The effect of coloured overlays and lenses on reading: a systematic review of the literature

Griffiths, P.G., [1], Taylor, R.H.,[2], Henderson, L.M.,[3] & Barrett, B.T.[4]

[1]Department of Ophthalmology, St Bernard's Hospital, Gibraltar

[2]Department of Ophthalmology, York Teaching Hospital Foundation Trust, York, UK

[3]Department of Psychology, University of York, UK

[4]School of Optometry & Vision Science, University of Bradford, Bradford, UK

Corresponding author:
Brendan Barrett,
School of Optometry & Vision Science,
University of Bradford,
Richmond Road, Bradford,
West Yorkshire BD7 1DP,
UK.

Abstract

Purpose: There are many anecdotal claims that coloured lenses and overlays improve reading performance and there is a substantial literature on the topic of whether reading performance is enhanced through the use of colour. Here we present the results of a systematic review of this literature and examine the quality of the evidence concerning the assertion that reading can benefit from use of coloured overlays or lenses.

Methods: We systematically reviewed the literature concerning the effect of coloured lenses or overlays on reading performance by searching the PsychInfo, Medline and Embase databases. Our searches revealed 51 published items (containing 54 data sets). Different systems are in use for issuing coloured overlays or lenses and we reviewed the evidence under four separate system headings (*Intuitive, Irlen, Harris/Chromagen* and *Other*). We classified each published item using the Cochrane *Risk of Bias* tool.

Results: Although the different colour systems have been subjected to different amounts of scientific scrutiny, the results do not differ according to the system type, or whether the sample under investigation have been classified as having visual stress (or a similarly defined condition), reading difficulty, or both. The majority of studies are subject to 'high' or 'uncertain' risk of bias in one or more key aspects of study design or outcome. Studies at lower risk from bias offered less support for the benefit of colour on reading ability. Whilst many studies report improvements with colour, the effect size is generally small and/or similar to the improvement found with a placebo condition. We discuss the strengths and shortcomings of the published literature and, whilst acknowledging the difficulties associated with conducting trials of this type, offer some suggestions about how future trials might be conducted.

Conclusions: Consistent with previous reviews and advice from several professional bodies, we conclude that the use of coloured lenses or overlays to ameliorate reading difficulties cannot be endorsed and that any benefits reported by individuals in clinical settings are likely to be the result of placebo, practice or Hawthorne effects.

Introduction

In 1980, Olive Meares, a school teacher from New Zealand, described visual perceptual difficulties reported by some children when learning to read. These difficulties were apparently alleviated by placing sheets of coloured plastic, such as Perspex, over text(1). Separately, American psychologist, Helen Irlen, documented the use of coloured overlays and lenses to the same effect(2). The set of symptoms described in these publications became known as 'Irlen syndrome', 'Meares-Irlen syndrome', or 'scotopic sensitivity syndrome', the symptoms of which include (but are not restricted to) subjective reports of movement or blurring of print, doubling of letters, illusions of colour, glare from printed material, headaches, and eye-strain during reading.

In the UK, research into this area has been led by Professor Arnold Wilkins, who coined the term 'visual stress' (used throughout the present paper*) and developed one of the coloured filter systems reviewed here. It has been claimed by Wilkins and colleagues that visual stress may be one cause of reading difficulties(3). Wilkins acknowledges the considerable overlap in symptoms between visual stress and asthenopia that arises due to other reasons including uncorrected refractive error, and oculomotor/binocular vision anomalies (4). Once basic optometric needs have been addressed, it has been argued that the use of coloured overlays and lenses can have a positive impact on the symptoms of visual stress in affected individuals, which, in turn, may lead to better reading performance, in particular higher reading speed(5).

Despite 35 years having elapsed since the initial description, neither the International Classification of Disease (ICD-10; World Health Organisation) nor the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association) list visual stress as a recognised disorder. Similarly, neither of these widely used diagnostic manuals make any reference to visual-perceptual distortions as being associated with reading impairment. The ability of coloured filters to improve reading performance in individuals who report symptoms of visual stress has been contested(6)(7)(8)(9)(10)(11) and the practice has even been listed among 'neuromyths in education'(12). The two narrative reviews that are broadly supportive of the use of colour adopt conflicting viewpoints. A review by Wilkins 2002 argues that precisely the right colour is required(4) whereas a review by Stein 2014

*We have used the term 'Visual Stress' throughout this review to signify the symptoms which are 'treated' using coloured lenses or overlays. Although this term was introduced by Prof. Wilkins and is associated with the Intuitive colour system, here we use "Visual Stress" (VS) more generally to describe any visual symptoms which may respond to colour intervention, regardless of the colour system (Irlen, Intuitive etc). In the absence of precise definitions for the conditions which the various colour systems purport to treat/manage, we could have adopted one of the other terms in common usage, e.g. Meares-Irlen syndrome, Irlen-syndrome or Scotopic Sensitivity Syndrome. Thus, our use of 'Visual Stress' does not signify support for the Intuitive system (or any system) over another system.

argues that blue or yellow overlays suffice(13). Taken together these narrative reviews raise doubt over whether the available evidence supports their widespread use. Indeed, it has been suggested that any improvement in reading performance seen in individuals who usecoloured overlays or lenses is the result of a placebo effect, whereby belief that a product will help is enough to give the user the impression of improvement(14).

Despite suspicions about the true effectiveness of coloured overlays and lenses, these ‘reading aids’ have received widespread media exposure and their use is regularly accepted in schools and higher education institutions. The use of coloured overlays and lenses continue to be endorsed by some dyslexia charity websites(15). The issuing of coloured overlays has become embedded, to a greater or lesser extent, into the practice of a range of professionals in the UK including teachers, Educational Psychologists, optometrists, and NHS orthoptic departments. Furthermore, there is now an array of colour systems on offer (e.g., Irlen, Intuitive, ChromaGen/Harris), with proponents of each system claiming that their system provides an effective testing and management approach(16)(2).

What is the status of the evidence to support use of coloured overlays or lenses (spectacles or contact lenses) for the purposes of helping an individual to read, or to learn to read? We have brought together a multidisciplinary team spanning psychology, optometry and ophthalmology to conduct a systematic search of the literature to address this question. A systematic review of the literature by Albon et al. 2008 concluded that the available evidence was of too low a quality to reach firm conclusions about the effectiveness, and cost effectiveness of coloured lenses for reading disability(6). Similarly a systematic review Galuschka et al in 2014 could not prove any positive effect of coloured lenses or overlays on literary achievement(17). This review is the first attempt to systematically and separately evaluate the quality of the evidence for each of the available colour systems: Irlen, Intuitive, ChromaGen/Harris and other systems that have received less attention, or are less widely used.

Method

Literature Searches

We conducted our searches using Medline, PsycInfo and Embase. The searches utilised three ‘concepts’ identified during a preliminary search of the literature: [Concept 1] colour *(('Colour'/exp OR colour) OR (coloured OR colored) OR (lens OR lenses) OR (overlay OR overlays) OR (filter OR filters) OR (tint OR tinted))*, [Concept 2] reading *(reading OR text OR print OR printed OR words OR word OR writing OR write)*, [Concept 3] reading difficulties/visual stress terms *(Irlen OR ‘Meares Irlen syndrome’ OR ‘Irlen syndrome’ OR Meares OR 'visual stress' OR dyslexia OR 'dyslexia acquired' OR ‘Learning disorders’ OR 'specific learning disability' OR 'specific learning disorder' OR 'specific language disorder' OR 'specific language disability' OR 'specific language impairment' OR 'specific learning impairment' OR 'specific learning disorder’ OR 'specific learning disability’ OR magnocellular OR 'scotopic sensitivity syndrome' OR misvis)*. Figure 1 shows a Prisma flow chart of the number of published papers identified by these searches, and from other sources (e.g. through contact with published authors in the topic area and searching reference lists of the search-identified papers).

Inclusion & Exclusion Criteria

Figure 1 shows how the final number of studies that we examined in detail in this review was arrived at (n=51). We included experimental studies, which featured primary data on at least one measure of reading ability or reading-related activity (e.g. reading unconnected words), and which described the effect on such measures when filters (i.e., coloured spectacle lenses, contact lenses or overlays) were worn or used. Studies were only included if they incorporated a control group, and employed either a crossover design (where participants undergo a number of different treatments or exposures and thus act as their own controls) or a parallel design (where the treatment group were compared to another group, for example, a group that did not receive colour, or received a non-optimal colour). Studies of both adults and children were included. There were no restrictions on the baseline reading ability of the study sample, or on how the sample was identified. Some samples comprised ‘poor’ readers identified in remedial settings (e.g. classroom settings where the students were present because of known reading difficulties) whereas others comprised unselected samples (e.g., where all children in a school/year group participated). We noted whether the study looked at the ‘overlap group’ (i.e., individuals diagnosed with ‘visual stress’ *and* reading difficulties), or whether only ‘visual stress’ or reading difficulties were diagnosed in the sample under test. The review was not restricted to any one colour ‘system’ but instead included all studies where the effect of colour on reading performance had been examined provided this was in the form of overlays and lenses. The results for each system were analysed separately. Studies were included irrespective of whether or not symptoms, or changes in symptoms, associated with use of colour were reported. The literature searches also revealed six unpublished PhD theses (i.e. from the so-called 'grey literature') that we included in our review.

Two authors (PG and LH) independently reviewed the 244 items identified and excluded papers that did not fit the above-mentioned inclusion criteria. Although the Medline, PsychInfo and Embase databases are American and the vast majority of the journals they index are English-language journals, non-English language journals are also indexed. Of the 244 items, only two were not in English. While there was substantial overlap in the choices made, any paper that was selected by only one reviewer was also included. This exercise reduced the list to 49 items and neither of the non-English articles met the inclusion criteria (Figure 1). Individuals who had made a significant contribution to research on colour and visual stress or reading difficulties were invited (by e-mail) to view the list of 49 papers and to make suggestions for additional papers that should be included. Finally, we examined the reference lists of the 49 papers identified by the search-engines. Through these means we identified a further four papers which had not appeared in our search-engine results: Ray, Fowler & Stein(18), Loew et al(19) Monger, Wilkins & Allen(20) and Rosenfeld et al(21). Peer reviewers for this review suggested a further four papers: Wilkins et al(22), Wilkins & Lewis(23), Northway(24) and Allen(25). A total of 57 papers was therefore reviewed (Figure 1).

Procedures for Review

The four authors worked in pairs (LH & RT; PG & BB). The list of 57 papers was split based upon an alphabetical listing of the first author of each published article (A-K, LH & RT; L-Z, PG & BB). For each item, the pairs completed a form which gathered the following information from each item: a brief description of the study and design; whether or not there was a control group; which colour systems had been employed and whether lenses or overlays had been used; what the independent and dependent variables were (the latter had to feature some measure of reading in order for the study to be included) and which measures of reading had been used

Each published item was evaluated according to threats of internal and external validity, in accordance with the Cochrane Collaboration's tools for assessing bias(26). Internal validity refers to the risk of bias resulting from study design and reporting. External validity refers to the degree to which the results, even if at low risk of internal bias, can be generalised to different settings and populations.

The domains of selection bias (e.g., judgements on the method for random sequence generation and whether intervention allocations could have been foreseen before or during enrolment), performance bias (e.g., when participants and personnel have knowledge of the intervention (e.g. experimental or placebo tint) used during the study), detection bias (e.g., when assessors have knowledge of the allocated intervention), attrition bias (e.g., bias arising from loss of participants from the study) and reporting bias (e.g., when only selective outcome measures are reported; the existence of a pre-trial protocol serves as evidence of 'low' risk of reporting bias). In keeping with advice from the Cochrane Collaboration, we did not sum the risk judgments to derive a global 'risk of bias' score for each study. This is because a study may be at serious risk of bias if the bias judgements are low in all but one area.

