Abstract

This study investigated the effects of using Irlen coloured overlays on reading rate, accuracy, fluency and comprehension under regular class conditions, with class teachers instigating the study and conducting the assessment. All subjects in grade 3 at Whitney and McKinley Elementary schools were screened for symptoms of Irlen syndrome, with 31 subjects identified at Whitney and 40 subjects identified at McKinley. The effects on reading achievement were investigated under conditions of immediate use and delayed use, with one group (Whitney) provided optimum coloured overlays for 3 months and the other group (McKinley) delayed treatment for 3 months. The effects were then assessed for a further three months with both groups.

After three months of use of overlays, the Whitney group demonstrated a significant improvement in reading achievement with mean gains in grade equivalence scores of between 1 year 2 months and 1 year 7 months. However, between the three to six month period of use, the gains for the Whitney group reached a plateau, with no significant improvement in reading achievement. The McKinley group had negligible gains in reading achievement during the first 3 months without the use of overlays, but significant gains during the 3 month to 6 month phase with the use of overlays, which ranged from 1 year 8 months to 2 years 8 months. It was suggested that the reported reduction in print and background distortions may improve accuracy of word recognition and allow attention to be directed more to the meaning of what is being read than to word recognition thus enhancing reading comprehension. The plateau effect identified for the Whitney group between 3 and 6 months could be related to the fact that after 3 months of overlay use, many students had reached grade level in reading achievement.
In the last decade, there has been increasing recognition that people with literacy problems may have visual processing difficulties (Booth, Perfetti, MacWhinney, & Hunt, 2000; Helenius & Salmelin, 2002; Skottun, 2000; Stein, 2001; Terepocki, Kruk, & Willows, 2002). One area of investigation has centred upon the proposal by Irlen (1991a) of a specific visual–perceptual dysfunction, which has been called Irlen Syndrome (IS) or visual discomfort (Conlon, Lovegrove, Chekaluk, & Pattison, 1999). Symptoms of IS 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 focussing for an extended period of time (Irlen, 1991a; Meares, 1980). These symptoms are claimed to be reduced by the use of coloured filters (Irlen, 1991a) and do not appear to be related to regular optometric problems, although binocular and accommodative anomalies may occur in conjunction with the syndrome (Evans, Patel, Lighthouse, Eperjesi, Speedwell et al., 1999; Evans, Wilkins, Brown, Busby, Wingfield, Jeane, & Bald, 1996; Evans, Wilkins, Busby, & Jeanes, 1996; Scott, McWhinny, Taylor, Stevenson, Irons, Lewis et al., 2002; Simmers, Gray, & Wilkins, 2001). While these optical anomalies were not considered to be the underlying physiological basis of the condition (Scott et al., 2002; Simmers et al, 2001), some people who respond to the use of coloured filters, also require ocular motor treatment, suggesting the symptoms may sometimes be in part due to accommodation inefficiency (Evans et al., 1999; Scott et al., 2002).

Irlen (1991a) hypothesised that individually diagnosed coloured overlays or coloured lenses would filter out the wavelengths and frequencies of the white light spectrum to which the person is sensitive. This claim of hypersensitivity to specific frequencies of light (colours) has been supported by Parker and Henson-Parker (2003) who used spectronic data analysis to examine the performance of coloured transparencies reported to reduce visual distortions. They found the coloured transparencies significantly suppressed photon energy to the eye for specific photon bands (light frequency bands), and claim that specific frequencies of light (photon energies) may cause electrochemical abnormalities in the eye's photoreceptors. The uncorrected eye is hypothesised to be transmitting distorted, high energy electronic signals to the brain when it encounters light frequencies to which it is sensitive. The authors liken this experience to the auditory acoustic feedback that occurs when a microphone is placed close to a speaker, causing the sound amplifier circuits to be overdriven. They conclude that visual anomalies may be corrected by selectively depressing specific frequencies of light (colours) to which the person's visual system is hypersensitive. The authors also claim that optic nerve time series data from a study by Lewine (1999) showed a higher power spectrum content for people with IS when not using coloured filters, which is consistent with their theory. Lewine (1999) found that visual evoked responses for subjects with IS showed an organised dipolar pattern with coloured filters reported to reduce symptoms and a more complex field pattern without them. It has also been hypothesised that the symptoms of IS could be related to a deficit in the magnocellular visual neurological pathway (Demb, Boynton, Best, & Heeger, 1998), which may cause the visual images between consecutive eye fixations to overlap when reading (Boden & Brodeur, 1999). Information is transmitted from the eye to the brain by two parallel pathways, the magnocellular (M) pathway and the parvocellular (P) pathway. These two pathways are claimed to have specific roles in reading, with the M pathway guiding eye movements and the P pathway providing detailed information at each focus point or fixation (Williams & Lovegrove, 1992). The movement of eyes from one focus point to the next (saccades) last approximately 20 to 40 milliseconds. During this eye movement, visual information does not appear to be processed (Rayner & Pollatsek, 1992). Detail is extracted during the focus or fixation phase, which lasts at least 200 to 300 milliseconds, but could be longer if there are difficulties with word identification (Whiteley & Smith, 2001). Information up to 15 character spaces is extracted during fixation and it is information to the far right of this visual field, transmitted by the M pathway, which helps guide eye movements to the next word or word cluster important for gaining meaning from text (Whiteley & Smith, 2001). The M pathway is not only claimed to guide eye movements, but may also be involved in suppressing the potential overlap of images between consecutive eye fixations (saccadic suppression; Hussey, 2002). It has been suggested that accumulating evidence indicates deficits in the M pathway may occur in approximately 70% of cases of dyslexia (Whiteley & Smith, 2001).

