Fluency / Accuracy / Speed
	Conclusions		conclusions as reported by the author(s)
	Limitations		Limitations as outlined by the author(s)

Appendix C: Completed Data Extraction Form Example

Data Extraction Form

	Item	Data
General Information	Date form completed	17th July, 2023
	Reference citation	Joseph, H., & Powell, D. (2022). Does a specialist typeface affect how fluently children with and without dyslexia process letters, words, and passages?. Dyslexia, 28(4), 448-470.
	Study Title	Does a specialist typeface affect how fluently children with and without dyslexia process letters, words, and passages?
	Year of Publication	2022
	Study author(s)	Holly Joseph and Daisy Powell
	Location	United Kingdom
	Document source	British Education Index (EBSCO)
Study Overview	Research Questions	1. Do children with and without dyslexia differ in their naming speed of letters in Dyslexie versus Calibri typeface, and of 'regular' and 'confusable' letters? 2. Do children with and without dyslexia differ in their reading efficiency of single words in Dyslexie versus Calibri typeface? 3. Do children with and without dyslexia differ in their reading times on target words embedded in text as a function of typeface and difficulty?
	Study design	Quantitative non-randomized. A 2 x 2 x 2 design. Two typefaces, Dyslexie and Calibri were compared, and this was manipulated within participants.
	Data type	Quantitative data
	Language of publication	English
	Language of Study Intervention	English

Intervention	Type of task	rapid letter naming task, a word reading efficiency task, and a passage reading task in two typefaces
	Typeface(s) used	Dyslexie compared with Calibri
	Control	Non-dyslexic children (n = 34)
	Attrition	Ten participants were excluded for the following reasons: did not complete all tasks (n = 1); did not comply with task instructions (n = 4); slow reading (timed out, n = 2); and tracker loss (n = 3).
Participants	Sample size	71 children, 37 of whom had a diagnosis of dyslexia and 34 did not have dyslexia.
	Age	8–12 years
	Language background	N//A
Outcomes	Outcome name	Reading fluency
	Conclusions	Results showed no differences in word or passage reading between the two typefaces, but letter naming did appear to be more fluent when letters were presented in Dyslexie rather than Calibri text for all children. The results suggest that a typeface in which letters are designed to be distinctive from one another may be beneficial for letter identification and that an intervention in which children are taught letters in a specialist typeface is worthy of consideration.
	Limitations	1) Dyslexic group more advantageous socioeconomic backgrounds than average and from the control group. 2) Did not have a fully crossed design for our RAN task, the only task in which we observed an effect. 3) Lack of effects of difficulty manipulation in the passage reading task was unexpected and may suggest that other confounding factors were at play. Although the study-controlled target words on a large number of lexical characteristics known to influence word processing, as the main manipulation was straightforward and a condition for observing any effect of typeface, it may be that it was failed to consider other influences. It is possible that had the manipulation been more successful, a benefit of Dyslexie typeface may have been observed and it is worth taking this into account in future studies.

Appendix D: Mixed Methods Appraisal Tool (MMAT)

What is the MMAT?
The MMAT is a critical appraisal tool that is designed for the appraisal stage of systematic mixed studies reviews, i.e., reviews that include qualitative, quantitative and mixed methods studies. It permits to appraise the methodological quality of five categories to studies: qualitative research, randomized controlled trials, non-randomized studies, quantitative descriptive studies, and mixed methods studies.

How was the MMAT developed?
The MMAT was developed in 2006 (Pluye et al., 2009a) and was revised in 2011 (Pace et al., 2012). The present version 2018 was developed on the basis of findings from a literature review of critical appraisal tools, interviews with MMAT users, and an e-Delphi study with international experts (Hong, 2018). The MMAT developers are continuously seeking for improvement and testing of this tool. Users' feedback is always appreciated.

What the MMAT can be used for?
The MMAT can be used to appraise the quality of empirical studies, i.e., primary research based on experiment, observation or simulation (Abbott, 1998; Porta et al., 2014). It cannot be used for non-empirical papers such as review and theoretical papers. Also, the MMAT allows the appraisal of most common types of study methodologies and designs. However, some specific designs such as economic and diagnostic accuracy studies cannot be assessed with the MMAT. Other critical appraisal tools might be relevant for these designs.

What are the requirements?
Because critical appraisal is about judgment making, it is advised to have at least two reviewers independently involved in the appraisal process. Also, using the MMAT requires experience or training in these domains. For instance, MMAT users may be helped by a colleague with specific expertise when needed.