Many of the studies we reviewed were crossover studies which are low risk of confounding due to problems with random sequence generation, allocation concealment and similarity of groups at baseline. We recorded these studies as being at low risk of bias in these domains even if a detailed account was not given for the method of sequence generation and allocation concealment. We considered studies that used disconnected text rather naturalistic text of the sort encountered in everyday life to have limited external validity also studies that recruited participants from specialist clinics such as those at the Institute of Optometry or Dyslexia Research Trust were recorded as hiving high or uncertain external bias because they may not be representative of the general population of poor readers and may have been attracted to those clinics because of a prior belief in the effectiveness of coloured lenses and overlays.

As well as making risk-of-bias judgments for each data set, we gathered information from each of the items we reviewed about the fall-off in the use of colour over time where such information was provided. The form was initially completed by one member of each team and then reviewed by the second member. In the event that a pair was not in agreement, the study in question was referred to the second pair for discussion and agreement. To ensure consistency between the pairs, each pair selected the three of the papers from their list which had generated the most discussion about the risk-of-bias judgments and invited the other pair to reach their own, independent judgements about the risk of bias. No systematic differences in the application of criteria for the bias judgments were identified.

From the list of 57, four items were excluded following review because they did not feature a control group(27)(28)(29)(30) Another was excluded as it featured the use of coloured light rather than lenses or overlays(19) and the paper by Wilkins and Neary(31) was also excluded as no formal measure of reading was included.

Where items contained several studies, we viewed them as separate data sets and included them only if they satisfied our criteria. There were two items where this occurred, Jeanes et al.1997 Studies 4 and 6 (32), Lightstone et al. 1999 studies 1 and 2(33) and Wilkins and Lewis 2001; Studies 1-3 (34). Therefore, there were additional participant data sets for Jeanes et al (+1) and Wilkins and Lewis et al (+2) and Lightstone et al. (+1). Based on the same criteria, it was noted that two publications by Robinson & Forman 1999 reported on the same samples, so we considered them as one for the purposes of assessing bias making one less data set(35)(36). Overall, therefore, there were three additional data sets in the 51 items that we reviewed, leaving a grand total of 54 data sets.

Figure 1: Flow diagram showing our search strategy and how this review came to examine 51 items of literature (but 54 data sets) concerning the impact of coloured overlays or lenses on reading.

Results

A synopsis of the number of items we identified by our searches and of our reasons for excluding a proportion of these is provided in Figure 1. In total, 54 datasets including 2690 participants were analysed; 23 studying the Intuitive system, 15 studying Irlen, 4 studies of Chromagen/Harris and 12 of non-commercial filters (referred to here as 'Other'). Of these, 55% defined a 'condition' that was under evaluation in the study (or they referred to a definition), but many of these definitions were inconsistent from one publication to another. Eighteen of the studies were carried out in a single experimental session; 23 were carried out over two sessions (ranging from 1-36 weeks where the data were provided, mean 11 weeks) and 6 were carried out over 3 sessions; in one study (Blaskey, 1990) the timing is unclear(37). Our evaluation of the items we identified that were relevant to our question is presented separately for the different colour systems (Intuitive, Irlen, Chromagen and 'Other'). Tables 1 & 2 contain summary features and 'risk of bias' judgements, respectively, for the separate systems.

Intuitive Overlays and Lenses

Background

Influenced by the work of the Irlen Foundation, Professor Arnold Wilkins developed Intuitive Overlays (IOO Sales Ltd., London, UK). Although superficially similar to the Irlen system, it has been claimed that Intuitive overlays sample the colour space more evenly, systematically and efficiently(32). Unfortunately, however, the graphs displayed to support this assertion plot the chromaticity coordinates for Irlen and Intuitive overlays on different scales that make comparison difficult (32 pages 534-535).

Intuitive Overlays consist of nine coloured overlays and one grey overlay. Overlays are selected by making pair-wise comparisons until the optimal single tint is found. Wilkins subsequently developed the Intuitive Colorimeter (Cerium Optical Products, Tenterden Kent) to more precisely define the optimum colour required to ameliorate visual stress and for prescribing tinted spectacle lenses (known as Precision Tinted Lenses). The Intuitive Colorimeter consists of an illuminated chamber in which random letters arranged to resemble text can be viewed through an aperture. The hue, saturation and brightness can all be varied independently by the examiner according to the subjective responses of the person being tested. Glasses can be ordered based on this tint which may be different to that required for overlays(33). Precision tinted lenses are generally only made up following sustained use of overlays. A detailed account of the technique for determining optimum tints for both Intuitive Overlays and Precision Tinted Lenses is given by Allen, Evans and Wilkins(38).

The Wilkins Rate of Reading Test (WRRT) is a test that is frequently used to assess the benefit of coloured overlays and lenses(22). The test comprises passages of randomly ordered, high-frequency words, printed in a small font and the outcome measure is the number of words correctly read per minute. It is not intended to be a test of reading ability *per se* but rather a measure of the extent to which colour can influence reading rate. Reading the WRRT 5%, 8%, 10% faster with the chosen overlay has been used as the criterion for a clinically significant improvement with colour(39). A recent article has suggested that an increase in reading speed of 15% or more may be required to be confident that there is a genuine improvement with coloured overlays(5), although we aware of no studies that use this diagnostic criterion for visual stress. There appears to be no objective reasoning behind the various cut-offs that have been suggested, and even the most stringent criteria remain completely arbitrary. These criteria also need to viewed in the context of the test-retest variability of the WRRT. For example, one study reported that 5% of children read more than 25% faster using Intuitive Overlays(34). However the same paper also reports the test-retest variability of the WRRT. Reading from Figure 2 in that paper (page 47), it appears that, without the use of any overlay, 3.8% of participants read more than 25% faster and 12.3% read more than 15% faster when tested on a second occasion. As a result, even the more rigorous criterion of reading the WRRT 15% faster may still produce large numbers of false positives. It is notable that in any given population (including those with and without visual stress), there is substantial variability in baseline reading performance prior to intervention (Table 1). For example, 20-145 words per minute in children 7-12 years of age(39) and 105-245 in adults(40). Wilkins et al acknowledge this variability, indicating that the highest rate of reading (>160 words per minute) can be more than four times the slowest rate (<40 wpm) in children with similar scholastic reading ability(5). Consequently, we do not know what the normal reading rate is for a particular age group and it is not possible to determine what a ‘normal’ range of improvement would be with an overlay. In line with this evaluation Wilkins et al. state “a confusing array of criteria have been applied and further analysis of WRRT data is required”(5).

It has been argued that changes in reading should be seen first with the WRRT (which is designed to be crowded and visually aversive) and only later using naturalistic text, which depends on higher order reading skills. We found no evidence to support this claim. Furthermore, it could be argued using the same logic that changes would also be seen immediately in psychophysical tests that use gratings at the spatial frequencies that are said to be aversive in visual stress. This has not been observed(41).

General features of trials

Nineteen papers including 23 trials of Intuitive Overlays or Precision Tinted Lenses fit our selection criteria. In most cases these were exploratory studies that contained a crossover trial as part of wider study investigating the use of overlays. All studies were at high risk of bias in at least one (often multiple) domains (Table 2). The diagnostic criteria for visual stress were not consistent, even in papers by the same authors. Criteria included subjective reports of perceptual distortions while reading(42); immediate subjective benefit from overlays(43)(20); reported distortions on viewing a 3 cycles per degree square wave grating(44)(45); subjective reports on the Visual Processing Problems Inventory (VPPI)(46); computerised visual search test performance under visually stressful conditions(47), voluntary sustained use of overlays(32)(33)(48), or reading rate on the WRRT that was 5, 8 or 10% faster with a chosen overlay than without(39). In those studies that used the voluntary sustained use criterion, the duration of overlay use ranged from 3 to 12 weeks(48)(24).

In general, studies using disconnected or naturalistic texts comparing a chosen colour with a placebo colour reported improvements in reading for both conditions compared to baseline, but crucially no significant difference between the placebo and the experimental condition(48)(42). One argument that has been put forward to explain this effect is that improvements in reading naturalistic text are only observed after prolonged periods of reading. This is based on a study by Tyrell et al.(49). However this exploratory study used Irlen not Intuitive overlays and is at high risk of bias in a number of domains discussed in detail in the section on the Irlen tinting system.. Another argument that has been put forward is that some studies, for example Wilkins et al. (48), compared the chosen colour with a closely related colour and it is for this that reason no difference was found. However this calls into doubt the need for precision tinting that is claimed by some study authors(5). Furthermore studies comparing chosen tint with a complementary colour did not find any significant improvement in reading naturalistic text(32)(42).

The majority of trials of the Intuitive system have utilised a crossover design which compared reading with a chosen overlay compared to reading with a clear overlay or without an overlay(32)(50)(34)(20)(40)(47)(44)(45)(22)(23)(24)(25)(51), and were therefore at a high risk of bias due to the lack of a placebo control condition. Moreover, it is not made clear if the authors assume that the small improvements in the rate of reading on the WRRT generalise to the reading of naturalistic text.

One head-to-head study compared Intuitive Overlays with another system (i.e., Reading Rulers, Crossbow Education Ltd., UK). This crossover study showed that participants read the WRRT significantly faster with Intuitive overlays (mean: 80.5 words per minute (wpm), SD 27 compared to Eye Level Rulers (75.7wpm,SD 25.8), which consist of only five tints(16). However, this was an unmasked study, the diagnosis of visual stress was based on group reading tests and the clinical significance of small differences in the WRRT is questionable.

Most studies showed a high rate of decay in the use of overlays and lenses over time (32)(34) which brings into question the practicability of a technique that demands the assessment of all poor readers, issuing overlays to as many as 60% for a prolonged period of time in order to identify a subset with clinically significant visual stress who may or may not benefit in terms of reading naturalistic text.

Specific studies

Three studies were primarily designed as clinical trials and had some features of masked randomised control trials (RCTs). These will now be discussed in detail.

In 1994, Wilkins and colleagues published a double masked, placebo-controlled trial using a crossover design(48). The criterion for diagnosing visual stress was voluntary sustained use of overlays for at least 3 weeks. Sixty-eight children aged 11-12 years with dyslexia were recruited. Participants used tinted spectacle lenses prescribed using the optimum colorimeter setting as well as a placebo colour that was just outside the range reported to improve perception. Because the participants did not see the actual lens and an interval of one month was left between testing and receiving the experimental or placebo lens, effective masking was maintained. Participants wore each set of tinted lenses for one month and were tested using the Neale Analysis of Reading Ability(52) at the end of each period. Participants also kept symptom diaries throughout the study. The study was hampered by a high rate of drop-out and a failure to analyse the data on an intention-to-treat basis(53)(54). There was no improvement in reading rate, accuracy or comprehension for 45 out of the 68 participants (66.2%) for whom data were available. Although analysis of the symptom diaries appeared to show a small benefit in favour of the optimum tint, data were only available for 36 out of 68 (52.9%) participants meaning that this study is at high risk of bias. The study also contains some contrary evidence: Overall 22 participants preferred the experimental lenses, but 26 preferred the placebo control lenses. There was also evidence of novelty effects: 31 children preferred the first pair of glasses, whereas only 17 preferred the second pair and 4 expressed no preference.

Bouldoukian et al adopted a different method, this time using overlays(43). The optimum tint was compared with a pale yellow, placebo filter that was manifestly different from the Intuitive Overlays. As result, neither participants nor experimenter were masked to the intervention being used. The placebo overlay was described to the participants as ‘a new filter from the United States where it was thought to be a wonderful discovery' and marked with the words *'Research Model A16 Anti UV IR Filter. Made in the USA'*. The authors reported that by pooling all of the data, there was a 4% increase in speed on the WRRT in the participants when reading with the prescribed overlay as compared to the placebo (an increase of 4 wpm). The assumption that it is possible to match the placebo effect of the experimental intervention with an enhanced placebo is unfounded and as a result this study is at high risk of bias. The problems associated with enhanced placebos are discussed in more detail in the general discussion.