Studies have identified a diminished or delayed visual evoked potential for poor readers along the M pathway in response to moving stimuli (Brannan, Solan, Ficarra, & Ong, 1998; Romani, Conte, Caliecco, Bergamaschi, Versino, Lanzi et al., 2001). Investigations of poor motion sensitivity (Slaghuis & Ryan, 1999; Talcott, Hanson, Assoku, & Stein, 2000) have also found a reduced activation of the V5/MT area of the visual cortex, which is sensitive to visual motion and is dominated by magnocellular input. Colour filtering is claimed to influence the functioning ability of the M pathway (Edwards, Hogben, Clark, & Pratt, 1996), with Chase, Ashourzadeh, Kelly, Monfette, and Kinsey (2003) conducting a number of studies which found that red light suppressed M-cell activation, thus affecting M-pathway activities such as perception of motion, perception of global shape and possibly reading performance. According to the authors, using blue filters should improve M-pathway activities because longer wavelengths (red) would then be absent and therefore unable to suppress M-cell activation (Chase et al., 2003).

Numerous controlled studies have also reported improvements in reading with the use of coloured filters. These studies have reported improvements in reading when using coloured plastic overlays or coloured computer monitors (Bouldoukian, Wilkins, & Evans, 2002; Croyle, 1998; Jeanes, Busby, Martin, Lewis, Stevenson, Pointon et al., 1997; Northway, 2003; Scott et al., 2002; Solan, Ficarra, Brannan, & Rucker, 1998; Tyrrell, Holland, Dennis, & Wilkins, 1995; Wilkins & Lewis, 1999; Wilkins, Lewis, Smith, Rowland, & Tweedie, 2001;
Robinson & Foreman, 1999a). Additional reading tuition (Kyd, Lovegrove, & McNicol, 1993), reported positive results (Cotton & al., 1994; Wilkins & Lewis, 1999). Brown, Busby, Wingfield, Jeanes et al., 1999a; Wilkins, Evans, Brown, Busby, Wingfield, Jeanes et al., 1994; Wilkins & Lewis, 1999). Not all studies, however, have reported positive results (Cotton & Evans, 1990; Martin, Mackenzie, Lovegrove, & McNicol, 1993), which is to be expected, as reported improvements in print clarity may make word recognition easier, but is unlikely to lead to the development of word recognition skills without additional reading tuition (Kyd, Sutherland, & McGettrick, 1992; Robinson & Foreman, 1999a).Identifying the specific colour lens combination which is reported to most reduce symptoms of IS usually requires a lengthy (two hour) diagnostic-prescriptive session, involving a person who has considerable experience and training (Irlen, 1991a). These time and expertise requirements can limit the opportunity for children with IS in regular classes to be identified and assisted. There is thus a need to investigate the use of less detailed and time consuming methods of identification and intervention.

In order to provide more access to colour filtering, Irlen (1991a) has developed a range of tinted plastic overlays, which require far less time and expertise for assessment and intervention than the use of coloured lenses. She also developed a method of training to use these overlays called Screener training, which could be conducted in regular school situations, without the need for an extensive time commitment. While these overlays do not have the range and density of colours usually required to most effectively reduce symptoms, they do provide a method for identifying and assisting the potentially large numbers of children in regular class situations who may not have sufficient symptoms to seriously delay reading, but who still find reading tasks difficult. Such children may not need the intensive diagnostic-prescriptive analysis required for people with more serious symptoms. They may also benefit from the use of coloured plastic filters as an interim measure, while deciding whether to progress to the use of coloured lenses.

Estimates of incidence of IS in the general population range from 12% to 20% (Evans, Patel, Wilkins, Lightstone, Eperjesi, Speedwell et al., 1999; Irlen, 1997; Jeanes et al., 1997), with 12% considered to be more accurate (Evans et al., 1999). Large numbers of children with symptoms are thus likely to be present in regular classes, where the only assistance likely to be available will come from the regular class teacher and where a less intensive form of intervention, such as the use of coloured overlays, is the most feasible method of support. There is thus a need to investigate the effectiveness of the use of coloured filtering in the regular classroom through the use of coloured plastic overlays.