How to use the MMAT?
This document comprises two parts: checklist (Part I) and explanation of the criteria (Part II).

1. Respond to the two screening questions. Responding 'No' or 'Can't tell' to one or both questions might indicate that the paper is not an empirical study, and thus cannot be appraised using the MMAT. MMAT users might decide not to use these questions, especially if the selection criteria of their review are limited to empirical studies.
2. For each included study, choose the appropriate category of studies to appraise. Look at the description of the methods used in the included studies. If needed, use the algorithm at the end of this document.
3. Rate the criteria of the chosen category. For example, if the paper is a qualitative study, only rate the five criteria in the qualitative category. The 'Can't tell' response category means that the paper do not report appropriate information to answer 'Yes' or 'No', or that report unclear information related to the criterion. Rating 'Can't tell' could lead to look for companion papers, or contact authors to ask more information or clarification when needed. In Part II of this document, indicators are added for some criteria. The list is not exhaustive and not all indicators are necessary. You should agree among your team which ones are important to consider for your field and apply them uniformly across all included studies from the same category.

How to score?
It is discouraged to calculate an overall score from the ratings of each criterion. Instead, it is advised to provide a more detailed presentation of the ratings of each criterion to better inform the quality of the included studies. This may lead to perform a sensitivity analysis (i.e., to consider the quality of studies by contrasting their results). Excluding studies with low methodological quality is usually discouraged.

How to cite this document?
Hong QN, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, Gagnon M-P, Griffiths F, Nicolau B, O'Cathain A, Rousseau M-C, Vedel I. Mixed Methods Appraisal Tool (MMAT), version 2018. Registration of Copyright (#1148552), Canadian Intellectual Property Office, Industry Canada.

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018

Category of study designs	Methodological quality criteria	Responses			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?				
	S2. Do the collected data allow to address the research questions?				
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
1. Qualitative	1.1. Is the qualitative approach appropriate to answer the research question?				
	1.2. Are the qualitative data collection methods adequate to address the research question?				
	1.3. Are the findings adequately derived from the data?				
	1.4. Is the interpretation of results sufficiently substantiated by data?				
	1.5. Is there coherence between qualitative data sources, collection, analysis and interpretation?				
2. Quantitative randomized controlled trials	2.1. Is randomization appropriately performed?				
	2.2. Are the groups comparable at baseline?				
	2.3. Are there complete outcome data?				
	2.4. Are outcome assessors blinded to the intervention provided?				
	2.5 Did the participants adhere to the assigned intervention?				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?				
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?				
	3.3. Are there complete outcome data?				
	3.4. Are the confounders accounted for in the design and analysis?				
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?				
4. Quantitative descriptive	4.1. Is the sampling strategy relevant to address the research question?				
	4.2. Is the sample representative of the target population?				
	4.3. Are the measurements appropriate?				
	4.4. Is the risk of nonresponse bias low?				
	4.5. Is the statistical analysis appropriate to answer the research question?				
5. Mixed methods	5.1. Is there an adequate rationale for using a mixed methods design to address the research question?				
	5.2. Are the different components of the study effectively integrated to answer the research question?				
	5.3. Are the outputs of the integration of qualitative and quantitative components adequately interpreted?				
	5.4. Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?				
	5.5. Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?				