Mitchell et al (2008) used a parallel-groups design(42). Participants had dyslexia and reported visuo-perceptual distortions. Seventeen children received no lenses; 17 children received lenses based on the optimum intuitive colorimeter setting; 15 received lenses of a colour complementary to the chosen colour. Participants were pre- and post -tested using the Irlen Differential Perceptual Schedule and the Neale Analysis of Reading Ability test. All groups showed improvements in reading but there was no significant difference between the placebo and experimental lenses for reading speed, accuracy or comprehension. This study could be criticised for treating dyslexia rather than visual stress. However all subjects reported visuo-perceptual distortions as well as reading difficulties and chose a colour on the Intuitive Colorimeter that minimised those distortions.

Another study of interest did not look at reading directly, but instead examined the effect of colour on spatiotemporal contrast sensitivity and contrast increment thresholds in response to square wave gratings, including the spatial frequencies (2-8 cycles per degree) that are said to be aversive in visual stress(41). Twenty individuals with visual stress who had successfully worn Precision Tinted Lenses were enrolled and compared to 20 control participants without visual stress. There was no significant difference between the control group and the visual stress group in spatiotemporal contrast sensitivity or contrast increment thresholds. Furthermore, in the visual stress group there was no significant difference in the results with and without their chosen tints.

Summary

The results of the exploratory studies of the intuitive system have to be viewed with caution(32)(34). Multiple statistical comparisons and a flexible post-hoc approach to data interpretation leaves these studies at high risk of producing false positive results(55)(56). As a result they should be seen as generating - rather than testing hypotheses. Furthermore, the three studies with features of masked RCTs(48)(42)(51) were each prone to bias. Improvements have been reported with prescribed overlays/lenses, but similar improvements are also found with placebo colours, questioning the need for precision in tinting lenses and overlays.

Irlen

Background

In 1983 Helen Irlen argued that a perceptual disorder not diagnosed by conventional eye examination, and termed (but not limited to) "Irlen Syndrome$_{TM}$" (IS), "scotopic sensitivity syndrome" (SSS), Meares-Irlen Syndrome (MIS) or "visual discomfort", was one of the reasons why some children have difficulties learning to read. Symptoms of IS during reading include a blurring and shadowing of letters and words, a doubling, merging or movement of print, eye strain and fatigue, a restricted span of focus and problems of focussing for an extended period of time(2)(1). It was claimed that the symptom complex of IS could be ameliorated by modifying the wavelength of light reaching the eyes using coloured overlays or lenses, such that the wavelengths and frequencies of the white light spectrum to which the individual is sensitive are filtered out. The reduction of these symptoms is, in turn, hypothesised to lead to an improvement in reading(2).

The Irlen testing procedure involves a series of questions to probe for perceptual difficulties during reading (e.g., distortions and movement of text, light sensitivity, headaches, eyestrain, tiredness, loss of concentration etc.), followed by a series of visual tasks involving counting lines/symbols within high-contrast pictures. Finally, one of 10 (or a combination of) overlays is selected by the individual, following a series of pair-wise comparisons between coloured overlays placed over text. Irlen in her book, *The Irlen Revolution,* argues that only coloured overlays and lenses provided by the Irlen Institute are effective in treating IS(57). However, no scientific evidence supports this claim.

General features of Irlen trials

Nine studies were identified that had features of RCTs(35)(36)(37)(58)(59)(60)(61)(49)(62)(63) we have counted two publications by Robinson & Foreman 1999 which report on the same trial as a single study(35)(36). A consistent definition of Irlen syndrome was applied across these studies and the diagnostic procedures consistently involved the use of the Irlen proprietary testing materials delivered by Irlen trained staff. The Irlen Reading Perceptual Scale (IRPS) and Irlen Differential Perceptual Schedule (IDPS) are frequently used to assess for IS, and consists of three sections: (i) the Irlen Reading Strategies Questionnaire (32 questions, 16 of which are related to reading strategies and reading behaviours such as skipping/re-reading lines, misreading words, losing place in the text, poor comprehension and slow reading; the remaining 16 questions relate to eye strain and fatigue whilst reading, including headaches, dry/itchy/burning eyes, blinking and squinting); (ii) a series of visual tasks (e.g., counting squares in gridded rows; answering questions about distortions whilst observing visual images): and (iii) an assessment of the extent to which performance on these visual tasks and on reading is improved by the use of coloured plastic overlays. It is claimed that the subsections of this assessment have high internal validity and reliability although the evidence to support such claims comes from the unpublished literature on the Irlen website(64).

Where available, the classification rates of Irlen syndrome in various samples of participants are displayed in Table 1. These rates ranged from 46.2%(30) to 96.2%(65) with a median of, 66%.

Two publications contained information about sustained use. Cotton et al studied 60 participants, of whom 38 chose an overlay and 22 (36.7% of the original cohort) were stillusing their overlay 6 weeks later(58). Ritchie et al studied 61 children, 47 were diagnosed with Irlen Syndrome (77%), and of these, 22 (36% of the original cohort) were still using their filter a year later(63).

.*Specific studies of the Irlen system*

Blaskey et al carried out a 'pilot' RCT to ascertain whether Irlen lenses improve comfort when reading and are an effective treatment for improving reading performance when compared to traditional optometric intervention(37). Forty participants were recruited (aged 9-51 years) following a feature about Irlen overlays on CBS news, all of whom tested positive for both 'scotopic sensitivity syndrome' and visual problems, most often binocular or accommodative disorders, based on optometric assessment. Participants were randomly allocated to an IS group (*n*=11), a vision therapy group (*n*=11, with 3 dropping out) or an untreated control group (*n*=8, with 5 dropping out). Neither the participants nor the experimenters were blind to group status. The Irlen syndrome group tried both a prescribed set of coloured lenses and a 'placebo tint' (selected by the optical service) for two weeks, and then selected the lenses that made them feel 'most comfortable and subjectively improved their reading ability' for a further two weeks. Three participants supposedly chose the placebo, but this is not reflected in Table 3 of the paper, where it is stated that eight participants chose the placebo lenses. Symptom scores improved from pre- to post-test for the Irlen syndrome group, but not for the placebo group. Improvements were reported for speeded letter reading and word reading, as measured via the Woodcock Reading Mastery Test, but these were only found in three participants without reading difficulties; the eight participants who had reading difficulties did not show improvements. No improvements were reported for reading comprehension or reading rate. All 11 participants in the Irlen syndrome group continued to show visual anomalies based on the optometric assessment after wearing the lenses, and there were no significant improvements in these measures from preto post-test. In contrast, for the vision therapy group, visual problems were resolved in 7of the 8 cases and significant improvements were found for the optometric and scotopic sensitivity symptom questionnaire scores, and for the Gray Oral Reading Test scores. Thus, in some cases, Irlen syndrome may be effectively alleviated via optometric interventions. No significant changes were reported in the untreated control group. On the basis of these pilot data, the authors conclude that Irlen lenses have negligible effects on reading, even in individuals who have already expressed an interest.

Noble et al examined the effects of Irlen overlays on reading rate, accuracy, fluency and comprehension, via teacher-led screening and assessment(66). Participants were screened from two Grade 3 mainstream schools, rather than being referred or self selected. Seventy-one participants were identified as having Irlen syndrome and competent reading ability as measured via the Woodcock-Johnson III Tests of Achievement(67), although reading ability varied considerably within treatment and control groups. Children from one school (*n*=31) were provided with coloured overlays for 3 months; a waiting list control group children from the other school (*n*=40) received overlays after 3 months. Three children dropped out of the treatment group, however it is unclear at what point this drop-out occurred. Furthermore, "..there were a small number of students in the study who did not use their overlays consistently" (p. 19), but no details are provided on how many children this applied to, or which group they were from. Significant improvements in reading rate, accuracy, fluency and comprehension were reported after three months of use (grade-equivalent score increases of 1 year 2 months to 1 year 7 months), but no further improvements were reported at a 6 month follow-up. The waiting control group showed no significant improvements during the first three months of the study (without overlays), but showed significant gains (grade-equivalent score increases of between 1 year 8 months and 2 years 8 months) during the second three months of the study (with their overlays). Improvements in symptoms were also reported when overlays were used. However, these results are subject to high levels of bias as a consequence of no group allocation concealment, no blinding of participants and experimenters, no blinding of the outcome assessment, and no placebo or control intervention group (Table 2).

Martin et al found a lack of convincing evidence for coloured overlays and lenses as form of treatment for reading difficulties(60). They examined the impact of coloured overlays on a array of "process variables" (i.e., visual processing, phonological processing and working memory) as well as reading accuracy, reading comprehension, non-word reading, sentence comprehension. Three-hundred Tasmanian children, aged 11-12 years, were screened; 58 were selected with reading difficulties and 62 were selected who were reading at a level consistent with their age and IQ. Sixty-two percent of the poor readers were diagnosed with Irlen syndrome and prescribed coloured lenses. Subsequent analysis focused on children with reading difficulties who were diagnosed with Irlen syndrome and prescribed tinted lenses (n=20), children with reading difficulties who were not diagnosed with Irlen syndrome (n=20), and 20 children without reading difficulties or Irlen syndrome. Children were tested prior to intervention, at a 6 month post-test and at a follow-up session after 12 months. There were no significant differences in any of the outcome measures. There were a number of limitations that make these results difficult to interpret: small sample sizes (coupled with a long list of dependent variables); no Irlen syndrome control group and no placebo lens group; no random allocation to groups and no blinding to group status.

O'Connor et al also studied the use of coloured lenses in children with poor reading ability(59). Out of a total of 600 mainstream school children (aged 8-12 years), teachers selected 105 children who were "reading at least 18 months below grade level and whom they considered to have reading ability well below their abilities in other areas" (p. 599). Students who displayed symptoms on the IDPS and "a marked improvement" in reading with an overlay were classified as "scotopic" (n=67/105, 64%) and 25 students (24%) were classified as "non-scotopic"; the remaining 13 children (12%) showed scotopic signs but no improvement with an overlay and were dropped from the study. This suggests a remarkably high prevalence rate of Irlen syndrome amongst poor readers of 64%. Scotopic children were randomly assigned to one of four groups via a stratified randomisation procedure to ensure similarity at baseline: Group A (n=17) received the prescribed coloured overlay; Groups B (n=17) and D (n=16) received a transparent overlay; Group C (n 17) received a randomly coloured (non-prescribed) overlay. Non-scotopic children were randomly assigned to one of two groups: Group E (n= 12) were given transparent overlays and Group F (n=13) were given a random (i.e. a non-prescribed) colour. Children were told that the overlays would "make reading a little easier for them". Pre- and post-tests were administered one week apart and involved the Neale Analysis of Reading Ability(52) and the Formal Reading Inventory(68). The pre-test was omitted for Group D, to control for repeated testing effects. Children who received their chosen colour showed significant improvements of 6.6 months in reading rate, 6.9 months in reading accuracy, and a substantial 19.35 months in reading comprehension; the other groups appeared to decline in their reading performance over the week of the experiment possibly signalling a "nocebo" effect. This study is prone to bias, because participants were not blinded to group status and there is also question as to how representative the "poor readers" were, given the small group sizes and the narrow definition of reading difficulties relative to "..abilities in other areas".

A double-masked RCT was carried out by Ritchie et al, following a 'dyslexia friendly schools' initiative by Inverclyde Council(62). The Irlen Institute agreed to facilitate this study in a mainstream primary school. Ritchie et al. examined 75 children with below average reading ability who were screened by an Irlen trained practitioner. Fourteen children were unable to complete the Irlen screening tasks and were excluded. Of the remaining 61 children, 47 were diagnosed with Irlen syndrome (77%), again suggesting an extremely high prevalence. The first part of the study used a crossover design and compared reading rate (as measured via the WRRT) in 43 children with Irlen syndrome and in 14 without Irlen syndrome. Three of the children with Irlen syndrome who were not masked because they were aware that their optimum colour was for assisting with reading were excluded from the main analysis. All children were tested using the prescribed overlay, a placebo overlay of a complementary colour and a clear overlay. For both the Irlen syndrome and the non-Irlen syndrome group, overlays had no significant effect on reading rate. Twenty-eight percent of children in the Irlen syndrome group read >5% faster with their prescribed overlay than without an overlay, but 32% read more than 5% slower with their overlay. Indeed the data revealed high test-retest variability for the WRRT. For the 3 children who were non-masked showed significant improvements in the WRRT (78%, 58% and 15%), indicative of a placebo effect. However, this latter analysis has been criticised as selective and non-representative(64). Forty-four children with Irlen syndrome were enrolled into a parallel-groups study where 22 were randomised to receive their optimum overlay while 22 received a colourless overlay. There was no significant difference between the two groups for any of the measures of reading or reading comprehension on the Gray Oral Reading Test (GORT)(69). Ritchie et al followed up the same cohort one year later and found that 22 (30%) were still using Irlen overlays or lenses, and of these 18 were available for follow up(63). Ten of the group without IS were available for follow up. Both groups showed deterioration in their Oral Reading Quotients over 12 months. In summary, this is one of the strongest RCTs on Irlen overlays published to date, but nevertheless, is not without limitations. For example, the null result reported here could be a consequence of an inadequately powered design. Furthermore, given the exclusion of three children who showed positive effects, this study may be subject to reporting bias.