There have been studies of the use of coloured plastic overlays to improve reading achievement in the United Kingdom, with positive results reported in a number of cases (Jeanes et al., 1997; Kyd et al, 1992; Tyrrell et al., 1995; Wilkins & Lewis, 1999; Wilkins et al., 2001; Bouldoukian et al., 2002). Some of these studies involved the use of coloured overlays in regular class conditions (Jeanes et al., 1997; Kyd et al., 1992; Tyrrell et al., 1995; Wilkins & Lewis, 1999; Wilkins et al., 2001), but identification of symptoms of IS was undertaken by researchers, not regular class teachers, and the studies did not assess a range of reading skills, only rate of reading (with the exception of Kyd et al., 1992). Investigations of the use of coloured lenses have found there can be differential effects for different reading skills, with improvements in reading comprehension and accuracy, but not rate of reading in some studies (Robinson & Conway, 1990; Robinson & Foreman, 1999) and improvements in comprehension and rate of reading, but not accuracy in another study (Robinson & Conway, 1994). There have also been positive investigations of the use of coloured plastic overlays in the United States, with Williams, Le Cluyse, and Rock-Fauschau (1992) and Solan et al. (1998) finding improvements in both rate of reading and reading comprehension. However, these studies were conducted in controlled clinical conditions, not regular class situations. There is thus a need to further assess the effects of coloured plastic overlays over a range of reading skills and also in regular class conditions, with assessment of symptoms of IS and identification of the optimum coloured overlay being conducted by regular class teachers.

This study investigated the effects of the use of Irlen coloured overlays on reading rate, accuracy, fluency and comprehension under regular class conditions, with class teachers (trained as screeners) instigating the study and conducting the assessment. The effects on reading achievement were investigated under conditions of immediate use and delayed use. One group (immediate intervention) was provided with optimum coloured overlays for 3 months, while the other group (delayed intervention) was not given overlays for 3 months. The effects were then assessed for a further 3 months with both groups.

**METHOD**

**Subjects**

The study initially involved all students in grade 3 at Whitney and McKinley Elementary Schools in the Yakima school district, Washington State, which had similar reading programs (six grade 3 classes in total). These students were all screened for symptoms of IS and for competence in basic word attack skills using the reading comprehension sub-test of the Woodcock-Johnson III Tests of Achievement (Mather & Woodcock, 2001). Screening for symptoms of IS was undertaken by teachers in the school system who had been trained over 2 days as Irlen screeners. Students who met the criteria of symptoms of IS and competence in basic word attack skills (N=71) were then divided into an immediate intervention group with overlays (Whitney Elementary School) and a delayed intervention group without overlays (McKinley Elementary School). All subjects were then pre-tested on the Gray Oral Reading Test – IV (GORT-IV).
Immediate and Delayed Intervention Groups on Pre-Testing

Table 1: Mean Grade Levels on the GORT-IV for Immediate and Delayed Intervention Groups on Pre-Testing

<table>
<thead>
<tr>
<th>GORT-IV</th>
<th>Immediate (N=31)</th>
<th>Delayed (N=40)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Male=15; Female=16)</td>
<td>(Male=31; Female=10)</td>
</tr>
<tr>
<td>Rate</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Accuracy</td>
<td>2.8</td>
<td>1.3</td>
</tr>
<tr>
<td>Fluency</td>
<td>2.7</td>
<td>1.2</td>
</tr>
<tr>
<td>Comprehension</td>
<td>3.7</td>
<td>1.8</td>
</tr>
</tbody>
</table>

(Wiederholt & Bryant, 2001). Table 1 shows the mean grade level of subjects on the four sub-tests of the GORT-IV.

No subjects with symptoms of IS who met the grade achievement criteria were eliminated from the study to ensure that conditions were as close as possible to those of the regular school. As a consequence, while the majority of students were from a regular education setting (N=52) some were in transition from a Spanish reading to English reading (N=11) and some were in special education (N=8). A one way ANOVA of mean score differences between the three education categories of students revealed no significant differences for all four measures of reading at McKinley at pre-testing. There were also no significant differences on subsequent testing occasions, which suggests that the effects of coloured overlay use were similar for students from each education category. At Whitney there were significant differences at pre-testing between the three education categories of students for rate, accuracy and fluency, with students in regular education having higher means than students in transition education who, in turn, had higher mean scores than students in special education. These differences, however, remained similar on subsequent testing occasions, which suggests that effects of coloured overlay use were the same for each education category of students (with the possible exception of Fluency). The differences at pre-testing were: Rate (F 2,25=5.98; p=.008), Accuracy (F 2,25=4.78; p=.017) and Fluency (F 2,25=5.26; p=.012). At final testing the differences were: (Rate: F 2,25=8.27; p=.002; Accuracy: F 2,25=5.13; p=.014; Fluency: F 2,25=6.46; p=.005).

