A NEW SCREENING PROGRAM FOR THE DETECTION OF AMBLYOPIA AND OTHER EARLY VISUAL DISORDERS IN TODDLERS AND PRESCHOOL CHILDREN

JAMES R. DROVER

Running Head: A NEW SCREENING PROGRAM

A New Screening Program for the Detection of Amblyopia and Other Early Visual Disorders in Toddlers and Preschool Children

©James R. Drover, M.Sc

A thesis submitted to the School of Graduate

Studies in partial fulfillment of the

Requirements for the degree of

Doctor of Philosophy

Department of Psychology

Memorial University of Newfoundland

July 2006

St. John's Newfoundland

Abstract

Several studies indicate that permanent visual dysfunction can be easily prevented if amblyopia or amblyogenic factors are detected sufficiently early in life. Although this highlights the importance of mass vision screening of young children, the implementation of preschool vision screening programs is lacking in most industrialized countries for at least three reasons. First, the detection of subtle amblyogenic factors is often quite difficult using current techniques of assessment. Second, the screening of young children is often challenging due to the attentional demands of the tests. Third, there is a lack of consensus among researchers/clinicians as to what constitutes an effective vision screening program.

In order to address these issues, we implemented a population-based screening program to assess toddlers and preschoolers in the St. John's, NL, Canada, metropolitan area. The program was one of the most comprehensive conducted to date, as we attempted to assess 954 children on up to five separate aspects of functional vision (visual acuity, ocular alignment/motility, stereoacuity, refractive error, and contrast sensitivity). The screening battery included commonly used tests to assess toddlers and preschoolers such as the Teller Acuity Cards, Randot E Stereotest, and the cover-uncover test along with promising, new tests such as the Lea symbols, the Randot Preschool Stereoacuity Test, the contrast sensitivity cards, and the Welch-Allyn SureSight autorefractor. Children who failed at least one test were sent to an optometrist for an optometric gold standard exam. To determine which tests should be included in an effective preschool vision screening program, four measures of validity (sensitivity, specificity, positive

predictive value, and negative predictive value) were calculated for tests of each visual function and 29 different combinations of tests.

Results suggested that although all individual tests possessed relatively high positive predictive values, they generally possessed low sensitivity, specificity, and negative predictive value. Combinations of tests tended to yield high sensitivity and positive predictive value, but relatively low specificity and negative predictive value. The most effective combinations were those that included autorefraction and the ocular alignment/motility tests. Also, prevalence estimates of vision disorders based on the present study were in agreement with those from other areas of Canada and other industrialized nations. Finally, completion times and completion rates revealed that of all tests in the screening battery, autorefraction was the easiest for children to complete, whereas contrast sensitivity was the most difficult.

Based on the data analyses, we have made several important recommendations that may be instrumental in improving the quality of preschool vision screening: (1) vision screening should be conducted with a combination of three to four tests; (2) the program should implement relatively lenient referral criteria; (3) autorefraction should be included as part of the combination in order to assess spatial vision; (4) the combination should include either stereoacuity tests or alignment/motility tests; (5) three-test combinations are preferable to four-test combinations as they are more cost-effective.

Acknowledgements

I would first like to offer my sincerest appreciation to my supervisor Dr. Russ Adams. Russ's guidance, intelligence, and wisdom have provided an immeasurable contribution to this thesis and to my years as a graduate student. Equally important is Russ's easy-going nature which is a great form of stress relief as he never panics and always seems to put things in perspective. The Christmas parties and hockey games were also very therapeutic. I would also like to thank Drs. Mary Courage, Cathy Penney, and John Evans for their support; financial, intellectual, and otherwise. I am deeply indebted to each of them as I could always count on them when things were most difficult. Also appreciated is our team optometrist, Dr. Patrick Kean, who always receives rave reviews from our referrals. Special thanks to Dr. Chuck Malsbury for his assistance and thoughtful insights on this thesis. I would also like to express my gratitude to Christina Dove, Kaitlin Penney, Wendy House, and Tia Osmond for their assistance with the data collection. This research was supported by fellowships from the Newfoundland and Labrador Council for Applied Health Research (NLCAHR) and The Canadian National Institute for the Blind (CNIB).

Finally, there are several special people who must be mentioned. My family and friends have been instrumental throughout my university years and without their support, none of this would have been possible. None of them found it unusual that a person could spend so many years as a student. Of course, I would like to offer my deepest, sincerest gratitude to my lovely wife, Jacqueline Blundell. Her kindness, support, and

good nature were always essential and appreciated. I look forward to our time together in Dallas.

Table of Contents

Abstract ii
Acknowledgements iv
Table of Contents vi
List of Tables viii
List of Figures xi
List of Appendices xii
Introduction 1
Recommendations for Early Vision Screening and Tests of Functional Vision 6
Visual Acuity 9
Ocular Alignment/Motility 15
Stereopsis/Stereoacuity 17
Refractive Error 18
Contrast Sensitivity 22
Current State of Vision Screening 23
Evaluation of Vision Screening 25
Validity of Vision Screening Programs 27
The Present Study 64

Method 65
Participants 65
Materials and Procedure 66
Visual Acuity 67
Preschoolers 67
Toddlers 78
Ocular Alignment/Motility 80
Preschoolers 80
Toddlers 81
Stereoacuity 82
Preschoolers 82
Toddlers 86
Contrast Sensitivity 89
Preschoolers 89
Toddlers 98
Autorefraction 98
Counterbalancing 99
Referral Criteria 100
Visual Acuity 100
Ocular Alignment/Motility 104
Stereoacuity 105
Contrast Sensitivity 107
Autorefraction 109

Optometric Gold Standard Exam....110

Results....114
Study Population....114
Inter-Tester Agreement....115
Completion Times and Completion Rates....118
Retests, Referrals, Yield, and Prevalence of Visual Deficits....124
Progressive Validity Analyses....128
Validity of Each Test....129
Calculation of Validity by Reducing Criteria and Combining Results From Two Screening Tests....134
Calculation of Validity: Three-Test Combinations....140
Calculation of Validity: Four-Test Combinations....142
Calculation of Validity: Five Test Combinations....143

Discussion....146
Comparison of Prevalence Rates to Those of Other Studies....147
Comparison of Completion Times and Completion Rates....150
Comparison of the Validity of Tests of Each Visual Function to Other Studies....153
Visual Acuity....153
Ocular Alignment/Motility....157
Stereoacuity....159
Contrast Sensitivity....162
Autorefraction....165
Comparison of Combinations of Tests of Visual Functions to Other Studies....168
Cost-Effectiveness of Government Funded Vision Screening....171
Calculation of Costs....172
Effectiveness....173
Cost-Effectiveness Ratios....174
Other Major Issues and Directions for Preschool Vision Screening....177
Methodological Limitations of the Present Study....181
Recommendations for Preschool Vision Screening....185
The Future of Our Vision Screening Program....187
Conclusions....188

References....190
Appendix A....228
Appendix B....232

List of Tables

Table 1. Current screening recommendations by North American vision and pediatric associations. 8

Table 2. The four possible outcomes of a single screening test. 26

Table 3. Validity of screening studies for preschoolers and school-aged children conducted over the last four decades. 31

Table 4. Mean validity estimates and percentage of studies judged effective which employed specific screening tests or combinations of tests 52

Table 5. Summary of the validity of the Vision in Preschoolers (VIP) Study Group screening studies 55

Table 6. Card dimensions and optotype size of the Patti Pics two-alternative-forced-choice isolated symbols test (Patti Pics cards). 78

Table 7. Contrast sensitivity (in contrast sensitivity units) and spatial frequencies of the contrast sensitivity cards. 91

Table 8. Contrast sensitivity (in contrast sensitivity units) and spatial frequencies of the gratings composing the contrast sensitivity booklet. 95

Table 9. Referral criteria for visual acuity testing as recommended by North American pediatric and vision organizations. 102

Table 10. Visual acuity referral criteria for the present study. 103

Table 11. Referral criteria for tests of ocular alignment/motility. 105

Table 12. Recommended stereoacuity referral criteria of North American pediatric/vision organizations. 106

Table 13. Stereoacuity referral criteria for the present study. 107

Table 14. Referral criteria in CS units for each spatial frequency and each age group for the contrast sensitivity cards. 108

Table 15. Referral criteria in CS units for each spatial frequency and each age group for the contrast sensitivity booklet 109

Table 16. Autorefraction referral criteria of the present study 110

Table 17. Visual functions and tests of the gold standard exam. 112

Table 18. Disease/disorder criteria for the optometric gold standard exam..... 113

Table 19. Results of inter-tester agreement analysis for all visual functions..... 116

Table 20. The completion rate and percentage (in bold) for each test implemented in the screening program categorized by age group. 120

Table 21. Mean completion time for each test in the screening program categorized by age group.. 123

Table 22. Number and percentage of children retested and referred for the gold standard exam. 125

Table 23. Data from children referred for gold standard exams categorized by age group and visual function failed 126

Table 24. Frequency and estimated prevalence of the present study categorized by type of vision disorder. 127

Table 25. Summary of validity of tests of each visual function..... 130

Table 26. Sensitivity of tests of each visual function by category of disorder. 132

Table 27. Lenient referral criteria of tests of visual acuity and stereoacuity. 135

Table 28. Validity of visual acuity and stereoacuity tests with lenient referral criteria..... 136

Table 29. Validity of combinations of tests measuring two visual functions..... 139

Table 30. Validity of three vision test combinations.. 141

Table 31. Validity of four vision test combinations..... 143

Table 32. Validity of five vision test combinations. 145

Table 33. Prevalence estimates of the present study, other areas of Canada, and other industrialized countries. 149

Table 34. Mean completion times and completion rates across all tests and age groups for each visual function. 152

Table 35. Summary of the validity of visual acuity tests from vision screening studies. 155

Table 36. Summary of the validity of tests of alignment/motility from the present study and the VIP Study Group (2004). 158

Table 37. Summary of the validity of stereoacuity tests from various screening studies. 161

Table 38. A comparison of the validity of CS to the validity of visual acuity tests. 163

Table 39. Summary of the validity of autorefraction screening studies....166

Table 40. A comparison of the validity of test combinations from the present study to previous vision screening studies. 169

Table 41. Simulated cost-effectiveness ratios of several combinations of tests. 176

List of Figures

Figure 1. Photograph of a standard Snellen chart. 11

Figure 2. An example of a letter optotype surrounded by crowding bars. 15

Figure 3. Photograph of the Landolt C visual acuity test. 69

Figure 4. Photograph of the Lea Symbols linear optotype test. 71

Figure 5. Photographs of the Lea Isolated Symbols Book. 72

Figure 6. Photograph of the Patti Pics linear optotype test. 75

Figure 7. Photograph of the Patti Pics Isolated Symbols Book. 76

Figure 8. Photograph of the Patti Pics two-alternative-forced-choice isolated symbols test (Patti Pics cards). 77

Figure 9. Photograph of a Teller Acuity Card. 79

Figure 10. Photograph of a preschooler being tested with the Randot E Stereotest. 83

Figure 11. Photograph of the Randot Preschool Stereoacuity Test. 86

Figure 12. Photograph of the Randot Stereo Smile test. 88

Figure 13. Photograph of a contrast sensitivity card. 90

Figure 14. Photograph of a page of the contrast sensitivity screening booklet. 95

Figure 15. Photograph of a preschooler being tested with the Welch-Allyn SureSight autorefractor. 99

Figure 16. Number of children attempting the screening program classified by age group. 115

Figure 17. Hypothetical, fitted receiver operator characteristics (ROC) curves. 181

List of Appendices

Appendix A. Early consent form for the present screening program. 228
Appendix B. Later consent form for the present screening program.....232

A New Screening Program for the Detection of Amblyopia and Other Early Visual Disorders in Toddlers and Preschool Children

The first decade of life represents a period of substantial development in the human visual system (Atkinson, 1984; Boothe, Dobson, & Teller, 1985; Hickey & Peduzzi, 1987; Wilson, 1988; Yuodelis & Hendrickson, 1986). In particular, the first seven years is a sensitive period during which the eyes compete against each other in order to form permanent connections in the central nervous system (CNS), namely within the visual cortex (Cuiffreda, Levi, & Selenow, 1991; Maurer, Lewis, & Brent, 1989). The normal formation of these cortical connections in turn, yields the highly developed and acute visual functions shown by human adults (Atkinson, 1984). However, if at least one of the eyes is deprived of normal visual experience due to early eye disease (sometimes, even subtle forms), these cortical connections develop poorly and may even regress within those CNS areas subserving the unaffected eye (Maurer et al., 1989; Mills, 1999; Odom, Hoyt & Marg, 1981). If left untreated, this deprivation may lead to a permanent deficit in functional vision, the most common and serious of which is *amblyopia.*

Amblyopia, or "lazy eye", refers to a condition in which one or both eyes possess a substantial reduction of vision (usually defined as a reduction in visual acuity) in the absence of any detectable optical or retinal abnormalities (Cuiffreda et al., 1991; Simon & Kaw, 2001; U.S. Public Health Service, 1994). Amblyopia is the most common cause of vision loss in children with prevalence estimates in industrialized countries ranging from 0.2 to 5.5%, and it is also the leading cause of irreversible monocular vision loss in

those over the age of 20 (Appelboom, 1985; Barry & König, 2003; Bolger, Stewart-Brown, Newcombe, & Starbuck; 1991; Cuiffreda et al., 1991; Friendly, 1993; Junghans & Crewther, 2003; Köhler & Stigmar, 1978; Kvarnström, Jakobsson, & Lennerstrand, 2001; Mills, 1999; Moseley, 1998; Rubin & Nelson, 1993; Simons, 1996; U. S. Public Health Service, 1994; Williams, Harrad, Harvey, Sparrow, & The ALSPAC Study Team, 2001). In Canada, the prevalence of amblyopia ranges from 0.8 to 5.6% (Feightner, 1994; Robinson, Bobier, & Martin, 2000; Ross, Murray, & Stead, 1977).

There are several types of amblyopia, each of which is classified by its amblyogenic factor, i.e., the cause of the infantile deprivation, which results in the visual cortex receiving degraded information from one or both eyes. The most common form is strabismic amblyopia, which is a result of the misalignment of the eyes (strabismus) early in life. The deviating eye can be turned outward (exotropia) or inward (esotropia), the latter being the most common cause of strabismic amblyopia (Ciuffreda, et al., 1991). The second type is *anisometropic amblyopia*, which is due to early anisometropia, a condition in which the degree of optical refractive power[1] differs significantly between

[1] Refractive power refers to the extent to which an optical system bends light to focus it on a specific point (Sekuler & Blake, 1994).

the young child's two eyes (defined typically as a difference of at least 1.5 dioptres[2]). In most cases, the visual system possesses a weak eye that transmits a blurred image and a strong eye that transmits a relatively clear image. The third kind of amblyopia is *image degradation amblyopia*, which is due to an early optical obstruction that prevents the formation of a sharp, clear image in at least one eye. These obstructions commonly include cataracts (the most frequent), corneal opacities, or congenital ptosis, (i.e., a drooping eyelid). Resulting amblyopia may be bilateral (if both eyes are obstructed), or unilateral (if only one eye is obstructed). The final type of amblyopia is *ametropic amblyopia* which occurs bilaterally and is caused by substantial uncorrected refractive error (myopia, hyperopia, and/or astigmatism) in both eyes (Kushner, 1998).

Although amblyopia can occur bilaterally, it is most often unilateral, the effects of which are far more severe (Ciuffreda et al., 1991). During development, the unilateral amblyogenic factor causes the two retinas to be stimulated by differing images, one of which is severely inadequate. For instance, in both anisometropic and image degradation amblyopia, the deprived eye receives an image that lacks the clarity of that received by the unaffected eye. In the case of strabismic amblyopia, the portion of the deprived retina

---

[2] Dioptres are a measure of the refractive power of a lens and are equal to the reciprocal of the focal length in meters. Focal length refers to the distance between the lens and the point where the image is focused. In the human visual system, the image is focused on the fovea of the central retina. Given that the average focal length of the human eye is 17 mm, the total refractive power required is approximately 60 D. In such a case, the eye is emmetropic, i.e., images are focused directly on retina. In many cases however, the eye is ametropic, i.e., it possesses a refractive error. The refractive errors may be spherical, caused by a mismatch between the focusing power of the cornea and lens, and the length of the eye. These errors include myopia (i.e., nearsightedness), in which image is focused in front of the retina, and hyperopia (i.e., farsightedness), in which the image is focused behind the retina. Also, the refractive error may be cylindrical, caused by astigmatism, a condition in which the cornea is misshapen causing different degrees of refractive power along the various meridians of the cornea (Vaughan, Asbury, & Riordan-Eva, 1992). As a result, the retinal image is distorted (Sekuler& Blake, 1994). Note that refractive errors are reported relative to the 60 D norm. For instance, if a subject's eye has a total refractive requirement of 61 D, his/her refractive error is +1 D.

corresponding to that of the unaffected eye actually receives a completely different image from the visual field as it is out of alignment. In most cases, the image falls on the fovea of the aligned eye and on the peripheral retina of the strabismic eye and can cause diplopia (double vision). In order to prevent confusion, the central nervous system accepts only the more detailed foveal image from the unaffected eye and in turn, inhibits or suppresses the image from the affected eye (von Noorden, 1990). This inhibition or suppression eventually becomes permanent, leading to anatomical and physiological changes within the visual cortex. These changes include both reduced activation within Areas 18 and 19 of the visual cortex corresponding to the misaligned eye, and a concomitant reduction in the number of binocular cortical cells (Hubel, 1988; Imamura, Richter, Fischer, Lennerstrand, Franzen, Rydberg, Andersson, Schneider, Onoe, Watanabe, & Langstrom, 1997; Kushner, 1998). Importantly, these binocular cells are critical to the process of stereopsis, the neurological basis of depth perception (Ciuffreda et al., 1991; Cool, 1979; Hubel, 1988). The result is permanently "dysmorphic" functional vision characterized by reductions in visual acuity, contrast sensitivity, and depth perception which persisting beyond approximately 6 or 7 years of age, is very difficult to reverse optically, surgically, or pharmacologically (Ciuffreda et al., 1991; Cooper & Feldman, 1978; Kani, 1978).

However, the prognosis for amblyopic children need not be so poor. Fortunately if detected sufficiently early and before the anatomical/physiological changes have become permanent, amblyopia can be treated to allow substantial or even full recovery of vision loss (Birch, Stager, & Wright, 1986; Cheng, Hiles, Biglan, & Pettapiece, 1991;

Drummond, Scott & Keech, 1989; Lennerstrand, Jakobsson, & Kvarnström, 1995; Lloyd, Dowler, Kriss, Speedwell, Thompson, Russell-Eggett, & Taylor, 1995; Maurer & Lewis, 1993; Maurer et al., 1989; Maurer, Lewis, Brent, & Levin, 1999; Mohindra, 1977; Wali, Leguire, Rogers, & Bremer, 1991). Treatment may involve a number of measures, depending upon the amblyogenic cause. These include removing obstructions, surgery to correct deviations in ocular alignment, refractive correction of affected eye(s), and/or occlusion/penalization of the unaffected eye with an adhesive patch or pharmacologic agents[3] to force use of the amblyopic eye. In all cases, the goal of the treatment is to allow the affected eye to regain or initiate cortical connections (Cashell & Durran, 1980; Cuiffreda et al., 1991; Kushner, 1998, Maurer & Lewis, 1993; Maurer, et al., 1989). More recently, occlusion therapy has been augmented with or replaced by oral doses of levodopa/carbidopa[4]. Studies indicate that this treatment leads to short-term improvement in visual acuity (just hours after the ingestion of a capsule) and some long-term improvement even after cessation of treatment, at least in patients between 4 and 15 years of age (Leguire, Rogers, Bremer, Walson, & McGregor, 1993; Leguire, Walson, Rogers, Bremer, & McGregor, 1993; 1995; Mohan, Dhankar, Sharma, 2001).

Importantly, the success of any treatment depends on three critical factors, notably the *depth of imbalance between the two eyes*, the *age of onset*, and the *duration of deprivation*. The latter two factors highlight the necessity of early detection of amblyopia

---

[3] A recently developed treatment is to "penalize" the unaffected eye with a pharmacologic agent such as atropine. This agent inhibits accommodation and thus, prevents the formation of a sharp image on the retina of the fellow eye (Pediatric Eye Disease Investigator Group, 2002; 2003; Repka & Ray, 1993)

[4] Note, levodopa (also referred to as L-Dopa) is a precursor of the neurotransmitter dopamine, and is commonly used to treat Parkinson's patients (Pinel, 1997). Carbidopa prevents the breakdown of levodopa allowing it to produce a greater therapeutic effect (Gottlob, Weghaupt,Vass, & Auff, 1989; Leguire, Walson, Rogers, Bremer, & McGregor, 1995).

and/or the predisposing amblyogenic factors as these conditions must be treated during the sensitive period of visual development. It is well-established that detection and treatment earlier in the sensitive period leads to better outcome of functional vision (Cashell & Durran, 1980; Ciuffreda, et al., 1991; Kushner, 1998, Mills, 1999; Simon & Kaw, 2001). Thus, the implementation of mass screening programs to effectively detect amblyopia and other vision disorders in young children is clearly critical and is the focus of the present study. In the sections below, I will first review the current vision screening recommendations that have been made by several major vision and pediatric associations, and describe the existing clinical tests of functional vision. Following this, I will provide a critical review of early screening studies conducted over the past four decades, and based on this review, we will implement an experimental, comprehensive vision screening program to optimize the detection of visual dysfunction in toddlers and preschoolers.

Recommendations for Early Vision Screening and Tests of Functional Vision

The importance of early vision screening has been acknowledged by a number of vision and pediatric organizations worldwide, all of whom have made recommendations regarding both the ages at which screening should be conducted and which visual functions should be tested. The recommendations of major North American vision and pediatric organizations are presented below in Table 1. In all, the Table indicates that vision screening should begin by three years of age (The Canadian Pediatric Society advocates that vision screening begin at birth) and that children should have their visual acuity assessed regularly. Note however, there is no clear consensus on what other visual

functions should be tested. Specifically, the organizations in the Table do not agree on whether ocular alignment and stereopsis/stereoacuity should be included in a vision screening program. Furthermore, a growing number of researchers and clinicians point out that a more comprehensive and/or efficient vision assessment may be obtained by including the measurement of other visual functions, namely refractive error and contrast sensitivity (Adams, Hall, Drover, Dalton, Vernescu, & Courage, 2001; Adams, Mercer, Courage, & van Hof-van Duin, 1992; Drover, Earle, Courage, & Adams, 2002; Freedman & Preston, 1992; Kennedy, Sheps, & Bagaric, 1995; Kushner, 1998; Simons, 1996; Tong, Bassin, Enke-Miyazaki, Macke, Tielsch, Stager, Beauchamp, Parks, & the National Children's Eye Care Foundation Vision Study Group, 2000). In light of these collective suggestions, the five critical components of visual evaluation mentioned above (visual acuity, ocular alignment, stereoacuity, refractive error, and contrast sensitivity) are discussed in the following paragraphs, with particular focus on the methods for assessing these functions in a young pediatric patient.

Table 1. Current screening recommendations by North American vision and pediatric associations

Organization	Age	Aspect of Vision	Test*
Canadian Pediatric Society	Birth - 12 months	External	External Exam
		Clarity	Red Reflex
		Alignment	Corneal Reflex
	3 - 5 years	External	External Exam
		Visual Acuity	Recognition Acuity
	6 - 18 years	Visual Acuity	Not Specified
Canadian Task Force on Periodic Health Examination	3 - 5 years	Visual Acuity	Not Specified
American Academy of Pediatrics American Academy of Ophthalmology American Association of Pediatric Ophthalmology and Strabismus Maternal and Child Health Bureau	3 - 5 years	Visual Acuity	Snellen Letters, Snellen Numbers, Tumbling E, HOTV, Pictures Tests, Allen Figures, or Lea Test
		Alignment	Cover Test
		Stereopsis/ Stereoacuity	Randot E Stereotest
American Optometric Association	2 - 6 years	Visual Acuity	Not Specified,
		Alignment	Corneal Reflex or Cover Test
		Color Vision	Ishihara Plates
Head Start Program	3 years	Visual Acuity	Tumbling E
		Alignment	Cover Test or Corneal Reflex
U.S. Public Health Service	3 - 4 years	Visual Acuity	Snellen Letters, Snellen Numbers, Tumbling E, HOTV, Allen Figures, or Lea Symbols
		Stereopsis/ Stereacuity	Random Dot E Stereotest

* Note, the majority of these tests are described below.

Visual Acuity

There are two broad categories of visual acuity tests. The first, *recognition acuity*, is the classic and most widespread measure of visual acuity. It refers to the smallest, easily recognized visual target (optotype) that can be correctly identified. Recognition acuity is estimated with either single or linear optotype tests. In single optotype testing, designed mainly for young children (2 – 3 years), the visual target is presented in isolation. In linear optotype testing, the subject is presented with multiple visual targets of the same size, usually in the form of a horizontal line or row. In adults and literate, verbal children, recognition acuity is most often estimated with linear optotype charts composed of letters such as the standard Snellen ("Big E") chart shown in Figure 1 below. The subject stands a fixed distance (most often 20 feet/6 m) from a chart that contains rows of letters which become progressively smaller as one reads from top to bottom. Beginning at the top, the subject reports the letters one row at a time until he/she confronts a row with targets that are too small to recognize. The last row of letters correctly identified provides an estimate of visual acuity. In a normal adult, or an older child (above 6 years of age) Snellen visual acuity is 20/20.[5]

Although the standard Snellen test is often attempted, other methods must be used to test recognition acuity in preschool children who are not literate and/or not sufficiently

[5] For traditional or historical purposes, Snellen visual acuity is expressed in relation to the test distance (normally 20 feet, or 6 meters) and in comparison to a person with normal vision (Sekuler & Blake, 1994). If one is able to identify at a distance of 20 feet, the letters that a person with normal sight can identify at the same distance, he/she possesses a visual acuity of 20/20. However, visual acuity can be better or worse than 20/20. For example, a visual acuity of 20/60 implies that one can identify at a distance of 20 feet, the letters that a person with normal sight can identify at 60 feet. Conversely, a visual acuity of 20/15 means that one can identify at a 20 feet, the letters that a person with normal sight can identify at 15 feet. It is important to note that test distance is not varied during the modern Snellen test and testing is usually conducted from a distance of 20 ft (6 m). Instead, the denominator in the Snellen visual acuity fraction represents letter size which is correlated with test distance.

verbal to complete the Snellen test. Fortunately, a number of simpler alternatives exist. For instance, recognition acuity can be measured with a limited set of letters such as the HOTV, STYCAR, or Sheridan-Gardiner tests (Jarvis, Tamhe, Thompson, Francis, Anderson, & Colver, 1991; Kushner, 1998; Newman & East, 1999; Wagner, 1998; Wormald, 1991). These letter optotype tests consist for four to seven letters arranged in linear or single optotype format, and may be used to estimate visual acuity with very young subjects (e.g. 3 to 5 years old). The child is presented with one letter at a time (e.g. H) and given a card to hold which contains all the relatively easily identifiable letters included in the test. Instead of naming the target, the child can simply match it to his/her card, a task that can be completed by a child who has even yet to learn letter names. Second, preschoolers may be assessed with tests that contain a single optotype arranged in various orientations. For instance, visual acuity may be estimated with the Landolt C test in which the subject must locate the position of the C's opening or gap located at one of four clock positions, either 12:00, 3:00, 6:00, or 9:00. When the eye cannot resolve the gap in the C, it appears as an "O" shape, and the subject performs erroneously. The Landolt C test is available in both single and linear optotype formats. Similarly, preschoolers can be assessed with the Illiterate E or Tumbling E test. This test is essentially the same as the Landolt C test, except that the optotypes are formed by arranging the letter E in one of four clock orientations. The preschooler must identify the direction in which the E is facing. This test is also available in single or linear optotype forms. A third alternative to estimate visual acuity in younger, illiterate patients is to use "picture" optotypes such as in the Lea test (available in linear or single optotype forms)

or Kay Picture test (available in linear optotype forms). These tests consist of symbols (i.e., houses, squares, boots, fish, hearts, etc.) of various sizes instead of letters (Kushner, 1998; Wagner, 1998; Wormald, 1991).

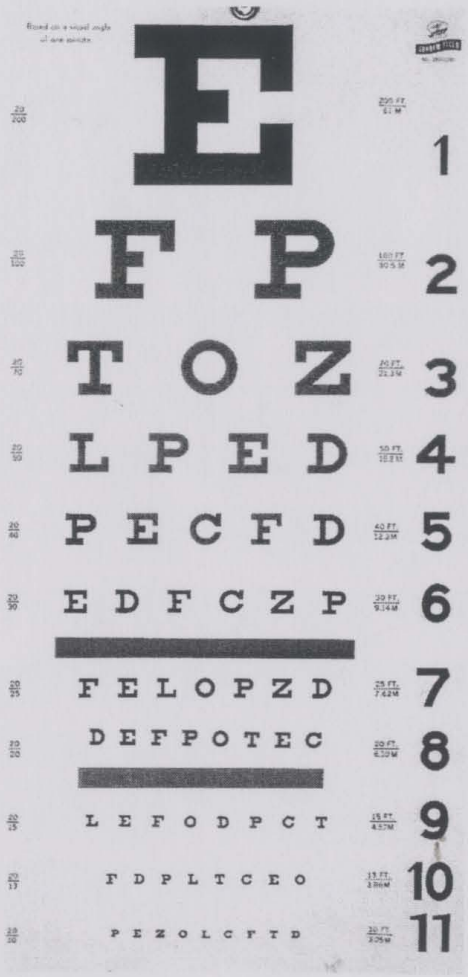

Figure 1. Photograph of a standard Snellen chart.