Robinson and Foreman (1999a and b) are frequently cited in support of the use coloured overlays as a treatment for reading difficulties(35)(36); however, the results of this study do not support this claim. One hundred and thirteen subjects aged 9-13 years with poor reading and IS (according to the Scotopic Sensitivity Screening Manual) were randomly allocated to one of three experimental groups: an optimum (diagnosed) tint group, a blue tint group, or a placebo tint group (i.e., a similar colour to the optimum tint but which did not ameliorate visual symptoms). These participants were recruited via referrals to educational, optical and medical personnel to the ‘Special Education Centre’ at the University of Newcastle, Australia, for reading or study problems. There was also a no-treatment control group of 35 children with poor reading but without IS. In contrast to the IS group, these children were recruited from two local schools, introducing a potential recruitment bias.

Although this study is described as a “long-term placebo-controlled study” lasting for 20 months, it was only placebo-controlled for the first 3-4 months, after which all participants used their optimum tint. Robinson and Foreman (1999a) reported the effects of colour on reading speed, accuracy and comprehension, as well as on perception of academic ability(35); Robinson and Foreman (1999b) reported the effects of long-term effects of using coloured filters on the frequency and type of errors in oral reading(36). The study was a pre-test, post-test parallel groups study. At the start of the study the 95% confidence intervals for all measures of reading overlapped in the optimum tint, blue tint and placebo tint groups, although the comprehension scores for the optimum tint group were lower. At the end of the initial 3 month period all groups had improved but the 95% confidence intervals for all measures of reading still overlapped. Although attention has been drawn to the improvement in comprehension in the optimum tint group taken in isolation, it is not statistically correct to make within-group comparisons in a parallel groups study(70). Furthermore the baseline comprehension in the optimum tint group was lower allowing more room for improvement.

Finally, Tyrell et al published an exploratory study of 60 children aged 8-16 years(49). The sample were introduced as an array of small subgroups, based on performance on the Group Reading Test: 10 above average readers; 18 average readers; 12 below-average readers; 6 well-below average readers; 8 reading-age controls; 6 chronological-age controls (for the well-below-average readers). However, subsequent statistical analysis focussed on comparing children who ‘chose coloured overlays’ against children who ‘chose clear overlays’. Participants were tested reading aloud naturalistic text of their own choosing for 15 minutes. There was no immediate effect of coloured overlays; however, after ten minutes, the children who chose a coloured overlay read significantly more syllables per minute, with their overlay than without and reported more symptoms of visual discomfort and tiredness when reading without their overlay. However, these differences were very small: during the final 5 minutes of reading, children who chose a coloured overlay read 133.7 syllables per minute with their overlay and 123 syllables per minute without their overlay. It is debatable whether this is clinically significant. There was no placebo control group and it is not clear if the division of the reading task into 5 minute segments was a post-hoc analysis or a pre-defined variable. At best, therefore, this study should be seen as hypothesis-generating rather than hypothesis-confirming.

Unpublished data

An additional six unpublished postdoctoral theses were identified via our searches(71)(30)(72)(65)(73)(74) (Figure 1). These represent items from the ‘grey’ literature. Anderson (1997) is not considered here as it was purely observational and there was no control group(30). The results from the other five grey items are now considered.

Donovan studied 83 children with ages from 10-15 years (average ~12 years) who were diagnosed as reading disabled(71). Each child tested positive for IS according to criteria set by the Irlen Institute. The pattern of results obtained was extremely mixed with the prescribed overlay improving some performance measures but reducing others, or having beneficial effects in readers of a certain level but negative effects in readers of a different level. Interestingly there was no significant interaction between IS-level (mild, moderate or severe) and the effect of the overlay, even for those variables for which the overlay appeared to have an effect

Mason (2000) studied 30 university students who demonstrated low reading ability and whose symptoms of Irlen syndrome were in the severe range. The participants were self-referrals to the University's learning assistance centre. Participants were divided into 3 groups; ten received coloured overlays, ten received reading instruction and the remainder received no treatment. There was no difference between the groups in relation to the 'change in reading rate' they exhibited from pre- to post-treatment testing. Indicating that coloured overlays were no more beneficial than reading instruction and no better than no intervention at all. The author acknowledged the study was underpowered owing to the small number of participants per group.

Faraci examined 26 children with an average age of 9 years(65) who tested positive for scotopic sensitivity syndrome. The children were divided in two groups; one group received overlays while the other group did not. The overlay group were asked to use their overlays for all school- and home-based reading and homework activities. Both groups received the same instruction in reading. After 3 months, reading performance was compared in the two groups. Reading Fluency was significantly higher amongst the overlay group, but no statistically significant difference was evident for phonics or for reading accuracy comprehension. A major drawback of this study is that the author assumed that the two groups (overlay and no overlay) were matched in their baseline reading performance.

Morrison examined whether individuals diagnosed with Irlen syndrome showed differences in reading fluency and eye movements when they read with and without coloured overlays(73). Participants (n=24) were mainly undergraduate psychology students with an average age of 22 who did not report a reading problem or a reading disability but who were Irlen syndrome-positive. The results revealed no difference in reading fluency of curriculum-based material, or in the eye movements they exhibited when reading this material when the optimum coloured overlay was compared to a clear or randomly coloured overlay.

Adams (2012) compared reading on a computer screen in 32 children (aged 12 to 14 years of age) with a clear overlay compared to with a chosen overlay(74). The children were selected for the study on the basis that they reported perceptual distortions and that the chosen overlay removed the distortions. Around three-quarters (78%) of the sample did not show improved scores with the chosen overlay and over half (59%) did not read faster with the chosen overlay. Also the differences in scores for the clear versus chosen overlay conditions were not statistically significant.

Overall, the ‘grey’ literature does not support the use of coloured overlays and lenses to improve reading performance. Generally the study quality was found to be acceptable but many of the studies are underpowered as too few participants were recruited or the participants were divided into too many groups.

Summary of Irlen Studies

The use of Irlen lenses and overlays to improve reading in individuals with Irlen syndrome cannot be endorsed on the basis of the studies in the peer-reviewed literature or ‘grey’ literature. The two trials at lowest risk of bias failed to show any improvement in reading when using chosen coloured overlays and lenses(35)(36). One point of particular note is that the use of Irlen procedures has led to high percentages of both normal readers and poor readers being diagnosed with Irlen syndrome and related perceptual phenomena across studies. Although these prevalence rates are consistent with Irlen’s original predictions(2) they have been criticised for being vastly over-inclusive(14).

ChromaGen/Harris Lenses

Background

ChromaGen$_{TM}$ spectacles or contact lenses(75) were developed by David Harris as a treatment for congenital colour vision disorders to allow the subjective appreciation of a wider range of colours. On the back of anecdotal reports from patients with colour vision deficiency that the lenses were improving the clarity of text, and claims that colour improves reading performance of individuals with visual stress(1) they were applied to the treatment of dyslexia. When used in colour vision deficiency, one lens (usually a contact lens) is worn on the non-dominant eye. In reading difficulties, the right and left eyes are assessed independently so that subjects may receive different coloured lenses (contact lenses or spectacles) for each eye. Thus, although the original set of spectacle lenses comprised eight colours (substantially less than Intuitive and Irlen systems), there is obviously a much larger number of combinations because the optimal colour for the two eyes may differ. Indeed, it appears that around 50% of individuals fitted with ChromaGen lenses are prescribed different colours for each eye(76). Harris Lenses are similar to ChromaGen lenses (i.e., they involve the same number of colours and are prescribed via the same procedures) but Harris Lenses have a surface mirror coating that reflects light more evenly across the spectrum. Consequently, they appear more natural to an outside viewer while preserving transmission qualities(77).

The mechanism by which ChromaGen or Harris filters work to help reading is not well established, nor is the reason why the two eyes may require a different colour. In relation to the latter, one suggestion is that different coloured lenses may differentially affect the rate of neurological transmission in the two eyes, akin to the use of neutral density filters in the Pulfrich phenomenon(78).

General Features

By comparison with other colour systems, the ChromaGen/Harris system has not been subject to the same volume of scientific scrutiny; only four papers were identified in the peer-reviewed literature that assessed some measure of reading. The studies in this area have compared the ChromaGen system to placebo lenses, where participants are typically told there is an invisible tint, or to control (no lens) conditions. In head-to-head trials published in the peer-reviewed literature, the ChromaGen system has only been compared to the Dyslexia Research Trust (DRT) system which comprises blue and yellow lenses(79)(80).

Specific Studies

Following a pilot investigation of ten participants (published in the Optical Press, rather than in the peer-reviewed literature) in which it was claimed that ChromaGen lenses out-performed coloured lenses from the Intuitive Colorimeter, Harris and MacRow-Hill compared the ability of ChomaGen contact lenses and placebo contact lenses carrying a ‘light blue’ handling tint to improve reading fluency in adults with dyslexia(81). The study is described as a ‘double-masked’ trial; however, it is highly likely that participants (who had responded to media interest) were aware of the difference between the two types of contact lenses. Nevertheless, the research team who carried out the outcome assessments were blind to group status. To be included, participants needed to have a formal diagnosis of dyslexia from an educational psychologist and be willing to wear contact lenses. Fifty-three participants started the trial but six failed to complete the study, because they were unable to tolerate contact lenses, unable to fulfil the minimum reading requirement or because they were unwilling to complete the testing. All of the testing took place on the same day. It is not clear if participants suffered from visual stress.

Irrespective of whether comparison were made in all participants were restricted to those who reported distortion (both colour-normals and colour deficient individuals), or to those with both distortion and colour deficiency, a broadly similar pattern of results was obtained. For example, across all participants, there was an increase of 12.2 wpm (a 14.6% increase) relative to the baseline reading rate, compared to 6.5 wpm with the placebo lens (a 7.7% increase). The improvement in reading rate with the Chromagen lenses was statistically significant relative to both the baseline rate of reading and the improvement seen with the placebo lenses. However, the improvement seen with the placebo lenses relative to baseline was also statistically significant. One important aspect of the results was that participants who received the ChromaGen lenses before the placebo lenses showed a statistically larger improvement in reading rate compared to those who received the placebo lenses first. This suggests that novelty effects may have exerted an influence on the results. There are other serious issues of external validity. For example, participants were recruited in response to publicity in the media about the possible benefits of colour.

Cardona et al compared ChomaGen spectacle lenses with placebo lenses in 56 teenage children(76). The placebo lenses were clear but, as in the study by Harris and MacRow-Hill(81), children were informed that lenses had a new invisible tint that provided the same effect as coloured lenses. It seems unlikely that this measure would have controlled for placebo effects associated with coloured lenses. Overall placebo lenses and ChromaGen lenses improved reading relative to the control condition where no lenses were worn. However, there was no improvement in reading speed with ChromaGen lenses over that seen with the placebo lenses. The results for reading accuracy showed a borderline significant benefit of the ChromaGen lenses over the placebo, although the magnitude of the effect size was not stated.