MEASURES

Woodcock-Johnson III Tests of Achievement (Mather & Woodcock, 2001)

The Woodcock-Johnson III Tests of Achievement have two parallel forms (A and B), divided into two batteries – Standard and Extended. The 12 test Standard Battery provides a broad set of scores, with the 10 tests in the Extended Battery giving more in-depth diagnostic information on specific academic strengths and weaknesses. The reading comprehension sub-test used in this study involved orally supplying the missing word removed from a sentence or very brief paragraph (e.g., "Wood", said the ------, biting the hand that fed it.).

The battery of tests was normed between 1996 and 1999 on a large representative national sample of people from Kindergarten to University level, involving 8,818 children and adults. For school age children, the norm group included public, private and home-schooled students, as well as students with disabilities who were included at least part-time in regular classes. Internal consistency reliability was calculated using a Rasch approach and using split-half reliabilities. These estimates were uniformly high and most often in the .80s and .90s for individual tests and in the .90s for test clusters. Stability reliability (test-retest) was calculated over one to three year intervals and ranged between the .70s to .90s. Extensive confirmatory factor analysis provided validating evidence for the internal structure of the battery. Scores on the achievement measures correlated in the range of .50 to .80 with corresponding tests on other well-established measures of achievement. Differential item functioning was assessed across gender, ethnic and racial groups, with factor analysis results suggesting that the same factors are measured across groups.

Irlen Reading Perceptual Scale (Irlen, 1991b)

The Perceptual Scale consists of three sections: (1) the Irlen Reading Strategies Questionnaire; (2) a series of visual tasks; and (3) an assessment of the extent to which performance on these visual tasks and on reading is improved by the use of coloured plastic overlays. The Reading Strategies Questionnaire has 32 questions. Sixteen of these questions are related to reading strategies and reading behaviours (skipping or re-reading lines, misreading words, losing place, poor comprehension and slow reading). The Questionnaire also has 16 questions related to eye strain and fatigue while reading (headaches, dry, itchy or burning eyes, fatigue, blinking and squinting).

The second section of the Perceptual Scale involves a series of visual tasks, including counting the number of squares in specific rows on grid patterns superimposed on cubes and answering questions about the nature of any visual distortions. Subjects are also required to answer questions about possible distortions while observing other visual images such as lines of music and print. The third section involves placing coloured plastic overlays over the cubes, other visual images and print, to ascertain whether symptoms may be reduced by a particular colour. The criteria used to identify moderate symptoms of the syndrome is a score of 8 or more out of 32 on questionnaire items related to reading difficulties, eye strain and fatigue, and a score of 4 or more out of 14 on each of the visual tasks.

Validity studies of the Irlen manual have been undertaken (Haag, 1984; Miller, 1984; Tyrrell et al., 1995). Haag (1984) found significant differences on scores of all sections between reading disabled and normally achieving students. Miller (1984) also identified significant differences in scores between students of low and high reading ability, while Tyrrell et al. (1995) found a significant association between subjects with poor scores on the Irlen screening manual and below average reading achievement. Gray (1999)
reported high internal validity of subsections of the manual, as well as significant relationships between scores on the manual and standardised measures of reading achievement, spelling achievement, and visual processing.

High test-retest reliability of the coloured lenses chosen to most reduce visual distortions has been documented by Robinson and Foreman (1999a) over periods of 8 to 20 months of use. An estimate of the consistency of preferences for coloured overlays reported to most reduce visual distortions was undertaken by Jeanes et al. (1997) and Wilkins et al. (2001). They found that such preferences for school age children were more consistent than could reasonably be expected on the basis of chance alone. Wilkins (1997) also assessed the reliability of choice of coloured filters for children and adults over periods of six months to two years of use, with much higher consistency again being obtained than could be expected on the basis of chance.

The Gray Oral Reading Test – IV (Wiederholt & Bryant, 2001)

The Gray Oral Reading Test has two parallel forms, each consisting of 14 developmentally sequenced reading passages which decrease in print size. For each reading passage, there are five multiple choice (four item) questions following each story. Four measures of reading skill are assessed: (1) Reading rate (time taken to read each passage in seconds); (2) Accuracy (number of deviations from print); (3) Fluency (a combination of rate and accuracy); and (4) Comprehension (the number of multiple choice questions answered correctly). The test was normed on a sample of more than 1600 children aged 7- to 18-years-old in the United States, using a nationally representative sample stratified according to key demographic variables (race, gender, ethnicity and geographic region). Internal consistency reliability for each of the 4 sub-tests was above 0.90. Test-retest reliability for each of the sub-tests ranged from .62 to .92. Validity estimates were extensive, with correlations between this test and other recognised tests of reading ranging from .38 to .67. Internal consistency of the sub-tests was .58 to .88, with the test able to identify poor readers who scored significantly below the mean on independent analyses. Bias studies show an absence of bias based on gender and ethnicity.