Part II: Explanations

1. Qualitative studies	Methodological quality criteria
"Qualitative research is an approach for exploring and understanding the meaning individuals or groups ascribe to a social or human problem" (Creswell, 2013b, p. 3). Common qualitative research approaches include (this list if not exhaustive): **Ethnography** The aim of the study is to describe and interpret the shared cultural behaviour of a group of individuals. **Phenomenology** The study focuses on the subjective experiences and interpretations of a phenomenon encountered by individuals. **Narrative research** The study analyzes life experiences of an individual or a group. **Grounded theory** Generation of theory from data in the process of conducting research (data collection occurs first). **Case study** In-depth exploration and/or explanation of issues intrinsic to a particular case. A case can be anything from a decision-making process, to a person, an organization, or a country. **Qualitative description** There is no specific methodology, but a qualitative data collection and analysis, e.g., in-depth interviews or focus groups, and hybrid thematic analysis (inductive and deductive). Key references: Creswell (2013a); Sandelowski (2010); Schwandt (2015)	1.1. Is the qualitative approach appropriate to answer the research question? Explanations The qualitative approach used in a study (see non-exhaustive list on the left side of this table) should be appropriate for the research question and problem. For example, the use of a grounded theory approach should address the development of a theory and ethnography should study human cultures and societies. This criterion was considered important to add in the MMAT since there is only one category of criteria for qualitative studies (compared to three for quantitative studies).
	1.2. Are the qualitative data collection methods adequate to address the research question? Explanations This criterion is related to data collection method, including data sources (e.g., archives, documents), used to address the research question. To judge this criterion, consider whether the method of data collection (e.g., in depth interviews and/or group interviews, and/or observations) and the form of the data (e.g., tape recording, video material, diary, photo, and/or field notes) are adequate. Also, clear justifications are needed when data collection methods are modified during the study.
	1.3. Are the findings adequately derived from the data? Explanations This criterion is related to the data analysis used. Several data analysis methods have been developed and their use depends on the research question and qualitative approach. For example, open, axial and selective coding is often associated with grounded theory, and within- and cross-case analysis is often seen in case study.
	1.4. Is the interpretation of results sufficiently substantiated by data? Explanations The interpretation of results should be supported by the data collected. For example, the quotes provided to justify the themes should be adequate.
	1.5. Is there coherence between qualitative data sources, collection, analysis and interpretation? Explanations There should be clear links between data sources, collection, analysis and interpretation.