Two studies compared Harris lenses with blue or yellow lenses from the DRT in a head-to-head fashion(79)(80). These studies report results from the same group of participants which comprised 73 delayed readers who said that a filter (Harris or DRT) helped them see text more clearly. A positive feature of these studies was the lack of external bias because subjects were recruited from mainstream state primary schools. Unfortunately, because of a prior assumption that the treatment works, there was no placebo control group. Treatment fidelity is also questionable, as there was no mention of whether all of the children who chose a filter continued to use it for the full three month period. The groups were well matched on spelling and reading at baseline. After 3 months, both groups had improved their reading and, to a lesser extent, their spelling: statistical analysis revealed no significant difference in the improvement in reading or spelling scores between the two groups. The DRT group did improve their speed of reading non-words more than the Harris lens group, but neither group improved in their ability to read irregular words. Although the conclusions of these papers was that both systems improved reading ability in children with reading delay, the added time and effort required to decide upon the optimal tint using the Harris filters compared to the DRT filters (where either a yellow or blue tint is issued) was considered to give the DRT system advantage over the Harris lens system. Because no control group or placebo lenses were used in these studies, it is impossible to know the part that placebo effects played in the improvements seen.

Summary of ChromaGen/Harris Lens Studies

The results from the small number of studies assessing the effectiveness of the ChromaGen/Harris filter system for patients with reading difficulty collectively suggest that the system may deliver better reading performance than placebo lenses. However, it is unlikely that the participants in these studies were well masked to treatment groups because they would have known the difference between the two types of lenses (i.e., tinted versus clear). In head-to-head studies with the DRT lens system, Chromagen/Harris lenses lead to comparable changes in reading performance to the DRT system. However, the value of the 'benefits' that were claimed in these studies is difficult to establish because of the absence of placebo-lens or no lens (control) groups.

Non-mainstream (other) studies of colour

Background

This section evaluates studies of less well known or non-commercial coloured overlays or filters Three publications used one colour one blue(82); two yellow(83)(18), one used two colours (blue and red)(84) and the remaining used less than 10 colours except one which used 15 colours(85). Two publications that compared blue or yellow overlays with ChromaGen are considered under the Chromagen section(79)(80)

General Features

Of the 12 publications, 7 utilised a crossover design, and 5 used a between subject parallel design. All were identified as having at least one design aspect that was judged to be at 'high' risk of bias (Table 2).

Specific Studies

Christenson et al examined the effect of a blue filter in 16 children with dyslexia(82). Participants were randomly allocated to a blue lens group or a clear (placebo) lens group. No statistically significant difference was found in reading accuracy, fluency or comprehension scores between the two groups.

Palomo-Alverez and Puell studied the use of a yellow filter(79). Poor readers aged 9 to 11 years old, (defined as "poor readers without dyslexia and minimal refractive error") were randomly assigned to a yellow filter group (46) or a no treatment, control group (36). Lenses were worn for 3 months for school and homework. At the end of this period, there was no statistically significant difference between the two groups in accommodation, symptom scores or reading speed.

A paper published by Ray, Fowler and Stein in 2005 is difficult to evaluate because the publication appears in conference proceedings and insufficient information is presented to assess the risk of bias, in particular in relation to random sequence generation and allocation concealment(18). The study examined the use of a single, yellow overlay on word reading accuracy in 38 "severely disabled readers" aged 7-14 years. Some of the 38 "severely disabled readers" used the yellow filter and some used a cardboard sheet with a rectangle cut out that revealed one line of text only. There was also a blue overlay placebo control that was not reported. The children were assessed using the British Ability Scales II, Word Reading subtest(86) after 3 months. Detailed quantitative data such as means, standard deviations, confidence intervals and effect sizes are lacking. Although this study is sometimes described as being 'double masked'(13)(80) it was at high risk of bias primarily because children would have been aware of the which intervention they were receiving (Table 2). There are also issues of external validity because children were recruited from the Dyslexia Research Trust, which promotes the use of blue and yellow overlays as an intervention for reading impairment.

Iovino et al.(84) studied 60 children in total, comprising 4 groups of fifteen, categorised as reading/spelling/arithmetic disabled, reading/spelling disabled, arithmetic disabled and those with ADHD (86). Each group viewed text through a blue overlay, red overlay and no overlay in a single session. There was no difference in reading rate or accuracy between the three environments. The authors claimed a significant improvement in comprehension accuracy. From their table (page 798) the difference was 6 wpm between no colour and a blue overlay in the reading/spelling/arithmetic group (baseline 82), and 5 wpm in the reading/spelling group (baseline 92). It is also reported that 22/60 (37%) children showed a “decline” in reading comprehension, the majority of whom showed between 1 and 9 standard score points lower than baseline. An additional 9 children showed a considerable decline (between 10 and 28 score points). Thirty-four were “better” but of these, 13 had “minimal improvement” meaning a 1-9 score point increase, and 21 had a sizeable improvement (increase of 11-32 score points).. There was no significant group-to-colour interaction in reading rate.

Sawyer et al. studied 86 pupils from 7-15 years of age from their caseload of specific learning disorders(88). From a cohort of approximately 300, 110 reported a positive reaction from 4 coloured overlays (red, green, blue and yellow). One-hundred and eighty five similar students from a nearby town served as a control group. After one and half school terms there was no significant improvement in confidence in reading, interest in reading, or in the amount read between the two groups.

Gole et al. recruited 24 students with ‘dyslexia’(89). Thirteen were allocated to the treatment group on the basis of their positive subjective response to 6 coloured lenses presented in random order. The remaining 11, all of whom had a negative subjective response, acted as controls and received a clear lens for one term followed by a randomly selected tinted lens for two terms. Reading was assessed using the Neale Analysis of Reading Test(52) at the start of the study and at the end of each school term. There was no statistical difference between the absolute value or change in reading ages for rate, comprehension, or accuracy of reading in the treatment and control groups.

Menacker studied 24 children (8-12 years) with dyslexia(90). All children read passages of naturalistic text using 4 coloured lenses. Half of the children used 0.12 log unit density lenses and the other half 0.3 log unit density lenses. All children read similar passages using the 4 coloured lenses, a neutral density filter and with no filter. The San Diego Quick Assessment reading test(91) was used and 6 passages appropriate to ability were selected, one for each testing condition. The authors counted words read per minute and the number of errors made. One-way analysis of variance of reading performance showed neither improvement nor deterioration attributable to lens colour or density when applied to error or reading rates.

Saint-John and White studied 11 children (aged 11-12 years) with specific reading difficulty and 11 controls who had no difficulty(92). The children chose 1 of 6 coloured overlays, which was cut and placed into spectacles. The dependent variables were reading accuracy and speed on four passages of text (specifically, four 11-line passages from “Let the balloon go” by Ivan Southall). All children read with their selected colour, with the polaroid and with no lens. Colour transparencies did not improve reading any more than a polaroid or an empty frame.

Evans et al. explored the hypothesis that the effect of coloured overlays was mediated by treating pattern glare(93). They described two studies. In the first, they asked 151 optometry students to look at a pattern glare stimulus consisting of a high-contrast striped pattern with a spatial frequency of 4 cycles/degree. Symptoms such as bending of lines, blur, diamond shaped lattices, fading, flickering, shimmering or wobbling, glare or dazzle, or colours were present in 149. Five of these individuals with high scores and 6 individuals with low scores were assessed on a Simulated Reading Visual Search Task (SRVST), with and without 8 coloured overlays (supplied by Lee filters). There was no statically significant difference on the search time on this SRVST task in the with- versus without- overlay conditions.

Vidal-Lopez studied 54 children aged 12-14 years(85). Twenty-seven were diagnosed with visual stress according to the Wilkins' pattern glare test and assessment questions based on the Irlen Questionnaire. The remainder acted as a control group. The visual stress group selected the coloured filter that ameliorated their perceptual distortions while the control group was given a filter chosen randomly from among the 15 supplied by Panoptical (Delt Orgaz, S.L., Barcelona, Spain). All subjects read single Spanish words and a Spanish equivalent of the WRRT whilst wearing the coloured lens and a clear lens. Both groups read slightly faster with the coloured filter but the difference was only statistically significant in the non-visual stress group (word reading time with coloured lens 99.16 seconds SD 24.77; with clear lens 94.21 seconds SD21.37). The authors went on to determine whether this improvement might be due to improved motivation by measuring response criterion to a psychophysical test, again with a coloured lens and a clear lens. They found that increases in reading speed were associated with changes in response criterion suggesting that participants had become less conservative observers. However, this difference was not statistically significant. The authors argue that because the improvement in reading was greater in the group without visual stress and because it was associated with changes in response criterion, placebo effects were the most likely explanation. According to the visual stress hypothesis, greater improvements would be expected in the group with visual stress and these would not be associated with changes in response criterion; this was the opposite of what was observed. The study could usefully be repeated with a larger sample size.

Francis et al used blue red green and yellow overlays in 35 reading disabled children (10-14 years) and compared them to 27 children who received no intervention(94). Of the 35 children, 23 (66%) continued to use them for a whole term. The other 12 were resistant to using the device or the teachers felt it was having an adverse effect. The authors used the Salford Reading Test before the trial and at the end of the 3 month period(95). The Neale reading test was used with and without the Dex Frame at the start and end of the trial; at each test-point, both passages of text were used twice, with and without the frame in a counterbalanced order to neutralise practice effects. There were no significant group differences on the Salford Reading Test; both groups made roughly 4.5 months progress during the study.

Summary of Non-Mainstream Studies of Colour

None of these studies contain strong evidence that the use of coloured overlays or lenses leads to benefits in the in measures of reading in individuals with reading difficulties and/or visual stress.

General Discussion

Randomised controlled trials and systematic reviews of those trials are considered the best available form of evidence for therapeutic interventions. The key feature of a systematic review is that all studies are appraised according to the same template for assessing the risk of bias but only those at low risk of bias are included in the final analysis. In general, we found the studies we reviewed were at high or uncertain risk of bias but in order to appraise the literature as it currently stands we adopted an inclusive approach. As result we have included numerous studies excluded by other systematic reviews. For example the systematic review by Galushka et al. 2014(17) only included two studies one by Mitchell et al. 2014(42) and the other by Robinson & Foreman 1999(35)

One approach to reviewing this literature would have been to perform a simple summation or averaging of studies, or a relative score of studies that offer support and those that do not, irrespective of the risk of bias. However, this will result in a conclusion at high risk of bias and hence it is not an approach that features in systematic reviews. It is true to say that data from studies at low risk of bias are sometimes combined. The latter is described as a meta-analysis, and it involves combining the data that are derived from a systematic-review. Thus, every meta-analysis should be based on an underlying systematic review, but not every systematic review leads to a meta-analysis(96). Based upon our view that a large majority of the literature we have reviewed is at high risk of bias it is not clear that this field of research is ready for meta-analysis.

Many of the studies we reviewed relied on p-values to support their outcomes for example, claiming a result was statistically significant if p was <0.05. However, it is important to remember that the p-value is just the final step in the design and execution of a study. In practice, decisions made earlier in experimental design or in the analysis of the data are more important to the outcome and the idea of the ‘risk of bias’ tool is to give greater weight to the behaviour and practices that lead to the statistics. If there are problems with those behaviours and practices earlier in the chain (which there were with almost all of the studies we reviewed) a p-value with an arbitrary value of 0.05 or less adds nothing useful(97)(98). For these reasons we have not quoted p-values.

This systematic review of the literature leads to the conclusion that there is little evidence to support the use of coloured filters (overlays, spectacle lenses or contact lenses) to improve reading. Although each of the colour systems has been subject to varying levels of empirical evaluation, the results of this review do not differ according to system-type, and the majority of the limitations identified apply to all systems. Our results are consistent with the results of previous literature reviews, including a recent review by Albon et al.(6), which concluded that the available evidence was too low in quality to reach firm conclusions about the effectiveness of coloured filters for reading disability. Similarly, Uccula et al. concluded that the issue remains ‘controversial’ and ‘not settled’(9). Handler and Fierson found a ‘continued lack of definitive evidence of the effectiveness’ of coloured lenses and filters(8). A 2009 review prepared for the New Zealand Ministry for Health concluded that “there is not a sufficient body of evidence as yet to state that coloured filters or overlays improve the reading ability of those with reading difficulties”(99). An evidence and consensus based clinical practice guideline recently published in the German literature concluded that Irlen lenses should not be used in the treatment of reading and spelling disorders in children or adolescents(100).