Note however, that although young, normal, preliterate subjects may be assessed with the simpler alternatives described above, these tests can not be used to estimate

recognition acuity in nonverbal, or "non-instructable" subjects such as infants, young toddlers, or the multiply handicapped. These subjects can be assessed with the second category of visual acuity termed *resolution acuity*. Resolution acuity refers to the subject's ability to distinguish a pattern from a uniform field of equal average luminance, or "blank" (Schwartz, 1999). In contrast to recognition acuity, the subject need not recognize the target, but merely detect its presence. Formal tests to measure resolution acuity include the Teller Acuity Cards or the Wright Cards (McDonald, Dobson, Sebris, Baitch, Verner, & Teller, 1985; Raina, 1998). Both of these tests consist of a series of rectangular cards that contain square wave gratings (Teller Acuity Cards) or checkerboard patterns (Wright Cards) at one end of the card and an unpatterned stimulus of equal average luminance (i.e., a "blank") at the other end. Gratings consist of repeating black and white stripes of a specific thickness or *spatial frequency,* whereas checkerboard patterns consist of alternating black and white checks of a specific spatial frequency. Spatial frequency (SF) is a measure of the size of the elements (in this case, the size of the stripes in a grating, or checks in a checkerboard), and is defined as the number of cycles of the elements (i.e., one black stripe and one white stripe, or one white check and one black check) that repeat within 1 degree of visual space (c/deg). Thus, gratings/checkerboards of low SF (e.g., 1 c/deg) consist of relatively thick stripes/checks, and gratings/checkerboards of high SF (e.g., 10 c/deg) consist of much thinner stripes/checks. To perform these tests, the experimenter follows the forced-choice preferential looking method (FPL) which is based upon the pioneering work of Fantz (1965), who found that infants and toddlers prefer to fixate a patterned stimulus over an

unpatterned stimulus. Testing usually begins with cards containing gratings/checkerboards of low SFs and proceeds with cards containing gratings/checkerboards of progressively higher SFs until it is judged by the observer that the subject can not detect the grating/checkerboard. The grating/checkerboard with the highest SF detected by the subject is taken as an estimate of his/her visual acuity. Another option to estimate resolution acuity is the Cardiff cards, a relatively new picture optotype test (Adoh, Woodhouse, & Oduwaiye, 1992). This test consists of a series of 21 X 28 cm cards, each of which contains a picture optotype of a familiar object (eg., dog, fish, car, duck, etc.) on the top or bottom of the card whereas the opposite portion of the card is blank. The size of the optotypes ranges from 10 to 30 c/deg. Similar to the Teller Acuity Cards and the Wright cards, the Cardiff cards are presented following the FPL procedure. Note that no screening study to date has used the Cardiff cards.

As outlined above, there are a number of options to measure visual acuity in preschool children, however it is important that the test chosen for any screening program complies with the six generally agreed upon recommendations for screening for childhood vision disorders. First, it is recommended that visual acuity be assessed with a test of recognition acuity. Although gratings are useful for assessment of nonverbal or "non-instructable" subjects, they tend to overestimate visual acuity, and often fail to identify amblyopic children, the key target disease of any screening program (Fern & Manny, 1986; Friendly, Jaafar, & Morillo, 1990; Kushner, Lucchese & Morton, 1995; Mayer, 1986; Simons, 1983; Simpson, 1991). Although much less researched, checkerboard patterns, however, may hold more promise as they may be more sensitive

to visual deficits (Raina, 1998). Second, although single optotype tests are completed more easily by younger children (2 – 3 year-olds), linear optotype testing is recommended as the multiple targets tend to blend into each other to make the test more difficult (Simmers, Gray, & Spowart, 1997; Simons, 1983). This *crowding effect* is more apparent in amblyopic subjects than in normal subjects, thus linear optotype tests tend to be more sensitive for detecting amblyopia (Kushner, 1998; Simons, 1996). However, to counteract the lack of crowding effect, single optotype tests have recently been developed that possess crowding bars in which each optotype is surrounded completely by vertical and horizontal lines as shown in Figure 2. The use of these crowding bars ensures that all optotypes are subject to equal, sufficient crowding. Third, all optotypes within the test should be equally legible. This is important as it ensures that at least theoretically, size alone determines whether the optotype is recognized (Bailey & Lovie, 1976; Simons, 1983; Rosser, Laidlaw, & Murdoch, 2001). Of the tests described above, only the Landolt C and the Illiterate E tests satisfy this recommendation. Fourth, each line on a linear optotype test should contain the same number of optotypes. Fifth, the inter-optotype spacing should be proportionate to optotype size. These latter two recommendations ensure that each line is equal in visual demand when size is not taken into account and that the crowding effect is consistent as one progresses from line to line (Bailey & Lovie, 1976, Simmers, Gray, & Spowart, 1997). Finally, there should be a systematic progression of optotype size from line to line. If optotype size progression is not systematic, the scale of measure is not equal over the entire chart. Specifically, a difference of one or two lines near the top of the chart is not equal to a difference of one

or two lines near the bottom of the chart (Ricci, Cedrone, & Cerulli, 1998). Furthermore, systematic progression of optotype size guarantees that the scale measure is the same over the entire chart even if the testing distance is changed, a situation that is often necessary when testing subjects with low vision.

Figure 2. An example of a letter optotype surrounded by crowding bars.

Ocular Alignment/Motility

There are a variety of techniques to assess ocular alignment/motility. The simplest exam is to observe the Hirschberg corneal reflex in which the patient fixates a small target 40-50 cm away while a penlight is placed coincident with the target and shone into his/her eyes. The positions of the corneal reflections from the light are inspected carefully and any asymmetry within these reflections suggests misalignment, i.e., strabismus. Another more precise option is the cover-uncover test which allows one to detect slight strabismus that may not be revealed upon a simple examination of the

eyes or by the Hirschberg corneal reflex (Hall & Elliman, 2002). During this test the subject fixates, binocularly, a target from 3 m (distant cover-uncover test) or 40 cm (near cover-uncover test). One eye is repeatedly covered and uncovered with a plastic occluder while the unoccluded eye closely observed. If the unoccluded eye then shifts in order to find the target, it implies that it is not fixating the image properly and is therefore, out of alignment. The process is then repeated with the other eye. Another option is the alternate cover test, in which the occluder is moved quickly from one eye to the other without a period of binocular viewing. If an eye shifts while not occluded, this again is evidence that the eye is out of alignment.

There are also tests of ocular alignment/motility that are designed to detect even more subtle forms of strabismus, such as that due to nerve palsy, in which one of the extraocular muscles controlling alignment and movement of the eyes is underactive. This form of misalignment is manifested only when the eyes are fixated in a particular direction, and one can often compensate by tilting the head slightly (Olitsky & Nelson, 1998). Thus, this type of strabismus may be detected by inspecting either head posture, and/or by performing tests of ocular motility in which eye movements are closely examined in the nine cardinal directions (i.e., straight ahead, upward, downward, leftward, rightward, diagonally upward and rightward, diagonally downward and rightward, diagonally upward and leftward, and diagonally downward and leftward). A final test of ocular alignment/motility is the inspection of opposing eye movements of convergence/divergence as an object is moved toward/away from the subject's eyes.

Stereopsis/Stereoacuity

Tests of stereopsis/stereoacuity have been recommended for screening for visual deficits, most notably, poor binocular functioning due to persisting monocular deprivation (Kushner, 1998; Simons, 1996). Stereopsis refers to the perception of true depth without reliance on monocular or kinetic cues. Monocular cues (i.e., those that require only one eye) include the relative size of the object, texture, interposition, etc. Kinetic cues (i.e., motion cues) consist of movement of either the observer or the visual target. Instead, fine or "true" stereoscopic depth perception requires retinal disparity, i.e., the lateral difference seen between objects due to the slightly different positions of the left and right eye (Sekuler & Blake, 1994). These disparate images are then sent to binocular cells in the visual cortex where they are fused together to provide the sensation of depth (Hubel, 1988). Stereoacuity is an index of one's stereopsis and is defined as the minimum amount of disparity that one can use to detect depth. It is measured in seconds of arc (arc sec). The finer the disparity one can detect, the finer one's stereoacuity. Although a great deal of variability exists among studies, normal adult stereoacuity is generally less than 40 arc sec, whereas the stereoacuity of a normal 3-year-old is about 65 arc sec (Birch & Salamão, 1998; Birch, Williams, Hunter, Lapa, & the ALSPAC Study Team, 1997; Simons, 1981). In order to possess good stereoacuity, one must possess accurate, clear, balanced, and fully developed binocular vision, which is often impaired in patients who have suffered from visual deprivation due to strabismus, anisometropia, cataract, or other forms of monocular suppression (Cashell & Durran, 1980; Hall & Elliman, 2002).

Most often, stereopsis/stereoacuity is measured with a random dot stereogram, an array of dots that when viewed monocularly (or by a person with no stereopsis) appears to possess a uniform, patternless texture. However, the stereogram contains a portion that is slightly displaced laterally. Therefore, when viewed with polarized glasses and normal stereoscopic vision, the lateral displacement within this portion of the stereogram creates "artificial" retinal disparity as a different image is seen by each eye. As a result, the normal subject experiences the sensation of depth as the displaced portion of the target appears to either "float above" the rest of the stereogram (i.e., crossed disparity) or "lie below" it (i.e., uncrossed disparity; Birch, 1993; Millodot, 1986; Sekuler & Blake, 1994). By gradually reducing the amount of lateral displacement, and thus, the amount of retinal disparity within the stereogram, one can obtain an estimate of the subject's stereoacuity. There are a number of commercial versions of random dot stereograms available, many of which include pediatric targets suitable for young children (The Random Dot E Stereo Test, The Randot Preschool Stereoacuity Test, The Frisby Stereo Test, The Wirt Fly Test, The TNO, and The Randot Stereosmile Test), some even designed for children as young as one year.

Refractive Error

The measurement of refractive error is often advocated as a critical component of a screening program. The emphasis placed on this visual function appears to be validated by the finding that at least historically, it has been the greatest predictor of amblyopia (Taylor, 1987). There are three classes of techniques to measure refractive error in infants and very young children. The traditional technique is retinoscopy (also termed

skiascopy) in which a retinoscope (or skiascope) is used to shine a beam of light through the subject's optical system. By looking through a site-hole in the retinoscope, one can see the light reflected from the patient's pupil as well as a shadow at the edge of the pupil. This shadow is observed as a mirror inside the retinoscope is moved in various directions or, meridians. If the shadow moves in the direction opposite to the movement of the mirror, the subject is myopic. If the shadow moves in the same direction as the movement of the mirror, the subject is hyperopic. If the direction of movement, or thickness of the shadow, is different in two different meridians 90° apart, the subject has astigmatism. If the shadow does not move and is the same thickness in all meridians, the subject has perfect optics (i.e., is emmetropic). Retinoscopy is often conducted with the use of cycloplegic drops, pharmacological agents placed in the eye to prevent accommodation, the occurrence of which decreases the accuracy of "pure" refractive estimates (Repka, 1998). Although cycloplegic retinoscopy is considered the "gold standard", it is seldom used in screening procedures with young pediatric patients for a number of reasons. First, it requires a great deal of clinical expertise and time (Köhler & Stigmar, 1973; Kvarnström, Jacobsson, & Lennerstrand, 1998; Nordlöw & Joachimson, 1962). Second, it may require at least 40 minutes for cycloplegic agents to take effect and in many cases cycloplegic drops must be administered by the parent(s) at home prior to the examination (Repka, 1998). Third, the administration of cycloplegia may be distressing to the child and the parent(s), and may cause side effects such as an allergic reaction and vomiting (Barry & Loewen, 2001; Repka, 1998).

In response to these limitations, researchers developed a more child-friendly technique termed photoscreening (traditionally referred to as photorefraction[6]; see Atkinson & Braddick, 1983a; 1983b; Howland, Atkinson, Braddick, & French, 1978; Howland, Braddick, Atkinson, & Howland, 1983). With this method, a device called a photoscreener, which consists of a camera and a flash source fixation target, is used to take a photograph of the subject's eyes or more specifically, an image of the flashed light as it returns from its passage through the optical system. This image is analyzed, and based on the amount and position of the crescent-shaped light reflected from the subject's pupil as he/she fixated the target, refractive error can be determined. Although photoscreening may be performed with cycloplegic drops, it is generally performed without cycloplegia (see Freedman & Preston, 1992; Kennedy & Thomas, 2000; Kennedy et al., 1995; Morgan & Johnson, 1987; Tong et al., 2000). This technique of measuring refractive error holds great promise for the future of vision screening due to its relative objectivity. Also, because photoscreening requires minimal cooperation from the subject, it may be used to screen infants and other nonverbal subjects. Moreover, photoscreeners can also detect the presence of strabismus and media opacities (e.g., cataracts), and as it does not require the expertise of an optometrist, ophthalmologist, or orthoptist, it can be carried out by a trained technician (Freedman & Preston, 1992; Kennedy & Thomas, 2000; Kennedy et al., 1995; Simons, 1996). There are two types of photoscreeners and they are categorized based on the position of the flash source in relation to the optical axis of the camera. An on-axis photoscreener has a flash source

[6] This technology was initially termed photorefraction, but has become more frequently referred to as photoscreening as its use in vision screening studies has become more widespread.

located on the same axis as the camera lens source, whereas an off-axis photoscreener has a flash source on a slightly different axis in relation to the camera lens. Each type of photoscreener possesses advantages and disadvantages. For instance, on-axis photoscreeners are sensitive to astigmatism, but not to strabismus (Hamer, Norcia, Day, Haegerstrom-Portnoy, Lewis, & Hsu-Winges, 1992; Simons, 1996). Off-axis photoscreeners, on the other hand, are sensitive to strabismus, but have difficulty detecting *both* small and large refractive errors (Simons, 1996).

A third option for the measurement of refractive error in infants and young children is automatic refraction, or autorefraction. This technique was first established in the 1970s with the development of the automatic refractor, or autorefractor (Cornsweet, 1974; Cornsweet & Crane, 1970; McDevitt, 1977). Although these devices have changed significantly over the past 30 years, most current autorefractors measure refractive error using the same basic technology. A target is displayed to the patient that contains infrared ray beams that project to the back of the eye. The beams in turn, are reflected back from the subject's retina to the autorefractor which then determines the extent to which the image is out of focus, thus providing an estimate of refractive error (Hazel, Cox, & Strang, 2003; Wesemann & Rassow, 1987). This method can be performed with or without cycloplegia. Although accurate, traditional table-top autorefractors are quite large, and therefore cumbersome and very expensive. Furthermore, autorefraction requires that the subject view the pattern for up to a minute in order to obtain a single measurement. Both these factors make autorefraction unsuitable for screening young children in a preschool or elementary school setting. Recently however, portable,

handheld versions of the autorefractor have been developed which require much less time to obtain a measurement. These new autorefractors shine a small beam of infrared light or a low powered laser into the eye allowing rapid measurements of optical power and thus, the estimation of any refractive error. Importantly, this technique represents a substantial innovation as it can be completed in a matter of seconds and requires little expertise on the part of the examiner (Adams, Dalton, Murphy, Hall, & Courage, 2002; Adams et al., 2001; Barry & König, 2001, Cordonnier & Dramaix, 1998;).

Contrast Sensitivity

Contrast sensitivity (CS) measurement may also be beneficial as part of a screening program. CS estimates the minimum amount of contrast (i.e., the contrast threshold) required to detect sine wave gratings[7] at different SFs. Contrast refers to the difference in light intensity between an object and its surroundings (contrast is typically defined as $C = [I_{max} - I_{min}]/[I_{max} + I_{min}]$ where I_{max} and I_{min} refer to the brightest and darkest portions of the target, respectively). CS is measured in CS units which are simply the reciprocal of contrast. In infants, both psychophysical (e.g., the FPL procedure) and electrophysiological methods similar to those used to assess grating visual acuity, are used to measure CS (see Adams et al., 1992; Banks & Salapatek, 1981; Drover et. al., 2002; Norcia, Tyler, & Hamer, 1990; Pirchio, Spinelli, Fiorentini, & Maffei, 1977).

Although CS is similar in some ways to visual acuity, it provides a more complete description of one's visual environment by assessing the detection of objects of different

---

[7] Sine wave gratings refer to a series of black and white stripes in which the transition from black to white is gradual, i.e., it follows a sinusoidal pattern.

size *and* contrast simultaneously, arguably the two most important features of a visual stimulus (Banks & Dannemiller, 1987; Banks & Salapatek, 1981; Lennie & van Hemel, 2002; Mohn & van Hof-van Duin, 1991; Sekuler & Blake, 1994). On the other hand, tests of visual acuity estimate the limits of functional vision at a single level of high contrast (~ 95%). Although the measurement of CS has not yet been used as part of a screening program, it does indeed hold promise. For instance, it has been demonstrated that CS provides a more complete description of the visual losses suffered by subjects with amblyopia, as different types of amblyopia have specific effects on the different segments of the contrast sensitivity function (i.e., a graphical representation of one's CS at several SFs; Hess & Holliday, 1992; Kushner, 1998; Lennie & van Hemel, 2002).

Current State of Vision Screening

Despite the above recommendations which stress the importance of early, regular vision screening, and the wide variety of available tests, the implementation of comprehensive preschool vision screening has been limited mainly to certain Scandinavian countries (Köhler & Stigmar, 1973; 1978; Kvarnström et al., 1998; Nordlöw & Joachimson, 1962). Other industrialized nations such as Britain, Canada, and the United States lag behind in the development of screening programs (Ciner, Dobson, Schmidt, Allen, Cyert, Maguire, Moore, Orel-Bixler, & Schultz, 1999, Simons, 1996). For instance, it is estimated that in the United States, only 5 – 14% of all preschoolers undergo an eye exam before beginning kindergarten (Ciner, Schmidt, Orel-Bixler, Dobson, Maguire, Cyert, Moore, Schultz, 1998; Erlich, Reinecke, & Simons, 1983). Similarly, only 5/10 Canadian provinces (Prince Edward Island, New Brunswick, British

Columbia, Nova Scotia, and Manitoba) and two territories (Yukon and North West Territories) currently possess mass preschool screening programs. The failure to implement vision screening programs may be due partly to the position of some that amblyopia can not be adequately treated and therefore, screening is of little use (see Bray, Clarke, Jarvis, Francis, & Colver, 1996; Ingram, 1989; Stewart-Brown & Haslum, 1988; Stewart-Brown, Haslum, & Howlett, 1988). Second, the detection of the less obvious amblyogenic disorders (such as anisometropia, slight strabismus, monocular suppression, or a small cataract) is often difficult. Moreover, this is compounded by the problems inherent in testing infants and young children who often lack cooperation and attention. On a related note, many subtle amblyogenic factors can be detected only by "experimental" tests of functional vision which have been developed only recently (e.g., CS) and thus, are currently not included in standard pediatric eye exams. Finally, as shown in Table 1 (see Recommendations for Early Vision Screening and Tests of Functional Vision subsection above), there remains a general lack of consensus as to what constitutes an optimal vision screening program (Ciner et al., 1999; Erlich, et al., 1983; Simons, 1996). In order to provide the impetus for the establishment of clear screening guidelines, it is necessary to conduct a detailed, critical review of the existing vision screening studies. In the subsections below, the evaluation of screening studies is discussed, and those conducted over the last four decades are reviewed with particular attention focused on the validity or effectiveness of the measurement tools within these studies. It is hoped that this review will reveal which vision screening tools should be considered as part of a comprehensive vision screening program.

Evaluation of Vision Screening

To understand the evaluation of vision screening, the primary purpose of screening must be discussed. In general, screening employs a rapidly applied test that permits the presumptive identification of disease or defect (Last, 1988; Lennerstrand et al., 1995; MacPherson, Braunstein, La Roche, 1991). In other words, the purpose of screening is not diagnostic but to merely identify those who *might* have a disease or defect. Those who screen positive are referred to an eyecare specialist who then administers the "gold standard" test(s) in order to make a final diagnosis.

It is important to note therefore, that no single screening test is completely valid. Instead, different screening tests vary greatly in terms of their effectiveness. In light of this, before conducting a screening program, one should evaluate each test in the battery to ensure that it provides an acceptable level of effectiveness. To evaluate any clinical test used in a screening study, the four possible outcomes of a single test must be considered. These outcomes, presented below in Table 2, can be divided into correct and incorrect decisions. Correct diagnoses include those in which (1) the screener has concluded *correctly* that a patient has a disease (true positive); (2) or the screener concludes that a patient does not have a disease and he/she in fact, does not (true negative). The incorrect diagnoses includes those decisions in which the screener has either *incorrectly* concluded that a patient has a disease when he/she does not (false positive), or that a patient does not have a disease when he/she truly does (false negative). The frequency of each of these four decisions is used to calculate the validity of a screening test. In all, there are four measures of validity, the first two of which relate to

the effectiveness of the test in identifying those patients with the disease. *Sensitivity*, refers to the ratio of true positives based on the screening test to all patients who truly have the disease [i.e., A/(A + C), see Table 2]. Similarly, *positive predictive value* (PPV) refers to the ratio of true positives to those who have tested positive according to the screening test [i.e., A/(A + B)]. The final two measures of validity reflect the ability of a screening test to identify those who do not have a disease. The first, *specificity*, is the ratio of true negatives based on the screening test to all patients who do not truly possess the disease [i.e., D/(B + D)]. Second, *negative predictive value* (NPV) refers to the ratio of true negatives to those who have tested negative with the screening test (i.e., D/(C +D)].

Table 2. The four possible outcomes of a single screening test.

		Disease Status	
		Positive	Negative
Screening Results	Positive	A True Positive	B False Positive
	Negative	C False Negative	D True Negative

It should also be noted that the validity of any screening test, particularly sensitivity and specificity, are closely related to the referral criteria implemented by researchers/clinicians. Referral criteria refer to the range of scores that indicate abnormally low levels of visual functioning. If a subject scores within this range (i.e., fails the test), he/she is considered positive for a disorder and is referred for the gold standard exam. Conversely, if a subject scores outside this range (i.e., passes the test),

he/she is considered negative for a disorder and no exam is necessary. If the referral criteria are too strict (i.e., it is difficult to pass the test) sensitivity will be high, but specificity will be low, and thus, there will be an unacceptable number of false positives. As a result, there will be a large number of subjects unnecessarily referred for a complete gold standard examination (i.e., overreferrals). On the other hand, if the criteria are too lenient (i.e., it is relatively easy to pass the test) specificity will be high, but sensitivity will be low, and thus, there will be an unacceptable number of false negatives. In other words, a large number of subjects with vision disorders will be misdiagnosed as normal.

Validity of Vision Screening Programs

Table 3 summarizes the results of 70 vision screening studies conducted over the previous four decades[8]. The majority of these studies screened children for amblyopia or any amblyogenic factor (e.g., large refractive error, strabismus, anisometropia, poor ocular motility) that could potentially lead to poor performance on the particular visual function assessed[9]. The presence or absence of a vision disorder was then confirmed by a complete ophthalmological/optometric examination (the gold standard exam), which in most cases included cycloplegic refraction[10]. Note that each study was required to meet two important criteria to be included in the Table. First, the study had to include toddlers and/or preschool children as part of the patient/subject population. This criterion was

---

[8] Note, only 45 screening studies are reviewed here. However, several studies assessed the same subjects with different screening tools, referral criteria, or assessed different age groups and provide complete data for each tool, set of referral criteria, or age group. In each case, these data are treated as separate studies. Thus in all, there are 70 "studies" in the Table.

[9] A few exceptions apply and are noted on the Table.

[10] Some studies do not include cycloplegic refraction as part of the gold standard exam (Allen & Bose, 1992; Kennedy, Sheps, & Bagaric, 1995). Others studies do not specify whether cycloplegic refraction was part of the gold standard exam (Barry & König, 2001; Newman & East, 1999; Robinson, Bobier, Martin, & Bryant, 1999; Ruttum & Nelson, 1991; Spowart, Simmers, & Tappin, 1998).

considered important because screening of other age groups, particularly older, school-aged children, may provide high measures of validity that reflect the relatively advanced cognitive abilities of these children, *not* the superiority of the screening test/program. Second, each study was required to provide results on *all four* measures of validity or adequate data that allowed us to calculate these four measures. Quite often screening studies include only two measures of validity, namely sensitivity and specificity. Though informative, these measures do not provide direct information on important aspects of a screening program such as the percentage of children who are overreferrals that unnecessarily receive the gold standard exam and the percentage of children who are false negatives that do not receive necessary treatment. Both of these aspects are critical to determine whether a screening program is both effective and feasible.

The studies in the Table are divided into those deemed effective (i.e., those above the double dashed line) and those deemed ineffective based on whether they reach critical scores that have been set as *cut-off criteria* for each measure of validity (see below). The cut-off criteria were chosen by us and some of these criteria have since been validated by the recommendations of other researchers (see Büchner, Schnorbus, Grenzebach, & Busse, 2005). Importantly, these criteria reflect the goal of any screening study, that is, to detect most cases of a deficit while at the same time, limiting the number of overreferrals (i.e., false positives) and missed cases (i.e., false negatives). With this goal in mind, the cut-criterion for sensitivity is set high at $\geq$ 90% indicating that at least 9/10 subjects who truly possess an amblyogenic factor, including strabismus, hyperopia, myopia, astigmatism, opacities, poor stereoacuity, and anisometropia, were correctly

identified by a poor result on the screening test(s). The cut-off criterion for specificity has also been set at $\geq$ 90% indicating that at least 9/10 truly healthy subjects were correctly identified by the screening test(s). The cut-off criterion for PPV must be set at a relatively low critical value as it is often the lowest of all validity measures (as it was in 39/70 studies reviewed here). For this review, the cut-off criterion for PPV is set at $\geq$ 65%. In other words, at least 2/3 of those who screened positive actually possessed an amblyogenic factor (i.e., they are true positives), whereas 1/3 or less were overreferrals. Finally, NPV is often the highest of all four measures (as it was in 42/70 studies reviewed here). It is important to set this cut-off criterion high as it determines the acceptable proportion of false negatives. As is common in medical/clinical practice, it is important that the proportion of false negatives is kept very low as patients who have, or eventually develop permanent visual problems are less likely to seek further medical attention if they have received a negative screening result. In this case, the cut-off criterion for NPV is set at $\geq$ 95%, i.e., only a maximum of 5% of subjects were permitted to show a false negative result. Note that the studies in the Table are ranked out of the total of 70 studies on each of the four measures of validity. Also, average rank is determined for each study and studies are listed in ascending order of these mean ranks (i.e., the first study listed in the Table has the lowest average rank on all four validity measures and thus, is the highest ranked study overall).

Inspection of Table 3 leads to a number of interesting observations. First, only seven studies shown in the Table, i.e., those above the double horizontal line, were judged effective as they passed all four cut-off criteria for an effective program. Second,

visual acuity appears to be an essential vision screening tool as five of the seven (71%) studies[11] considered effective incorporated visual acuity as either the sole test within the study (three studies; Kvarnström et al., 1998, ranked 7th; [12] Nordlöw & Joachimson, 1962, ranked 1st; Raina, 1988, ranked 4th), or as part of a larger screening battery (two studies; Köhler & Stigmar, 1973, ranked 3rd; Wormald, 1991, ranked 2nd). Both of the latter studies also assessed ocular alignment/motility and stereoacuity. Conversely, the vast majority (76%) of those studies deemed ineffective, did not include tests of visual acuity. A third conclusion that can be reached from Table 3 is that among recognition acuity tests, the Boström Hooks test (similar to the Landolt C test) appears to be particularly effective for detecting amblyogenic factors, whereas the Stycar graded balls test appears to be ineffective (Kennedy et al., 1995, ranked 45th; Köhler & Stigmar, 1973, ranked 3rd; Nordlöw & Joachimson, 1962, ranked 1st). The acuity tests that use letter optotypes (i.e., Stycar, Sheridan-Gardiner, and HOTV) on the other hand, provide mixed results (DeBecker, MacPherson, LaRoche, Braunstein, Cottle, McIntyre, & Kozousek, 1992, ranked 56th; Kvarnström et al, 1998, ranked 7th; Wormald, 1991, ranked 2nd). The results of the relatively new Lea Symbols are also mixed. Programs that include this test (Barry & König, 2003, ranked 18th; Chui et al., 2004, ranked 50th and 70th; Shallo-Hoffman et al., 2004, ranked 28th) range greatly in terms of sensitivity (50-100%) and specificity (68-94%), while yielding low PPV (24-43%) and high NPV (90-100%). However, it is difficult to evaluate this test as in each program, tests of other aspects of

---

[11] Note, the other effective studies used photoscreening to assess children (Arnold, Arnold, Stark, Arnold, Leman, & Armitage, 2004; Kennedy & Sheps, 1998)

[12] In order to locate the studies in the Table, each study's overall rank on all four measures of validity is provided.