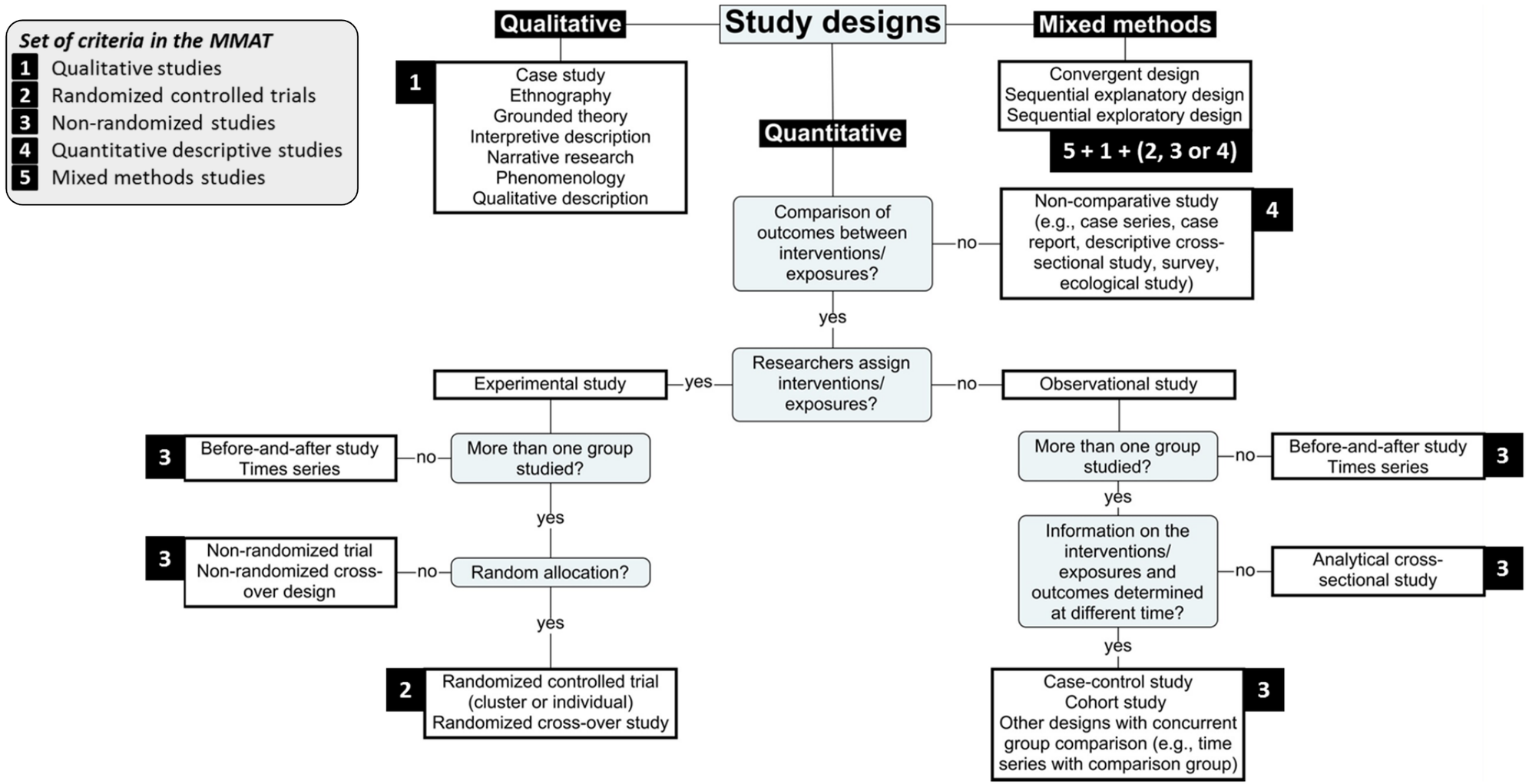
2. Quantitative randomized controlled trials	Methodological quality criteria
Randomized controlled clinical trial: A clinical study in which individual participants are allocated to intervention or control groups by randomization (intervention assigned by researchers). Key references: Higgins and Green (2008); Higgins et al. (2016); Oxford Centre for Evidence-based Medicine (2016); Porta et al. (2014)	2.1. Is randomization appropriately performed? Explanations In a randomized controlled trial, the allocation of a participant (or a data collection unit, e.g., a school) into the intervention or control group is based solely on chance. Researchers should describe how the randomization schedule was generated. A simple statement such as 'we randomly allocated' or 'using a randomized design' is insufficient to judge if randomization was appropriately performed. Also, assignment that is predictable such as using odd and even record numbers or dates is not appropriate. At minimum, a simple allocation (or unrestricted allocation) should be performed by following a predetermined plan/sequence. It is usually achieved by referring to a published list of random numbers, or to a list of random assignments generated by a computer. Also, restricted allocation can be performed such as blocked randomization (to ensure particular allocation ratios to the intervention groups), stratified randomization (randomization performed separately within strata), or minimization (to make small groups closely similar with respect to several characteristics). Another important characteristic to judge if randomization was appropriately performed is allocation concealment that protects assignment sequence until allocation. Researchers and participants should be unaware of the assignment sequence up to the point of allocation. Several strategies can be used to ensure allocation concealment such relying on a central randomization by a third party, or the use of sequentially numbered, opaque, sealed envelopes (Higgins et al., 2016).
	2.2. Are the groups comparable at baseline? Explanations Baseline imbalance between groups suggests that there are problems with the randomization. Indicators from baseline imbalance include: "(1) unusually large differences between intervention group sizes; (2) a substantial excess in statistically significant differences in baseline characteristics than would be expected by chance alone; (3) imbalance in key prognostic factors (or baseline measures of outcome variables) that are unlikely to be due to chance; (4) excessive similarity in baseline characteristics that is not compatible with chance; (5) surprising absence of one or more key characteristics that would be expected to be reported" (Higgins et al., 2016, p. 10).
	2.3. Are there complete outcome data? Explanations Almost all the participants contributed to almost all measures. There is no absolute and standard cut-off value for acceptable complete outcome data. Agree among your team what is considered complete outcome data in your field and apply this uniformly across all the included studies. For instance, in the literature, acceptable complete data value ranged from 80% (Thomas et al., 2004; Zaza et al., 2000) to 95% (Higgins et al., 2016). Similarly, different acceptable withdrawal/dropouts rates have been suggested: 5% (de Vet et al., 1997; MacLehose et al., 2000), 20% (Sindhu et al., 1997; Van Tulder et al., 2003) and 30% for a follow-up of more than one year (Viswanathan and Berkman, 2012).
	2.4. Are outcome assessors blinded to the intervention provided? Explanations Outcome assessors should be unaware of who is receiving which interventions. The assessors can be the participants if using participant reported outcome (e.g., pain), the intervention provider (e.g., clinical exam), or other persons not involved in the intervention (Higgins et al., 2016).
	2.5 Did the participants adhere to the assigned intervention? Explanations To judge this criterion, consider the proportion of participants who continued with their assigned intervention throughout follow-up. "Lack of adherence includes imperfect compliance, cessation of intervention, crossovers to the comparator intervention and switches to another active intervention." (Higgins et al., 2016, p. 25).