Previous reviews of this literature have been criticised for considering the literature without taking into account which particular colour system was under investigation and whether it was reading difficulty itself that was being treated or the co-morbid condition visual stress(5). In this review, we have considered the studies separately for each of the main colour systems and we find no evidence to support the use of any system to aid reading. Our analysis also shows that even if only published research using the inconsistent diagnostic criteria for visual stress are selected, there are no studies at low risk of bias to support the use of coloured overlays and lenses to aid reading.

It has been argued that coloured lenses and overlays are also being used to treat the symptom complex of visual stress. Although we specifically only searched the literature for studies of whether colour overlays or lenses impact on reading, some of the studies we reviewed also made reference to changes in symptoms. For example Wilkins et al. 1994 looked at symptom diaries but data were only available for 52.9% of study participants thus precluding any meaningful analysis of the data(48). Also, Mitchell et al. 2008 used the IDPS which contains questions about symptoms. In this parallel groups study there were significant improvements in the IDPS scores in both the chosen lens and the placebo lens group but no difference between the groups(42). While our study was focused on colour and reading, in none of the studies did we see evidence that colour impacts positively on symptoms. Ultimately, however, a large scale RCT using a validated symptom questionnaire would be required to answer this question.

It is important to stress that the lack of evidence which we and previous reviewers have identified does not in itself prove that colour has no effect on reading; lack of evidence is not evidence for a lack of effectiveness. On the surface this statement could be interpreted as tacit support for a continuation of the practice of issuing coloured filters and lenses while the necessary evidence is being gathered. However, our main finding, consistent with several previous reviewers is that the quality of the available evidence is sufficiently low that, despite the many anecdotal claims and often powerful testimony of patients, we have serious reservations about this practice. Published studies on the topic first appeared over 20 years ago so this field of research is not new. We believe the onus is on the proponents to increase their efforts to gather the evidence to support this clinical practice. Below we draw attention to the limitations of previous studies and we make some suggestions about how future studies might be conducted so as to avoid or counter such criticisms.

Most of the studies we reviewed were not well designed, there was little evidence of a pre-study protocol, studies were often under-powered and all had areas of bias that were either 'high' or 'uncertain' (Table 2). Many publications contained errors between tables and text, and suspect statistical analysis was frequently observed including absence of a pre-trial specified statistical approach, uncorrected multiple tests on the same datasets, effect sizes not reported and missing descriptive statistics.

There was a common failure to consider how research participation effects might influence results(101). For instance, many studies showed evidence of novelty effects. The latter refers to an intervention that is new and exciting and which, consequently, may improve motivation and produce initial positive effects that diminish over time(102). For example, in Wilkins et al (1994) 31 participants preferred their first filter whereas only 17 preferred the filter they received second(48). Similarly in a crossover trial of ChromaGen lenses, participants who received the experimental lenses before the placebo lenses showed a bigger improvement in reading rate(81). Furthermore, an uncontrolled field trial of Irlen lenses showed improvements in reading during the first 3 months of use but no improvements in the second 3 months(66). Although this result was attributed to participants reaching their grade level, another explanation is that participants were less enticed by their overlay with time. Arguably, the high rate of attrition observed in many studies may also reflect novelty effects.

The act of being observed by the experimenter may also enhance performance; this is known as the Hawthorne effect(103)(104). When participants in both the experimental group and placebo group improve more than would be expected due to normal maturation, Hawthorne effects may be the most plausible explanation. Related to this, it was striking that trials that were well masked showed no statistically significant improvement in reading with a chosen colour compared to a placebo condition(48)(35)(42)(62). On the other hand unmasked studies that compared chosen colour with no overlay or clear overlay(32)(33)(34)(40)(47)(44)(49)(66), or a card with a rectangular slot cut out(18) often reported significant effects on reading. This difference in outcome between masked and unmasked studies points strongly to placebo effects.

Three studies attempted to control for placebo effects using enhanced placebos rather than trying to mask participants and experimenters to the intervention(43)(81)(76). In such studies steps were taken to enhance the placebo effect of the control filter by describing it as a ‘special’ or ‘wonderful discovery’. Implicit in this, is the assumption that the placebo effect of the experimental intervention can be accurately quantitated and that the placebo effect of the control intervention can be precisely modified to match it. The placebo effect is not sufficiently well understood to allow this. Indeed selecting the chosen tint involves a more prolonged relationship with the practitioner and a richer therapeutic ritual, both of which are powerful drivers of the placebo effect(105)(106)(107). Hence, such ‘enhanced’ placebos are not recommended for future research. It is important to acknowledge that incorporating a well-masked placebo control condition that comprises identical diagnostic and therapeutic rituals is particularly difficult in trials of coloured lenses and overlays. Nevertheless, the use of the Intuitive Colorimeter has the potential to allow a good degree of masking because during the assessment, participants do not see the actual lens they will ultimately receive(48).

Some researchers have claimed that it would be unethical to include a placebo control group(79)(80). This reflects a prior assumption that treatment with colour *is* effective. Our review shows that in those studies that were well masked there was as much improvement in the placebo control group as the experimental group(48)(35)(42)(62) . As a result we do not consider it unethical to include a placebo control group. Indeed, the ethics of organising further trials that are at high risk of bias because of the lack of a placebo control group also needs to be considered.

One line of thinking is that, even if the benefit of coloured filters upon reading stems purely from the placebo effect, the most important aspect is that reading has improved and the source of that improvement is of lesser importance. While we understand this logic, we disagree with it because coloured overlay or lens therapy can have a substantial financial cost for the patient or their parents, and because it may delay the identification of appropriate means of support.

Most studies adopted a crossover or within-subject design. Since participants act as their own controls, such studies are less prone to confounding at baseline(108) and the paired data they produce add to the statistical power. Studies of this type are generally considered suitable for assessing short acting or temporary interventions for chronic conditions(109)(108). The principle drawback of crossover studies is their vulnerability to attrition because there has to be sufficient time to allow all participants to receive both treatments. For this reason, a longer-term, parallel-groups study might be more suitable for assessing the effect of coloured lenses and overlays on reading performance. Parallel-arms designs require a substantially larger number of participants because the different groups need to be carefully matched on variables such as age, gender, reading-skill at baseline and the rate of attrition. Nonetheless, the advantages of parallel arms studies far outweigh these practical disadvantages. Based upon pilot data using the particular reading test used and the test-retest variation of the test, power calculations can be conducted in advance of the study to establish the appropriate sample size, taking account the likely attrition.

The ability to generalise the data to the wider population (in other words the external validity) should also be considered. Participants recruited from specialist clinics may not be representative of the general population of poor readers and furthermore participants recruited from clinics such as those at the Institute of Optometry(43) or Dyslexia Research Trust(18) may have specifically sought treatment with overlays because of their prior belief that treatment with colour is an effective therapy. An additional problem is that they may know their preferred colour making masking difficult or impossible. In Ritchie et al it was striking that the two participants who showed big improvements in reading were aware of their chosen colour(62). Ideally, to ascertain the effect of colour on reading, participants should have no prior exposure to the use of coloured lenses or overlays, and be drawn from unselected samples of children and adults.

In terms of applicability of the results to real-world reading, the external validity of the reading tests themselves also needs to be considered. Even if the WRRT is a useful diagnostic test for visual stress, unless it can be shown that improvements in reading the WRRT translate to reading naturalistic text of the sort that is encountered in everyday life, it is, by itself, an unsuitable outcome measure. The reading tests should be standardised so that the average measures of performance amongst different populations is known. The test-retest repeatability of the test should also be well established so that the impact of any change following the use of colour can be considered alongside the normal variation in baseline measures of reading performance. Also, since there are different measures of reading skill, outcome measures of reading should not be restricted to any one particular aspect (e.g. speed, accuracy or comprehension); rather it is suggested that all of these aspects should be represented in the outcome measures used.

Since poor reading may have a variety of causes, studies of the impact of coloured lenses or filters on reading performance should rule out other possible causes in study participants. The College of Optometrists guidelines on examining patients with specific learning difficulties or visual discomfort (Guideline A85) state “Patients with specific learning difficulties may have conventional visual or orthoptic problems that require treatment(110). You should not use tinted lenses or vision therapy before you have excluded visual problems by means of a thorough eye examination”. Many of the studies we reviewed included eye examinations of their participants prior to the issuing of any colour intervention. This could be considered good scientific practice in order that the impact of colour on only the ‘target’ condition is assessed. Although there is ongoing debate about the frequency with which refractive and oculomotor anomalies account for poor reading performance(111)(112)(113) they are likely to be important confounding variables that need to be controlled for.

Many of the studies we reviewed sought to establish if colour makes an immediate difference to reading. We suggest that the effect of colour should be examined over a period of time that is not less than 3 months, and ideally for up to one year. This will enable researchers to observe any decay in the frequency with which the coloured overlay/lenses is/are used, and to assess any changes in reading performance over the same time period in relation to age-appropriate norms(114). Follow-up over longer periods will also help to eliminate the impact of novelty effect upon study outcome.

The proportion of cases diagnosed with visual stress ranged from 46-96% for Irlen and 13-88% for Intuitive (Table 1). One would expect that prevalence rates to be reasonably constant between unselected populations. Furthermore, based on the rapid discontinuation of use seen in many studies, it can be argued that the diagnostic procedures currently in use would appear to produce a large number of false positives. The lack of constant diagnostic criteria makes it difficult to be sure that the same condition is being investigated and treated across different studies. Researchers with an interest in visual stress need to agree on the diagnostic criteria for the condition that they are aiming to treat. This will not only enable robust epidemiological studies to ascertain the prevalence but also examination of whether coloured lenses or filters have an impact on reading performance in individuals who test positive for that condition.

To avoid claims that statistical analyses were conducted post-hoc, in addition to setting out, a priori, the outcome measures that will be used to establish whether colour has aided reading, researchers should set out in advance which statistical tests will be applied. This measure alone has been shown to reduce the number of trials for which a positive effect is reported(115). Although this statistical approach does not preclude the reporting of exploratory analyses post-hoc, the results of such analyses should be seen only as hypothesis-generating rather than hypothesis-confirming.

Proponents of the use of coloured lenses or filters to treat visual stress have attached importance to the results of fMRI studies(5)(116). A detailed account of these studies is beyond the scope of this review. However, ignoring the problems of interpretation of this kind of study which can be at high risk of producing false positive results(117)(118), ultimately it has to be shown that colour improves the behaviour in question, in this case reading(119). For this reason we have not considered the results of neuro-imaging studies here.

Conclusion

There are many anecdotal claims that reading can benefit from use of coloured overlays or lenses. There is also a substantial literature on the topic of whether reading performance is enhanced through the use of colour. We have systematically reviewed the literature concerning the effect that colour may have on reading. Whilst many studies report improvements with coloured lenses or filters, the effect size is generally small and/or similar to the improvement found with a placebo condition. The vast majority studies in each area are subject to high or uncertain risk of bias in one more key aspects of study design or outcome. Studies which are less at risk from bias generally offered less support for the benefit of colour on reading ability. For these reasons, in common with previous reviews of the literature, we conclude that the use of coloured overlays and lenses to ameliorate reading difficulties cannot be endorsed. From the evidence reviewed, placebo, Hawthorne and novelty effects provide the most likely explanation for the benefit which many individuals report.

Declaration: The authors declare that this review was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Meares, O. Figure/background, brightness/contrast and reading disabilities. Visible Lang. 1980;14:13–29.

2. Irlen H. Reading by the colors: overcoming dyslexia and other reading disabilities through the Irlen method. Garden City Park, N.Y: Avery Pub. Group; 1991. 195 p.

3. Wilkins AJ. Reading through colour: how coloured filters can reduce reading difficulty, eye strain, and headaches. Chichester, West Sussex, England ; Hoboken, NJ: John Wiley & Sons; 2003. 156 p.

4. Wilkins A. Coloured overlays and their effects on reading speed: a review. Ophthalmic Physiol Opt J Br Coll Ophthalmic Opt Optom. 2002 Sep;22(5):448–54.