Table 3. Validity of screening studies for preschoolers and school-aged children. Numbers in bold represent those passing the cut-off criteria for an effective vision screening study (i.e., sensitivity ≥ 90, specificity ≥ 90, PPV ≥ 65, NPV ≥ 95). Numbers in parentheses indicate that study's rank (out of 70) on that particular measure of validity. Note that each study is given an overall rank in the leftmost column based on its mean rank on all validity measures. Studies are listed in ascending order of overall rank.

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(1) Nordlöw & Joachimson (1962)	4 years	1166[3,4]	Visual Acuity	Bostrom Hooks[5] (5m; single)	Nurses	Acuity ≤ 20/24	**95** (8)	**99** (2)	**86** (15)	**100** (1)	6.5
(2) Wormald (1991)[6]	4 years	345[3,4]	Visual Acuity	Snellen (6m; linear), Sheridan Gardiner (6m; single), or Kay Picture (distance and optotype format not reported)	Orthoptist	Acuity < 20/30	**90** (23)	**99** (2)	**94** (6)	**99** (8)	9.75
			Alignment/ Motility	Cover Test		Movement					
				Prism Cover Test		Deviation					
				Convergence		Abnormal					
				Head Posture		Abnormal					
				Nine Positions of Gaze		Abnormal					
			Stereoacuity	Wirt Fly Stereo Test		Not reported					
(3) Köhler & Stigmar (1973)[7]	4 years	910[3,4]	Visual Acuity	Bostrom Hooks (5m; single)[3]	Nurse	Acuity ≤ 20/24	**96** (7)	**97** (10)	**84** (20)	**99** (8)	11.25
			Alignment/ Motility	Cover Test	Pediatrician	Movement					
				Motility		Abnormal					
			Stereoacuity	Wirt Fly Stereotest	Nurse	Not Reported					
			Ocular Health	External Eye Exam	Not Reported	Abnormal					
				Pupillary Light Reflex		Abnormal					

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(4) Raina (1998)[8]	4 to 12 years (mean = 7.3 years)	44[3]	Visual Acuity	Wright Cards	"Masked Examiner"	Acuity ≤ 20/40	**100** (1)	**95** (16)	**77** (33)	**100** (1)	12.75
(5) Arnold, Arnold, Stark, Arnold, Leman, & Armitage (2004)[8]	4 to 6 years	93[3]	Refractive Error Alignment Clarity	Gateway DV-S20 Photoscreener	School Nurses	Not Reported	**91** (18)	**98** (5)	**97** (2)	**95** (27)	13
(6) Kennedy & Sheps (1989)[8]	17 days to 6 years (mean = 1 year)	236[3]	Refractive Error Alignment Clarity	Otago Photoscreener (off-axis)	Technician	Not Reported Asymmetry Opacity	**94** (11)	**94** (21)	**92** (10)	**96** (24)	16.5
(7) Kvarnström, Jakobsson, & Lennerstrand (1998)[6,7,9]	4 to 5.5 years	3126[4]	Visual Acuity	HOTV (3m; linear)	Nurses	Acuity 4 yr < 20/25; 5.5 yr < 20/20; Anisometopia ≥ 2 line difference	**92** (15)	**97** (10)	**69** (40)	**99** (8)	18.25

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(8) Ladenvall (1988)	4 years	1459[4]	Visual Acuity	HOTV (3m; linear)	Nurses	Acuity < 20/25; Anisometropia ≥ 2 line difference	87 (29)	**97** (10)	**86** (15)	**98** (18)	18
(9) Eibschitz-Tsimhonie, Friedman, Naor, Eibschitz, & Friedman (2000)	1 to 2.5 years	808[4]	Refractive Error	Rapid Retinoscopy (without cycloplegia)	Ophthalmologist or Orthoptist	Not Reported	86 (31)	**99** (2)	62 (45)	**100** (1)	19.75
			Alignment/ Motility	Cover-Uncover Test		Movement					
				Alternate Cover Test		Movement					
				Hirschberg Test		Asymmetry					
				Fix and Follow		Abnormal					
				Ductions and Versions		Abnormal					
			Ocular Health	External Inspection		Abnormal					
(10) Guo, Jia, Guo, Xiao, Shen, Li & Zhang (2000)[8]	9 to 50 months-old (mean = 28 months)	300[3]	Refractive Error	Computer Photoscreener (off-axis)	Technicians	Not reported	**95** (8)	**90** (30)	**92** (10)	93 (35)	20.75
			Alignment			Asymmetry					
			Clarity			Opacity					
(11) Walraven & Janzen (1993)	4 to 18 years	730[4]	Stereoacuity	TNO	Not Reported	> 240 arc sec	**100** (1)	**93** (24)	44 (59)	**100** (1)	21.25

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(12) Arnold et al (2004)[8]	4 to 6 years	93[3]	Refractive Error Alignment Clarity	MTI Photoscreener (off-axis)	School Nurses	Not Reported	82 (38)	**98** (5)	**97** (2)	91 (42)	21.75
(13) Watts, Walker, and Betts (1999)	3 to 18 years (mean = 10 years)	37[3]	Refractive Error Strabismus Clarity	MTI Photoscreener (off-axis)	Not Reported	Sphere < -1.5D > 2.5D; Cylinder > 1.5D; Anisometropia > 1.5D; Asymmetry Opacity	**93** (14)	**90** (30)	**97** (2)	90 (45)	22.75
(14) Kennedy, Sheps, & Bagaric (1995)	5 years	270[3,10]	Refractive Error Alignment Clarity	Otago Photoscreener (off-axis)	Technician	Not reported Asymmetry Opacity	81 (40)	**98** (5)	**77** (33)	**98** (18)	24
(14) Williams, Lumb, Harvey, & Sparrow (2000)[8]	13 to 69 months (mean = 48 months)	189[3]	Refractive Error (Anisometropia)	Topcon PR2000 autorefractor	Orthoptist	Anisometropia > 1.25D	**91** (18)	**92** (25)	62 (45)	**99** (8)	24

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(16) Tong, Macke, Bassin, Everett, Enke-Miyazake, Tielsch, Stager, Parks, Beauchamp, & The National Children's Eye Care Foundation Vision Screening Study Group (2000)	1 to 47 months (mean = 22 months)	111[3]	Refractive Error (hyperopia only)	MTI Photoscreener (off-axis)	Orthoptist	0.5mm Crescents in Two Meridians	**100** (1)	88 (37)	40 (63)	**100** (1)	25.5
(16) Newman & East (1999)[6]	3.5 years	596[4]	Visual Acuity	Sheridan Gardiner (6m; single)	Orthoptist	Acuity ≤ 20/30	85 (33)	**96** (14)	60 (47)	**99** (8)	25.5
			Alignment/ Motility	Cover Test		Movement					
				Prism Cover Test		Deviation					
				Ocular Movements & Convergence		Abnormal					
			Stereoacuity	TNO		> 1980 sec arc					
(18) Barry & König (2003)	3 years (mean = 3.5 years)	1114[4]	Visual Acuity	Lea Isolated Symbols (3m; single)	Orthoptists	Acuity < 20/50; Anisometropia ≥ 2 line difference	**91** (18)	**94** (21)	25 (66)	**100** (1)	26.5
			Alignment/ Motility	Unilateral and Alternate Cover and Uncover Test		Movement					
				Ocular Motility		Abnormal					
				Head Posture		Abnormal					
			Ocular Health	Anterior Segment Exam		Abnormal					

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(19) Kennedy & Thomas (2000)	< 3 to > 13 years (median = 7 years)	423[3,4]	Refractive Error Alignment Clarity	iScreen Photoscreener (off-axis)	Trained Technicians	Not Reported	**92** (15)	89 (34)	**94** (6)	86 (54)	27.25
(19) Kennedy & Sheps (1989)[8]	17 days to 6 years (mean = 1 year)	236[3]	Refractive Error Alignment Clarity	Otago Photscreener (off-axis) with cycloplegia	Ophthalmologist	Not reported Asymmetry Opacity	**95** (8)	88 (37)	**85** (17)	89 (47)	27.25
(21) Miller, Dobson, Harvey, & Sherrill (2003)	36 to 63 months	167[3]	Refractive Error (Astigmatism only) Visual Acuity (Hybrid)[11]	Nidek KM500 autokeratometer Lea Symbols (3m; linear; MassVAT format)	Program Coordinator	Cylinder $\geq$ 2.25D Acuity < 20/63	**97** (5)	79 (56)	**80** (28)	**96** (24)	28.25
(22) Simon, Siegfried, Mills, Calhoun, & Garland (2004)	6 months to 5 years (mean = 3.3 years)	115[3]	Grating Acuity	Enfant II VEP Headband	Trained Technicians	96% CI based on ROC[12]	**97** (5)	81 (52)	**71** (39)	**98** (18)	28.5
(23) Allen & Bose (1992)[6]	3.5 years	284[4,9]	Visual Acuity	Stycar (3m; single)	Clinical Medical Officer	Not Reported	77 (45)	**96** (14)	50 (55)	**99** (8)	30.5

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(23) Enzenauer, Freeman, Larson, & Williams (2000)[8,9]	7 months to 20 years (mean = 6 years)	109[3]	Visual Acuity	HOTV (3m; linear)	"Certified Screeners"	Not Reported	89 (26)	**90** (30)	**67** (43)	**97** (23)	30.5
			Muscle Balance	Titmus Model OV-7 or Titmus II Stereoscope		Not Reported					
			Ocular Health	External Exam		Abnormal					
(23) Arnold et al (2004)[8]	4 to 6 years	93[3]	Visual Acuity	HOTV (distance not reported; single; crowding bars)	School Nurses	Not Reported	**91** (18)	86 (44)	**79** (29)	94 (31)	30.5
(26) Spowart, Simmers, & Tappin (1998)[6]	3.5 years	718[4]	Visual Acuity	Stycar (6m; single)	Nurses	Acuity $\leq$ 20/30	83 (36)	**95** (16)	40 (63)	**99** (8)	30.75
(27) Raina (1998)[8]	4 to 12 years (mean = 7.3 years)	44[3]	Visual Acuity	Teller Acuity Cards	"Masked Examiner"	Acuity $\leq$ 20/40	15 (70)	**100** (1)	**100** (1)	87 (52)	31

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(28) Shallo-Hoffmann, Coulter, Oliver, Hardigan, & Blavo (2004)	2 to 6 years	81[3]	Visual Acuity	Lea Symbols (3m; linear); HOTV (3m; linear)	Not Reported	Acuity 3 yr <20/40; 4 yr < 20/32	**100** (1)	79 (56)	24 (67)	**100** (1)	31.25
			Stereopsis	RDE		< 600 arc sec					
(29) Freedman & Preston (1992)	5 months to 23 years; (mean = 7 years)	202[3]	Refractive Error	Eyecor Photoscreener (off-axis)[13]	Pediatrician, Opthalmologist, or Orthoptist	Sphere < -1D; Hyperoperopia = age based; Cylinder ≥ 1.5D; Anisometropia > 1.5D	87 (29)	89 (34)	**93** (8)	80 (58)	32.25
			Alignment			Asymmetry					
			Clarity			Opacity					
(30) Ottar, Scott, & Holgado (1995)	6 to 59 months (mean = 29 months)	949	Refractive Error	MTI Photoscreener (off-axis)	Ophthalmologist or Orthoptist	Sphere ≤ -0.5D ≥ 2D; Cylinder ≥ 1.5D	82 (38)	**91** (27)	**69** (40)	**95** (27)	33
			Alignment			Asymmetry					
			Clarity			Opacity					

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(31) Cordonnier & Dramaix (1998)[8]	9 to 36 months (mean = 2 years)	220[3]	Refractive Error (Hyperopia only)	Retinomax Autorefractor	Orthoptist	Cylinder > 1.5D	70 (50)	**95** (16)	**79** (29)	92 (39)	33.5
(31) Cordonnier & Dramaix (1999)[8]	9 to 36 months	302[3]	Refractive Error (Astigmatism only)	Retinomax Autorefractor	Ophthalmologist or Orthoptist	Cylinder ≥ 1.75 D	70 (50)	**95** (16)	**74** (37)	94 (31)	33.5
(33) Cordonnier & Dramaix (1999)[8]	9 to 36 months	302[3]	Refractive Error (Astigmatism only)	Retinomax Autorefractor	Ophthalmologist or Orthoptist	Cylinder ≥ 2D	52 (67)	**98** (5)	**83** (22)	91 (42)	34
(34) Kennedy & Sheps (1989)[8]	17 days to 6 years (mean = 1 year)	236[3]	Refractive Error Alignment Clarity	Photoscreener (off-axis) with cycloplegia	Technician	Not reported Asymmetry Opacity	**90** (23)	83 (46)	**79** (29)	92 (39)	34.25
(35) Cordonnier & Kallay (2001)[9]	9 to 36 months (mean = 23 months)	302[3]	Refractive Error	Retinomax Autorefractor	Orthoptist	Sphere < -3D > 1.5D; Cylinder ≥ 2D; Anisometropia > 1.5D	61 (61)	**98** (5)	47 (57)	**98** (18)	35.25

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(36) Granet, Hoover, Smith, Brown, Bartsch, & Brody (1999)	9 months to 16 years (mean = 6 years)	206[3]	Refractive Error Alignment Clarity	Kodak DC - 120 Photoscreener (off-axis)	"Nonprofessional Volunteer"	Sphere < -1D > 3D; Cylinder > 1D; Anisometropia > 1D Strabismus Opacity	89 (26)	83 (46)	**91** (13)	80 (58)	35.75
(36) Cordonnier & Dramaix (1999)[8]	9 to 36 months	302[3]	Refractive Error (Astigmatism only)	Retinomax Autorefractor	Ophthalmologist or Orthoptist	Cylinder ≥ 1.5D	77 (45)	**91** (27)	**65** (44)	**95** (27)	35.75
(38) Kennedy & Sheps (1989)[8]	17 days to 6 years (mean = 1 year)	236[3]	Refractive Error Alignment Clarity	Photoscreener (off-axis)	Technician	Not reported Asymmetry Opacity	85 (33)	87 (40)	**82** (24)	89 (47)	36
(39) Büchner, Schnorbus, Grenzebach, & Busse (2005)	3.5 to 4.5 years	56[3]	Refractive Error	SureSight Autorefractor	Not Reported	Sphere ≤ -1D ≥ 3D; Cylinder ≥ 1D; Anisometropia ≥ 1.00D	**94** (11)	70 (64)	**82** (24)	89 (47)	36.5

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(40) Williams et al. (2000)[8]	13 to 69 months (mean = 48 months)	189[3]	Refractive Error (Astigmatism)	Topcon PR2000 autorefractor	Orthoptist	Cylinder > 1.25D	**94** (11)	75 (59)	43 (60)	**98** (18)	37
(41) Cooper, Bowling, Hall, Colville, Dortmans, Munch, & Gole (1996)[8]	11 to 44 months (mean = 26 months)	89[3]	Refractive Error Alignment Clarity	Dortmans Photoscreener (off-axis)	Orthoptist	Not Reported Asymmetry Opacity	70 (50)	**90** (30)	**93** (8)	62 (66)	38.5
(41) Schmidt (1994)[8]	2 to 3 years (mean = 3 years)	30[3,4]	Refractive Error	Noncycloplegic Retinoscopy	Not reported	Sphere ≤ -0.5D ≥ 2.5 D; Cylinder ≥ 1D; Anisometropia ≥ 1D	62 (59)	**94** (21)	**89** (14)	76 (60)	38.5
(43) Morgan & Johnson (1987)	3 months to 8 years	57[3]	Refractive Error Alignment Clarity	Visiscreen 100 Photoscreener (off-axis)	Technician	Not Reported Strabismus Opacity	**91** (18)	74 (62)	**84** (20)	85 (55)	38.75

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(44) Guo et al. (2000)[8]	9 to 50 months (mean = 28 months)	300[3]	Refractive Error	Non-cycloplegic Retinoscopy	Not Reported	Not Reported	86 (31)	81 (52)	**85** (17)	82 (57)	39.25
			Alignment			Asymmetry					
			Opacity			Opacity					
(45) Schmidt (1994)[8]	2 to 3 years (mean = 3 years)	30[3]	Stereoacuity	Randot E	Not Reported	> 168 arc sec	77 (45)	88 (37)	**83** (22)	83 (56)	40
(45) Kennedy et al. (1995)[8]	5 years	270[3,10]	Visual Acuity	Snellen or Stycar Graded Balls	Health Care Aide	Acuity < 20/40	33 (69)	**97** (10)	54 (50)	94 (31)	40
			Stereoacuity	Titmus Stereotest		> 80 arc sec					
(47) Hope & Maslin (1990)[8]	3 - 4 years	168[4]	Stereoacuity	Randot E	Nurses	> 250 arc sec	89 (26)	75 (59)	17 (68)	**99** (8)	40.25
(48) Olver (1988)	6 mos to 11 years (median = 2 years	93[3]	Refractive Error	Simplified Retinoscopy with cycloplegia	Orthoptists	Sphere < -1.5D > 3D; Cylinder > 0.5D; Anisometropia > 0.5D	**90** (23)	74 (62)	**72** (38)	91 (42)	41.25
(49) Hope & Maslin (1990)[8]	5 to 15 years	100[3,4]	Stereopsis	Randot E	Not Reported	> 250 arc sec	53 (65)	**92** (25)	**92** (10)	55 (68)	42

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(50) Chui, Fraser, Hoar, & LaRoche (2004)[8]	≥ 41 mos	112	Visual Acuity	Lea Isolated Symbols (3m; single)	Public Health Nurse	≤ 20/40	50 (68)	**95** (16)	43 (60)	**96** (24)	42
			Ocular Health	External Penlight Exam		Abnormal					
			Stereoacuity	Frisby Stereo Plates		> 600 sec arc					
(51) Tong, Bassin, Enke-Miyazaki, Macke, Tielsch, Stager, Beauchamp, Parks, & The National Children's Eye Care Foundation Vision Screening Study Group (2000)	1 to 47 months (mean = 22 months)	313[3]	Refractive Error	MTI Photoscreener (off-axis)	Orthoptist	Sphere ≤ -2D ≥ 2D; Cylinder ≥ 2D; Anisometropia ≥ 2D	65 (55)	87 (40)	**95** (5)	41 (70)	42.5
			Alignment			Asymmetry					
			Clarity			Opacity or Ptosis					
(52) Williams et al. (2000)[8]	13 to 69 months (mean = 48 months)	189[3]	Refractive Error (hyperopia/ myopia)	Topcon PR2000 autorefractor	Orthoptist	Sphere < -3.75D > 3.75D	80 (41)	82 (49)	51 (53)	**95** (27)	42.5

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(53) Cooper et al. (1996)[8]	11 to 44 months (mean = 26 months)	96[3]	Refractive Error	Otago Photoscreener (on-axis)	Orthoptist	Not Reported	70 (50)	82 (49)	**85** (17)	65 (63)	44.75
			Alignment			Asymmetry					
			Clarity			Opacity					
(54) Schmidt (1994)[8]	2 to 3 years (mean = 3 years)	30[3,4]	Visual acuity	Broken Wheel Test	Not reported	Acuity < 20/40; Anisometropia ≥ 2 line difference	**92** (15)	53 (69)	57 (49)	89 (47)	45
(55) Cordonnier & Dramaix (1998)[8]	9 to 36 months (mean = 2 years)	220[3]	Refractive Error (Hyperopia only)	Retinomax Autorefractor	Orthoptist	Sphere > 1.25D	72 (49)	86 (44)	59 (48)	92 (39)	45
(56) DeBecker, MacPherson, LaRoche, Braunstein, Cottle, MacIntyre, & Kozousek (1992)	4.5 to 5.5 years	193[3]	Visual Acuity	HOTV (3m; linear)	Public Health Nurse	Acuity ≤ 20/30	62 (59)	89 (34)	50 (55)	93 (35)	45.75
			Stereoacuity	Randot		> 200 sec arc					
			Ocular Health	Inspection		Abnormal					

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(57) Enzenauer et al. (2000)[8,9]	7 months to 20 years (mean = 6 years)	127[3]	Refractive Error	Eyecor Photoscreener (off-axis)[13]	"Certified Vision Screening Professionals"	Age-based	79 (43)	83 (46)	**79** (29)	63 (65)	45.75
			Alignment			Asymmetry					
			Clarity			Opacity					
(58) Hatch, Tibbles, Mestito, Read, Traveis, & Richman (1997)[8]	3 to 11 years (mean = 6.5 years)	138[10]	Refractive Error	MTI Photoscreener (off-axis)	Trained Examiners	Sphere ≤ -2D ≥ 2D; Cylinder ≥ 2D; Anisometropia ≥ 2D;	53 (65)	**91** (27)	45 (58)	93 (35)	46.25
			Alignment			Strabismus					
			Clarity			Opacity					
(59) Barry & König (2001)[8]	3.5 years	404[3]	Refractive Error	Retinomax Autorefractor	Orthoptist	Sphere < -1D > 3D; Cylinder > 1.5D; Anisometropia: > 1D	80 (41)	58 (68)	5 (69)	**99** (8)	46.5
(60) Cordonnier & Dramaix (1998)[8]	9 to 36 months (mean = 2 years)	220[3]	Refractive Error (Hyperopia only)	Retinomax Autorefractor (Hyperopia only)	Orthoptist	Sphere > 1D	79 (43)	79 (56)	51 (53)	93 (35)	46.75

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(61) Barry & König (2001)[8]	3.5 years	404[3]	Refractive Error	Retinomax Autorefractor	Orthoptist	Sphere < -3D > 1.5D; Cylinder ≥ 2D; Anisometropia ≥ 1.5D	70 (50)	60 (67)	4 (70)	**99** (8)	48.75
(62) Cooper, Gole, Hall, Colville, Carden, Bowling (1999)[8,9]	12 to 44 months (mean = 26 months)	103[3]	Refractive Error Alignment Clarity	Fortune Optical VRB-100 Photoscreener (off-axis)	Orthoptist	Not Reported Asymmetry Opacity	64 (56)	81 (52)	**81** (27)	64 (64)	49.75
(63) Hope & Maslin (1990)[8]	5 to 15 years	100[3,4]	Stereopsis	Randot E	Not Reported	> 125 arc sec	85 (33)	53 (69)	**75** (35)	69 (62)	49.75
(64) Cooper et al. (1999)[8,9]	12 to 44 months (mean = 26 months)	99[3]	Refractive Error Alignment Clarity	MTI Photoscreener (off-axis)	Orthoptist	Sphere ≤ -2D ≥ 2D; Cylinder ≥ 2D; Anisometropia ≥ 2D Asymmetry Opacity	61 (61)	82 (49)	**82** (24)	61 (67)	50.25
(65) Schmidt (1994)[8]	2 to 3 years (mean = 3 years)	30[3,4]	Visual Acuity	Teller Acuity Cards	Not reported	Acuity < 20/40; Anisometropia > 1 octave difference	64 (56)	81 (52)	**75** (35)	71 (61)	51

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(66) Ruttum & Neslon (1991)	3 to 4 years	58[3]	Stereopsis	Randot E	Volunteer Screeners	> 168 arc sec	54 (63)	87 (40)	54 (50)	87 (52)	51.25
(67) Hatch et al. (1997)[8]	3 to 11 years (mean = 6.5 years)	138[10]	Refractive Error	MTI Photoscreener (off-axis)	Trained Examiners	Sphere < -1D > 1.5D; Cylinder > 1D; Anisometropia > 1D	54 (63)	87 (40)	52 (52)	88 (51)	51.5
			Alignment			Strabismus					
			Clarity			Opacity					
(68) Robinson, Bobier, Martin & Bryant (1999)	3 to 6 years	1505[3]	Visual Acuity	Cambridge Crowding Cards (3m; single)	Nurse	Acuity ≤ 20/30	64 (56)	75 (59)	26 (65)	94 (31)	52.75
			Alignment	Hirschberg Test (later discontinued)		Asymmetry					
			Stereoacuity	Titmus Stereo Test (added late)		> 100 arc sec					
(69) Weinand, Graf, & Demming (1998)	6 to 48 months	102[3]	Refractive Error	MTI Photoscreener (off-axis)	Photos Analyzed by Pediatrician, Ophthalmologist, or Orthoptist	Sphere ≤ -2D ≥ 2D; Cylinder ≥ 2D Anisometropia ≥ 2D	83 (36)	62 (66)	**68** (42)	48 (69)	53.25
			Alignment			Strabismus					
			Clarity			Opacity					

Study	Age Group	N[1]	Visual Function[2]	Test	Examiner	Criteria	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value	Average Rank
(70) Chui, Fraser, Hoar, & LaRoche (2004)	< 41 mos	53[3]	Visual Acuity	Lea Isolated Symbols (3m; single)	Public Health Nurse	Acuity ≤ 20/40	75 (48)	68 (65)	41 (62)	90 (45)	55
			Ocular Health	External Penlight Exam		Abnormal					
			Stereoacuity	Frisby Stereo Plates		> 600 sec arc					

[1]Unless otherwise noted, N refers to the number of children who received the gold standard exam.
[2]Note, studies screened for any disorder that could impair the functions(s) in question unless otherwise stated.
[3]Children were at risk for deficits either because the sample consisted of patients at optometrists' or ophthalmologists' clinics or a relatively large sample of the children receiving the gold standard exam had suspected visual deficits.
[4]Not all children received cycloplegic refraction as part of the gold standard exam or it is not clear whether children received cycloplegic retinoscopy.
[5]The Boström Hooks Test is similar to the Landolt C test except the visual target is in the shape of a square as opposed to a ring.
[6]Refers to studies which were retrospective reviews of regional programs conducted by clinicians and/or researchers other than the author.
[7]Most children who screened positive were retested.
[8]Complete data were provided for all children implementing several different tests, several different criteria, or different age groups. Thus, each data set is evaluated as a separate study.
[9]Validity measures were calculated as weighted means across multiple observers, disease types, or screening locations.
[10]Gold standard exam did not include cycloplegic refraction.
[11]Only subjects who passed the autokeratometer exam received the visual acuity exam.
[12]Referral criteria were based on 96% confidence intervals for interocular difference.
[13]The Eyecor photoscreener is a prototype of the MTI photoscreener.

functional vision were also implemented to screen children. Thus, it can not be determined whether the results of these studies reflect the merits/pitfalls of the Lea Symbols, or the merits/pitfalls of the other tests. Finally, according to the Table, the specific expertise of the examiner may not be important as those with a variety of training (nurses, orthoptists, and ophthalmologists) were all primary examiners in effective screening programs.

Several points should also be mentioned about those programs that were not considered effective. First, the finding that the 90% (63/70) of the studies reviewed were not judged effective highlights the difficulty of conducting an effective screening program. In fact, there are a number of obstacles that make vision screening a formidable task, including time constraints, poor evaluative tools (e.g., the Stycar graded balls, Teller Acuity Cards, and Fortune Optical VRB-100 photoscreener generally provided disappointing results, see Cooper et al., 1999 ranked 62nd; Kennedy et al., 1995 ranked 45th; Raina, 1988 ranked 27th; Schmidt, 1994 ranked 65th), and expensive vision testing equipment.[13] A second point that should be mentioned is that some of these studies just narrowly missed the cut-off criteria outlined above and may be considered effective if the criteria were slightly less conservative. In fact, if the cut-off criteria of each validity measure were reduced by just 5%, eight additional studies would have been deemed effective (Arnold et al., 2004, ranked 23rd; Eibschitz-Tshimonie et al., 2000, ranked 9th; Enzenauer et al., 2000, ranked 23rd; Guo et al., 2000, ranked 10th; Ladenvall, 1988, ranked 8th; Newman & East, 1999, ranked 16th; Watts et al., 1999, ranked 13th; Williams

[13] For instance, the Teller Acuity Cards cost $3, 200 US, and the Nikon Retinomax costs approximately $16, 000 US.

et al., 2000, ranked 14th). Finally, it appears that, once again, the level of expertise of the examiner is not critical as clinical medical officers, technicians, nurses, pediatricians, orthoptists, and ophthalmologists were the primary examiners for these studies.

Due to the large size of Table 3, and to compare the effectiveness of the different types of screening tests, the results for studies shown in Table 3 are classified into six categories and re-presented in Table 4 below.[14] The Table provides the weighted means of sensitivity, specificity, PPV, and NPV for each category of screening study. Note, weighted means were calculated so that the data justly reflected the contributions of large-scale screening studies. The mean rank of each category of study across all measures of validity is provided along with the percentage of studies from each category that passed the cut-off criteria for an effective screening program. The Table highlights three important results. First, studies that assess visual acuity alone, and those that assess the combinations of visual acuity, stereopsis/stereoacuity, and/or ocular alignment/motility were those most likely to be judged effective (at 27% and 18.2%, respectively). Note however, studies that incorporate visual acuity alone yielded higher scores on all measures of validity, particularly sensitivity, specificity, and PPV. Second, studies that employed estimates of refractive error (photoscreening and autorefraction) or tests of stereopsis/stereoacuity were much less likely to be judged effective. Although photoscreening programs possess relatively high estimates on all four measures of validity, only two met the cut-off criteria (Arnold et al., 2004, ranked 5th; Kennedy & Sheps, 1989, ranked 6th). The PPVs of autorefraction and stereopsis/stereoacuity tests, on

[14] Note, the studies conducted by Miller et al., (2003) and Olver (1988) do not fit into either of the categories, and thus, are not included here.

the other hand, were low and no single study from either category passed all the cut-off criteria. The poor performance of autorefraction may be due to the invasive testing technique of the Nikon Retinomax (the most frequently used autorefractor of all autorefraction studies) which must be positioned just 6 cm from the eyes of the subject (Cordonnier & Dramaix, 1999). This requirement may be distressing to young children and may explain why in some studies, researchers were often unable to obtain reliable measures (see Barry & König, 2001). Moreover, it may compromise the accuracy of refractive error estimates (Suryakumar & Bobier, 2003). Nevertheless, the poor performance of studies that assess refractive error (i.e., photorefractor and autorefractor studies) or stereopsis/stereoacuity indicate that measurement of these visual functions alone (or at least measurement with currently available tests/instruments) do not allow for effective vision screening. Third, studies that implement noncycloplegic retinoscopy yield high estimates on all measures of validity. However, it should be mentioned that these estimates are based only on three studies. Furthermore, similar to cycloplegic retinoscopy, the procedure is difficult and requires the technical expertise of highly trained eyecare professionals.