PROCEDURES

All students in grade 3 at both schools involved in the study were pre-screened for Irlen Syndrome using the Reading Strategies Questionnaire section of the Irlen Reading Perceptual Scale (Irlen, 1991c). Students with at least moderate symptoms on this questionnaire were then screened on the Woodcock-Johnson III Tests of achievement if there are still deficiencies in the basic word attack skills (Kyd et al., 1992; Robinson & Foreman, 1999a).

Students who met the above two criteria were then further screened on the full Irlen Reading Perceptual Scale (Irlen, 1991c) by teachers at the schools who had undergone screener training, to confirm the original screening results using the questionnaire. This further screening was also used to determine each student's unique profile of symptoms of IS and to identify the specific Irlen coloured overlay for each student.

Table 2: Mean Grade Scores (Yrs and Mths) at Pre-Test 3 Months and 6 Months and Gain Scores for Immediate and Delayed Intervention Groups

<table>
<thead>
<tr>
<th></th>
<th>Immediate (N=31)</th>
<th>Delayed (N=40)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gain 3mths 6mths</td>
<td>Gain 3mths 6mths</td>
</tr>
<tr>
<td>GORT-IV</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate</td>
<td>2.7 3.9 4.1 1.2 0.4</td>
<td>2.7 2.8 4.6 0.1 1.8</td>
</tr>
<tr>
<td>Accuracy</td>
<td>2.8 4.3 4.4 1.7 0.1</td>
<td>2.7 2.8 5.5 0.1 2.8</td>
</tr>
<tr>
<td>Fluency</td>
<td>2.7 3.9 4.2 1.2 0.5</td>
<td>2.5 2.7 4.9 0.1 2.2</td>
</tr>
<tr>
<td>Comprehension</td>
<td>3.7 4.8 4.8 1.1 0.0</td>
<td>3.3 2.8 5.2 -0.5 2.6</td>
</tr>
</tbody>
</table>

Figure 1: Changes in the Average of Mean Grade Scores for Immediate and Delayed Intervention Groups on Sub-Tests of the GORT-IV at Pre-Test and 6 Months
the immediate intervention group (Whitney) receiving their overlays. Retesting occurred at 3 months on a parallel form of the Gray test, at which time the delayed intervention group (McKinley) received their preferred coloured overlay. This 3 month retesting allowed a comparison of 3 months reading performance with overlays at Whitney and without overlays at McKinley. Both groups were further reassessed after an additional 3 month interval on the Gray test which allowed a comparison of both groups’ reading performance with the use of preferred overlays (for a period of 6 months of use at Whitney and for 3 months of use at McKinley).

At the beginning of the study, third grade staff at both schools were given 2 hours in-service training on the nature of IS and the likely impact of IS on reading. They were also provided information on how to identify children likely to have IS and were made aware of the likely effects of using Irlen overlays in the classroom. Parents were also invited to an information meeting about the nature and implications of having symptoms of IS.

In order to ensure that the conditions of use of Irlen overlays were as close as possible to regular class conditions, no students with symptoms of IS and other forms of disability who met the criteria for inclusion were eliminated from the study. The study sample thus included students in special education situations (as discussed previously), as well as small numbers of unmedicated students with ADHD and students with emotional problems. In addition, there were a small number of students in the study who did not use their overlays consistently. The class teachers did not remind these students to use their overlays when they were not being used.

RESULTS
Mean grade scores on the four sub-tests of the GORT-IV for immediate and delayed groups at pre-testing, 3 months, and 6 months, as well as score gains at the 3 months and 3 to 6 months intervals are outlined in Table 2. A graph of changes in the average of mean grade scores on sub-tests of the GORT-IV at each testing occasion are shown in Figure 1. The immediate intervention group used their preferred coloured overlay for the whole of the 6 month study, with the delayed intervention group having their preferred overlay for the 3 to 6 month period, thus having 3 months of overlay use.