5. Wilkins A.J., Allen P.M., Monger L.J., Gilchrist J.M. Visual Stress and Dyslexia For the Practicing Optometrist. Optom Pract. 2016;17:103–12.

6. Albon E, Adi Y, Hyde C, West Midlands Health Technology Assessment Collaboration. The effectiveness and cost-effectiveness of coloured filters for reading disability: a systematic review. Studley: West Midlands Health Technology Assessment Collaboration, Dept. of Public Health and Epidemiology, University of Birmingham; 2008.

7. Fletcher JM CD. Vision Efficiency Interventions and Reading Disability [Internet]. Perspectives on Language — Winter Edition 2011. 2011. Available from: http://www.onlinedigeditions.com/display_article.php?id=625330

8. Handler S.M., Fierson W.M., Ruben J.B., Granet D.B., Blocker R.J., Bradford G.E., et al. Joint technical report - Learning disabilities, dyslexia, and vision. Pediatrics. 2011;127(3):e818–56.

9. Uccula A, Enna M, Mulatti C. Colors, colored overlays, and reading skills. Front Psychol. 2014 Jul 29;5:833–833.

10. Aleci Carlo. Colored Filters and Dyslexia. A Quick Gliding over Myth and (Possible) Reality. Neuro Ophthalmol Vis Neurosci. 2014;1(1):18–21.

11. Hyatt KJ, Stephenson J, Carter M. A review of three controversial educational practices: perceptual motor programs, sensory integration, and tinted lenses. Educ Treat Child. 2009;32(2):313–42.

12. Howard-Jones PA. Neuroscience and education: myths and messages. Nat Rev Neurosci. 2014 Oct 15;15(12):817–24.

13. Stein J. Dyslexia: the Role of Vision and Visual Attention. Curr Dev Disord Rep. 2014 Dec;1(4):267–80.

14. McIntosh RD, Ritchie SJ. Rose-tinted: The use of coloured filters to treat reading difficulties. In:Della Sala S, Anderson M, editors. Neuroscience in education: the good, the bad, and the ugly. Oxford ; New York: Oxford University Press; 2012. 385 p.

15. Henderson LM, Taylor RH, Barrett B, Griffiths PG. Treating reading difficulties with colour. BMJ. 2014 Aug 19;349:g5160–g5160.

16. Smith L, Wilkins A. How many colours are necessary to increase the reading speed of children with visual stress? A comparison of two systems. J Res Read. 2007 Aug;30(3):332–43.

17. Galuschka K, Ise E, Krick K, Schulte-Körne G. Effectiveness of treatment approaches for children and adolescents with reading disabilities: a meta-analysis of randomized controlled trials. Plos One. 2014 Feb 26;9(2):e89900–e89900.

18. Ray NJ, Fowler S, Stein JF. Yellow Filters Can Improve Magnocellular Function: Motion Sensitivity, Convergence, Accommodation, and Reading. In: Clinical and basic oculomotor research: In honor of David S Zee [Internet]. New York, NY, US: New York Academy of Sciences; 2005. p. 283–93. (Annals of the New York Academy of Sciences, Vol. 1039). Available from: http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2005-07863-028&site=ehost-live

19. Loew SJ, Rodríguez C, Marsh NV, Jones GL, Núñez JC, Watson K. Levels of Visual Stress in Proficient Readers: Effects of Spectral Filtering of Fluorescent Lighting on Reading Discomfort. Span J Psychol [Internet]. 2015 [cited 2015 Aug 12];18. Available from: http://www.journals.cambridge.org/abstract_S1138741615000591

20. Monger LJ, Wilkins AJ, Allen PM. Pattern glare: the effects of contrast and color. Front Psychol [Internet]. 2015 Oct 27 [cited 2016 Feb 18];6. Available from: http://journal.frontiersin.org/Article/10.3389/fpsyg.2015.01651/abstract

21. Rosenfield Mark, Rosenlicht Tracy, Jones Yujia Huang, Simon Julia. Effect of Coloured Overlays on Computer Vision Syndrome. Optom Vis Perform. 3(3):133 – y.

22. Wilkins AJ, Jeanes RJ, Pumfrey PD, Laskier M. Rate of Reading Test: its reliability, and its validity in the assessment of the effects of coloured overlays. Ophthalmic Physiol Opt J Br Coll Ophthalmic Opt Optom. 1996 Nov;16(6):491–7.

23. Wilkins A, Lewis E. Coloured overlays, text, and texture. Perception. 1999;28(5):641–50.

24. Northway N. Predicting the continued use of overlays in school children--a comparison of the Developmental Eye Movement test and the Rate of Reading test. Ophthalmic Physiol Opt J Br Coll Ophthalmic Opt Optom. 2003 Sep;23(5):457–64.

25. Allen P.M., Gilchrist J.M., Hollis J. Use of visual search in the assessment of pattern-related visual stress (PRVS) and its alleviation by colored filters. Invest Ophthalmol Vis Sci. 2008;49(9):4210–8.

26. Higgins JPT, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011 Oct 18;343(oct18 2):d5928–d5928.

27. Cotton MM, Evans KM. Parents' and children's expectations about Irlen (tinted) lenses. Percept Mot Skills. 1994 Jun;78(3 Pt 2):1387–90.

28. Robinson GL, Conway RN. The effects of Irlen colored lenses on students' specific reading skills and their perception of ability: a 12-month validity study. J Learn Disabil. 1990 Dec;23(10):589–96.

29. MacLachlan A., Yale S., Wilkins A. Open trial of subjective precision tinting: A follow-up of 55 patients. Ophthalmic Physiol Opt. 1993;13(2):175–8.

30. Anderson LL. Effects of Irlen Syndrome and visual dysfunctions on reading comprehension. [Internet]. [US]: ProQuest Information & Learning; 1997. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1997-95001-146&site=ehost-live

31. Wilkins A, Neary C. Some visual, optometric and perceptual effects of coloured glasses. Ophthalmic Physiol Opt J Br Coll Ophthalmic Opt Optom. 1991 Apr;11(2):163–71.

32. Jeanes R, Busby A, Martin J, Lewis E, Stevenson N, Pointon D, et al. Prolonged use of coloured overlays for classroom reading. Br J Psychol. 1997 Nov;88(4):541–8.

33. Lightstone A, Lightstone T, Wilkins A. Both coloured overlays and coloured lenses can improve reading fluency, but their optimal chromaticities differ. Ophthalmic Physiol Opt J Br Coll Ophthalmic Opt Optom. 1999 Jul;19(4):279–85.

34. Wilkins A, Lewis E, Smith F, Rowland E, Tweedie W. Coloured overlays and their benefit for reading. J Res Read. 2001 Feb;24(1):41–64.

35. Robinson GL, Foreman PJ. Scotopic sensitivity/Irlen syndrome and the use of coloured filters: a long-term placebo controlled and masked study of reading achievement and perception of ability. Percept Mot Skills. 1999 Aug;89(1):83–113.

36. Robinson GL, Foreman PJ. Scotopic sensitivity/Irlen syndrome and the use of coloured filters: a long-term placebo-controlled study of reading strategies using analysis of miscue. Percept Mot Skills. 1999 Feb;88(1):35–52.

37. Blaskey P, Scheiman M, Parisi M, Ciner EB, Gallaway M, Selznick R. The effectiveness of Irlen filters for improving reading performance: a pilot study. J Learn Disabil. 1990 Dec;23(10):604–12.

38. Allen P, Evans B.J.W, Wilkins A.J. Vision and Reading Difficulties Part 5: Clinical protocol and the role of the eye care practitioner. Optom Today. 2009 May;30–6.

39. Kriss I, Evans BJW. The relationship between dyslexia and Meares-Irlen Syndrome. J Res Read. 2005 Aug;28(3):350–64.

40. Evans BJW, Joseph F. The effect of coloured filters on the rate of reading in an adult student population. Ophthalmic Physiol Opt J Br Coll Ophthalmic Opt Optom. 2002 Nov;22(6):535–45.

41. Simmers AJ, Bex PJ, Smith FK, Wilkins AJ. Spatiotemporal visual function in tinted lens wearers. Invest Ophthalmol Vis Sci. 2001 Mar;42(3):879–84.

42. Mitchell C, Mansfield D, Rautenbach S. Coloured filters and reading accuracy, comprehension and rate: a placebo-controlled study. Percept Mot Skills. 2008 Apr;106(2):517–32.

43. Bouldoukian J, Wilkins AJ, Evans BJW. Randomised controlled trial of the effect of coloured overlays on the rate of reading of people with specific learning difficulties. Ophthalmic Physiol Opt J Br Coll Ophthalmic Opt Optom. 2002 Jan;22(1):55–60.

44. Allen P.M., Hussain A., Usherwood C., Wilkins A.J. Pattern-related visual stress, chromaticity, and accommodation. Invest Ophthalmol Vis Sci. 2010;51(12):6843–9.

45. Hollis J, Allen PM. Screening for Meares-Irlen sensitivity in adults: can assessment methods predict changes in reading speed? Ophthalmic Physiol Opt J Br Coll Ophthalmic Opt Optom. 2006 Nov;26(6):566–71.

46. Singleton C, Trotter S. Visual stress in adults with and without dyslexia. J Res Read. 2005 Aug;28(3):365–78.

47. Singleton C, Henderson L-M. Computerized screening for visual stress in children with dyslexia. Dyslexia Chichester Engl. 2007 May;13(2):130–51.

48. Wilkins AJ, Evans BJ, Brown JA, Busby AE, Wingfield AE, Jeanes RJ, et al. Double-masked placebo-controlled trial of precision spectral filters in children who use coloured overlays. Ophthalmic Physiol Opt J Br Coll Ophthalmic Opt Optom. 1994 Oct;14(4):365–70.

49. Tyrrell R, Holland K, Dennis D, Wilkins A. Coloured overlays, visual discomfort, visual search and classroom reading. J Res Read. 1995 Feb;18(1):10–23.

50. Lightstone A, Evans BJ. A new protocol for the optometric management of patients with reading difficulties. Ophthalmic Physiol Opt J Br Coll Ophthalmic Opt Optom. 1995 Sep;15(5):507–12.

51. Henderson LM, Tsogka N, Snowling MJ. Questioning the benefits that coloured overlays can have for reading in students with and without dyslexia. J Res Spec Educ Needs. 2013 Jan;13(1):57–65.

52. Neale MD, McKAY MF, Childs GH. THE Neale Analysis of Reading Ability - Revised. Br J Educ Psychol. 1986 Nov;56(3):346–56.

53. Fergusson D, Aaron SD, Guyatt G, Hébert P. Post-randomisation exclusions: the intention to treat principle and excluding patients from analysis. BMJ. 2002 Sep 21;325(7365):652–4.

54. Sedgwick P. Analysis by intention to treat. BMJ. 2011 Apr 13;342(apr13 2):d2212–d2212.

55. Ioannidis JPA. Why Most Published Research Findings Are False. PLoS Med. 2005 Aug 30;2(8):e124.

56. Simmons JP, Nelson LD, Simonsohn U. False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. Psychol Sci. 2011 Nov 1;22(11):1359–66.

57. Irlen H. The Irlen revolution: a guide to changing your perception and your life. Garden City Park, NY: Square One Publishers; 2010. 212 p.

58. Cotton MM, Evans KM. An evaluation of the Irlen lenses as a treatment for specific reading disorders. Aust J Psychol. 1990 Apr;42(1):1–12.

59. O’Connor PD, Sofo F, Kendall L, Olsen G. Reading disabilities and the effects of colored filters. J Learn Disabil. 1990 Dec;23(10):597.

60. Martin F, Mackenzie B, Lovegrove W, McNicol D. Irlen lenses in the treatment of specific reading disability: An evaluation of outcomes and processes. Aust J Psychol. 1993 Dec;45(3):141–50.

61. Fletcher J, Martinez G. An Eye-Movement Analysis of the Effects of Scotopic Sensitivity Correction on Parsing and Comprehension. J Learn Disabil. 1994 Jan 1;27(1):67–70.