Table 4. Mean validity estimates and percentage of studies judged effective which employed specific screening tests or combinations of tests. These data are based on studies summarized in Table 3.

Function(s)	Number of Studies	Mean Sensitivity	Mean Specificity	Mean PPV	Mean NPV	Mean Rank on Validity Estimates	Percentage of Studies Judged Effective
Visual Acuity	11	89	96	72	99	27.5	27
*Noncycloplegic Retinoscopy	3	85	94	69	94	32.5	0
Visual Acuity + at least one other function	11	78	89	48	97	33.6	18.2
Photoscreening	24	83	88	81	84	34.6	8.3
†Autorefraction	13	74	82	49	96	38.4	0
Stereoacuity	6	90	87	48	92	40.8	0

* Includes the Eibschitz-Tsimhoni et al. (2000) study which screened children using noncycloplegic retinoscopy along with tests of alignment and ocular health.
† The studies of Willians et al. (2000) which used the Topcon PR 2000 have been included in the autorefraction category and not the photoscreening category. This decision was taken because the device automatically provides a refractive error measure and does not involve the typical crescent measurement of traditional photoscreening procedures.

Perhaps the most ambitious, comprehensive screening program conducted to date has been carried out in a series of three studies by the Vision in Preschoolers (VIP) Study Group (2004, 2005a, 2005b). This group screened a total of 4040 3- to 5-year-old children enrolled in the Head Start daycare program in the United States. Children were screened with a number of tests assessing several different visual functions by licensed eyecare professionals (i.e., optometrists and pediatric ophthalmologists), nurses, or trained lay screeners. Following screening, all children underwent a complete, on-site gold standard exam that included cycloplegic retinoscopy. In each study, the referral criteria of most tests were set retrospectively (after the gold standard exams) at levels that

yielded specificity of 90% or 94% (there are exceptions noted in Table 5) and the resulting sensitivity estimates of the tests were compared to determine which tests were superior. In their final study, the VIP Study Group (2005b) compared the sensitivity estimates of several tests conducted by a nurse vs. the sensitivity of the same tests when conducted by a trained lay screener.

The results of the three VIP studies are provided in Table 5 below. It should be mentioned that PPV and NPV were not calculated for these studies[15] and therefore, are not included in the Table. Nevertheless, the Table reveals a number of surprising results, some of which appear to contradict the results of the studies reviewed above. First, although the specificity of each test in the Table is quite high due to the referral criteria that were implemented, no single test possesses adequate sensitivity to reach the cut-off criteria for an effective screening program. In fact, in comparison to the studies reviewed in Table 3 that possessed similarly high levels of specificity, the VIP studies attained relatively low sensitivity estimates for visual acuity tests (range = 36% to 61%), stereoacuity tests (range = 22% to 44%), and photorefraction (37%; to compare visual acuity studies, see Arnold et al., 2004, ranked 23rd; Enzenauer et al., 2000, ranked 23rd; Ladenvall, 1988, ranked 8th; Kvarnström, et al., 1998, ranked 7th; to compare stereoacuity studies, see Hope & Maslin, 1990, ranked 49th; Ruttum & Nelson, 1991, ranked 66th; Schmidt, 1994, ranked 45th; to compare photorefraction studies, see Arnold et al., 2004, ranked 12th; Kennedy & Thomas, 2000, ranked 19th; Ottar et al., 1995, ranked 30th; Tong

[15] In fact, it is for this reason that the VIP studies were not included in Table 3. Note also that sufficient data were not provided to allow us to calculate these measures.

et al., 2000, ranked 16th; Watts et al., 1999, ranked 13th). Note however, the sensitivity of visual acuity tests implemented by the VIP Study Group is relatively high compared to that of stereoacuity and alignment tests (16%), and photorefraction. Second, in the VIP studies, the sensitivity of the Nikon Retinomax and Welch-Allyn SureSight autorefractors was relatively high across all screeners (range = 51% to 68%) and both specificity levels (90% and 94%). This contrasts with the results of Tables 3 and 4 which indicate that autorefractors generally yield the poorest sensitivity of all tests. Yet, it should be mentioned that the absolute sensitivity estimates of autorefraction reported by the VIP Study Group is in many cases, similar to those in Table 3 above (see Cordonnier & Dramaix, 1998, ranked 31st; 1999, ranked 31st; Cordonnier & Kallay, 2001, ranked 35th). Third, a comparison of the first and third VIP studies in the Table indicates that there is little difference between the results of screening conducted by licensed eyecare professionals and nurses, a result similar to that reported above. However, the final study in the Table demonstrates that sensitivity levels are higher when screening is conducted by nurses as opposed to trained lay screeners. Finally, the first and third VIP studies suggest that when specificity is held constant at 90%, combining two tests of visual function (refractive error and alignment/motility or refractive error and stereoacuity) provides only modest improvement, if any, to sensitivity.

Table 5. Summary of the validity of the Vision in Preschoolers (VIP) Study Group screening studies.

Study	Age Group	N	Visual Function	Test	Examiner	Criteria	Sensitivity	Specificity
VIP Study Group (2004)	3 to 5 years	*2588	Visual Acuity	Lea Symbols; 3m; linear; MassVAT	Licensed Eye Care Professionals	Age 3: < 20/64 Age 4: < 20/40 Age 5: < 20/40	61	90
				HOTV test (3m; linear) MassVAT		Age 3: < 20/50 Age 4: < 20/50 Age 5: < 20/40	54	89*
			Stereoacuity	Randot E		Age 3: Nonstereo Card Age 4: > 550 arc sec Age 5: > 252 arc sec	42	90
				Stereo Smile II		Age 3: > 240 arc sec Age 4: > 240 arc sec Age 5: > 120 arc sec	44	91*
			Alignment/Motility	Cover-Uncover Test		Movement	16	98*
			Refractive Error	Noncylopegic Retinoscopy		Sphere: $\leq$ -2.75D; $\geq$ 2.75D Anisometropia: $\geq$ 1.5D Cylinder: $\geq$ 1.25D	64	90

Study	Age Group	N	Visual Function	Test	Examiner	Criteria	Sensitivity	Specificity
				Retinomax Autorefractor		Sphere: ≤ -2.75D; ≥ 1.5D Anisometropia: ≥ 2D Cylinder: ≥ 1.5D	63	90
				Retinomax Autorefractor		Sphere: ≤ -2.75D; ≥ 1.5D Anisometropia: ≥ 1.75D Cylinder: ≥ 1.5D	64	90
				SureSight Autorefractor		Sphere: ≤ -1D; ≥ 4D Anisometropia: ≥ 3D Cylinder: ≥ 1.5 D	63	90
				SureSight Autorefractor		Sphere: < -1D; > 2D Anisometropia: > 1D Cylinder: > 1D	85	62†
				Power Refractor II autorefractor		Sphere: ≤ -3D; ≥ 3.5D Anisometropia: ≥ 1.5D Cylinder: ≥ 1.5D	54	90

Study	Age Group	N	Visual Function	Test	Examiner	Criteria	Sensitivity	Specificity
			Refractive Error	Noncylopegic Retinoscopy		Sphere: ≤ -2.75D; ≥ 2.75D Anisometropia: ≥ 1.5D Cylinder: ≥ 1.25D	66	90
			and					
			Alignment/Motility			Movement		
VIP Study Group (2005a)	3 to 5 years	2588	Visual Acuity	Lea Symbols; (3m; linear) MassVAT	Licensed Eye Care Professionals	Age 3: < 20/64 Age 4: < 20/50 Age 5: < 20/40	49	94
				HOTV (3m; linear) MassVAT		Age 3: < 20/64 Age 4: < 20/64 Age 5: < 20/50	36	93*
			Alignment	Cover-Uncover Test		Movement	16	98*
			Stereoacuity	Randot E		Age 3: Nonstereo card Age 4: Nonstereo card Age 5: > 550 arc sec	22	92*
				Stereo Smile II		Age 3: > 480 arc sec Age 4: > 480 arc sec Age 5: > 240 arc sec	33	94

Study	Age Group	N	Visual Function	Test	Examiner	Criteria	Sensitivity	Specificity
			Refractive Error	Noncylopegic Retinoscopy		Sphere: ≤ -2.75D; ≥ 2.50D Anisometropia: ≥ 1.5D Cylinder: ≥ 2.D	57	94
				Retinomax Autorefractor		Sphere: ≤ -2.75D; ≥ 1.75D Anisometropia: ≥ 2.75D Cylinder: ≥ 2.00D	52	94
				Retinomax Autorefractor		Sphere: ≤ -2.75D; ≥ 2.5D Anisometropia: ≥ 2.5D Cylinder: ≥ 1.75D	52	94
				SureSight Autorefractor		Sphere: ≤ -1D; ≥ 4.25D Anisometropia: ≥ 3.5D Cylinder: ≥ 1.75D	51	94
				Power Refractor II		Sphere: ≤ -3.75D; ≥ 3.5D Anisometropia: ≥ 2.75D Cylinder: ≥ 2.25D	36	94
				iScreen Photoscreener		Not Reported	37	94

Study	Age Group	N	Visual Function	Test	Examiner	Criteria	Sensitivity	Specificity
				MTI Photoscreener		Not reported	37	94
VIP Study Group (2005b)	3 to 5 years	1452	Visual Acuity	Lea Symbols (linear; 3m) MassVAT	Nurse	Age 3: < 20/64; Age 4: < 20/50 Age 5: < 20/40	49	90
			Stereoacuity	StereoSmile II		Age 3: > 480 arc sec Age 4: > 120 arc sec Age 5: > 120 arc sec	45	90
			Refractive Error	Retinomax Autorefractor		Sphere: ≤ 3.25D; ≥ 1.75D Anisometropia: ≥ 2.75D Cylinder: ≥ 1.5 D	64	90
				SureSight Autorefractor		Sphere: ≤ 1D; ≥ 4D Anisometropia: ≥ 2.75D Cylinder: ≥ 1.75 D	68	90
			Refractive Error and Stereoacuity	Retinomax Autorefractor and StereoSmile II	Nurse	Not Reported	63	90

Study	Age Group	N	Visual Function	Test	Examiner	Criteria	Sensitivity	Specificity
			Refractive Error	SureSight Autorefractor		Not Reported	58	90
			and					
			Stereoacuity	StereoSmile II		Not Reported		
			Visual Acuity	Lea Symbols (linear; 3m) MassVAT	Nurse	Not Reported	47	90
			and					
			Stereoacuity	StereoSmile II		Not Reported		
VIP Study Group (2005b)	3 to 5 years	1452	Visual Acuity	Lea Symbols (linear; 3m) MassVAT	Lay Screener	Age 3: < 20/50 Age 4: < 20/50 Age 5: < 20/50	37	90
				Lea Symbols (single; 1.5m; crowded)		Age 3: < 20/50 Age 4: < 20/40 Age 5: < 20/40	61	91*
			Stereopsis	StereoSmile II		Age 3: > 240 arc sec Age 4: > 120 arc sec Age 5: > 120 arc sec	40	90

Study	Age Group	N	Visual Function	Test	Examiner	Criteria	Sensitivity	Specificity
			Refractive Error	Retinomax Autorefractor		Sphere: ≤ 3D; ≥ 1.50D Anisometropia: ≥ 2D Cylinder: ≥ 1.75D	62	90
				SureSight Autorefractor		Sphere: ≤ 1D; ≥ 4.5D Anisometropia: ≥ 2.25D Cylinder: ≥ 1.75 D	61	90
			Refractive Error	Retinomax Autorefractor	Nurse	Not Reported	65	90
			and					
			Stereoacuity	StereoSmile II		Not Reported		
			Refractive Error	SureSight Autorefractor		Not Reported	63	90
			and					
			Stereoacuity	StereoSmile II		Not Reported		

Study	Age Group	N	Visual Function	Test	Examiner	Criteria	Sensitivity	Specificity
			Visual Acuity	Lea Symbols (single; 1.5 m; crowded)	Nurse	Not Reported	47	90
			and					
			Stereoacuity	StereoSmile II		Not Reported		

*Denotes tests for which specificity can not be set at the desired level (i.e., 90% or 94%).
†Followed referral criteria suggested by manufacturer. Thus, specificity was not set at 90%.

In summary, although this review of the literature reveals somewhat mixed results, three general conclusions regarding vision screening in preschool and early school-aged children may be reached. First, it is unlikely that a screening study will be effective without measuring visual acuity. Second, it may be beneficial to also include tests of ocular alignment and stereoacuity to provide more comprehensive screening of early visual problems. The most effective of these tests appear to be the cover test of ocular alignment (Köhler & Stigmar, 1973, ranked 3rd; Wormald, 1991, ranked 2nd), and the TNO and RDE[16] tests of stereoacuity (Ruttum, 1988; Simons, 1981; Simons, 1996; Walraven & Janzen, 1993, ranked 11th). The inclusion of these tests may allow the detection of visual deficits missed by visual acuity testing alone, or more likely, reinforce concurrent visual acuity results. Third, the assessment of refractive error or stereoacuity should not be relied upon as the sole measure of visual functioning. This does not, however, preclude the measurement of refractive error or stereoacuity as part of a larger screening procedure. Importantly, estimation of refractive error may be the only method that can diagnose anisometropia (Adams et al., 2001; Adams et al., 2002). Furthermore, given the relative ease and quickness of the latest techniques to measure refractive error and stereoacuity, these measures could still be a valuable component of a vision screening program.

[16] Note, although the Wirt Fly Stereo Test was used in two effective screening studies, this test is generally considered ineffective as many children with abnormal vision pass the test (Köhler & Stigmar, 1973, Reinecke & Simons, 1974; Ruttum, 1988).

The Present Study

Based on the above review, the present thesis will attempt to develop and evaluate an effective program to screen for amblyogenic factors (which include anisometropia, strabismus, high refractive error, poor visual acuity, and poor stereoacuity) in toddlers and preschoolers. Importantly, this thesis represents a critical first step in the implementation of a long-term screening program to be conducted by our laboratory in the coming years. This study will be a unique and valuable addition to the literature for a number of clinical and practical reasons. First, we will conduct a large-scale, population-based program designed to assess 2-5 year-old children in all daycare centres in the St. John's metropolitan area. As a result, this study may be instrumental in reducing the incidence of amblyopia in the area as all children who screen positive will be referred to an optometrist and offered treatment. Furthermore, this will also allow us to estimate the prevalence of vision disorders in toddlers and preschoolers in the province of Newfoundland and Labrador as no estimates currently exist. Also, this study will be of national importance as only three population-based vision screening studies have been conducted to date in this country[17] (DeBecker et al., 1992, ranked 56th; Kennedy et al., 1995, ranked 14th and 45th ; Robinson et al., 1999, ranked 68th), neither of which was judged to be effective in the above review. Second, this program will be one of the most comprehensive conducted to date, as it will attempt to assess up to five separate visual functions (visual acuity, alignment/motility, stereoacuity, refractive error, and contrast sensitivity). Conversely, the studies listed above in Table 3 typically assess only one to

[17] In fact, Kennedy and Sheps (1989) photoscreening program was only Canadian study judged effective. However, this study was conducted with a relatively small sample (N = 236) of pediatric outpatients.

three visual functions. Therefore, it is likely that this screening program will be much more sensitive in comparison to those reviewed above. Third, promising new tests, such as the contrast sensitivity cards (Adams et al., 1992; Adams & Courage, 1996; Drover et al., 2002) and the Welch-Allyn SureSight autorefractor (Adams et al., 2001; Adams et al., 2002), will be included in the screening program. Contrast sensitivity testing has not yet been used in any screening study, and the Welch-Allyn SureSight autorefractor has only recently been used as a screening tool (Büchner et al., 2005; VIP Study Group 2004, 2005a, 2005b). Though little is known about these tests as part of a screening program, their efficiency and unsophisticated testing procedures make them promising vision screening tools which may eventually revolutionize vision screening. Finally, this study will be one of the first to evaluate both the validity of tests of a single visual function *and* tests of combinations of visual functions to determine which should be included to provide a comprehensive assessment of vision. As a result, this program may be the first to provide clear guidelines as to what constitutes an effective vision screening program.

Method

Participants

In all, 954 toddlers, preschoolers, and young school-aged children were tested between July 2003 and October 2005. The children were arranged into the following five age categories: 2 years (1.5 to 2.4 years), 3 years (2.5 to 3.4 years), 4 years (3.5 to 4.4 years), 5 years (4.5 to 5.4 years), and over 5 years (5.5 years and older). Children were recruited by sending consent forms to daycare centres in the St. John's, Newfoundland and Labrador, Canada, metropolitan region. We tested only those whose parents or

guardians provided signed consent (see Appendix for two versions of the consent form), and ethical approval for the study was received from the Memorial University Interdisciplinary Committee on Ethics in Human Research (ICEHR). There were no exclusion criteria as all children, if cooperative, were tested. All vision assessments took place in the child's daycare facility and were conducted by either the author (JRD) or by a research assistant who was rigorously trained over a period of 5 months. For the first two months of her training, the research assistant observed the author as he carried out the assessments and provided her with instructions concerning possible dysfunctions and how to detect them. Following this period, if a child was cooperative, the research assistant conducted a single test of one aspect of functional vision after he/she had been completely assessed by the author. Thus, the child was tested twice with one procedure.[18] This phase of the training was conducted to allow the calculation of inter-tester reliability to ensure that the research assistant was sufficiently trained to carry out assessments independently.

Materials and Procedure

Whenever possible, children were screened with the tests described below. Note however, the program was flexible and continuously evolved based on practicality, new knowledge, and technological advances. Thus, tests of each aspect of functional vision varied across subjects and over time for three reasons. First, if a child was unable to complete the *preferred test* (i.e., the test designed for the child's age and cognitive ability) for a specific visual function, he/she was screened with an easier to administer,

[18] This was not always the case. If a child was very cooperative and time was sufficient, the research assistant conducted two or more tests already completed by the preschooler.

alternative test. Second, due to their lower level of cognitive development, their limited attention spans, and their apprehension towards strangers, toddlers were generally screened with tests different from those implemented for preschoolers or were not screened on some visual functions at all (e.g., stereoacuity). Third, over time, if superior tests emerged (e.g., tests that were easier for children to complete, considered more accurate) they were incorporated into the program as replacement tests for initially preferred tests. Due to these reasons, the sections below contain descriptions of *several* tests to assess visual functions. Furthermore, these sections are also subdivided based on age groups.

Visual Acuity

Preschoolers. Initially, the preferred visual acuity test to screen preschoolers was the Landolt C linear optotype test (see Figure 3; Precision Vision, LaSalle, Ill., U.S.A.). This test consists of a 23 x 35.5 cm white chart containing 12 rows of five Landolt C optotypes (except for the first line which contains three optotypes and the second line which contains four optotypes[19]) ranging in size from 20/200 to 20/8 (0.9 to -0.3 logMAR units). The progression of optotype size from line to line follows a logMAR format. LogMAR is an acronym for $\log_{10}$ minimum angle of resolution and is equal to $\log_{10}$ of optotype size expressed in minutes of arc (Ricci et al., 1998). Importantly, under logMAR format, there is a systematic, equivalent, reduction in optotype size from line to line of 0.1 log MAR units.

---

[19] This is a problem common to 10 ft charts. These charts are smaller than their 20 ft counterparts and thus, can only contain a limited number of optotypes on the first two lines. Note however, the optotypes on this line are quite large (20/200 and 20/125) and are well below the referral criteria of the present study.

This test was selected as the preferred test as it is very similar to the highly effective Boström hooks used by Nordlöw and Joachimson (1962) and Köhler and Stigmar (1973). Moreover, this version of the Landolt C test follows all six recommendations for visual acuity testing listed above (see the Visual Acuity heading in the subsection under Recommendations for Early Vision Screening and Tests of Functional Vision). The translucent chart was mounted onto a 23 x 35.5 x 10 cm illuminator cabinet (Precision Vision, LaSalle Ill., USA) equipped with an 8 watt fluorescent bulb. The cabinet illuminated the chart to approximately 170 cd/m^2 as measured with a cal-Spot 400VF photometer (The Cooke Corporation, London, Ontario), thus ensuring the chart was well-illuminated even in dimly lit rooms.[20]

Each child was screened at a distance of 3 m as preschoolers tend to be more cooperative at this distance when compared to the standard distance of 6 m (Atkinson, Anker, Evans, Hall, & Pimm-Smith, 1988; Pickert & Wachs, 1980; Simmers et al., 1997; Simons, 1983). Furthermore, the 3 m distance was chosen for convenience because it is often difficult to find a room large enough to allow screening at 6 m. In fact, some researchers report that due to space restrictions, visual acuity testing in elementary schools is often conducted in bathrooms, dimly lit corridors, or assembly halls (Stewart-Brown & Haslum, 1988). Note, screening was conducted monocularly as the child wore a pair of "monocular" children's sunglasses, i.e., the lens of the glasses over the eye being screened was removed allowing the child to clearly view the chart. The lens of the fellow eye on the other hand, was covered with masking tape to ensure that the child

[20] The illuminator cabinet was also used with the linear optotype tests described below, and thus, illuminated these chart to the same approximate measure of 170 cd/m^2.

could not view the chart with this eye. All monocular testing described below (i.e. other visual acuity tests and CS tests) was conducted while the child wore these glasses. Order of eye occlusion was determined randomly. To complete a single line of the Landolt C test, the child had to correctly determine the position of the gap in at least 4/5 optotypes. This criterion was chosen because it was recommended by the Maternal Child Health Bureau and National Eye Institute Task Force on Vision Screening in the Preschool Child (2001)[21] for linear optotype testing. The lowest line (i.e., smallest size of optotypes) at which the child could detect the gap on 4 occasions was taken as an estimate of visual acuity.

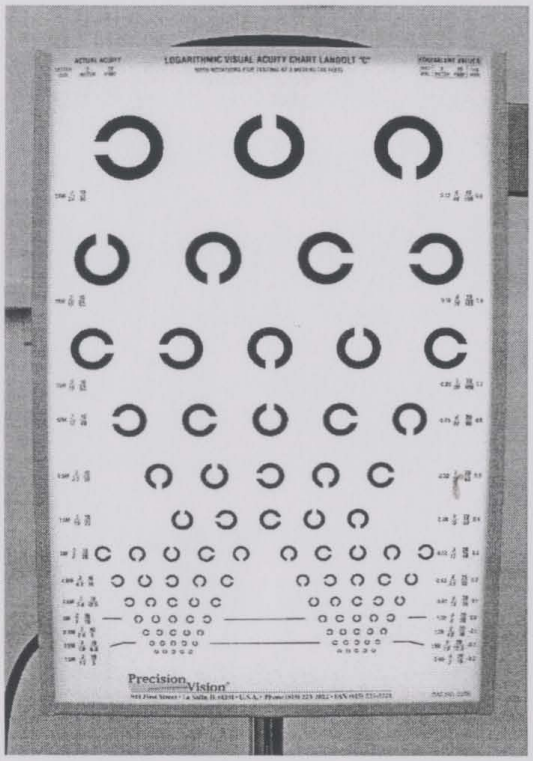

Figure 3. Photograph of the Landolt C visual acuity test.

[21] For this reason, this criterion was chosen for all linear optotype tests in the present study.

Despite its merits, the Landolt C test proved extremely difficult after only two older, cooperative participants. In fact, the second child was unable to complete the test. Thus whenever possible, the remaining preschoolers were screened with the linear optotype version of the Lea Symbols test (i.e., it was the preferred test; Precision Vision, LaSalle Ill., USA) shown in Figure 4 below. This test is similar to the Landolt C test described above as it contains the same number and size of optotypes and the progression of optotype size also follows a logMAR format. The chart was mounted onto the same illuminator cabinet described above. This test was considered to be a suitable alternative to the Landolt C test as it tends to lead to higher completion rates in comparison to other recognition acuity tests such as the HOTV test, the Illiterate E test, and the Landolt C test itself (Becker, Hübsch, Gräf, & Kaufmann, 2000; 2002; Hered, Murphy, & Clancy, 1997; Lennie & van Hemel, 2002). Moreover, the optotypes of the Lea Symbols test are roughly equally legible as recommended above, and therefore, tend to blur similarly when they cannot be resolved (Hered et al., 1997; Ottar, 1997).

As above, testing was conducted monocularly at a distance of 3 m. Once again, the order of eye occlusion was determined randomly. The child was required to either name the optotype, or if communication was difficult, he/she was given a card containing the four optotypes (i.e, circle, house, apple, and square) and asked to point to the one that was being presented. To complete a single line, the child had to identify correctly at least 4/5 optotypes. The lowest line (i.e., the smallest optotype size) at which the child could detect 4 optotypes was taken as a measure of visual acuity.

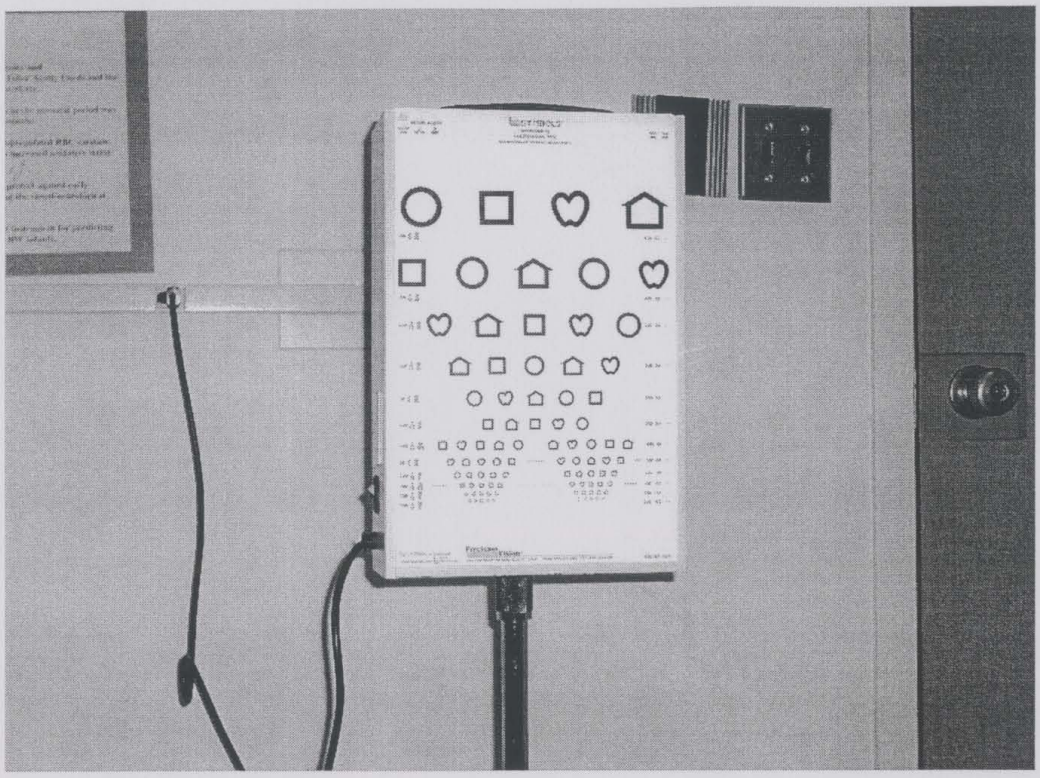

Figure 4. Photograph of the Lea Symbols linear optotype test.

Children who could not complete the Lea linear optotype test were screened with the alternative test, the Lea Isolated Symbols Book (Precision Vision, LaSalle Ill., USA). This test, shown in Figure 5, consists of 11 12.5 x 12.5 cm bound plastic pages containing Lea symbols ranging from 20/200 to 20/10 (0.9 to -0.2 logMAR units), which progress in size following the logMAR format. Each optotype size possesses four symbols (one apple, one house, one square, and one circle). Pages with optotypes ranging from 20/200 to 20/160 possess a single symbol, whereas those with optotypes ranging from 20/125 - 20/100 possess two symbols. All other pages (i.e., those with optotypes ranging from 20/80 to 20/10) possess four symbols. This isolated format was chosen as an alternative

visual acuity test because isolated optotype tests are less demanding than the linear optotype format (Simons, 1983), and therefore, younger children are more likely to complete the procedure.