3. Quantitative non-randomized studies	Methodological quality criteria
Non-randomized studies are defined as any quantitative studies estimating the effectiveness of an intervention or studying other exposures that do not use randomization to allocate units to comparison groups (Higgins and Green, 2008). Common designs include (this list if not exhaustive): **Non-randomized controlled trials** The intervention is assigned by researchers, but there is no randomization, e.g., a pseudo-randomization. A non-random method of allocation is not reliable in producing alone similar groups. **Cohort study** Subsets of a defined population are assessed as exposed, not exposed, or exposed at different degrees to factors of interest. Participants are followed over time to determine if an outcome occurs (prospective longitudinal). **Case-control study** Cases, e.g., patients, associated with a certain outcome are selected, alongside a corresponding group of controls. Data is collected on whether cases and controls were exposed to the factor under study (retrospective). **Cross-sectional analytic study** At one particular time, the relationship between health-related characteristics (outcome) and other factors (intervention/exposure) is examined. E.g., the frequency of outcomes is compared in different population subgroups according to the presence/absence (or level) of the intervention/exposure. Key references for non-randomized studies: Higgins and Green (2008); Porta et al. (2014); Sterne et al. (2016); Wells et al. (2000)	3.1. Are the participants representative of the target population? Explanations Indicators of representativeness include: clear description of the target population and of the sample (inclusion and exclusion criteria), reasons why certain eligible individuals chose not to participate, and any attempts to achieve a sample of participants that represents the target population.
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)? Explanations Indicators of appropriate measurements include: the variables are clearly defined and accurately measured; the measurements are justified and appropriate for answering the research question; the measurements reflect what they are supposed to measure; validated and reliability tested measures of the intervention/exposure and outcome of interest are used, or variables are measured using 'gold standard'.
	3.3. Are there complete outcome data? Explanations Almost all the participants contributed to almost all measures. There is no absolute and standard cut-off value for acceptable complete outcome data. Agree among your team what is considered complete outcome data in your field (and based on the targeted journal) and apply this uniformly across all the included studies. For example, in the literature, acceptable complete data value ranged from 80% (Thomas et al., 2004; Zaza et al., 2000) to 95% (Higgins et al., 2016). Similarly, different acceptable withdrawal/dropouts rates have been suggested: 5% (de Vet et al., 1997; MacLehose et al., 2000), 20% (Sindhu et al., 1997; Van Tulder et al., 2003) and 30% for follow-up of more than one year (Viswanathan and Berkman, 2012).
	3.4. Are the confounders accounted for in the design and analysis? Explanations Confounders are factors that predict both the outcome of interest and the intervention received/exposure at baseline. They can distort the interpretation of findings and need to be considered in the design and analysis of a non-randomized study. Confounding bias is low if there is no confounding expected, or appropriate methods to control for confounders are used (such as stratification, regression, matching, standardization, and inverse probability weighting).
	3.5 During the study period, is the intervention administered (or exposure occurred) as intended? Explanations For intervention studies, consider whether the participants were treated in a way that is consistent with the planned intervention. Since the intervention is assigned by researchers, consider whether there was a presence of contamination (e.g., the control group may be indirectly exposed to the intervention) or whether unplanned co-interventions were present in one group (Sterne et al., 2016). For observational studies, consider whether changes occurred in the exposure status among the participants. If yes, check if these changes are likely to influence the outcome of interest, were adjusted for, or whether unplanned co-exposures were present in one group (Morgan et al., 2017).