Table 3: Significance of Difference of Changes in Scores for Immediate and Delayed Intervention Groups Between Pre-Testing and 3 Month Testing on Sub-Tests of the GORT-IV

<table>
<thead>
<tr>
<th>GORT-IV</th>
<th>MS</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>18.57</td>
<td>17.21</td>
<td>0.000</td>
</tr>
<tr>
<td>Accuracy</td>
<td>31.64</td>
<td>18.76</td>
<td>0.000</td>
</tr>
<tr>
<td>Fluency</td>
<td>22.61</td>
<td>20.44</td>
<td>0.000</td>
</tr>
<tr>
<td>Comprehension</td>
<td>49.38</td>
<td>12.68</td>
<td>0.001</td>
</tr>
</tbody>
</table>

After 3 months use of overlays, the immediate intervention group at Whitney were above grade level for all reading measures on the GORT-IV, with mean gains in grade equivalence scores of between 1 year, 2 months and 1 year, 7 months. In comparison, the delayed intervention group at McKinley, without overlays, showed negligible gains, with a loss of 5 months in comprehension.

Between the 3 to 6 month period, the gains for the immediate intervention group tended to plateau, with gains ranging between 0 and 5 months. The delayed intervention group, however, had substantial gains during this period with the use of overlays. These gains ranged from 1 year, 8 months to 2 years, 8 months.

Differences in mean scores on the GORT-IV from pre-testing to 3 month testing between immediate and delayed intervention groups were analysed using MANOVA. These differences were significant: F(4,70)=4.6; p<.003. Changes in mean grade scores between immediate and delayed intervention groups on sub-tests of the GORT-IV between pre-testing and 3 month testing was also analysed using MANOVA, with the results being significant, as outlined in Table 3.

Differences in mean scores on the GORT-IV from testing at 3 months to testing at 6 months between immediate and delayed intervention groups using MANOVA also revealed a significant difference; F(4,70)=4.6; p=.003, with the delayed intervention group improving at a greater rate than the immediate intervention group. At this stage both groups were using their preferred overlays, with the immediate intervention group having 6 months of use and the delayed intervention group having 3 months of use. An analysis of changes in mean scores on the GORT-IV sub-tests between 3 months and 6 months using MANOVA confirmed the significant improvement for the delayed intervention group compared to the immediate intervention group, as shown in Table 4.

Table 4: Significance of Difference of changes in Scores for Immediate and Delayed Intervention Groups Between 3 Months Testing and 6 Months Testing on Sub-Tests of the GORT-IV

<table>
<thead>
<tr>
<th>GORT-IV</th>
<th>MS</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate</td>
<td>8.88</td>
<td>8.82</td>
<td>0.0004</td>
</tr>
<tr>
<td>Accuracy</td>
<td>23.22</td>
<td>17.12</td>
<td>0.000</td>
</tr>
<tr>
<td>Fluency</td>
<td>16.72</td>
<td>14.90</td>
<td>0.000</td>
</tr>
<tr>
<td>Comprehension</td>
<td>30.18</td>
<td>10.30</td>
<td>0.002</td>
</tr>
</tbody>
</table>

After 3 months use of overlays, the immediate intervention group at Whitney were above grade level for all reading measures on the GORT-IV, with mean gains in grade equivalence scores of between 1 year, 2 months and 1 year, 7 months. In comparison, the delayed intervention group at McKinley, without overlays, showed negligible gains, with a loss of 5 months in comprehension.

Between the 3 to 6 month period, the gains for the immediate intervention group tended to plateau, with gains ranging between 0 and 5 months. The delayed intervention group, however, had substantial gains during this period with the use of overlays. These gains ranged from 1 year, 8 months to 2 years, 8 months.

Differences in mean scores on the GORT-IV from pre-testing to 3 month testing between immediate and delayed intervention groups were analysed using MANOVA. These differences were significant: F(4,70)=4.6; p<.003. Changes in mean grade scores between immediate and delayed intervention groups on sub-tests of the GORT-IV between pre-testing and 3 month testing was also analysed using MANOVA, with the results being significant, as outlined in Table 3.

Differences in mean scores on the GORT-IV from testing at 3 months to testing at 6 months between immediate and delayed intervention groups using MANOVA also revealed a significant difference; F(4,70)=4.6; p=.003, with the delayed intervention group improving at a greater rate than the immediate intervention group. At this stage both groups were using their preferred overlays, with the immediate intervention group having 6 months of use and the delayed intervention group having 3 months of use. An analysis of changes in mean scores on the GORT-IV sub-tests between 3 months and 6 months using MANOVA confirmed the significant improvement for the delayed intervention group compared to the immediate intervention group, as shown in Table 4.

Paired sample t-tests were conducted of changes in mean grade scores on all sub-tests for the immediate intervention group, with the p=.01 level of significance used. This analysis was conducted for all phases of the study (pre-test to 3 months, pre-test to 6 months, 3 to 6 months). The immediate
interval group improved significantly in the initial 3 month phase, but then there was a plateau effect with no significant improvements for the next 3 months. These changes are outlined in Table 5.