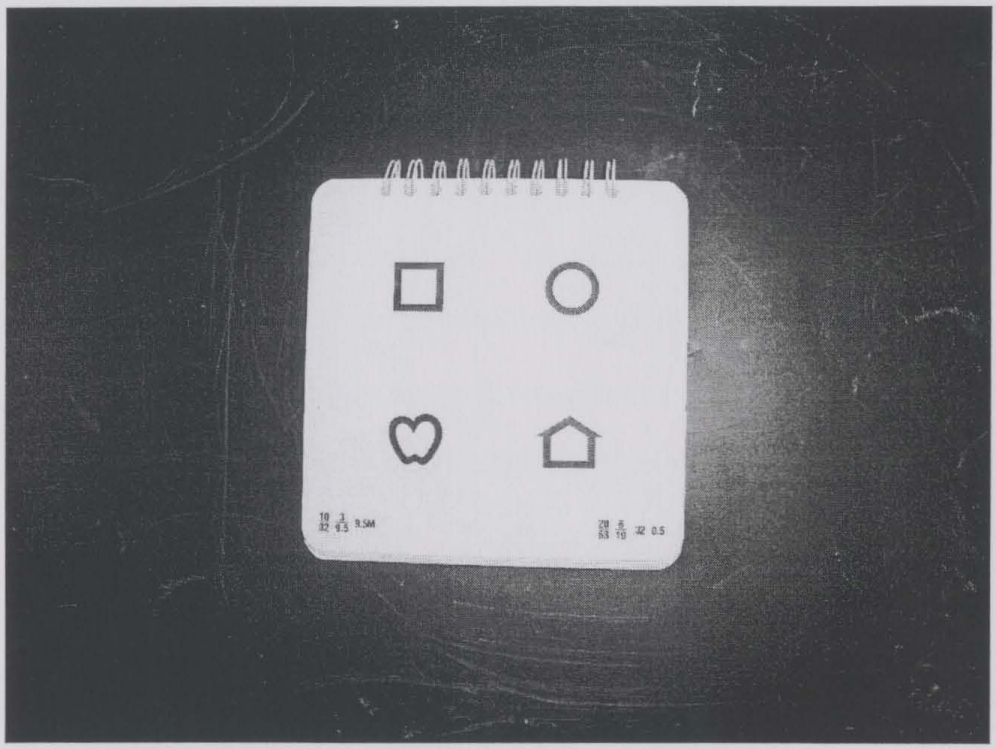

Figure 5. Photograph of the Lea Isolated Symbols Book. Note that each optotype is presented in isolation as the remaining optotypes are occluded with a plastic sheet.

As with the linear optotype tests, the single optotype test was carried out monocularly at a distance of 3m. Testing was conducted in a well-lit room under which conditions the average luminance of background of the optotypes was approximately 45

to 65 cd/m^2.[22] Each symbol was presented in isolation by occluding all others with an opaque plastic sheet. As with the Lea linear test, the child was required to name the symbols or was given a card containing the four optotypes and asked to point to the one that was being presented. To complete a single optotype size, the child had to correctly identify 3/4 optotypes. This criterion was chosen as it has been implemented in another screening program with a single optotype testing format (Barry & König, 2003). The smallest optotype size at which he/she could correctly identify at least 3 of the 4 symbols was taken as an estimate of visual acuity.

Beginning in May of 2005, the preferred test to measure visual acuity in preschoolers was the new Patti Pics linear optotype chart shown in Figure 6 below (Precision Vision, La Salle, Illinois, USA). The chart consists of 8 lines of optotypes ranging from 20/80 to 20/16 (i.e., 0.6 to -0.1 logMAR units) which progress following the logMAR format. The chart is essentially the same as the Lea Symbols chart, but possesses two important advantages. First, the optotype representing a house possesses a "chimney" and the optotype representing an apple possesses a "stem". Though subtle, these additions are important to reduce ambiguity in naming the optotypes. For instance, when tested with the Lea Symbols, children often referred to the apple as a butterfly or heart, whereas they often referred to the house as an arrow or triangle. Furthermore, children often used several references of an optotype during a single test. Second, the chart follows the new Massachusetts Visual Acuity Testing (MassVAT) format.

---

[22] Because the optotypes of the Patti Pics Isolated Symbols Book, and the Patti Pics cards (both are described below) are printed onto backgrounds of the same luminance as the Lea Isolated Symbols Book, the luminance of the background of these tests was also 45 to 65 cd/m^2 under testing conditions.

Following this format, lines of optotypes are spaced slightly further apart than under the traditional linear optotype format. Yet, each line of optotypes is *completely* surrounded by crowding contours in the form of a rectangle (See Figure 6). This ensures that all optotypes, including those at the beginning and end of each line, are subjected to equal crowding, either by the crowding contours or by other optotypes that are positioned above, below, to the left, and to the right. On the other hand, under the traditional linear optotype format (i.e., that of the Lea linear optotype test), optotypes at the beginning and end of each line are subjected to crowding by other optotypes positioned above, below, and in *one* lateral direction only (i.e., to the left or right). In light of this, it was expected that because the MassVAT format provides substantially more crowding, visual acuity scores would be slightly lower than with the traditional linear optotype format. Note, the Patti Pics chart possesses the same dimensions as the Landolt C chart and the Lea Symbols chart and thus, was mounted onto the same illuminator cabinet. As above, testing was conducted monocularly at a distance of 3 m and the child was required to correctly identify at least 4/5 optotypes to complete a line. The smallest optotype size at which the child identified 4/5 optotypes was taken as an estimate of visual acuity.

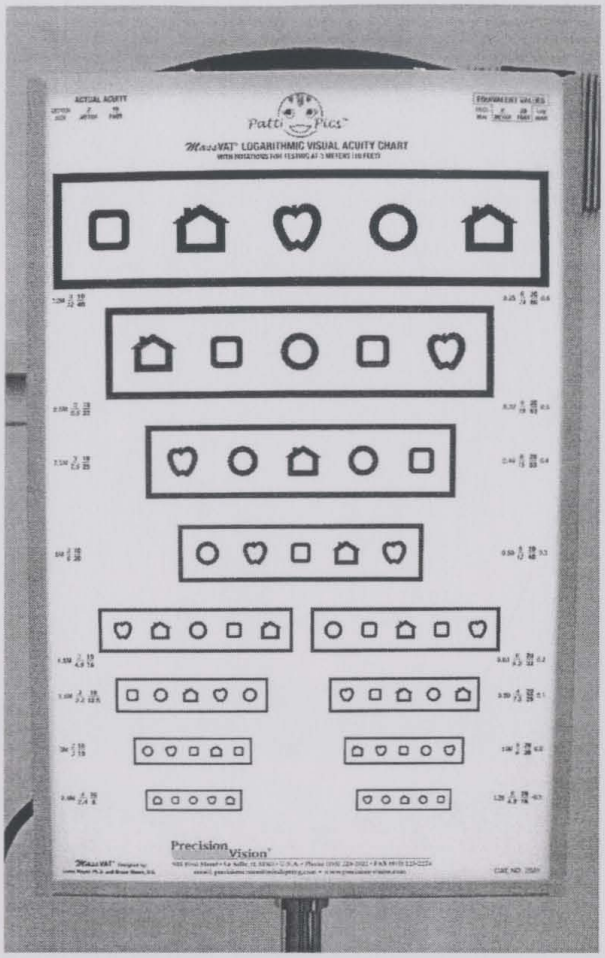

Figure 6. Photograph of the Patti Pics linear optotype test. Note that the house optotype possesses a "chimney" and the apple possesses a "stem" to prevent ambiguity. Also, the test follows the MassVAT format as each line of optotypes is surrounded by crowding contours.

Preschoolers who could not complete the Patti Pics linear optotype test were screened with the Patti Pics Isolated Symbols Book shown below in Figure 7. The book is identical to the Lea Isolated Symbols Book except once again, the optotype representing a house contains a "chimney" and the optotype representing an apple contains a "stem". Note that the presentation of the isolated Patti Pics Symbols was identical to the presentation of the Lea isolated symbols above.

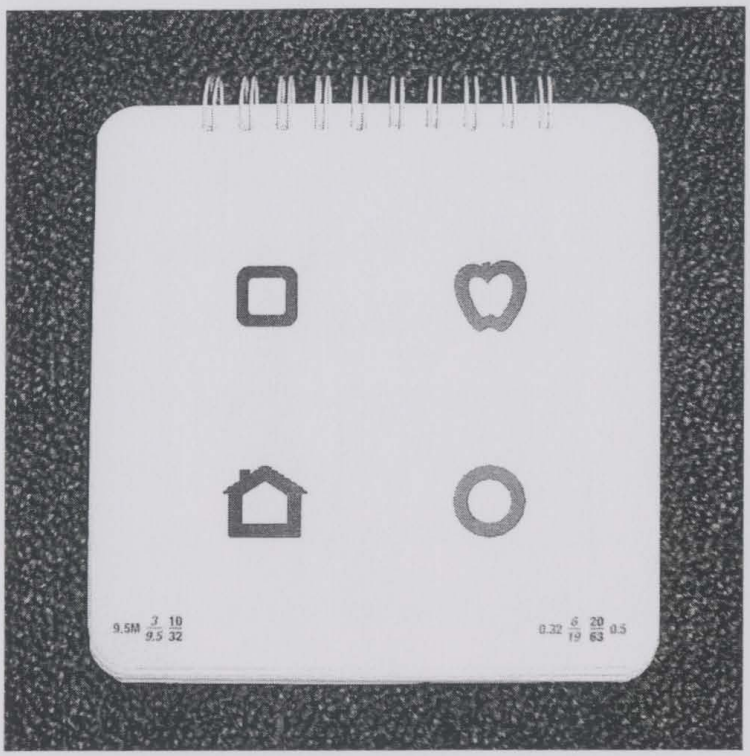

Figure 7. Photograph of the Patti Pics Isolated Symbols Book

Finally, children who could complete neither the Patti Pics linear optotype chart nor the Patti Pics Isolated Symbols Book were screened with the Patti Pics two-alternative, forced-choice cards (referred to as Patti Pics cards) shown in Figure 8. The test consists of 30 cards: each contains two optotypes, one each on the front and back. Optotype size ranges from 20/200 to 20/8 (1.0 to -0.3 logMAR units) and follows a logMAR progression. The dimensions of each card are based on the size of the optotype it possesses (see Table 6 below). This card dimension/optotype arrangement allows both easy access to each optotype size and easy progression of optotype size during testing. Screening began with the 20/200 optotypes, and followed the two-alternative-forced-

choice format with two different optotypes of the same size, (and thus, two cards) being presented at once. The child was required to point to one of the optotypes chosen by the tester. If the child would not point, he/she was told to simply look at the chosen optotype and his/her ability to detect it was based on his/her fixation. Four or five different combinations were shown at each optotype level and the child was required to correctly identify the optotype chosen by the tester at least four times. This criterion was chosen as it has been recommended by The Maternal and Child Health Bureau and National Eye Institute Task Force on Vision Screening in the Preschool Child (2001), for the similarly designed Randot E Stereotest which also follows this two-alternative forced-choice format.[23] The smallest optotype size at which the child correctly identified four optotypes provided an estimate of visual acuity. Note, as with the visual acuity tests described above, testing was conducted monocularly at a distance of 3 m.

Figure 8. Photograph of the Patti Pics two-alternative-forced-choice isolated symbols test (Patti Pics cards).

[23] For this reason, this criterion has also been chosen for RDE and contrast sensitivity testing (see below).

Table 6. Card dimension and optotype size of the Patti Pics two-alternative-forced-choice isolated symbols test (i.e., Patti Pics cards).

Card Dimensions (cm)	Optotype Size (Snellen Notation)
12.7 x 12.7	20/200; 20/160; 20/125
12.7 x 12.1	20/100; 20/80; 20/63
12.7 x 11.4	20/50; 20/40; 20/32
12.7 x 10.8	20/25; 20/20; 20/16
12.7 x 10.2	20/12; 20/10; 20/8

Toddlers. From July 2003 to May 2005, the preferred visual acuity test for toddlers was the Lea Isolated Symbols Book. Note however, if a toddler was extremely cooperative, he/she was assessed with the preferred test for preschoolers, i.e., Lea Symbols Chart. Yet, due to the attentional/cognitive demands of the test, this was very rarely the case. Importantly, neither of the above tests was attempted with a toddler until it was determined that he/she could match shapes. Specifically, the toddler was shown the largest optotypes (20/200) of the test and asked to match each to a card he/she was given that contained the same optotypes.

If a toddler could not complete either of these tests, he/she was screened monocularly with the Teller Acuity Cards shown in Figure 9 below (Vistech Consultants, Dayton, OH, U.S.A.). The test consists of 16 grey 25 x 58 cm rectangular cards, each of which contains a high contrast (83%) 12.5 x 12.5 cm black and white square wave grating located 7.5 cm to the left or right of a central peephole. When viewed from 55

cm, the test distance designed for this age group, the spatial frequencies (SFs) of the gratings range from 0.31c/deg to 38 c/deg in approximately 0.5 octave steps[24]. The average luminance of the gratings matches the background of the card, thus, if the child could not detect the grating, the card appeared to be a blank, isoluminant grey field.

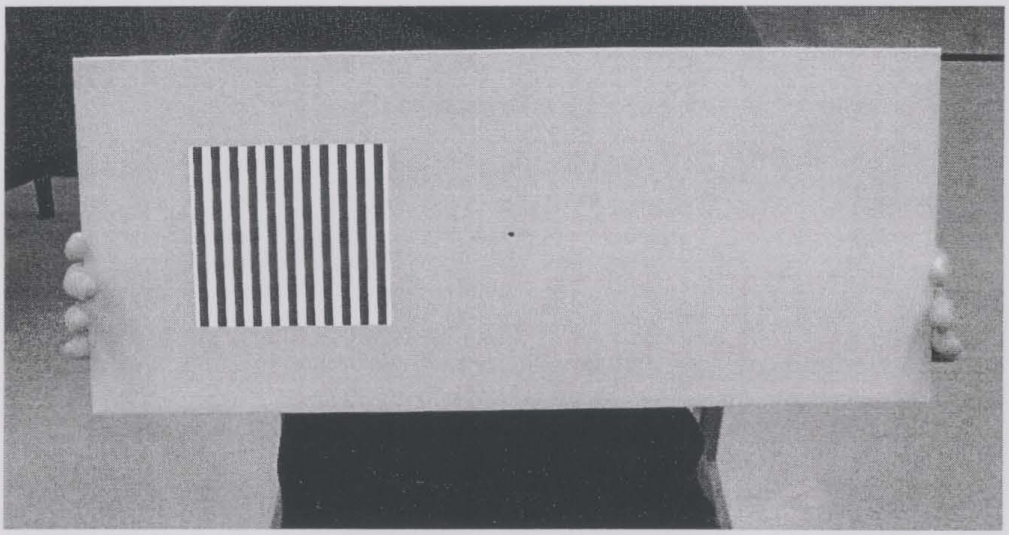

Figure 9. Photograph of a Teller Acuity Card.

Testing was conducted in a well-lit room under which conditions the average luminance of the cards was approximately 25 to 35 cd/m^2. Presentation of the cards followed the modified FPL procedure designed for the Teller Acuity Cards and used widely today for this and other card-based tests (see McDonald et al., 1985). To begin testing, the tester presented a card containing a low SF grating and the child was instructed to point in the direction of the grating. If he/she would not point, the tester observed the child's fixation under the assumption that if the grating could be detected, he/she would fixate it (Fantz, 1965). Note that the tester was never permitted to look at

[24] An octave, in the case, refers to the halving or doubling of spatial frequency.

the front of the card until after the child had pointed, or after a decision had been made regarding the position of the grating based on the child's fixation preference. The card was then rotated as many times as was necessary to conclude that the child could detect the grating. Screening then continued with cards containing gratings of progressively lower SF until the child could no longer point to the side of the card containing the grating, or the tester could not determine the position of the grating based on the fixation behavior of the child. The highest SF grating detected by the child provided an estimate of visual acuity.

Beginning in May 2005, the preferred visual acuity test for toddlers was the Patti Pics isolated symbols book. However, if a toddler was very cooperative, he/she was assessed with the preferred test for preschoolers, namely, the Patti Pics chart. Toddlers who could not complete either of these tests were tested with the Patti Pics cards described above. Note that if the child could not complete the Patti Pics cards, no other acuity testing was conducted. Once again, it should be noted that neither of these tests was attempted until it was determined that the toddler could match shapes using the procedure for the Lea Symbols described above.

Ocular Alignment/Motility

Preschoolers. Whenever possible, preschoolers were screened with the standard distance cover-uncover test and the Hirschberg corneal reflex described in the Introduction. Beginning in May 2005, preschoolers were also screened with the near cover-uncover test (see above). Following these tests, ocular motility and gaze in the nine cardinal directions was inspected. To conduct the motility/gaze test, a penlight was

positioned in front of the child, about 40 cm from his/her face. The light was turned off (and remained off throughout the test) and forward gaze was inspected. The child was then instructed to maintain his/her head position and to use only his/her eyes to follow the penlight as it was first moved upward and downward along the vertical plane, and then leftward along the horizontal plane. The child continued to follow these instructions as the penlight was moved diagonally upward to the left and diagonally downward to the left. The penlight was then moved to the right along the horizontal plane, diagonally upward to the right, and diagonally downward to the right. During the movement of the target, eye movements were inspected closely to ensure that ocular alignment was maintained at all times and that tracking movements were smooth. Also, at the end point of each direction of movement (i.e., the point at which upward/downward/diagonal/ lateral movement of the penlight was stopped), the ability of the child to fixate on the target with both eyes and maintain binocular alignment was examined. Finally, the children were screened with the convergence/divergence test. To begin this test, the penlight was held approximately 40 cm from the front of the child's face (note, the penlight remained off during this test) and then moved slowly towards his/her nose as convergence eye movements were examined. Next, the child's divergence eye movements were examined as the penlight was moved away from his/her nose.

Toddlers. Toddlers were screened with the Hirschberg corneal reflex. If the child appeared particularly cooperative, the cover-uncover, ocular motility, and convergence/divergence tests were also attempted in the identical manner used to test preschoolers.

Stereoacuity

Preschoolers. From July 2003 to March 2005, the preferred test of stereoacuity was the Randot E Stereotest shown in Figure 10 below. This test consists of two 8 x 10 cm random dot plates, a demonstrator plate, and a pair of polarized glasses. One of the random dot plates appears blank whereas the second possesses a "floating" E with a crossed level of retinal disparity that can only be detected if one possesses stereoacuity. This plate is calibrated so that the E subtends different disparities when held at different test distances. For instance, when held at distances of 0.5, 1 and 1.5 m, the E subtends a relative depth of 500, 250, and 168 arc sec, respectively. A third, "demonstrator" plate which is also included in the test, functions as a warm-up/training stimulus. This plate is a large embossed E that simulates the random dot stereogram and is shown first to each child to inform him/her what to look for when presented with the random dot plates.

To begin testing, the child was shown the demonstrator plate and asked to identify the letter. If the child could not do so, he/she was simply told it was an E. The polarized glasses were then placed on the child and the random dot plate that contained the E was presented from approximately 20 cm. The child was then asked whether he/she could see the E. If the response was negative, the position of the E was traced by the tester until it was confirmed that it could be detected. Next, the tester presented the two random dot plates (i.e., the "E" and the blank), one in each hand, and positioned them 50 cm from the child's eyes. The child was then instructed to point to the plate that contained the E. The plates were then shuffled and presented once again, and he/she followed the same instructions. In all, the plates were presented 4 or 5 times and if the child correctly

pointed the location of the E on at least four presentations, it was concluded that he/she could detect it. The tester then moved to the next test distance (1 m) and the same procedure was followed. Once again, if the child correctly identified the location of the E on 4/4 or 4/5 presentations, the exact same procedure was carried out at 1.5 m. The lowest disparity at which the child could correctly identify the E was taken as an estimate of stereoacuity. If the child could not detect the E at 50 cm, his/her score was recorded as > 500 arc sec.

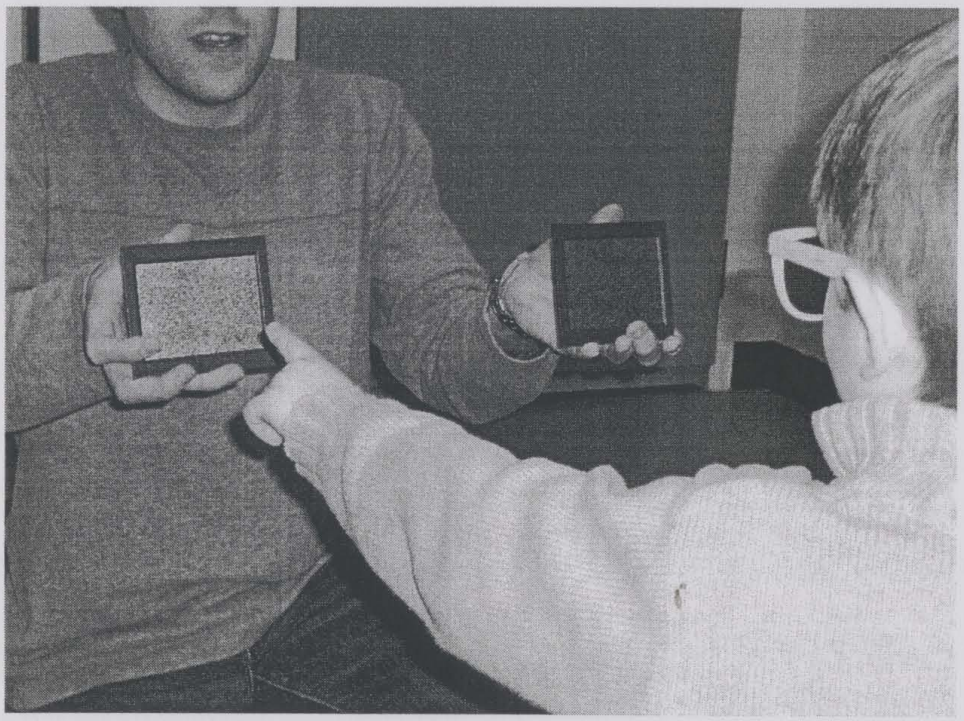

Figure 10. Photograph of a preschooler being tested with the Randot E Stereotest. Note that the child is pointing to an "E" of crossed disparity that can be detected when wearing polarized glasses.

The Randot E Stereotest was chosen for our battery as it is relatively sensitive to amblyopia and strabismus[25] and can be administered quickly and easily to preschool children (Reinecke & Simons, 1974; Rosner, 1978; Schmidt, 1994; Simons, 1981; Simons, 1994). Furthermore, this test possesses several important advantages over traditional techniques of stereopsis/stereoacuity screening such as the TNO, Randot Stereotest, Titmus test, etc. First, the Randot E stereotest does not generate monocular cues that allow spurious detection of the disparate target. Second, the test is conducted at several distances, thus allowing easy adjustment of disparity levels and thus, a more accurate measurement of stereoacuity levels. Third, the left/right position of the disparate target (i.e., the plate containing the E) can be changed at will by the tester, and therefore, cannot be memorized by the child. Finally, in comparison to other techniques, the Randot E Stereotest is relatively inexpensive.[26]

Beginning in April 2005, children were screened with the Randot Preschool Stereoacuity Test (Birch et al., 1997) presented in Figure 11. This test was chosen to replace the Randot E Stereotest because it provides a broader range of disparity levels with a relatively small step size between them. Thus, the test should provide more precise estimates at all ranges of stereoacuity. Moreover, Birch et al. (1997) have demonstrated that the test has higher rates of sensitivity and specificity (91 and 96%, respectively) in the detection of binocular abnormalities in preschoolers than do other tests of stereoacuity, including the Randot E stereotest.

The test consists of three booklets, the left page of each contains two sets of four

---

[25] There are exceptions. See Ruttum and Nelson (1991) in Table 3.
[26] The TNO costs approximately $330.00 US whereas the RDE costs approximately $130.00 US.

black and white, two-dimensional figures that can be easily identified by preschoolers (eg. duck, car, star, square, etc.). The right page contains two sets of four random dot patterns. Three of the four patterns possess stereofigures of crossed disparities that correspond to those on the left, but in a different order. The fourth pattern is a blank. Book 1 contains stereofigures of intermediate disparities (200 and 100 arc sec), whereas book 2 contains stereofigures of fine disparities (60 and 40 arc sec), and book 3 contains stereofigures of coarse disparities (800 and 400 arc sec). The testing procedure was identical to that of Birch et al. (1997). Testing began with book 1. To first determine whether the child could recognize the figures in general, he/she was asked to point at or name each of the easily visible black and white figures. Next, he/she was asked to identify the three stereofigures in the corresponding stereograms on the right page. At each disparity level, the child was required to correctly identify 2/3 stereofigures. If he/she correctly identified the stereofigures in book 1, testing continued with book 2. If the subject could not identify the stereofigures in book 1, testing continued with book 3. The finest disparity level at which the child correctly identified 2/3 stereofigures was taken as an estimate of stereoacuity. Note, if it appeared that the preschooler could not understand the instructions of the Randot Preschool Stereoacuity Test, he/she was assessed with the Randot E Stereotest.

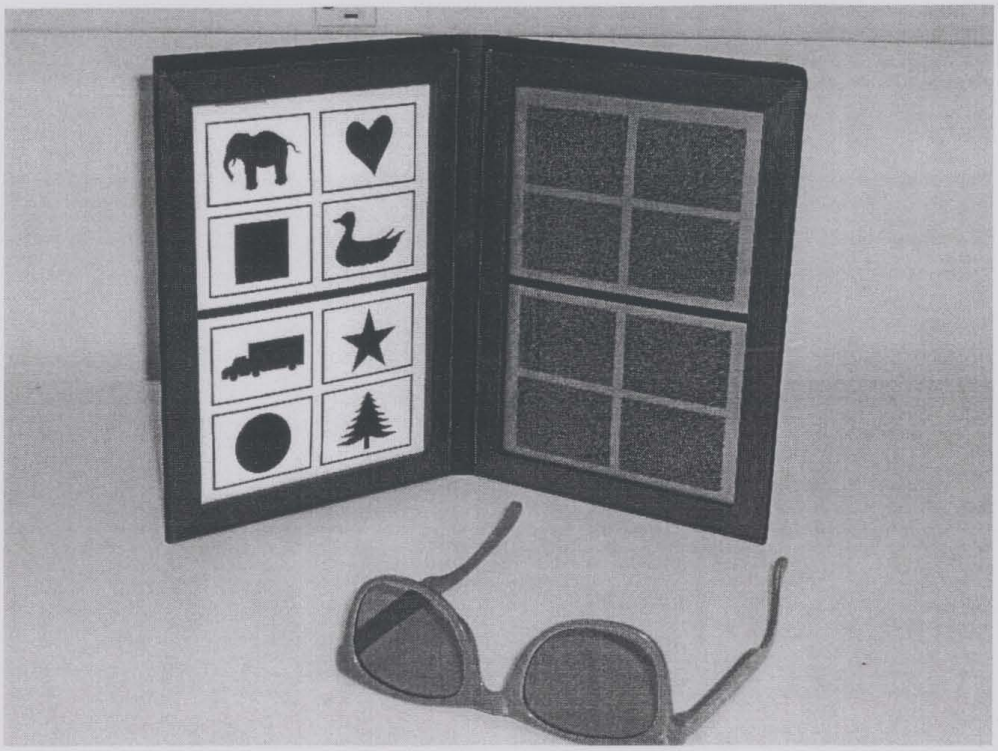

Figure 11. Photograph of one of the three books from the Randot Preschool Stereacuity Test.

Toddlers. At the onset of the study, it was intended for all toddlers to be screened with the Randot Stereo Smile Test shown in Figure 12. The test consists of three rectangular 24.5 X 54 cm cards covered completely with a random array (Stereo Optical Co Inc, Chicago, Ill, USA; Ciner, Schanel-Klitsch, & Herzberg, 1996). When viewed through polarized glasses by a subject with stereopsis, two of the cards contain a smiling face target of crossed disparity to the left or right of the center of the card. The reverse of these cards contains the exact same target, except that its location is on the opposite side of center. The smiling face targets measure 11 cm in diameter and subtend a crossed

disparity of either 480 arc sec (card 1) or 120 arc sec (card 2) when viewed at a distance of 55 cm. The third card, the training card, contains a highly visible, two-dimensional, embossed smiling face located on the left or right side of the card.

To begin the test, the child was shown the training card at a distance of 55 cm and he/she was instructed to point to the easily visible smiling face. If the child could not point to the target, he/she was told to simply look at it and as with the Teller Acuity Cards, the child's ability to detect the target was based on whether he/she could fixate the face. After a few trials, the training card was then turned over so that the smiling face was on the other side, and the child was given the same instructions. If he/she was able to successfully complete the pretest, the glasses were then placed on the toddler and the card with the coarsest disparity (i.e., 480 arc sec) was presented. As with the training card, the child was instructed to point to, or look at the target. The card was then flipped over and was presented and the same instructions were given. In all, the card was presented 4 or 5 times varying the location of the target randomly from trial to trial. If the child was correct on 4/4 or 4/5 trials, the next card (i.e., 120 sec arc of disparity) was presented following the same procedure. The finest disparity at which the child was correct on 4/4 or 4/5 trials was taken as an estimate of stereoacuity.