4. Quantitative descriptive studies	Methodological quality criteria
Quantitative descriptive studies are "concerned with and designed only to describe the existing distribution of variables without much regard to causal relationships or other hypotheses" (Porta et al., 2014, p. 72). They are used to monitoring the population, planning, and generating hypothesis (Grimes and Schulz, 2002). Common designs include the following single-group studies (this list if not exhaustive): **Incidence or prevalence study without comparison group** In a defined population at one particular time, what is happening in a population, e.g., frequencies of factors (importance of problems), is described (portrayed). **Survey** "Research method by which information is gathered by asking people questions on a specific topic and the data collection procedure is standardized and well defined." (Bennett et al., 2011, p. 3). **Case series** A collection of individuals with similar characteristics are used to describe an outcome. **Case report** An individual or a group with a unique/unusual outcome is described in detail. Key references: Critical Appraisal Skills Programme (2017); Draugalis et al. (2008)	4.1. Is the sampling strategy relevant to address the research question? Explanations Sampling strategy refers to the way the sample was selected. There are two main categories of sampling strategies: probability sampling (involve random selection) and non-probability sampling. Depending on the research question, probability sampling might be preferable. Non-probability sampling does not provide equal chance of being selected. To judge this criterion, consider whether the source of sample is relevant to the target population; a clear justification of the sample frame used is provided; or the sampling procedure is adequate.
	4.2. Is the sample representative of the target population? Explanations There should be a match between respondents and the target population. Indicators of representativeness include: clear description of the target population and of the sample (such as respective sizes and inclusion and exclusion criteria), reasons why certain eligible individuals chose not to participate, and any attempts to achieve a sample of participants that represents the target population.
	4.3. Are the measurements appropriate? Explanations Indicators of appropriate measurements include: the variables are clearly defined and accurately measured, the measurements are justified and appropriate for answering the research question; the measurements reflect what they are supposed to measure; validated and reliability tested measures of the outcome of interest are used, variables are measured using 'gold standard', or questionnaires are pre-tested prior to data collection.
	4.4. Is the risk of nonresponse bias low? Explanations Nonresponse bias consists of "an error of nonobservation reflecting an unsuccessful attempt to obtain the desired information from an eligible unit." (Federal Committee on Statistical Methodology, 2001, p. 6). To judge this criterion, consider whether the respondents and non-respondents are different on the variable of interest. This information might not always be reported in a paper. Some indicators of low nonresponse bias can be considered such as a low nonresponse rate, reasons for nonresponse (e.g., noncontacts vs. refusals), and statistical compensation for nonresponse (e.g., imputation). The nonresponse bias is might not be pertinent for case series and case report. This criterion could be adapted. For instance, complete data on the cases might be important to consider in these designs.
	4.5. Is the statistical analysis appropriate to answer the research question? Explanations The statistical analyses used should be clearly stated and justified in order to judge if they are appropriate for the design and research question, and if any problems with data analysis limited the interpretation of the results.

5. Mixed methods studies	Methodological quality criteria
Mixed methods (MM) research involves combining qualitative (QUAL) and quantitative (QUAN) methods. In this tool, to be considered MM, studies have to meet the following criteria (Creswell and Plano Clark, 2017): (a) at least one QUAL method and one QUAN method are combined; (b) each method is used rigorously in accordance to the generally accepted criteria in the area (or tradition) of research invoked; and (c) the combination of the methods is carried out at the minimum through a MM design (defined *a priori*, or emerging) and the integration of the QUAL and QUAN phases, results, and data. Common designs include (this list if not exhaustive): **Convergent design** The QUAL and QUAN components are usually (but not necessarily) concomitant. The purpose is to examine the same phenomenon by interpreting QUAL and QUAN results (bringing data analysis together at the interpretation stage), or by integrating QUAL and QUAN datasets (e.g., data on same cases), or by transforming data (e.g., quantization of qualitative data). **Sequential explanatory design** Results of the phase 1 - QUAN component inform the phase 2 - QUAL component. The purpose is to explain QUAN results using QUAL findings. E.g., the QUAN results guide the selection of QUAL data sources and data collection, and the QUAL findings contribute to the interpretation of QUAN results. **Sequential exploratory design** Results of the phase 1 - QUAL component inform the phase 2 - QUAN component. The purpose is to explore, develop and test an instrument (or taxonomy), or a conceptual framework (or theoretical model). E.g., the QUAL findings inform the QUAN data collection, and the QUAN results allow a statistical generalization of the QUAL findings. Key references: Creswell et al. (2011); Creswell and Plano Clark, (2017); O'Cathain (2010)	5.1. Is there an adequate rationale for using a mixed methods design to address the research question? Explanations The reasons for conducting a mixed methods study should be clearly explained. Several reasons can be invoked such as to enhance or build upon qualitative findings with quantitative results and vice versa; to provide a comprehensive and complete understanding of a phenomenon or to develop and test instruments (Bryman, 2006).
	5.2. Are the different components of the study effectively integrated to answer the research question? Explanations Integration is a core component of mixed methods research and is defined as the "explicit interrelating of the quantitative and qualitative component in a mixed methods study" (Plano Clark and Ivankova, 2015, p. 40). Look for information on how qualitative and quantitative phases, results, and data were integrated (Pluye et al., 2018). For instance, how data gathered by both research methods was brought together to form a complete picture (e.g., joint displays) and when integration occurred (e.g., during the data collection-analysis or/and during the interpretation of qualitative and quantitative results).
	5.3. Are the outputs of the integration of qualitative and quantitative components adequately interpreted? Explanations This criterion is related to meta-inference, which is defined as the overall interpretations derived from integrating qualitative and quantitative findings (Teddlie and Tashakkori, 2009). Meta-inference occurs during the interpretation of the findings from the integration of the qualitative and quantitative components, and shows the added value of conducting a mixed methods study rather than having two separate studies.
	5.4. Are divergences and inconsistencies between quantitative and qualitative results adequately addressed? Explanations When integrating the findings from the qualitative and quantitative components, divergences and inconsistencies (also called conflicts, contradictions, discordances, discrepancies, and dissonances) can be found. It is not sufficient to only report the divergences; they need to be explained. Different strategies to address the divergences have been suggested such as reconciliation, initiation, bracketing and exclusion (Pluye et al., 2009b). Rate this criterion 'Yes' if there is no divergence.
	5.5. Do the different components of the study adhere to the quality criteria of each tradition of the methods involved? Explanations The quality of the qualitative and quantitative components should be individually appraised to ensure that no important threats to trustworthiness are present. To appraise 5.5, use criteria for the qualitative component (1.1 to 1.5), and the appropriate criteria for the quantitative component (2.1 to 2.5, or 3.1 to 3.5, or 4.1 to 4.5). The quality of both components should be high for the mixed methods study to be considered of good quality. The premise is that the overall quality of a mixed methods study cannot exceed the quality of its weakest component. For example, if the quantitative component is rated high quality and the qualitative component is rated low quality, the overall rating for this criterion will be of low quality.