**DISCUSSION**

The results of this study confirm the results of previous positive studies on the use of coloured overlays in classroom conditions (Kyd et al., 1992; Tyrrell et al., 1995), especially when used for extended periods (Jeans et al., 1997; Wilkins et al., 2001). This study, however, found significant improvement in a range of reading skills, just not speed of reading as had occurred with the previous studies cited above. The significant increases in a range of reading measures using coloured overlays corresponds to results by Williams et al. (1992) and Solan et al. (1998). Studies using coloured lenses have also found improved reading performance in more than one reading skill (Harris & McRown-Hill, 1990; Irving & Irving, 1997; Robinson & Conway, 1999a, b, c). The conditions of use for this study were designed to approximate the regular classroom situation as closely as possible. Teachers were instructed not to provide special attention to students with overlays and in fact stated that they did not want to be responsible for reminding students to use their overlays. The classes also included children from special education settings and children in transition from special education to regular education settings. Children from both regular and special settings in this study were thus able to make significant improvements with the use of coloured overlays.

The significant improvement in all reading skills assessed in this study could be in part attributed to the exclusion of children with reading grade levels of below 1.5. This criterion was included as reported improvements in print clarity when using coloured filters are less likely to lead to improved reading achievement if there are deficiencies in basic word attack skills (Kyd et al., 1992; Robinson, 1994). In particular, problems with letter-sound analysis and synthesis (phonics) are common in children with reading problems in the English language (Maughan, 1995; Shany & Biemiller, 1995; Wagner, 1988). Improved print clarity is unlikely to result in the spontaneous development of these skills, especially due to the high degree of irregularity between sounds and symbols in the English language (Landerl, Wilmer, & Trinity, 1997). Wilkins et al. (2001) found that phonological reading strategies were not related to the visual symptoms of IS or to the benefit received from the use of coloured overlays. Previous studies of coloured lens use which assessed multiple reading skills, but which did not exclude children with deficiencies in the basic word attack skills, did not find improvement in all reading skills assessed (Robinson & Conway, 1990, 1994; Robinson & Foreman, 1990).

If subjects with IS and proficiency in basic word attack skills report improved print clarity when using coloured overlays, they may identify words more quickly and automatically, thus directing more attention towards comprehension. The oral reading of students with reading difficulties is usually characterised by hesitations and mispronunciations (Lyon & Chhabra, 1996). This is also typical of students with IS (Irlen, 1991a), suggesting that effort is allocated to decoding at the expense of comprehension (Perletti, 1992; Stanovich, 1994). People with symptoms of IS also frequently report a reduction in span of recognition (Irlen, 1991a; Robinson & Conway, 1995). This could lead to a slower and less accurate word-by-word reading by restricting exposure to words in peripheral vision (Jennings & Underwood, 1984; Rayner & Pollatsek, 1989). Increases in comprehension reported in this study may also provide greater access to semantic patterns in print as an aid to word identification (Denckla, 1996; Stanovich, 1986). Students with literacy problems may rely more on semantic cues to identify unknown words than skilled readers (Brennitz, 1997; Fink, 1996), and if such cues are less available, then fluency and accuracy may be restricted. Prior, Sanson, Smart, and Oberklaid (1991) found that poor readers tend to make wild guesses based on a few letters in a word, with Valdoiz, Gerard, Vanault, and Dugas (1995) hypothesising that such children have a limited window over which attention is focussed, causing them to guess at letters outside this focus.

While significant increases in reading achievement occurred for all reading skills assessed for both groups in this study when they were provided their chosen coloured overlay, there was a plateau effect for the immediate intervention group for the 3 month to 6 month period of overlay use (see Figure 1). A similar plateau effect was identified in an immediate intervention group by Martin et al. (1993) for 4 subjects with IS (Irlen, 1991a), suggesting that effort is allocated to decoding at the expense of comprehension. The oral reading of students with reading difficulties is usually characterised by hesitations and mispronunciations (Lyon & Chhabra, 1996). This is also typical of students with IS (Irlen, 1991a), suggesting that effort is allocated to decoding at the expense of comprehension (Perletti, 1992; Stanovich, 1994). People with symptoms of IS also frequently report a reduction in span of recognition (Irlen, 1991a; Robinson & Conway, 1995). This could lead to a slower and less accurate word-by-word reading by restricting exposure to words in peripheral vision (Jennings & Underwood, 1984; Rayner & Pollatsek, 1989). Increases in comprehension reported in this study may also provide greater access to semantic patterns in print as an aid to word identification (Denckla, 1996; Stanovich, 1986). Students with literacy problems may rely more on semantic cues to identify unknown words than skilled readers (Brennitz, 1997; Fink, 1996), and if such cues are less available, then fluency and accuracy may be restricted. Prior, Sanson, Smart, and Oberklaid (1991) found that poor readers tend to make wild guesses based on a few letters in a word, with Valdoiz, Gerard, Vanault, and Dugas (1995) hypothesising that such children have a limited window over which attention is focussed, causing them to guess at letters outside this focus.