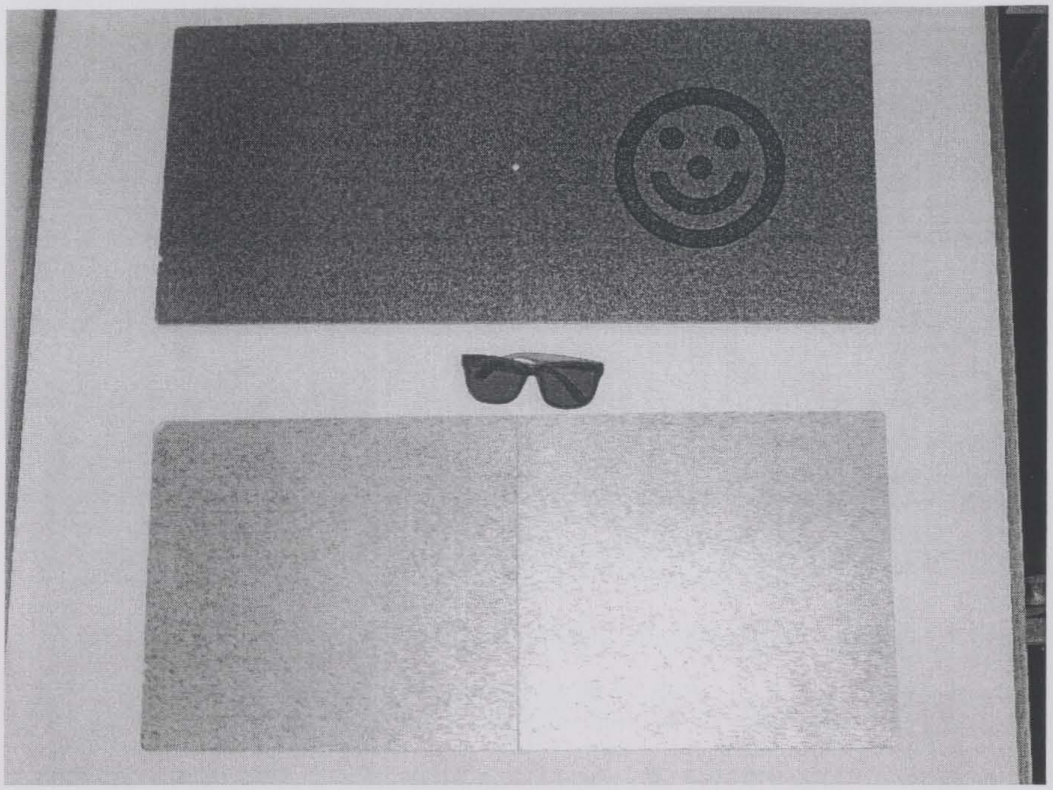

Figure 12. Photograph of the Randot Stereo Smile Test. The top card is the demonstrator card. The bottom card contains a "happy face" of crossed disparity that can be detected when wearing polarized glasses.

Despite its similarity to the Randot E Stereotest, screening with the Randot Stereo Smile Test proved very difficult after attempting to test only seven toddlers. Furthermore, it was obvious an examination of binocular function could be obtained much more easily implementing the Hirschberg corneal reflex. In light of this, the test was discontinued. Note that from this point on in the program, stereoacuity testing was not attempted with a toddler unless the child was very cooperative. In such a case, the toddler was assessed with either the Randot E Stereotest or the Randot Preschool Stereoacuity Test following successful completion of a training phase. Specifically,

before attempting the Randot E Stereotest, the child was presented with the demonstrator plate and the plate containing the blank stereogram and asked to point to the "E". Before attempting the Randot Preschool Stereoacuity Test, the examiner showed the toddler several two-dimensional black and white figures and told him/her the correct names of each of these figures. The examiner then provided the name of the stereofigures one at a time and the toddler was required to point at each in turn.

Contrast Sensitivity

Preschoolers. Beginning in February 2005, children, if cooperative, were screened with the contrast sensitivity (CS) cards designed by the author, Dr. Russell J. Adams, and Avery E. Earle (all of the Psychology Department at Memorial University). The cards, presented in Figure 13 below, consist of 20, 22 x 56 cm rectangular cards. Each card contains two large circles located 8 cm to the left and right of a central 2 mm peephole. The circles have a diameter of 17.5 cm and subtend a visual angle of 16.3° at a viewing distance of 60 cm. One circle is the *test* grating, which consists of a vertical, sine wave grating of a given spatial frequency and contrast. The other circle, the *control* grating, is a vertical, sine wave grating with the same spatial frequency (SF), but with a contrast of 0% (i.e., all stripes are of equal luminance). Thus, the control stimulus appears as a blank/subliminal field with luminance equal to the average luminance of the test grating and the background of the card, and to adults, is indiscriminable from the background of the card. A subthreshold grating was used as the control stimulus (vs. leaving one side of the card blank) to ensure that the child could not detect the test grating by relying on an edge/grating artifact (e.g., a slight brightness difference existing

on the outer edges of the grating). Therefore, if any artifact existed, it would be present on both sides of the peephole and would not reveal the location of the test grating. Even with this precaution, all adult observations of the test have yet to detect an edge artifact on any card to date.

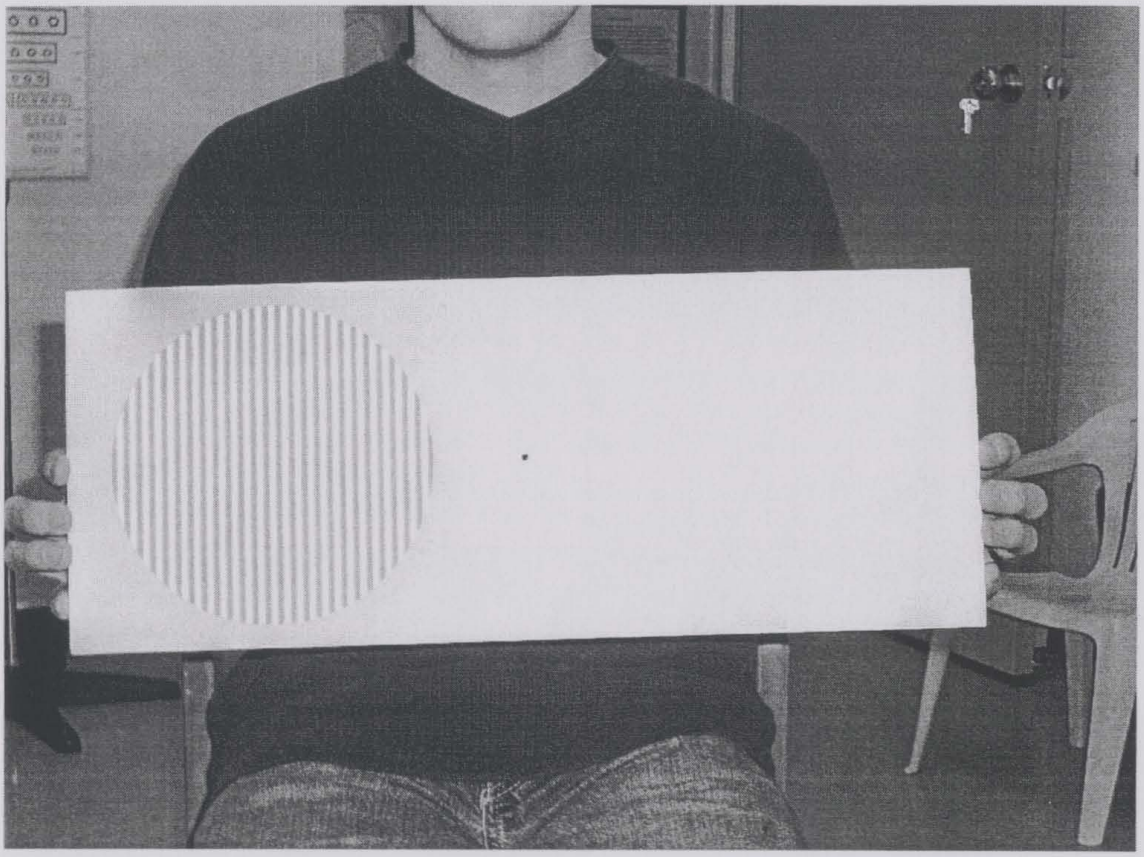

Figure 13. Photograph of a contrast sensitivity card from the 3.0 c/deg spatial frequency set.

All gratings were generated by composing suitable programs in PostScript programming language (Adobe Systems Incorporated, 1986; 1990). The gratings were then printed onto acid-free Hewlett-Packard Everyday Matte Photopaper with an Epson

Stylus 2200 Photoprinter. Matte photopaper was chosen as it portrays a photograph-quality, sharp image without the glare associated with glossy photopaper. The paper was then heat pressed onto 3.2 mm thick acid-free foamcore. Within the lighting conditions of the daycare centres, the average luminance of each grating and the background of the card ranged from 23 cd/m^2 to 35 cd/m^2.

The CS cards are divided into five sets (each containing 3 to 5 cards) based on the SF (0.75, 1.5, 3.0, 6.0, and 12.0 c/deg) of the test grating in each set. The SF, CS value, and contrast of each card are listed in Table 7. The Table also shows that each SF set includes a high contrast (48 to 57%) warm-up card that is presented to capture the attention of the child at the onset of testing with that set. In all, contrast ranges from 57 to 2.6%.

Table 7. Contrast sensitivity values (in contrast sensitivity units) and spatial frequencies of the contrast sensitivity cards. Note that numbers in parentheses represent percent contrast.

	Card Number				
Spatial Frequency	1	2	3	4	5
0.75	1.8 (57)	4.4 (22.7)	17.8 (6.4)	20.8 (4.8)	
1.5	2.1 (48.1)	4.4 (22.7)	17.8 (6.4)	20.8 (4.8)	27.8 (3.6)
3	2.1 (48.1)	4.4 (22.7)	16.9 (5.9)	38.5 (2.6)	
6	2.1 (48.1)	3.2 (31.7)	11.4 (8.7)	38.5 (2.6)	
12	1.8 (57.1)	3.2 (31.7)	4.4 (22.7)		

Initially, as the test is new for screening purposes, a pilot study was conducted with the first 24 children tested with the CS cards. This study was run to ensure that children could understand the procedure and that the referral criteria were adequate. Moreover, it was essential to determine whether the test could be carried out quickly as it was to be added to a battery that already required up to 10 minutes to complete. Testing was conducted monocularly from a distance of 60 cm. To test each SF set, the card containing the highest percent contrast (i.e., the warm-up card) was presented first and the child was instructed to point to the side that contained the test grating. As with the other two-alternative-forced-choice tests described above, if the child could not or would not point to the grating, he/she was told simply to look at it, and the tester decided its location based on his/her fixation. The card was then rotated several times and he/she was instructed to point to or look at the grating once again. Note, the tester was blind to the position of the grating and was not permitted to look at the front of the card until after the child had pointed, or after a decision had been made regarding the position of the grating based on the child's fixation. In all, the card was presented 4 or 5 times, and the child was required to point to or look at the grating at least 4 times in order for it to be judged that it could be detected. Screening then proceeded with cards containing gratings of lower contrast until the child could not point to or look at the correct side of the card at least four times. The lowest contrast grating detected was taken as an estimate of contrast threshold. This procedure was then continued with the remaining SF sets and the order of presentation was counterbalanced.

During the pilot study, two important points became clear. First, the procedure took a relatively long time (mean = 4.3 min) to complete. This is an important drawback as a lengthy procedure reduces the likelihood that the test and/or entire battery can be completed by a child. Second, CS was generally higher than expected. In fact, there appeared to be a ceiling effect at the three lowest SFs (i.e., 0.75, 1.5, and 3 c/deg) as even the youngest children could detect the lowest contrast gratings. Note however, there did not appear to be a ceiling effect at the two highest SFs (i.e., 6 and 12 c/deg). Thus, it was judged that the test could still potentially be used to detect cases of visual disorders.

In light of these findings, when CS measurement was formally included in the screening program in February 2005, the procedure was modified slightly to make it more time-efficient. For instance, SF order was no longer counterbalanced. Instead, screening began at 0.75 c/deg and then continued in order of increasing SF. Also, instead of presenting the warm-up card of each SF set, the lowest contrast card was generally presented first. Thus, the first card presented during the test was the lowest contrast card at 0.75 c/deg. This was justified by the pilot study which revealed that this card could be detected rather easily and thus, could be used to ensure that the child possessed an adequate understanding of the instructions thereby rendering warm-up card unnecessary. Furthermore, these changes generally guaranteed that at least one less card per SF set was presented during the procedure and therefore, test time was substantially reduced. As in the pilot study, the child was required to point to or look at the side of the card containing the test grating on 4/4 or 4/5 presentations. If he/she could not detect this grating, a card containing a higher contrast grating was then presented. The lowest contrast grating

detected was taken as an estimate of contrast threshold. Testing then continued with the remaining SF sets in a similar fashion.

Beginning in September of 2005, children were tested with the *CS screening booklet* developed by our lab. The booklet, presented in Figure 14, consists of a total of 168, 21.6 x 27.9 cm grey pages within a 3-ring binder. Each contains a sine wave grating located 0.9 cm to the left or right of a centrally located, 1 mm thick white line. The sine wave grating measures 12 cm in diameter, and subtends a visual angle of 16.7° when presented at the test distance of 40 cm. Like the CS cards, the opposite side of each page contains a control grating of the same SF with a contrast of 0% and equiluminant to the background of the page. Within the lighting conditions of the daycare centres, the average luminance of each grating and the background of the card ranged form 23 cd/m^2 to 35 cd/m^2. The test is divided into 5 sets of 32-40 pages based on the SF of the sine wave grating (0.75, 1.5, 3.0, 6.0, 12.0 c/deg). Within the SF sets, contrast ranges from 52.6% to 0.9 %, and each contrast level contains four pages (see Table 8). This allowed the test procedure to follow the 4/4 or 4/5 detection criteria described above for other two-alternative-forced-choice tests. The location of the grating (i.e., to the left or right of the white line) was determined by a random numbers table. Note however, for a single contrast level, the grating could not be located in the same position for all four pages. Thus, the possible outcomes of grating location for a given contrast level relative to the center of the page was 1 left/3 right, 2 left/2 right, or 3 left/1 right.

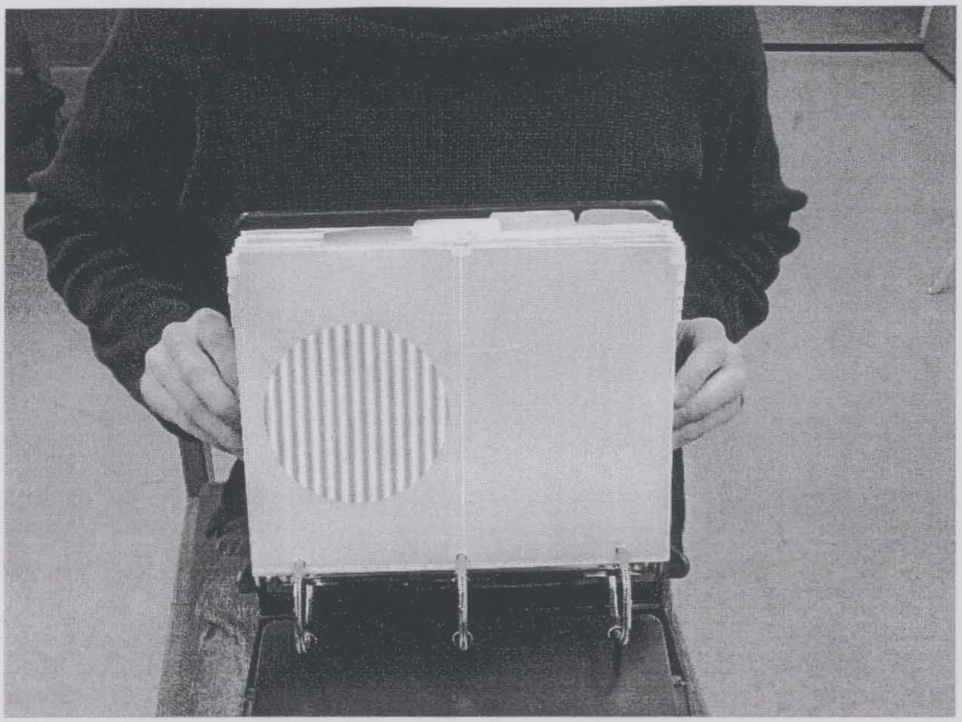

Figure 14. Photograph of a page of the contrast sensitivity screening booklet.

Table 8. Contrast sensitivity values (in contrast sensitivity units) and spatial frequencies of the gratings composing the contrast sensitivity booklet. Note that numbers in parentheses represent percent contrast.

	Page Number									
Spatial Frequency	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32	33-36	37-40
0.75	2.1 (47.6)	4.6 (21.8)	8.7 (11.5)	11.5 (7.9)	20.0 (5)	27.8 (3.6)	37.0 (2.7)	71.4 (1.4)		
1.5	2.1 (47.6)	4.8 (21)	9.6 (10.4)	13.7 (7.3)	20.4 (4.9)	27.0 (3.7)	41.7 (2.4)	83.3 (1.2)	111.1 (0.9)	
3	2.1 (47.6)	4.6 (21.8)	9.3 (10.7)	13.5 (7.4)	18.9 (5.3)	26.3 (3.8)	40.0 (2.5)	52.6 (1.9)	83.3 (1.2)	111.1 (0.9)
6	1.9 (52.6)	3.4 (29.8)	6.4 (15.7)	12.0 (8.3)	20.8 (4.8)	30.3 (3.3)	40.0 (2.5)	71.4 (1.4)		
12	1.9 (52.6)	3.3 (30.3)	4.9 (20.3)	9.6 (10.4)	20.8 (4.8)	40.0 (2.5)	71.4 (1.4)			

As with the CS cards, all gratings were generated using PostScript programming language (Adobe Systems Incorporated, 1986; 1990), and were printed with an Epson Stylus 2200 Photoprinter. However, unlike the CS cards, the gratings were printed onto Epson Heavyweight Matte Photopaper. This paper, also of high quality, was chosen because it is 1.4 times thicker than the Hewlett-Packard photopaper, and is therefore more durable. This is important because the gratings of the CS screening booklet are not mounted and must be directly handled by the tester, thus durability is necessary.

This booklet format was chosen as an alternative to the CS cards for two reasons. First, the foamcore background of the cards was not durable. This was a serious drawback as the cards were sometimes accidentally dropped by the tester, usually resulting in minor damage (bending of the card, especially around the corners). However, with the present format, the gratings are protected within the 3-ring binder and no longer need to be mounted onto foamcore. Second, the CS screening booklet contains test gratings of very low contrast (e.g., 0.9%) to counteract the ceiling effect that was found with the CS cards.

As with the CS cards, a pilot study was conducted with the first 20 children to determine whether the test was time-efficient and whether the referral criteria were adequate. Testing was conducted monocularly at 40 cm following a two-alternative-forced-choice procedure. To begin, the child was presented with a page containing a high contrast grating from 0.75 c/deg SF set and told to touch the grating with a paintbrush to prevent marking of the page. The three remaining pages for that contrast level were then presented and the child followed the same instructions. If the child correctly touched the

grating on 4/4 trials, it was concluded that he/she could detect the grating. If the child could detect the grating on 3/4 trials, the tester then re-presented a grating of the same contrast from one of the three previous pages. If the child could touch this grating, he/she correctly touched the grating on 4/5 trials and thus, it was concluded that he/she could detect the grating. Screening then proceeded with sheets containing gratings of lower contrast. The lowest contrast grating detected was taken as an estimate of contrast threshold. This procedure was continued with the remaining SF sets in order of increasing SF.

Importantly, the pilot study revealed that the CS booklet for the most part, remedied the ceiling effect that was apparent with the CS cards. However, the procedure of presenting gratings of progressively lower contrast within each SF until the child's threshold was reached was time-consuming (mean = 8.0 min). This represents a serious shortcoming in the present study as the test was part of an extensive battery that included tests of four other visual functions. As a result, when the CS booklet was formally added to the screening program, it was utilized following a *pass/fail* procedure. Testing began with the presentation of the four pages containing the highly visible, high contrast "warm-up" grating (48.7%) in the 0.75 c/deg SF set to determine whether the child could understand the procedure.[27] The child was then shown the pages containing the grating that was chosen as the *cut-off contrast level* (see Referral Criteria subsection below), a relatively low contrast grating that was chosen as a pass/fail cut-off point for that SF. If the child could detect this grating following the 4/4 or 4/5 detection criteria, he/she

[27] The high contrast "warm-up" grating was only presented in the 0.75 c/deg SF set.

"passed" the requirements for that SF set. However, if the child could not detect this grating, he/she "failed" the requirements for that SF set and thus, would be later retested or referred. In either case, testing then proceeded with the presentation of the four pages containing the cut-off contrast level grating for the 1.5 c/deg SF set following the same pass/fail procedure. Testing then continued with the progressively higher SF sets following the same procedure used with the 1.5 c/deg set.

Toddlers. CS testing was not generally attempted with toddlers except in rare cases in which the child was very attentive and cooperative. If this was the case, the toddler was tested with the CS cards only after successful completion of a training phase in which he/she consistently pointed to or looked at the test grating of the high contrast warm-up card at 0.75 c/deg. Note, testing was not attempted on any toddler using the CS booklet.

Autorefraction

All toddlers and preschoolers were tested with the Welch-Allyn SureSight hand-held autorefractor (see Figure 15; Welch-Allyn, Skaneateles, N.Y., U.S.A.), a relatively new wave-front based instrument that provides rapid estimates of refractive error. This particular autorefractor was chosen because its testing procedure is far less invasive than that of other portable devices (e.g., the Nikon Retinomax, see Refractive Error subsection in the Introduction). To use the SureSight, the tester placed the device in front of the child's face while looking through an aperture and pointing the autorefractor at the child's pupil. The tester was guided to the 35 cm test distance by the device's audible feedback system. Using infra-red light beamed into and then reflected from the eye, the

device took 5 to 7 rapid measurements of the child's eyes. In a few seconds, the device provided estimates of spherical refractive error (a measure of hyperopia or myopia), cylindrical refractive error (a measure of astigmatism), the axis of astigmatism, and the reliability for the set of measurements. All children were tested without cycloplegia and if possible, completed two measurements of each eye. The average of the two measurements was then taken as an estimate of refractive error.

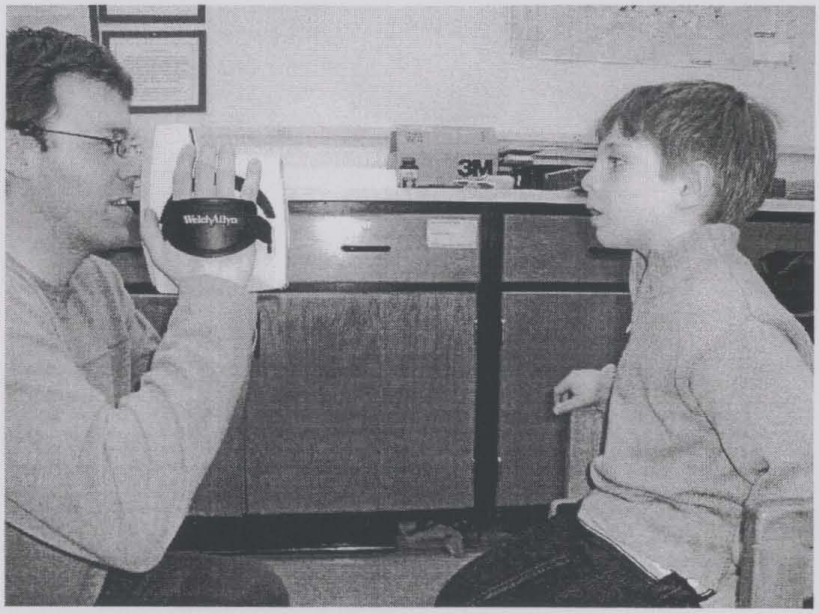

Figure 15. Photograph of a preschooler being tested with the Welch-Allyn SureSight autorefractor.

Counterbalancing

An attempt was made to counterbalance the order of tests throughout the study. However, this attempt was abandoned if a child appeared especially shy or timid, as was often the case with 2- and 3-year-olds. With such children, testing generally began with

autorefraction as it is least invasive (i.e., does not require spectacles) and most rapid of all tests in the battery. Testing then continued in an order following the tester's discretion.

Referral Criteria

The referral criteria for each screening test are listed below. If a child's performance met these criteria on any single test (i.e., he/she failed the test), he/she was retested completely, usually within a week. If the child failed any single test once again, he/she was referred to an optometrist who conducted a full optometric gold standard exam. A concerted effort was made to achieve a balance between strict and lenient referral criteria. Specifically, it was essential that the criteria were strict enough to identify even those with subtle cases of vision disorders (e.g., mild anisometropia, latent strabismus, etc.) so that few of these children would be misdiagnosed as normal. At the same time, it was considered critical that the criteria be lenient enough to avoid overreferrals. This consideration was especially important in the present study as the vast majority of optometric gold standard examinations were conducted by a single optometrist who was part of our research team. Thus, for each visual function below, referral criteria were chosen carefully based on recommendations from pediatric/vision organizations, results of previously successful vision screening studies, and/or developmental norms gathered by leading researchers in the field.

Visual Acuity. The recommended visual acuity referral criteria of several North American pediatric/vision organizations are shown below in Table 9. Although these criteria vary greatly, they are approximately < 20/40 for 3 and 4 year-olds, < 20/30 for 5 year-olds, and an interocular difference of 2 lines for all age groups. Yet, it is difficult to

follow these criteria as most of the organizations do not specify whether they are recommended for linear or single optotypes tests. This distinction is important as different test formats present subjects with varying levels of difficulty due to differences in the crowding of the optotypes. Specifically, the single optotype format (without crowding bars) is generally less demanding than the traditional linear optotype format, resulting in slightly higher acuity scores (Morad, Werker, & Nemet, 1999). Moreover, traditional linear optotype tests may be less demanding than the newer linear optotype tests which follow the MassVAT format (see the Visual Acuity subsection of the Methods section).

The visual acuity referral criteria for the present study are presented in Table 10 below. Although slightly more lenient than suggested by some of the organizations above, they are generally in agreement with The American Academy of Pediatrics, The American Academy of Ophthalmology, The American Optometric Association, and Prevent Blindness America. Importantly, these referral criteria take into account the varying levels of difficulty of the different visual acuity tests. For instance, at all ages, stricter referral criteria were applied for the isolated symbols tests (i.e., Lea isolated symbols book, Patti Pics isolated symbols book, and the Patti Pics cards) than for the Lea linear optotype test. Specifically, children were required to score the equivalent of one line higher[28] on isolated symbols tests than on the Lea linear optotype test in order to pass (i.e., test negative). Similarly, the referral criteria for the two linear optotype tests

[28] Note, the referral criteria of the isolated symbols optotype tests were actually stricter than those of the Patti-Pics linear optotype by two lines as the latter test follows the MassVAT format which is more difficult than traditional linear optotype tests.

used in the present study were different. That is, the referral criteria for the Lea linear optotype test were stricter than those for the Patti Pics linear optotype test by the equivalent of one line at all ages. Table 10 also shows that the present study employed an interocular difference referral criterion of two or more lines as suggested by the majority of vision/pediatric organizations (The American Academy of Pediatrics, The American Academy of Ophthalmology, and Prevent Blindness America).

Table 9. Referral criteria for visual acuity testing as recommended by North American pediatric and vision organizations.

Organization	Age	Referral Criteria for Visual Acuity	Referral Criteria for Interocular Difference
Canadian Pediatric Society	3 - 5 years	< 20/30	Not reported
Canadian Task Force on Periodic Health Examination.	3 - 5 years	< 20/30	Not reported
American Academy of Pediatrics	3 - 4 years	< 20/40	≥ 2 lines
	5 - 6 years	< 20/30	
American Academy of Ophthalmology	3 - 4 years	< 20/50	≥ 2 lines
	5 years	< 20/30	≥ 2 lines
American Optometric Association	< 5 years	≤ 20/50	Not reported
	≥ 5 years	≤ 20/40	Not reported
* Maternal and Child Health Bureau and National Eye Institute Task Force on Vision Screening in the Preschool Child	3 years	< 20/40	Not reported
	4 years	< 20/30	Not reported
Prevent Blindness America	3 years	≤ 20/50	≥ 2 lines
	4 - 6 years	≤ 20/40	≥ 2 lines
Head Start Program	3 - 5 years	≤ 20/50	Not reported
	6 years	≤ 20/40	Not reported

* These criteria are suggested for linear optotype tests or for single optotype test with crowding bars.

Table 10. Visual acuity referral criteria for the present study.