Algorithm for selecting the study categories to rate in the MMAT*



.

*Adapted from National Institute for Health Care Excellence. (2012). *Methods for the development of nice public health guidance*. London: National Institute for Health and Care Excellence; and Scottish Intercollegiate Guidelines Network. (2017). *Algorithm for classifying study design for questions of effectiveness*. Retrieved December 1, 2017, from http://www.sign.ac.uk/assets/study_design.pdf.

Appendix E: Completed MMAT Form for Individual Studies (only relevant section included)

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018
Joseph & Powell (2022)

Category of study designs	Methodological quality criteria	Responses			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?	Yes			
	S2. Do the collected data allow to address the research questions?	Yes			
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?	Yes			
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?	Yes			
	3.3. Are there complete outcome data?	Yes			
	3.4. Are the confounders accounted for in the design and analysis?	Yes			The researchers do acknowledge some limits to their study design, however they may not have accounted for all the possible confounders, which is impossible
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?	Yes			

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018

Malherbe (2017)

Category of study designs	**Methodological quality criteria**	**Responses**			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?	Yes			
	S2. Do the collected data allow to address the research questions?	Yes			
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?		No		Very small sample size of only 7 participants.
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?	Yes			offers results on all the variables that are being studied.
	3.3. Are there complete outcome data?	Yes			Attrition reasons reported and complete data provided for other participants.
	3.4. Are the confounders accounted for in the design and analysis?		No		
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?	Yes			

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018

Duranovic, Senka, & Babic-Gavric (2018)

Category of study designs	**Methodological quality criteria**	**Responses**			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?	Yes			
	S2. Do the collected data allow to address the research questions?	Yes			
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?	Yes			Suffiently big sample size with dyslexic population
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?	Yes			Reading rate and accuracy measured using 5 conditions of reading with conditions being a within subject factor.
	3.3. Are there complete outcome data?	Yes			
	3.4. Are the confounders accounted for in the design and analysis?	Yes			Every sentence appeared in every condition for all the children to avoid confounding the sentences with the font/spacing conditions.
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?	Yes			

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018
Bachmann & Mengheri (2018)

Category of study designs	**Methodological quality criteria**	**Responses**			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?	Yes			
	S2. Do the collected data allow to address the research questions?	Yes			
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?	Yes			Large sample size with 533 participants
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?	Yes			
	3.3. Are there complete outcome data?	Yes			Attrition of 131 participants but reasons clearly stated.
	3.4. Are the confounders accounted for in the design and analysis?		No		Third version of the stimuli using Times New Roman with expanded spacing is missing.
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?	Yes			

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018

Powell & Trice (2020)

Category of study designs	**Methodological quality criteria**	**Responses**			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?	Yes			
	S2. Do the collected data allow to address the research questions?	Yes			
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?	Yes			Sufficient sample size with 36 students. Age and gender fairly distributed.
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?	Yes			Recorded time and mistakes made in reading 3 stories each in a different typeface to measure reading rate and accuracy. Modeled after: Flesch-Kincaid Grade Level Scale and Qualitative Reading Inventory.
	3.3. Are there complete outcome data?		No		Complete data and scores not available
	3.4. Are the confounders accounted for in the design and analysis?			Can't tell	Confounders not clearly mentioned or explained.
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?	Yes			