62. Ritchie SJ, Della Sala S, McIntosh RD. Irlen colored overlays do not alleviate reading difficulties. Pediatrics. 2011 Oct;128(4):e932–8.

63. Ritchie SJ, Della Sala S, McIntosh RD. Irlen colored filters in the classroom: A 1-year follow-up. Mind Brain Educ. 2012 Jun;6(2):74–80.

64. Research Papers | Coloured Lenses and Visual Stress [Internet]. [cited 2016 Jan 20]. Available from: http://colouredlensesandvisualstress.com/research-papers/

65. Faraci ME. Does the Irlenrtm method bring about an increase in reading scores on a specific test of reading for students found to have Scotopic Sensitivity Syndrome? [Internet]. [US]: ProQuest Information & Learning; 2010. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2010-99130-231&site=ehost-live

66. Noble J, Orton M, Irlen S, Robinson G. A controlled field study of the use of coloured overlays on reading achievement. Aust J Learn Disabil. 2004 Jun;9(2):14–22.

67. Vaughn-Blount K, Watson ST, Kokol AL, Grizzle R, Carney RN, Rich SS, et al. Woodcock-Johnson III Tests of Achievement. In: Goldstein S, Naglieri JA, editors. Encyclopedia of Child Behavior and Development [Internet]. Boston, MA: Springer US; 2011 [cited 2015 Dec 23]. p. 1575–6. Available from: http://www.springerlink.com/index/10.1007/978-0-387-79061-9_3095

68. Wiederholt L. Formal reading inventory: a method for assessing silent reading comprehension and oral reading... [S.l.]: Pro-Ed; 1985.

69. Wiederholt J.L., Bryant B.R. GORT 4: Gray Oral Reading Tests. 4th ed. Pro Ed; 1986.

70. George BJ, Beasley TM, Brown AW, Dawson J, Dimova R, Divers J, et al. Common scientific and statistical errors in obesity research: Common Statistical Errors in Obesity Research. Obesity. 2016 Apr;24(4):781–90.

71. Donovan MA. Reading effectiveness of Irlen Overlays. [Internet]. [US]: ProQuest Information & Learning; 1995. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1995-95024-094&site=ehost-live

72. Mason CA. Scotopic Sensitivity/Irlen Syndrome and reading in college-level students. [Internet]. [US]: ProQuest Information & Learning; 2000. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2000-95003-188&site=ehost-live

73. Morrison RF. The effect of color overlays on reading efficiency. [Internet]. [US]: ProQuest Information & Learning; 2012. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2012-99170-193&site=ehost-live

74. Adams TA. Improving scores on computerized reading assessments: The effects of colored overlay use. [Internet]. [US]: ProQuest Information & Learning; 2013. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2013-99090-005&site=ehost-live

75. ChromaGen | Cantor+Nissel [Internet]. [cited 2015 Nov 14]. Available from: http://dyslexia-help.co.uk/

76. Cardona G, Borràs R, Peris E, Castañé M. A placebo-controlled trial of tinted lenses in adolescents with good and poor academic performance: reading accuracy and speed. J Optom. 2010 Apr;3(2):94–101.

77. Hodd, Nigel Burnett. The Cromagen method for colour vision deficiency and special learning difficulties. Optom Today. 2000;November 17.

78. Morgan MJ, Thompson P. Apparent motion and the Pulfrich effect. Perception. 1975;4(1):3–18.

79. Hall R, Ray N, Harries P, Stein J. A comparison of two-coloured filter systems for treating visual reading difficulties. Disabil Rehabil. 2013;35(26):2221–6.

80. Harries P, Hall R, Ray N, Stein J. Using coloured filters to reduce the symptoms of visual stress in children with reading delay. Scand J Occup Ther. 2015 Mar;22(2):153–60.

81. Harris D, MacRow-Hill SJ. Application of ChromaGen haploscopic lenses to patients with dyslexia: a double-masked, placebo-controlled trial. J Am Optom Assoc. 1999 Oct;70(10):629–40.

82. Christenson GN, Griffin JR, Taylor M. Failure of blue-tinted lenses to change reading scores of dyslexic individuals. Optom J Am Optom Assoc. 2001 Oct;72(10):627–33.

83. Palomo-Álvarez C, Puell MC. Effects of wearing yellow spectacles on visual skills, reading speed, and visual symptoms in children with reading difficulties. Graefes Arch Clin Exp Ophthalmol Albrecht Von Graefes Arch Für Klin Exp Ophthalmol. 2013 Mar;251(3):945–51.

84. Iovino I, Fletcher JM, Breitmeyer BG, Foorman BR. Colored overlays for visual perceptual deficits in children with reading disability and attention deficit/hyperactivity disorder: are they differentially effective? J Clin Exp Neuropsychol. 1998 Dec;20(6):791–806.

85. Vidal-López J. The role of attributional bias and visual stress on the improvement of reading speed using colored filters. Percept Mot Skills. 2011 Jun;112(3):770–82.

86. British Abilities Scales (3rd Revision) — The Psychometrics Centre [Internet]. [cited 2015 Nov 14]. Available from: http://www.psychometrics.cam.ac.uk/services/psychometric-tests/gl-assessment

87. Dyslexia Research Trust - Vision & Coloured Filters [Internet]. [cited 2015 Dec 27]. Available from: http://www.dyslexic.org.uk/research/vision-coloured-filters

88. Sawyer C, Taylor S, Willcocks S. Transparent coloured overlays and specific learning difficulties. AEP Assoc Educ Psychol J. 1994 Jan;9(4):217–20.

89. Gole GA, Dibden SN, Pearson CC, Pidgeon KJ, Mann JW, Rice D, et al. Tinted lenses and dyslexics--a controlled study. SPELD (S.A.) Tinted Lenses Study Group. Aust N Z J Ophthalmol. 1989 May;17(2):137–41.

90. Menacker SJ, Breton ME, Breton ML, Radcliffe J, Gole GA. Do tinted lenses improve the reading performance of dyslexic children? A cohort study. Arch Ophthalmol Chic Ill 1960. 1993 Feb;111(2):213–8.

91. La Pray M., Ross R. The graded word list: Quick guage of reading ability. J Read. 1969;12(4):305–7.

92. Saint-John LM, White MA. The effect of coloured transparencies on the reading performance of reading-disabled children. Aust J Psychol. 1988 Dec;40(4):403–11.

93. Evans BJ, Cook A, Richards IL, Drasdo N. Effect of pattern glare and colored overlays on a stimulated-reading task in dyslexics and normal readers. Optom Vis Sci Off Publ Am Acad Optom. 1994 Oct;71(10):619–28.

94. Francis M, Taylor S, Sawyer C. Coloured lenses and the Dex frame: New issues. Support Learn. 1992 Feb;7(1):25–7.

95. McCarty CT. New Salford Sentence Reading Test manual. London: Hodder Education; 2012.

96. Rudnicka AR, Owen CG. An introduction to systematic reviews and meta-analyses in health care: Systematic reviews and meta-analyses. Ophthalmic Physiol Opt. 2012 May;32(3):174–83.

97. Leek JT, Peng RD. Statistics: P values are just the tip of the iceberg. Nature. 2015 Apr 28;520(7549):612–612.

98. Nuzzo R. Scientific method: Statistical errors. Nature. 2014 Feb 12;506(7487):150–2.

99. Malins C, Mind Matters Psychology, New Zealand, Ministry of Health, Irlen Screening Research Project. The use of coloured filters and lenses in the management of children with reading difficulties a literature review commissioned for the Irlen Screening Research Project, Ministry of Health [Internet]. [Wellington, N.Z.]: Mind Matters Psychology; 2009 [cited 2016 Feb 15]. Available from: http://www.moh.govt.nz/moh.nsf/pagesmh/10456/$File/use-of-coloured-filters-and-lenses-children-learning-difficulties-nov2010.pdf

100. Galuschka K, Schulte-Körne G. The Diagnosis and Treatment of Reading and/or Spelling Disorders in Children and Adolescents. Dtsch Ärztebl Int. 2016 Apr 22;113(16):279–86.

101. McCambridge J, Kypri K, Elbourne D. Research participation effects: a skeleton in the methodological cupboard. J Clin Epidemiol. 2014 Aug;67(8):845–9.

102. Love A, Kavazis AN, Morse A, Mayer Jr. KC. Soccer-Specific Stadiums and Attendance in Major League Soccer: Investigating the Novelty Effect. J Appl Sport Manag Res Matters Vol 5 No 2 2013 [Internet]. 2013; Available from: http://js.sagamorepub.com/jasm/article/view/3294

103. Sedgwick P. The Hawthorne effect. BMJ. 2011 Jan 4;344(jan04 2):d8262–d8262.

104. McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne Effect: a randomised, controlled trial. BMC Med Res Methodol. 2007;7(1):30.

105. Kaptchuk TJ, Kelley JM, Conboy LA, Davis RB, Kerr CE, Jacobson EE, et al. Components of placebo effect: randomised controlled trial in patients with irritable bowel syndrome. BMJ. 2008 May 3;336(7651):999–1003.

106. Meissner K, Fässler M, Rücker G, Kleijnen J, Hróbjartsson A, Schneider A, et al. Differential Effectiveness of Placebo Treatments: A Systematic Review of Migraine Prophylaxis. JAMA Intern Med. 2013 Nov 25;173(21):1941.

107. Benedetti F, Dogue S. Different Placebos, Different Mechanisms, Different Outcomes: Lessons for Clinical Trials. Eldabe S, editor. PLOS ONE. 2015 Nov 4;10(11):e0140967.

108. Sedgwick P. Bias in randomised controlled trials: comparison of crossover group and parallel group designs. BMJ. 2015 Aug 7;h4283.

109. Ding H, Hu GL, Zheng XY, Chen Q, Threapleton DE, Zhou ZH. The Method Quality of Cross-Over Studies Involved in Cochrane Systematic Reviews. Wright JM, editor. PLOS ONE. 2015 Apr 13;10(4):e0120519.

110. Principles of examining patients with specific learning difficulties - The College of Optometrists [Internet]. [cited 2016 Feb 11]. Available from: http://guidance.college-optometrists.org/guidance-contents/knowledge-skills-and-performance-domain/examining-patients-with-specific-learning-difficulties/principles-of-examining-patients-with-specific-learning/?searchtoken=A85

111. Creavin AL, Lingam R, Steer C, Williams C. Ophthalmic Abnormalities and Reading Impairment. PEDIATRICS. 2015 Jun 1;135(6):1057–65.

112. Motsch S, Mühlendyck H. Frequency of reading disability caused by ocular problems in 9- and 10-year-old children in a small town. Strabismus. 2000 Dec;8(4):283–5.

113. Lambert SR. Should Glasses Be Prescribed for All Children with Moderate Hyperopia? Ophthalmology. 2016 Apr;123(4):676–8.

114. Duff FJ, Clarke PJ. Practitioner Review: Reading disorders: what are the effective interventions and how should they be implemented and evaluated?: Effective reading interventions. J Child Psychol Psychiatry. 2011 Jan;52(1):3–12.

115. Kaplan RM, Irvin VL. Likelihood of Null Effects of Large NHLBI Clinical Trials Has Increased over Time. Garattini S, editor. PLOS ONE. 2015 Aug 5;10(8):e0132382.

116. Society for Coloured Lens Prescribers [Internet]. [cited 2015 Nov 22]. Available from: http://www.ioo.org.uk/s4clp/code-of-conduct.htm

117. David SP, Ware JJ, Chu IM, Loftus PD, Fusar-Poli P, Radua J, et al. Potential reporting bias in fMRI studies of the brain. PloS One. 2013;8(7):e70104.

118. Vul E, Harris C, Winkielman P, Pashler H. Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition. Perspect Psychol Sci J Assoc Psychol Sci. 2009 May;4(3):274–90.

119. Bishop DVM. Research Review: Emanuel Miller Memorial Lecture 2012 - Neuroscientific studies of intervention for language impairment in children: interpretive and methodological problems: Research Review: Intervention and neuroscience. J Child Psychol Psychiatry. 2013 Mar;54(3):247–59.