Test	Age	Referral Criteria For Visual Acuity	Refer Criteria for Interocular Difference
Teller Acuity Cards	2	< 6.5 c/deg	≥ 1 octave difference
Patti Pics Cards		< 20/50	≥ 2 line difference
Lea/Patti Pics Isolated Symbols			
Lea Linear Optotype		<20/80	
Patti Pics Linear Optotype		<20/80*	
Lea Linear Optotype	3	< 20/50	≥ 2 line difference
Patti Pics Linear Optotype		< 20/64	≥ 2 line difference
Patti Pics Cards		< 20/40	≥ 2 line difference
Lea/Patti Pics Isolated Symbols			
Lea Linear Optotype	4	< 20/40	≥ 2 line difference
Patti Pics Linear Optotype		< 20/50	≥ 2 line difference
Patti Pics Cards		< 20/32	≥ 2 line difference
Lea/Patti Pics Isolated Symbols			
Lea Linear Optotype	5	< 20/32	≥ 2 line difference
Patti Pics Linear Optotype		< 20/40	≥ 2 line difference
Patti Pics Cards		< 20/25	≥ 2 line difference
Lea/Patti Pics Isolated Symbols			
Lea Linear Optotype	6 +	< 20/25	≥ 2 line difference
Patti Pics Linear Optotype		< 20/32	≥ 2 line difference
Patti Pics Cards		< 20/20	≥ 2 line difference
Lea/Patti Pics Isolated Symbols			

* Note that this criterion was used for the Patti Pics chart as it is the lowest possible score. If a 2-year-old scored less than 20/80, he/she was immediately tested with the Patti Pics isolated symbols book, or Patti Pics cards.

Unfortunately, it is difficult to set grating acuity referral criteria for 2-year-olds as no pediatric/vision organization has outlined acuity card referral criteria for this age group. However, several studies have collected normative monocular data which reveal a mean grating acuity that ranges from 9.6 to 20.9 c/deg in 2-year-olds (Kohl & Samek,

1988, Mayer, Beiser, Warner, Pratt, Raye, & Lange, 1995; McDonald, Ankrum, Preston, & Sebris, 1986; Salamão, & Ventura, 1995). Based on these studies, 90% tolerance limits were estimated following the assumption that children falling below the lower limit (i.e., the lowest 5% of the population) likely possess a subnormal acuity and should be referred. This lower tolerance limit is approximately 6.5 c/deg and thus, a referral criterion of < 6.5 c/deg (< 20/90 in Snellen notation) was chosen for the present study. An interocular difference of at least 1 octave was also chosen as the referral criteria for imbalance between the eyes. A similar criterion of < 20/80 was chosen for children who completed the isolated symbols tests (i.e., Lea isolated symbols book, Patti Pics isolated symbols book, and the Patti Pics two-alternative-forced-choice isolated symbols).

Ocular Alignment/Motility. The referral criteria for the tests of ocular alignment/motility are presented in Table 11 below. These criteria are widely accepted by both pediatric and vision organizations (American Academy of Ophthalmology, American Academy of Pediatrics, American Academy of Pediatric Ophthalmology and Strabismus, American Optometric Association, Committee on Practice and Ambulatory Medicine, Head Start Program, and Maternal and Child Health Bureau) and researchers/clinicians (Cashell & Durran, 1980; Eibschitz-Tsimhoni et al., 2000; Olitsky & Nelson, 1998; Robinson et al., 1999; von Noorden, 1990; Wormald, 1991), and are considered adequate for detecting disorders of alignment and ocular motility, notably strabismus, nerve palsy, Duane's Syndrome, and nystagmus.

Table 11. Referral criteria for tests of ocular alignment/motility.

Test	Referral Criteria
Cover-Uncover Test	Any abnormal eye movement
Hirschberg Corneal Reflex	Any asymmetry in corneal reflex
Ocular Motility	Any detectable anomaly such as vertical/horizontal deviation; nystagmus
Convergence/Divergence	Convergence/Divergence insufficiency; Ocular misalignment

Stereoacuity. The recommendations of North American pediatric/vision organizations regarding stereoacuity testing are presented in Table 12. Because several organizations recommend stereoacuity testing, and a number of studies have included the measurement of stereoacuity as part of a screening program, it is considered important for the assessment of binocular function. However, these organizations and screening programs typically recommend or implement referral criteria of 600 – 1980 arc sec (see Chui et al., 2004; Newman & East, 1999; Shallo-Hoffman et al., 2004). Given that normal stereoacuity levels of preschoolers may be as low as 40 – 60 arc sec (Birch et al., 1997), these criteria may be considered extremely lenient and render the test unlikely to detect subtle cases of binocular dysfunction. This notion is supported by findings that some amblyopes actually possess coarse stereopsis (Holopigian, Blake, & Greenwald, 1986; Simons, 1996; Wood, Fox, & Stephenson, 1978; Wood & Stephenson, 1981) and thus, could be incorrectly diagnosed as normal under the above criteria.

Table 12. Recommended stereoacuity referral criteria of North American pediatric/vision organizations.

Organization	Age	Test	Referral Criteria
American Academy of Pediatrics American Academy of Ophthalmology American Association of Pediatric Ophthalmology and Strabismus Maternal and Child Health Bureau	3 – 5	Randot E	$\geq$ 630 arc sec
Prevent Blindness America U.S. Public Health Service	3 – 4	Randot E	Not Reported

The referral criteria of the present study are provided in Table 13. Note, slightly different referral criteria were implemented for the Randot E Stereotest and the Randot Preschool Stereoacuity Test simply because the two procedures yield different stereoacuity scores.[29] Yet for each group, the referral criteria chosen for the Randot Preschool Stereoacuity Test were the values closest to those chosen for the Randot E Stereotest.[30] These criteria were considered strict enough so that the test could detect subtle cases of binocular dysfunction such as mild strabismus or anisometropia. At the same time, the criteria were considered lenient enough to prevent false positives, especially when testing children in less than optimal conditions (e.g., in dim lighting). These criteria are validated by other studies that employ similar criteria and report high sensitivity and moderate to high specificity (Hope & Maslin, 1990; Manny, Martinez, &

[29] The RDE yields scores of 500, 250, and 168 arc sec whereas the Randot Preschool Stereoacuity Test yields scores of 800, 400, 200, 100, 80, and 40 arc sec.

[30] Note, the referral criteria of the RDE were chosen earlier in the program as it was initially the preferred test to measure stereoacuity.

Fern, 1991; Molgaard, Biering-Sorenson, Michelson, Elmer, & Rydberg, 1984; Reinecke & Simons, 1974, Walraven, 1975; Walraven & Janzen; 1993).

Table 13. Stereoacuity referral criteria for the present study.

Age	Test	Referral Criteria
2 and 3	Randot E	> 500 arc sec
	Randot Preschool Stereoacuity Test	> 400 arc sec
4 and 5	Randot E	> 250 arc sec
	Randot Preschool Stereoacuity Test	> 200 arc sec
6 +	Randot E	> 168 arc sec
	Randot Preschool Stereoacuity Test	> 100 arc sec

Contrast Sensitivity. Despite its promise as a potential screening tool, no vision screening study has measured CS in preschoolers, nor has any North American vision/pediatric organization recommended measurement of CS as part of a vision screening program. Thus, it was necessary to base the referral criteria for the CS cards on monocular data gathered from previous studies (Adams & Courage, 1996; Scharre, Cotter, Stein-Block, & Kelly, 1990; Richman & Lyons, 1994). However, the establishment of these criteria was not an easy task for two reasons. First, a great deal of variation exists among the stimulus parameters of these previous studies (e.g., grating size, average luminance level, contrast levels, etc.) which may ultimately lead to large differences on scores of visual functioning (Banks, Geisler, & Bennett, 1987; Drover et al., 2002; Haegerstrom-Portnoy, Brabyn, Schneck, & Jampolsky, 1997; Rovamo, Mustonen, & Näsänen, 1994; Sheedy, Bailey, & Raasch, 1984; Sturr, Kline, & Taub,

1990; Waugh & Levi, 1993). Second, very few monocular data exist for preschoolers, particularly for 2- and 3-year-olds. Nevertheless, referral criteria were established and then evaluated as the cards were included in the screening program as a part of a pilot study to measure CS in 24 children. However, it became apparent that these criteria were too lenient and therefore, would be inadequate for detecting subtle cases of vision disorders. As a result, stricter criteria were then established based upon this pilot study and these are presented in Table 14 below. In order to pass the test (i.e., test negative), the child was required to detect at least the CS of the grating given below for each SF. If he/she could not detect this grating for one or more of the SFs tested, he/she was retested/referred.

Table 14. Referral criteria in CS units for each spatial frequency and each age group for the CS cards. Each referral card's number in that particular spatial frequency set is provided in parenthesis.

Age	**Spatial Frequency (c/deg)**				
	0.75	1.5	3	6	12
2 years	20.8 (card 4)	27.8 (card 5)	16.9 (card 3)	3.2 (card 2)	1.8 (card 1)
3 years	20.8 (card 4)	27.8 (card 5)	38.5 (card 4)	11.4 (card 3)	3.2 (card 2)
4 years	20.8 (card 4)	27.8 (card 5)	38.5 (card 4)	38.5 (card 4)	4.4 (card 3)
5 year & Older	20.8 (card 4)	27.8 (card 5)	38.5 (card 4)	38.5 (card 4)	4.4 (card 3)

A stricter set of referral criteria were established for the CS booklet, specifically for the low to mid SFs (0.75, 1.5, 3.0 c/deg). This measure was taken as it became apparent that there was a ceiling effect at these SFs when children were assessed with the CS cards. These criteria are provided in Table 15 below. As with the CS cards, the child was required to detect the grating of the CS value provided for each SF, or he/she would be retested/referred.

Table 15. Referral criteria in CS units for each spatial frequency and each age group for the CS booklet. Note that each referral sheet's number in that particular spatial frequency set is provided in parenthesis.

Age	Spatial Frequency (c/deg)				
	0.75	1.5	3	6	12
2 years	27.8 (pages 21-24)	41.7 (pages 25-28)	52.6 (pages 29-32)	6.4 (pages 9-12)	3.3 (pages 5-8)
3 years	37 (pages 25-28)	83.3 (pages 29-32)	83.3 (pages 33-36)	20.8 (pages 17-20)	3.3 (pages 5-8)
4 years	71.4 (pages 29-32)	111.1 (pages 33-36)	111.1 (pages 37-40)	40 (pages 25-28)	4.9 (pages 9-12)
5 year & Older	71.4 (pages 29-32)	111.1 (pages 33-36)	111.1 (pages 37-40)	40 (pages 25-28)	4.9 (pages 9-12)

Autorefraction. It was difficult to set autorefraction referral criteria for the present study because the technology is relatively new. As a result, pediatric/vision associations have yet to make recommendations for referral criteria within autorefraction

screening programs. This is further complicated by the fact that only a few autorefraction screening studies have been conducted to date, and none have been judged effective (see Table 4). Furthermore, the majority of these studies have used the Nikon Retinomax, an autorefractor different from the Welch-Allyn SureSight used in the present study. Nevertheless, referral criteria were chosen based on data collected by Courage, Drover, Vernescu, Keough, & Adams (2001). These criteria are presented in Table 16.

Table 16. Autorefraction referral criteria of the present study.

Disorder	Referral Criteria
Hyperopia	> 3.50 D
Myopia	> 1.00 D
Astigmatism	> 1.50 D
Anisometropia	> 1.75 D

Optometric Gold Standard Exam

As mentioned previously, if a child tested positive on any test based on the criteria provided above, he/she underwent a second, complete screening (i.e., a retest). If the child again tested positive on any single test, he/she was referred to a pediatric optometrist who was part of our research team for the optometric gold standard exam which included assessment of ocular health, alignment, motility, visual acuity, stereoacuity, and refractive error (see Table 17). This exam was conducted to enable

children with potential disorders to receive treatment and to allow the tabulation of the number of true and false positives. Table 18 provides clinical disease/disorder criteria for a positive diagnosis based on this exam. Importantly, these criteria are essentially the same as those of the VIP Study Group (2004) and are generally in agreement with those implemented by other researchers/clinicians who have developed vision screening programs worldwide (Chui et al., 2004; DeBecker et al., 1992; Freedman & Preston, 1992; Granet et al., 1999; Kennedy & Thomas, 2000; Köhler & Stigmar, 1973; Kvarnström et al., 1998; Miller et al., 2003; Morgan & Johnson, 1987; Ottar et al., 1995; Shallo-Hoffmann et al., 2004) and thus, were considered appropriate for the present study. If a child met these criteria, he/she was deemed to possess abnormal vision and therefore, classified as a true positive. However, if a child did not meet these criteria, he/she was deemed to possess normal vision and was classified as a false positive. It should also be noted that parents were also free to bring their child to see an optometrist/ophthalmologist that was not part of our research team. In such a case, it was not possible to obtain the child's medical records. Thus, a child was classified as true positive if he/she received treatment including corrective lenses, patching, orthoptic exercises, etc. If a child did not receive treatment, he/she was classified as a false positive.

Table 17. Visual functions and tests of the gold standard exam.

Visual Function	Test
Ocular Alignment	Near and Distant Cover-Uncover Tests
Ocular Motility	Broad H Test
Ocular Health	Anterior Segment Exam
Visual Acuity	Topcon Symbols
Stereoacuity	Randot Animals
Refractive Error*	Topcon Table Top Autorefractor/ Cycloplegic Retinoscopy

* Subjects were initially assessed with the autorefractor and then assessed with cycloplegic retinoscopy if a problem was suspected.

In order to ensure that children who tested negative on all tests of the battery typically possess normal vision, and to allow the tabulation of true and false negatives, a sample of children (n = 145) who passed all tests were invited to visit the optometrist to undergo the same gold standard exam for free. If a child scored within the range of the disease/disorder criteria listed above, he/she was judged to possess abnormal vision and was classified as a false negative. If a child scored outside this range, he/she was judged to possess normal vision and was classified as a true negative. Based on the tabulation of true positives, false positives, true negatives, and false negatives, sensitivity, specificity, PPV, and NPV, were calculated, and the validity of the screening program was evaluated (see Results section below). It is important to note that the vast majority of the children invited to receive the free gold standard exam completed all age-appropriate tests. This measure was taken to allow validity calculations for each test. Also, it should be

mentioned that the parents of *all* children who received the gold standard exam (i.e., those who screened positive or negative) were instructed not to inform the optometrist of their child's screening result to ensure that the optometrist was masked to each child's screening status.

Table 18. Disease/disorder criteria for the optometric gold standard exam.

Deficit/Disorder*	Definition
Amblyopia	
Unilateral	> 1 line difference in acuity and presenting a unilateral amblyogenic factor (e.g., strabismus)
Bilateral	
2- and 3-year-olds	< 6/18 in one eye, < 6/12 in the contralateral eye, and presenting bilateral amblyogenic factor (e.g., significant refractive error)
4-year-olds and older	< 6/12 in one eye, < 6/9 in the contralateral eye, and a presenting a bilateral amblyogenic factor
Reduced Visual Acuity	
Bilateral	
3-year-olds	< 6/18 in one eye; < 6/12 in the contralateral eye; no bilateral amblyogenic factor.
4-year-olds and older	< 6/12 in one eye; < 6/9 in the contralateral eye; no bilateral amblyogenic factor.
Unilateral	
2- and 3-year-olds	< 6/18 in one eye; > 1 line difference; no unilateral amblyogenic factor
4-year-olds and up	< 6/12 in one eye; > 1 line difference; no unilateral amblyogenic factor
Reduced Stereoacuity	
2- and 3-year-olds	< 400 arc sec
4-year-olds and older	< 200 arc sec
Strabismus	Any tropia
Significant Refractive Error	
Hyperopia (2 – 5 years)	> 3.0 D (sphere)
Hyperopia (6 years +)	> 2.0 D (sphere)
Myopia	< -1.0 D (sphere)
Astigmatism	> 1.5 D (cylinder)
Anisometropia	> 1.5 D (sphere and/or cylinder)

Results

Study Population

Overall, consent was received from parents of 972 children. Of these children, 18 (1.9%) refused to participate and thus, are classified as *refusals*. Specifically, they would not enter the area of the daycare in which the testing was conducted. Because these children did not attempt a single test, they are not included in any analyses, including completion rates and completion times (see Completion Times and Completion Rates subsection below). An additional 21 children with known visual disorders participated in the program and are classified as *confirmations*. These children are included in calculation of completion rates and completion times (see below) but not in analyses of validity (see Progressive Validity Analyses subsection below).[31]

Testing was attempted on a total of 954 children (526 males, 428 females) who are categorized by age in Figure 16 below. The children ranged in age from 1.6 to 11.6 years with a mean age of 4.2 years (SD = 1.1 years). As the Figure shows, the majority of children who attempted testing (75%) were 4 years of age and older. Also, the Figure reveals that relatively few toddlers were tested, a finding that is expected due to the low toddler enrollment in daycare programs in the St. John's metropolitan area. Note that children aged 6 years and older (i.e., 5.5 years and older) are represented in the Figure. However, these children were included in the 5 year-old age group for all subsequent analyses.

---

[31] Note, confirmations were tested while wearing optical correction. If they screened positive they were referred. Based on whether additional treatment was required, they were classified as true or false positives and included in the validity analyses.

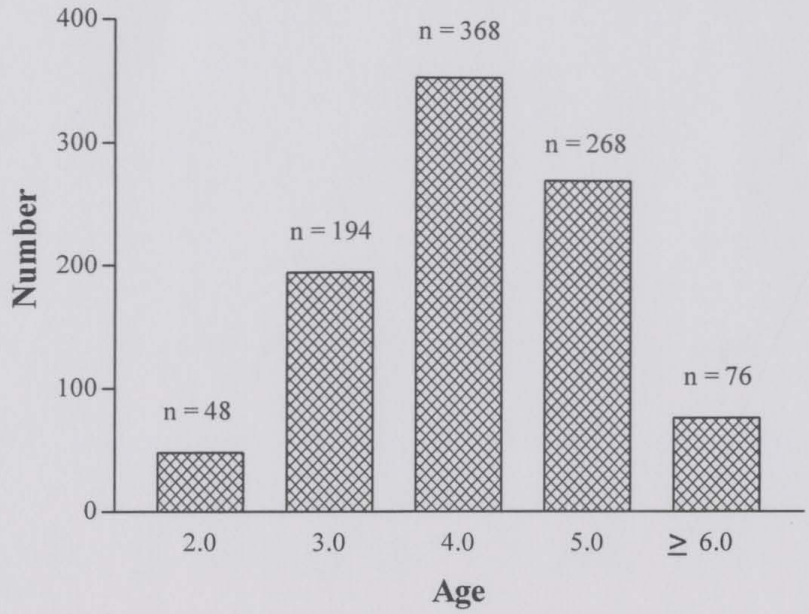

Figure 16. Number of children attempting the screening program classified by age group.

Inter-Tester Agreement

As noted above, children were tested by either the author (JRD) or by a research assistant who underwent a rigorous training procedure prior to testing independently (see Method section). Specifically, 20 children who were considered especially cooperative were tested by both the author *and* the research assistant. The inter-tester agreement of these data is calculated and is presented in Table 19 below. Whenever possible, agreement of pass/fail decisions was investigated by calculating Cohen's kappa for each visual function (see exceptions at the bottom of the Table). Furthermore, Pearson r correlations were calculated for each visual function (see exceptions at the bottom of the Table) between the data obtained by the author and the research assistant.

Table 19. Results of inter-tester agreement analysis for all visual functions.

Visual Function Test/Refractive Error/Spatial Frequency	N	Mean Age	Intertester Agreement (Cohen's Kappa)	Correlation Coefficient (Pearson r)	Number Referred	Percentage Of Cases Agreed Upon
Visual Acuity						
Patti Pics Chart	20	4.5	NA†	.78*	0	95
Ocular Alignment‡	20	4.6	1.00*	NA§	4	100
Stereoacuity						
Randot Preschool Stereoacuity	20	4.6	0.77*	0.85*	1	95
Autorefraction						
Sphere	20	4.1	0.77*	0.84*	2	95
Cylinder	20	4.1	1.00*	0.97*	2	100
Contrast Sensitivity						
0.75 c/deg	20	4.8	NA†	NA†	0	100
1.5 c/deg	20	4.8	NA†	NA†	0	100
3.0 c/deg	20	4.8	NA†	NA†	0	100
6.0 c/deg	20	4.8	1.00*	0.69*	0	100
12 c/deg	20	4.8	1.00*	1.00*	1	100

† Note, the author judged that none of 20 children tested should be referred on the basis of the test in question. Thus, there was no variation in the author's and/or researcher assistant's decisions (i.e., all children passed), a finding which precludes the calculation of kappa.

* $p < .01$.

‡ The results of the ocular alignment tests are combined. Therefore, the battery of alignment/motility tests (i.e., cover tests, Hirschberg corneal reflex, ocular motility, and convergence/divergence) are treated as a single test.

§ Pearson's r could not be calculated for ocular alignment/motility tests as the data are binomial (i.e., pass/fail).

The results reveal that there was strong agreement between the pass/fail decisions of the author and research assistant for each visual function. In fact, out of 100 decisions (i.e., 20 per each visual function), the testers agreed on 97%. Furthermore, for ocular alignment/motility, autorefraction (cylinder), and contrast sensitivity (at 6 and 12 c/deg) there is perfect agreement between the testers. Correlational analyses also reveal strong agreement between scores obtained when children were tested by the author and by the research assistant. Note, however, although a correlation coefficient could not be calculated for contrast sensitivity (CS) scores at 0.75, 1.5, and 3 c/deg due to a ceiling effect and subsequent lack of variation within the CS scores, there is perfect agreement between the two testers.

In spite of these positive results that indicate strong agreement between the two testers, four important points must be mentioned. First, due to their limited numbers and shyness, only one toddler was tested twice and included in the analysis.[32] As a result, with the exception of autorefraction, the average age of children taking part in this analysis is slightly higher than that of children participating in the study overall. Second, inter-tester agreement of CS was analyzed for the CS cards only. The CS booklet was not included because it was developed and implemented into the program following the analysis. Third, inter-tester agreement was calculated only for the preferred tests for visual acuity (i.e., Patti Pics chart) and stereoacuity (Randot Preschool Stereoacuity Test). Yet, the exclusion of alternative tests (e.g., Patti Pics symbols) is considered acceptable as the preferred tests were usually the most difficult to administer.

---

[32] This toddler completed two tests of visual acuity.

In other words, given that the testers showed strong agreement on the more challenging, preferred tests, it is likely they would also show strong agreement on the easier to administer, alternative tests (e.g., Patti Pics isolated symbols book and Randot E Stereotest). Fourth and most importantly, in approximately half of the cases in which children were tested twice, the research assistant was present while the child was tested by the author and thus, was aware of the child's screening result. This measure was justified as the primary purpose of this phase of the experiment was to train the research assistant. Therefore, her presence was considered necessary as the experimenter often provided important instructions as he was testing.

As a further measure to ensure consistency, the first 30 children tested by the research assistant after the training phase were evaluated under the close supervision of the author. Following this supervision period, the research assistant was considered sufficiently trained to test children independently. Note that overall, 95% of the children (N = 906) in the final sample were tested by the author, whereas the remaining 5% (N = 48) were tested by the research assistant.

Completion Times and Completion Rates

In all, 946 (99.2%) children who participated in the screening program completed at least one test. The remaining 8 children agreed to participate but were unable to complete a single test, usually because they became fussy or frightened shortly after testing began. The overall completion rate of each test and the entire screening battery is provided in Table 20 below. Importantly, these completion rates *do not* reflect the true testability of children with these procedures for two reasons. First, the testers were

primarily concerned with assessing children on as many visual functions as possible. Thus, if it was anticipated that the child would have difficulty completing the preferred test, that test was not attempted in order to avoid distressing the child, and causing him/her to refuse further testing. Instead, the tester would attempt an easier to complete alternative, or avoid testing that aspect of functional vision altogether. Second, those children attempting the alternative tests (e.g., Lea and Patti Pics books, Teller Acuity Cards, etc) represent a somewhat biased sample, namely children who are extremely shy, distressed, or possess disorders/deficits that make them difficult to test (e.g., autism, Down's Syndrome).

Despite these caveats, the Table reveals a number of important findings. For instance, the percentage of children able to complete most tests increased with age. Two important exceptions are autorefraction and the Hirschberg corneal reflex, both of which yielded high completion rates for all age groups. In fact, completion rates for autorefraction were the highest of all tests as at least one estimate of refractive error was obtained in each eye for 98% of all children attempted. Conversely, the overall completion rates for the cover tests (81% and 77% for distant and near cover tests, respectively), contrast sensitivity (81% and 63% for the CS cards and CS booklet, respectively), and Randot Preschool Stereoacuity (73%) were quite low, likely due to the attentional demands of these procedures. Finally, the Table reveals that the completion rate for the entire battery of tests was 78%, and for the most part, it increased with age. Note however, completion rates actually decreased from 2 to 3 years of age, a result likely explained by the limited number of tests required for 2-year-olds to complete the

Table 20. The completion rate and percentage (in bold) for each test implemented in the screening program categorized by age group.

	Age				
Function/Test	**2**	**3**	**4**	**5**	**Total**
Visual Acuity					
Lea Chart	3/23 **13**	73/133 **55**	235/248 **95**	232/239 **97**	543/643 **84**
Lea Book	6/20 **30**	24/60 **40**	7/13 **54**	2/7 **29**	39/100 **39**
Teller Acuity Cards	10/14 **71**	15/36 **42**	2/6 **33**	1/5 **20**	28/61 **46**
Patti Pics Chart	2/25 **8**	30/61 **49**	103/120 **86**	99/105 **94**	234/311 **75**
Patti Pics Book	4/23 **17**	6/31 **19**	2/17 **12**	1/6 **17**	13/77 **17**
Patti Pics Cards	7/19 **37**	9/25 **36**	0/15 **0**	0/5 **0**	16/64 **25**
Total Visual Acuity	32/48 **67**	157/194 **81**	349/368 **95**	335/344 **97**	873/954 **92**
Alignment/Motility					
Distant Cover	7/48 **15**	109/194 **56**	328/368 **89**	330/344 **96**	774/954 **81**
Near Cover	1/25 **4**	29/63 **46**	118/134 **88**	116/122 **95**	264/344 **77**
Hirschberg	46/48 **96**	173/194 **89**	355/368 **97**	335/344 **97**	909/954 **95**
Ocular Motility	9/48 **19**	130/194 **67**	351/368 **95**	331/344 **96**	821/954 **86**
Convergence/ Divergence	9/48 **19**	131/194 **68**	349/368 **95**	332/344 **97**	821/954 **86**

Total Alignment/ Motility	7/48 **15**	106/194 **55**	328/368 **89**	329/344 **96**	770/954 **81**
Stereoacuity					
Randot E	6/40 **15**	103/154 **67**	198/212 **93**	167/176 **95**	474/582 **81**
Randot Preschool Stereoacuity	1/28 **4**	33/92 **36**	155/196 **79**	168/181 **93**	357/497 **73**
Total Stereoacuity	*9/48 **19**	136/194 **70**	353/368 **96**	335/344 **97**	833/954 **87**
Refractive Error					
Autorefractor	45/48 **94**	181/194 **93**	352/368 **96**	332/344 **97**	910/954 **95**
†Autorefractor at least 1 measure	47/48 **98**	187/194 **96**	359/368 **98**	341/344 **99**	934/954 **98**
Contrast Sensitivity					
CS Cards	2/22 **7**	34/72 **44**	144/163 **88**	158/162 **98**	338/419 **81**
CS Booklet	0/8 **0**	7/19 **37**	26/33 **79**	17/19 **89**	50/79 **63**
Total Contrast Sensitivity	2/30 **6**	41/91 **45**	170/196 **87**	175/181 **97**	388/498 **78**
Entire Battery	‡29/48 **58**	90/194 **46**	304/368 **83**	320/344 **93**	743/954 **78**

* Includes two toddlers who completed testing with the Randot Preschool Stereoacuity Test

† Represents the number and percentage of children who completed at least one measure of autorefraction with each eye.

‡ Assessment of a 2-year-old was considered complete if he/she finished any visual acuity test, the Hirschberg corneal reflex, and autorefraction.

assessment. Specifically, a 2-year-old was judged to have completed the battery if he/she completed one visual acuity test, the Hirschberg corneal reflex, and autorefraction. On the other hand, 3-year-olds (and all older children) were required to complete one visual acuity test, all ocular alignment/motility tests, one stereoacuity test, autorefraction, and one CS test.

The mean time required to complete each test implemented in the screening program is provided in Table 21 below. Note, as with completion rates above, those children who completed the alternative tests represent a biased sample. As a result, mean completion times for several alternative visual acuity tests were much higher than those for the preferred tests. Overall, approximately 10.6 to 13.6 minutes was required to complete the entire program. Note, however, that the mean completion time for 2-year-olds is much faster at 5.3 minutes as fewer tests were required to complete the battery for this age group. As shown in Table 21, mean completion time is lowest for autorefraction (1.4 min), followed by the tests of stereoacuity (1.6 to 2.0 min). The relatively short time and limited attention required for autorefraction likely explains its high completion rates shown in Table 20 above. On the other hand, CS had the longest mean completion time of any visual function (3.3 min and 5.6 min for the CS cards and CS booklet, respectively), which may in part explain its low completion rate (see Table 20 above). Note that for most visual functions, completion times decreased with age. The most notable exception is visual acuity, in which completion times for all tests actually increased slightly from ages 2 to 3, likely because 3-year-olds possess higher acuities than 2-year-olds, and thus, are presented with more optotypes during testing.

Table 21. Mean completion time (minutes) for each test in the screening program categorized by age group. Standard deviations are provided in parentheses. Note that in cases in which no standard deviation is provided, only a single child was screened with the test.