### Table 5: Paired Sample t-tests of Differences Between Pre and 3 Months, Pre and 6 Months and 3 and 6 Months Testing for the Immediate Intervention Group

<table>
<thead>
<tr>
<th></th>
<th>Pre to 3 Months</th>
<th>Pre to 6 Months</th>
<th>3 to 6 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>t</td>
<td>df</td>
<td>Sig</td>
</tr>
<tr>
<td></td>
<td>(2-tailed)</td>
<td>(2-tailed)</td>
<td>(2-tailed)</td>
</tr>
<tr>
<td>Rate</td>
<td>-8.03</td>
<td>30</td>
<td>0.000</td>
</tr>
<tr>
<td>Accuracy</td>
<td>-8.78</td>
<td>30</td>
<td>0.000</td>
</tr>
<tr>
<td>Fluency</td>
<td>-9.76</td>
<td>30</td>
<td>0.000</td>
</tr>
<tr>
<td>Comprehension</td>
<td>-3.76</td>
<td>30</td>
<td>0.001</td>
</tr>
</tbody>
</table>

developed the basic reading skills, the reported improved print clarity may have allowed them to use these established skills more effectively. Therefore, only a short amount of time was required to reach an acceptable grade level. Once grade level was reached, improvement slowed to a rate more like that of readers in the grade already performing at grade level. An inspection of grading levels for the immediate intervention group after 3 months of use of coloured overlay use (see Table 2) shows them to be above grade 3 level for all reading skills assessed. As the subjects for this study were grade 3 students, this suggests they had reached an acceptable grade level of reading and any further rate of development may be at normal grade expectations rather than at accelerated levels. In previous studies of the effects of coloured filters, a criteria for inclusion of efficiency in word attack was not used, with subjects having reading ages far below their chronological age (Kyd et al., 1992; Robinson & Conway, 1990, 1994; Robinson et al., 1999). It would be likely that many students in these earlier studies had deficiencies in the basic word attack skills and would take much longer to reach acceptable grade levels.

There remains the question of whether the significant improvements in reading skills for both groups when using preferred coloured overlays could be due to a placebo effect. It could be argued, however, that if children continue to use coloured overlays for a sustained period (with teachers instructed not to remind students to use them), then it is likely they are receiving a benefit beyond what could be attributed to increased motivation or expectation. The large gains in reading skills by both groups is also likely to be beyond what might be expected from increased parent and teacher interest and support, or the novelty of a new method. A number of studies of the use of coloured filters that used placebo controls have found improved reading performance (Bouldoukian et al., 2002; Wilkins et al., 2002; Wilkins & Lewis, 1999; Robinson et al., 1999).

This study aimed to investigate the effects of coloured overlay use when instigated and implemented by regular class teachers in regular class conditions. However, children with reading grade levels of below 1.5 were excluded. Further field studies are required to investigate the effects of coloured filters for students with symptoms of IS who may have lower reading levels and may lack the basic word attack skills necessary to make quick improvements in reading once print clarity has been improved. While such children may report that print is clearer, additional reading tuition may be necessary if improved print clarity is to lead to improved reading achievement. It would also be useful in further field studies to identify the frequency of choice of coloured overlays. The transient system theory, which has been highlighted as a possible causal mechanisms for IS (Whiteley & Smith, 2001), predicts a more likely choice of blue coloured overlays (Chase et al., 2003) and an analysis of frequency of choice of coloured overlays may help to support this theory. It would also be useful in further studies to keep a record of the consistency of use of overlays to see if students who used the overlays less consistently had a poorer reading performance.

The results of this study offer further support for the claim that the visual processing problems as identified by IS may be a significant causal factor in reading difficulties. The study also supports the results of previous investigations which found that the use of coloured plastic overlays in regular classroom conditions can be an effective intervention option for students with IS. Identification of symptoms of IS can be undertaken by regular class teachers without the need for extensive diagnostic training. The large number of students identified with symptoms of IS in this study also suggests that there may be many students in regular classrooms who have visual processing problems which restrict reading development, but who are not receiving specific assistance. Such children do not have any "visible" signs of a difficulty and while they may be working below their full academic potential, their reading skills may not be sufficiently delayed to receive special assistance. In this study at least 27% of students in 6 third grade classes were identified as having symptoms of IS. It needs to be emphasised, however, that there are usually complex patterns between cause and effect in the area of literacy development. Learning to read involves many cognitive processes and a breakdown in any one of these may affect ability to read. While we need to be cautious of simple linear explanations between cause and effect, the results of this study suggest that the use of coloured overlays should be considered as one possible option for students with literacy problems.

REFERENCES