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018
Wery & Diliberto (2017)

Category of study designs	**Methodological quality criteria**	**Responses**			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?	Yes			
	S2. Do the collected data allow to address the research questions?	Yes			
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?		No		Small sample size of only 12 from 1 school only
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?	Yes			
	3.3. Are there complete outcome data?	Yes			Attrition of 1 student reported and reason given.
	3.4. Are the confounders accounted for in the design and analysis?			Can't tell	
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?	Yes			

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018
Ramsey (2014)

Category of study designs	Methodological quality criteria	Responses			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?	Yes			
	S2. Do the collected data allow to address the research questions?	Yes			
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?	Yes			34 students through grades 8-12 and even gender distribution.
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?	Yes			Reading performance assessment tools used in schools: curriculum-based measurements (CBM)
	3.3. Are there complete outcome data?		No		Statistical analysis of data not reported.
	3.4. Are the confounders accounted for in the design and analysis?		No		Individuals with reading disabilities vary greatly in skills and symptom presentation that were not controlled for.
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?	Yes			

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018
Galliussi, Perondi, Chia, Gerbino, & Bernardis (2020)

Category of study designs	**Methodological quality criteria**	**Responses**			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?	Yes			
	S2. Do the collected data allow to address the research questions?	Yes			
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?	Yes			
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?			Can't tell	
	3.3. Are there complete outcome data?	Yes			
	3.4. Are the confounders accounted for in the design and analysis?	Yes			The repetition of the same text across the 8 conditions was avoided in order to prevent learning effects.
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?	Yes			

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018
Damiano, Gena, & Venturini (2019)

Category of study designs	**Methodological quality criteria**	**Responses**			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?	Yes			
	S2. Do the collected data allow to address the research questions?	Yes			
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?	Yes			26 participants: all with dyslexia
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?		No		The effect of the font was not measured in a isolated way but together with many other factors
	3.3. Are there complete outcome data?	Yes			
	3.4. Are the confounders accounted for in the design and analysis?		No		Confounders not mentioned or accounted for.
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?	Yes			

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018
Heiner & Tanja (2017)

Category of study designs	**Methodological quality criteria**	**Responses**			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?	Yes			
	S2. Do the collected data allow to address the research questions?	Yes			
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?	Yes			47 participants all with dyslexia
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?	Yes			Eye tracking to measure speed of reading.
	3.3. Are there complete outcome data?		No		Not enough data to calculate effect size
	3.4. Are the confounders accounted for in the design and analysis?			Can't tell	
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?	Yes			

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018

Kuster, Sanne, Van Weerdenburg, Marjolijn, Gompel, Marjolein, & Bosman, Anna. (2018)

Category of study designs	**Methodological quality criteria**	**Responses**			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?	Yes			
	S2. Do the collected data allow to address the research questions?	Yes			
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
2. Quantitative randomized controlled trials	2.1. Is randomization appropriately performed?	Yes			Participants were randomly assigned to one of three Font-order conditions
	2.2. Are the groups comparable at baseline?	Yes			
	2.3. Are there complete outcome data?	Yes			
	2.4. Are outcome assessors blinded to the intervention provided?			Can't tell	No information provided
	2.5 Did the participants adhere to the assigned intervention?			Can't tell	No information provided

Part I: Mixed Methods Appraisal Tool (MMAT), version 2018

Md Mahtar (2018)

Category of study designs	**Methodological quality criteria**	**Responses**			
		Yes	No	Can't tell	Comments
Screening questions (for all types)	S1. Are there clear research questions?	Yes			
	S2. Do the collected data allow to address the research questions?	Yes			
	Further appraisal may not be feasible or appropriate when the answer is 'No' or 'Can't tell' to one or both screening questions.				
3. Quantitative non-randomized	3.1. Are the participants representative of the target population?	Yes			36 participants all with varying extents of dyslexia severity.
	3.2. Are measurements appropriate regarding both the outcome and intervention (or exposure)?	Yes			Eye tracing to measure reading speed (fixation duration and number of fixations), other studies have used similar measurements.
	3.3. Are there complete outcome data?	Yes			
	3.4. Are the confounders accounted for in the design and analysis?	Yes			Erratic eye gaze accounted for
	3.5. During the study period, is the intervention administered (or exposure occurred) as intended?	Yes			

Appendix F: PRISMA Flow Diagram (Completed)