Visual Function Test	Age 2	3	4	5	Total
Visual Acuity					
Lea Chart	2.6 (0.4)	3.6 (1.4)	2.8 (0.9)	2.4 (0.7)	**2.8 (1.0)**
Lea Book	4.2 (1.7)	4.8 (2.1)	4.6 (1.4)	2.4	**4.5 (1.8)**
Teller Acuity Cards	3.2 (0.8)	3.7 (1.7)	7.3		**3.8 (1.7)**
Patti Pics Chart	2.5 (0.4)	3.1 (1.2)	2.5 (1.1)	2.2 (0.6)	**2.4 (1.0)**
Patti Pics Book	2.5 (0.9)	2.7 (0.1)	4.8 (1.7)		**3.2 (1.3)**
Patti Pics Cards	3.8 (1.8)	4.1 (1.8)	3.0		**3.9 (1.7)**
Total Visual Acuity	3.3 (1.3)	3.6 (1.5)	2.8 (1.0)	2.3 (0.7)	**2.7 (1.1)**
Alignment/Motility	1.7 (0.1)	1.7 (0.2)	1.7 (0.3)	1.6 (0.3)	**1.6 (0.3)**
Alignment/Motility including near cover test*	2.4	2.4 (0.7)	2.4 (0.6)	2.2 (0.6)	**2.3 (0.7)**
Stereacuity					
Randot E	2.3 (1.5)	1.7 (0.5)	1.6 (0.3)	1.6 (0.4)	**1.6 (0.4)**
Randot Preschool Stereoacuity	3.1	2.1 (0.8)	2.1 (0.8)	1.9 (0.6)	**2.0 (0.8)**

Total Stereoacuity	2.4 (1.5)	1.9 (0.6)	1.8 (0.7)	1.8 (0.5)	**1.8 (0.6)**
Refractive Error					
Autorefractor	1.5 (0.7)	1.6 (1.0)	1.4 (0.8)	1.2 (0.6)	**1.4 (0.8)**
Contrast Sensitivity					
CS Cards	3.9	3.8 (1.0)	3.8 (1.0)	3.1 (0.9)	**3.3 (0.9)**
CS Booklet			6.4 (0.9)	6.0 (1.9)	**5.6 (2.0)**
Total Contrast Sensitivity	3.9	4.1 (1.6)	3.5 (1.1)	3.1 (0.9)	**3.4 (1.2)**
Entire Battery†	5.3	13.6	11.9	10.6	

* All tests of ocular alignment are combined for completion time calculations.
† Completion time for the entire battery was calculated by adding mean completion time across tests of each visual function for each age group. Two-year-olds are the lone exception. To calculate completion time, mean completion times were added for tests of visual acuity and autorefraction, and 30 seconds were added as the estimated time required to complete the Hirschberg test.

Retests, Referrals, Yield, and Prevalence of Visual Deficits

The number and percentage of children who required a retest and those who were referred for the gold standard exam are presented in Table 22 below. The Table indicates that 152 children required a retest. A total of 100 children failed the retest and were referred whereas an additional 20 children were referred following the initial screening because a retest was not possible (the child no longer attended daycare, was not at the daycare during all subsequent visits, was too timid to be retested, already had an appointment to see an ophthalmologist/optometrist, etc). Therefore, a total of 120

children were referred for the gold standard exam. To date however, only 53 of these children have completed the gold standard exam with the team optometrist, whereas 23 have completed a gold standard exam with other optometrists/ophthalmologists in the St. John's metropolitan area. In addition, 34 children who screened negative for vision disorders (i.e., passed all screening tests) received the gold standard exam with the team optometrist.

Table 22. Number and percentage of children retested and referred for the gold standard exam.

Category	Number	Percentage
Required a retest	152	16.1
Referred	120	12.7
Attended appointment with team optometrist	53	5.6
Attended appointment with another optometrist/ophthalmologist	23	2.4
Negatives who received gold standard exam	34	3.6
Total who received gold standard exam	110	11.6

The screening results of children referred for a gold standard exam are broken down by visual function and age group in Table 23. According to the Table, children were most likely to be referred due to failure of visual acuity or stereoacuity tests, or autorefraction. Interestingly, children were less likely to fail the alignment/motility or contrast sensitivity tests. However, it should be mentioned that the low number of failures on contrast sensitivity was expected as this test was added approximately half

way through the study. The Table also indicates that a greater number of 4- and 5-year-olds failed screening tests compared to 2- and 3-year-olds. This is predictable as 75% of the children tested were 4 years of age or older.

Table 23. Data from children referred for gold standard exams (N = 120) categorized by age group and visual function failed. Whenever possible, data are provided for retest results, not the original screening. Note that total N does not add up to 120 as many children failed tests of more than one visual function.

Age Groups	Test Failed				
	Visual Acuity	Alignment/ Motility	Stereoacuity*	Autorefraction	Contrast Sensitivity*†
2 years	4	1	1	4	0
3 years	8	3	7	15	1
4 years	30	14	26	28	13
5 years and older	29	16	20	23	12

* Number of 2-year-olds failing stereoacuity and contrast sensitivity tests is relatively low as few toddlers were assessed with these tests.
† Number of 2- to 5-year-olds failing contrast sensitivity tests is relatively low because these tests were not added until approximately mid-way through the study.

Thus far, the yield of the study, i.e., the number of previously undetected cases of visual deficits/disorders identified by screening and confirmed by the gold standard exam, is 58. The projected yield of the present study, including children who have been referred but have not yet received the gold standard exam, is 92 (9.7% of all children tested). The yield is broken down into categories of deficits/disorders following the criteria outlined in Table 18 and now presented in Table 24 below. Based on both the projected yield of the present study, and cases of confirmation, the projected prevalence

rates of several categories of deficits/disorders have been estimated and are also presented in the Table. Note, the most commonly detected disorders were significant refractive errors, the vast majority of which were cases of hyperopia or astigmatism. In contrast, cases of myopia, reduced visual acuity, and reduced stereoacuity (in the absence of amblyogenic factors) were rare and have low prevalence rates in the program. Finally, the Table reveals that 22 children had suspected amblyopia and the prevalence rate was estimated to be 4.3%.

Table 24. Frequency and estimated prevalence of the present study categorized by type of vision disorder.

Vision Disorder	Number of Cases Detected*	Estimated Prevalence (percent)
Reduced Visual Acuity	3	0.6%
Strabismus or Motility/ Fixation Disorders	12	3.1%
Reduced Stereoacuity	1	0.2%
Significant Refractive Error	40	7.1%
Astigmatism	17	3.1%
Hyperopia	23	4.2%
Myopia	3	1.0%
Anisometropia	7	1.4%
Suspected Amblyopia	22	4.3%
Overall Disorders	58	†14.4%
Undetected Disorders	58	‡12.5%

* Values do not add up to overall disorders as several children were diagnosed with more than one disorder.
† This estimate accounts for potential false negatives and includes confirmations.
‡ Excludes cases of confirmation.

Progressive Validity Analyses

One of the primary goals of the present study was to design an effective preschool vision screening program, i.e., one that correctly identified both children who possessed visual dysfunction and those who possessed normal vision. Thus, it was essential to determine which test or combination of tests provided the most effective screening of toddlers and preschoolers. Therefore, for each test[33] and numerous combinations, the four measures of validity described above were calculated as indices of effectiveness (see Evaluation of Vision Screening subsection of the Introduction): (1) sensitivity; (2) specificity; (3) PPV; (4) NPV. Estimates on these measures of validity were then compared to the same pre-set, cut-off values (see the Validity of Vision Screening Programs subsection) that were implemented as cut-off criteria for effectiveness in the literature review above. In order to be considered effective, the test or combination was required to reach the cut-off criteria for *all* measures of validity. These criteria were chosen based on the goal of any screening program, i.e., to detect most cases of a deficit while at the same time, limiting the number of overreferrals (i.e., false positives) and missed cases (i.e., false negatives). The cut-off criteria for both sensitivity and specificity were set at 90% to ensure that the screening program correctly identified 9/10 children who possessed vision disorders (i.e, sensitivity), and 9/10 children who possessed normal vision (i.e., specificity). PPV was set at 65% so that at least 2/3 of those who screened positive actually possessed a disorder (i.e., they were true positives), whereas 1/3 or less were overreferrals. Finally, NPV was set at 95% to guarantee that only a maximum of

[33] In this case, and in the remainder of the thesis, "each test", "a single test", or "test" refers to the one or more tests that were used to assess a single visual function. This is done for the sake of simplicity.

5% of children who passed the screening program possessed a disorder, i.e., they were false negatives.

Note that in essence, *progressive analyses* of the validity data for the screening program were conducted as the validity of each test in the battery was first calculated. However, the data were also re-analyzed to estimate validity for combinations of a progressively increasing number of tests. It should also be mentioned that as with all screening studies reviewed above, no inferential statistics (i.e., comparisons of means, variability, frequencies) were calculated based on the data. Yet this is justified as it is standard practice for vision screening studies.

Validity of Each Test

As the first step of the progressive analyses, the estimates of validity for each screening test individually are presented in Table 25 below. It should be mentioned that for visual functions that were assessed with more than one screening test (i.e., visual acuity, ocular alignment/motility, stereoacuity, and CS), the results of all tests were combined, usually due to the limited number of children who were assessed with each test. Also, because relatively few children were assessed on CS, children from the pilot study are included in the analysis. Table 25 indicates that no screening test on its own was effective as each fell below the cut-off criteria on sensitivity and NPV. Furthermore, three of the tests failed to reach the cut-off criterion for specificity. Note however, that all tests possessed relatively high PPV indicating that few of the children who were referred based on the screening (i.e., those who screened positive), were classified as overreferrals (or false positives) according to the optometric gold standard exam. Also,

as shown in the Table, the Welch-Allyn SureSight autorefractor possessed the highest estimates on three of four measures of validity. Despite this favorable comparison, the NPV and sensitivity of autorefraction were still below cut-off criteria, indicating that children with visual deficits were not correctly identified by the screening test as having a disorder. This result precludes the use of the autorefractor as the lone test in a vision screening program. Closer analysis reveals that the device had difficulty detecting disorders of alignment/motility in the absence of amblyopia (see Table 26 below). However, this result is not surprising given that the autorefractor is designed to assess refractive error only and thus, is more sensitive to disorders that affect *this* visual function.

Table 25. Summary of validity of tests of each visual function. The numbers in bold represent those that reached the cut-off criterion for an effective vision screening program based on that measure of validity.

Visual Function	N*	Sensitivity	Specificity	PPV	NPV
Visual Acuity	103	67	84	**82**	69
Alignment/Motility	108	28	88	**73**	52
Stereoacuity	103	56	84	**78**	65
Autorefraction	110	64	**94**	**93**	70
Contrast Sensitivity	56	56	**93**	**88**	70

* N includes only subjects that completed the test in question.

To better segregate the data and to determine whether the other screening tests are also specialized in terms of the type of visual deficit they can detect, vision disorders

identified in the present study were divided into two mutually exclusive categories: (1) disorders that presumably affect spatial vision (i.e., reduced acuity, amblyopia, and significant refractive error)[34]; (2) disorders of ocular alignment/motility (i.e., strabismus, convergence insufficiency, unsteady fixation/nystagmus). Note that cases of strabismic amblyopia and strabismus paired with reduced acuity were excluded as they can be classified as both spatial vision and ocular alignment/motility disorders. The sensitivity of the screening tests for detecting category-specific disorders is presented in Table 26. The results show that the visual acuity tests and CS tests were similar to the autorefractor. Specifically, each was sensitive to spatial vision disorders, and insensitive to alignment/motility disorders. In the case of CS however, these data must be considered preliminary as relatively few children were tested with this procedure. In contrast, ocular alignment/motility tests were obviously sensitive to alignment/motility disorders, but insensitive to spatial vision disorders. Note however, that the stereoacuity tests were somewhat sensitive to each category of disorder, though their sensitivity to alignment/motility disorders is slightly superior. This finding is predictable however, as stereoacuity requires both acute vision in each eye and proper alignment. Consequently, it will likely be affected by either category of disorder.

Collectively, the above data indicate that testing a single aspect of functional vision poses two problems. First, because the screening tests are specialized in terms of the disorders they can detect, several children who possessed disorders according to the optometric gold standard exam were not identified during screening and thus, classified

[34] From this point on, these disorders will be collectively referred to as spatial vision deficits.

as false negatives. As a result, the sensitivity and NPV of each test was well below the cut-off criteria.[35] Second, because many children with normal vision were incorrectly classified during screening as having a vision disorder, i.e., they were classified as false positives, the specificity estimates were relatively low (i.e., below the cut-off criterion) for three of five screening tests.[36]

Table 26. Sensitivity of tests of each visual function by category of disorder.

	Sensitivity	
	Disorder	
Visual Function	**Spatial Vision Disorders** (N = 39)	**Alignment/Motility Disorders** (N = 11)
Visual Acuity	74	10
Ocular Alignment/ Motility	12	64
Stereoacuity	47	55
Autorefraction	79	9
Contrast Sensitivity	53*	20†

* Based on a limited number of children (N = 15).
† Based on a limited number of children (N = 5).

[35] This is because the number of false negatives is in the denominator of both the sensitivity equation (i.e., sensitivity = true positives/[true positives + false negatives]) and the NPV equation (i.e., NPV = true negatives/[true negatives + false negatives]). Therefore, the greater the number of false negatives, the lower the sensitivity and NPV.

[36] This is because the number of false positives is in the denominator of the specificity equation (i.e., specificity = true negatives/[false positives + true negatives]). Therefore, the greater the number of false positives, the lower the specificity.

The first problem can be remedied by re-analyzing the results to include combinations of two or more screening tests, most importantly one that assessed spatial vision (i.e., CS, visual acuity, and autorefraction) *and* one that assessed ocular alignment/motility (i.e., stereoacuity and alignment/motility). These combinations should increase the likelihood that children who truly possess either category of disorder according to the optometric gold standard exam, would be identified by the screening tests (i.e., more true positives and fewer false negatives), thereby increasing sensitivity and NPV.[37] However, combining screening tests will have the opposite effect on specificity. In essence, when screening tests are combined, the specificity of the overall combination can only be equal to, or lower than, the specificity of that test which possesses the lowest specificity in the combination.[38] This is because the number of false positives resulting from a combination of two or more screening tests can only be equal to, or greater than, the number of false positives from the test with the most false positives. However, the specificity of a single screening test can be increased by

---

[37] Note that including combinations of vision screening tests will not necessarily improve NPV. Although this measure will decrease the number of false negatives, it may, in some cases, substantially reduce the number of true negatives as combining tests increases the likelihood that a child with normal vision will screen positive. In other words, children who were previously true negatives, will now be false positives. As a result, NPV may remain unchanged or even decrease.

[38] Although theoretically correct, there are circumstances under which specificity of combinations may be slightly higher than that of individual tests. First, to be included in validity calculations of tests of individual functions, each child was required to complete that test. However, to be included in validity calculations of the combinations presented below, each child was required to complete only one of the tests in the combination. Thus, validity calculations include children who did not complete one of the tests in the combination, but completed the other(s) and were classified as true negatives. As a result, the number of true negatives in a combination will likely be higher than that of any test alone. Because the number of true negatives is the numerator of the specificity equation (i.e., true negatives/(true negatives + false positives), specificity may therefore, increase slightly in comparison to that of single tests. Second, in combinations that include CS measurement, only those children who received the gold standard exam after CS measurement was added to the battery were considered in the validity calculations (N = 62). Thus, if the specificity of a test was higher during this latter part of the study, the specificity of the combination could be higher than that of the test(s) combined with CS measurement.

reducing the referral criteria of that test (i.e., implementing more lenient referral criteria), as fewer children will be classified as false positives. Thus, the reduction in specificity when combining tests can be counteracted by reducing the referral criteria of screening tests that possess low specificity. Although this action will likely lower the sensitivity and NPV of each test (as the number of false negatives will increase), these measures should improve when tests are combined. The results of attempts to increase sensitivity, NPV, and specificity are discussed below.

Calculation of Validity by Reducing Criteria and Combining Results from Two Screening Tests

As mentioned above, the specificity of any combination of vision screening tests is limited by the poor specificity of the individual tests. Thus, Table 25 indicates that visual acuity and stereoacuity tests will likely place the greatest limitations on a combination of tests as they yielded the poorest specificity (specificity = 84% for both). Therefore, the validity data were re-analyzed implementing more lenient referral criteria for these tests as shown in Table 27 below. Note that the referral criteria for 2-year-olds' visual acuity were not changed as specificity is already high for this age group. Also, the referral criterion for 3-year-olds' scores on the Randot E Stereotest were not changed as the current criterion (500 arc sec) was the maximum measure of stereoacuity attempted and thus, a higher score (other than > 500 arc sec) could not possibly be attained. Furthermore, the criterion of the Randot Preschool Stereoacuity Test was not changed for the same age group as the current criterion (400 arc sec) is the score closest to that of the Randot E (500 arc sec) for 3-year-olds.

Table 27. Lenient referral criteria of tests of visual acuity and stereoacuity.

Visual Function/Test	Age	Previous Referral Criteria	New Referral Criteria
Visual Acuity			
Lea Linear Optotype Chart	3	< 20/50	< 20/64
Patti Pics Linear Optotype Chart		< 20/64	< 20/80
Patti Pics Cards; Lea/Patti Pics isolated symbols book		< 20/40	< 20/50
Lea Linear Optotype Chart	4	< 20/40	< 20/50
Patti Pics Linear Optotype Chart		< 20/50	< 20/64
Patti Pics Cards; Lea/Patti Pics isolated symbols book		< 20/32	< 20/40
Lea Linear Optotype Chart	5 and	< 20/32	< 20/40
Patti Pics Linear Optotype Chart	older	< 20/40	< 20/50
Patti Pics Cards; Lea/Patti Pics isolated symbols book		< 20/25	< 20/32
Stereoacuity			
Randot E Stereotest	4 and	> 250	> 500 arc sec
Randot Preschool Stereacuity Test	older	> 200	> 400 arc sec

* Note, referral criteria for interocular difference remain unchanged and thus, are not presented in the table.

Thus, as a second step of the progressive validity analyses, the data have been re-analyzed following more lenient referral criteria for the visual acuity and stereoacuity screening tests. The results of this re-analysis are shown in Table 28. This Table also shows again, the validity measures of each test from Table 25. Two expected results were confirmed. First, the sensitivity and NPV with lenient criteria for both the visual acuity and stereoacuity tests were lower as more children with disorders were incorrectly classified as negative during screening (i.e., false negatives). Second, the specificity of

both tests was higher as fewer children were incorrectly classified as positive during screening (i.e., false positives). In fact, the reduction of referral criteria increased specificity of visual acuity testing to well above the cut-off criterion for effectiveness (98%). However, despite this increased specificity, it must be pointed out that no combination of tests that assesses spatial vision/refractive error *and* ocular alignment/motility will likely surpass the cut-off criterion for specificity as each combination will include at least one test that yields specificity below this level (i.e., alignment/motility and stereoacuity even under lenient criteria). As stated above, the specificity of any combination of tests can only be as high as the lowest specificity in that combination (but see footnote 38).

Table 28. Validity of visual acuity and stereoacuity tests with lenient referral criteria. For comparison, the validity of all tests, including visual acuity and stereoacuity with original criteria are presented again from Table 25. Numbers in bold represent those that reached the cut-off criteria for an effective vision screening program based on that measure of validity.

Visual Function	N	Sensitivity	Specificity	PPV	NPV
Visual Acuity	103	67	84	**82**	69
Visual Acuity (LC)*	103	54	**98**	**97**	66
Alignment/Motility	108	28	88	**73**	52
Stereoacuity	103	56	84	**78**	65
Stereoacuity (LC)*	103	38	88	77	58
Autorefraction	110	64	**94**	**93**	70
Contrast Sensitivity	56	56	**93**	**88**	70

* LC denotes lenient referral criteria.

As a third step of the progressive analyses, the validity estimates of combinations of two vision screening tests are presented in Table 29. Importantly, to maximize sensitivity, each combination within the Table possessed a screening test which assessed spatial vision/refractive error (visual acuity, autorefraction, or CS), and a screening test that assessed ocular alignment/motility (ocular alignment/motility or stereoacuity) as discussed above. In addition, each combination that included visual acuity and/or stereoacuity tests were analyzed with both original, strict referral criteria, and the lenient referral criteria in order to compare specificity. It should also be noted that children had to complete at least one of the two tests to be included in these validity calculations. This inclusion criterion was chosen in order to follow the same protocol that was implemented for the overall battery and therefore, to treat each combination as a separate and complete screening battery.

Inspection of the Table demonstrates that as expected, combinations of two screening tests yielded higher sensitivity and NPV than either test used in isolation (see Table 28), thus indicating that the combinations correctly detected more children with disorders during screening. Furthermore, implementation of lenient referral criteria for both visual acuity and stereoacuity tests had the anticipated effect of increasing specificity of combinations while decreasing sensitivity. In fact, the specificity of the combination of ocular alignment/motility tests and visual acuity tests under lenient referral criteria reached the cut-off criterion for an effective program. Interestingly, the combination of CS and ocular alignment/motility tests also reached the cut-off criterion for specificity. However, it should be pointed out that no single combination provided in

Table 29 surpassed the cut-off criteria for sensitivity or NPV. Nevertheless, combinations that include autorefraction as the index of spatial vision yielded the highest estimates of sensitivity, PPV, and NPV, along with moderate estimates of specificity. This result is not surprising given that of any single test, autorefraction generally yielded the highest validity estimates (see Table 25). On the other hand, combinations that utilized visual acuity or CS as the index of spatial vision were less sensitive (CS and visual acuity under both original and lenient criteria) and yielded lower NPVs (CS and visual acuity under both criteria). Finally, comparisons between combinations that employed ocular alignment/motility or stereoacuity tests reveal mixed results. However, the Table indicates that combinations that included either alignment/motility tests or stereoacuity tests under lenient referral criteria generally yielded similar sensitivity and specificity.

Despite improvement over single test measurements, two-test combinations failed to reach the cut-off criteria for sensitivity and NPV. It is apparent that at least three tests are required to sufficiently improve validity and thus, re-analysis of data was undertaken for three-test combinations. Once again, each combination included an index of spatial vision as well as an index of alignment/motility. In addition, in order to maximize specificity of all remaining combinations, validity re-analyses were conducted implementing stereoacuity and visual acuity testing under lenient referral criteria only.

Table 29. Validity of combinations of tests measuring two visual functions. Note that the numbers in bold represent those that reached the cut-off criteria for an effective vision screening program on that measure of validity.

Visual Functions	N*	Sensitivity	Specificity	PPV	NPV
Visual Acuity + Ocular Alignment/Motility	107	79	73	**77**	73
Visual Acuity (LC) + Ocular Alignment/Motility	107	64	**90**	**88**	68
Visual Acuity + Stereoacuity	105	78	72	**75**	75
Visual Acuity (LC) + Stereoacuity	105	72	82	**81**	73
Visual Acuity Stereoacuity (LC)	104	76	74	**76**	74
Stereoacuity (LC) + Visual Acuity (LC)	105	65	86	**84**	69
Stereoacuity + Autorefractor	110	86	80	**84**	84
Stereoacuity (LC) + Autorefractor	110	85	82	**85**	82
Stereoacuity + Contrast Sensitivity	57	63	83	**77**	71
Stereoacuity (LC) + Contrast Sensitivity	57	63	87	**81**	72
Contrast Sensitivity + Ocular Alignment/Motility	58	68	**90**	**86**	75
Autorefractor + Ocular Alignment/Motility	110	80	80	**82**	77

* In combinations that included CS measurement, N includes only children who received the gold standard exam once CS measurement was added to the battery.

Calculation of Validity: Three-Test Combinations

As the next step of the progressive validity analyses, the validity estimates for three-test combinations are shown in Table 30 below. Although no single combination was completely effective, a clearer picture emerges as to which tests may be the best to include in a sensitive screening program. Specifically, combinations that included both autorefraction and ocular alignment/motility tests yielded relatively high sensitivity. In fact, the sensitivity of three of these combinations was above the cut-off criterion. These results imply that the majority of children who possessed undetected vision disorders were correctly identified by the screening tests. This is consistent with the earlier findings that autorefraction was the most sensitive of all tests to spatial vision deficits and ocular alignment/motility tests were the most sensitive to alignment/motility disorders (see Table 26). Table 30 also indicates that combinations that included both autorefraction and ocular alignment/motility tests yielded the highest NPVs, though each fell below the cut-off criterion. This suggests that compared to other combinations, most children who passed all tests in these combinations did indeed possess normal vision as determined by the optometric gold standard exam.

Table 30. Validity of three vision test combinations. Combinations above the horizontal line in the visual functions column contain two indices of spatial vision and one index of alignment/motility. Those below the line possess two indices of alignment/motility and one index of spatial vision. Numbers in bold represent those estimates surpassing the cut-off criteria for an effective program.

Visual Functions	N*†	Sensitivity	Specificity	PPV	NPV
Autorefraction + Visual Acuity + Alignment/Motility	110	**92**	78	**83**	89
Autorefraction + Visual Acuity + Stereoacuity	110	88	80	**84**	85
Autorefraction + CS + Alignment/Motility	62	**90**	**90**	**90**	90
Autorefraction + CS + Stereoacuity	62	81	87	**86**	82
Visual Acuity + CS + Alignment/Motility	60	77	**90**	**88**	79
Visual Acuity + CS + Stereoacuity	58	71	87	**83**	76
Alignment/Motility + Stereoacuity Autorefractor	110	**95**	76	**82**	93
Alignment/Motility + Stereoacuity + Visual Acuity	108	76	80	**81**	74
Alignment/Motility + Stereoacuity + CS	60	73	87	**85**	76

* Note, to be included in validity calculations, children were required to complete at least one of the tests in the combination.

† Combinations that include CS measurement only include those children who were tested once CS measurement was added to the screening battery.

Calculation of Validity: Four-Test Combinations

As the fifth step in the progressive analyses, the validity data were re-analyzed for four vision test combinations. The results of this re-analysis are provided in Table 31 below. For the most part, these data confirmed the conclusions reached regarding three-test combinations (see Table 30). Specifically, to maximize sensitivity, a screening program should include autorefraction and tests of alignment/motility. Those combinations that did not include these tests yielded sensitivity and NPV estimates that were markedly lower. The Table also indicates that although the combination of autorefraction, visual acuity, stereoacuity, and alignment/motility tests possessed relatively low specificity (75%), it had the highest sensitivity of any combination analyzed thus far (98%). Furthermore at 97%, it was the only combination yet to surpass the cut-off criterion for NPV. Finally, the Table also reveals that only one combination reached the cut-off criterion on specificity (autorefraction + visual acuity + CS and alignment/motility). However, the relatively low specificity of the remaining combinations is expected as the addition of extra tests to a combination generally decreases specificity.

Table 31. Validity of four vision test combinations. Numbers in bold represent those estimates surpassing the cut-off criteria for an effective program.

Visual Functions	N*†	Sensitivity	Specificity	PPV	NPV
Autorefraction + Visual Acuity + CS + Alignment/Motility	62	**94**	**90**	**91**	93
Autorefraction + Visual Acuity + CS + Stereoacuity	62	84	87	**87**	84
Autorefraction + Visual Acuity + Stereoacuity + Alignment/Motility	110	**98**	75	**82**	**97**
Autorefraction + CS Stereoacuity + Alignment/Motility	62	**94**	87	**88**	93
CS + Visual Acuity + Stereoacuity Alignment/Motility	60	83	87	**86**	84

* Note, to be included in validity calculations, children were required to complete at least one of the tests in the combination.
† Combinations that include CS measurement only include those children who were tested once CS measurement was added to the screening battery.

Calculation of Validity: Five Test Combinations

As the final step in the progressive analyses, Table 32 below presents the validity of the combination of all tests within the battery. For comparison, the Table includes three variations of the five-test combination. The first variation contains data for children tested under the lenient referral criteria for stereoacuity and visual acuity testing, and only includes children who received the optometric gold standard exam after CS measurement was included in the program (N = 62). This variation was considered in order to provide a more accurate reflection of the contribution of CS testing to the battery and therefore, a more accurate reflection of a *true* five-test combination. The second variation also

presents the data from children tested under the lenient referral criteria, but does so for *all* children who received the optometric gold standard exam (N = 110). Thus, it provides the validity for the entire study under conditions to maximize specificity. The final variation includes all children who received the optometric gold standard exam (N = 110), but follows the original criteria set for the tests of stereoacuity and visual acuity in order to determine the effect of strict referral criteria on the specificity of the entire battery.

The Table indicates that all variations of the program possessed extremely high sensitivity as for each one, only one child who possessed a disorder (a subtle strabismus) was incorrectly identified as negative during screening. In addition, all variations possessed high NPV and reached the cut-off criterion of 95%. A comparison of the three variations reveals a couple of interesting results. For instance, a comparison of the first two variations of the program demonstrates that sensitivity was essentially the same after the introduction of CS to the program, although specificity improved from 75% to 87%. However, this improvement is not the result of the addition of CS testing as it is impossible to increase specificity by adding a test unless it replaces a test that possesses lower specificity. Instead, it is more likely that the specificity of other tests increased in this latter stage of the program as the testers became more experienced and thus, increased the overall specificity of the battery. Comparison of the second and third variations demonstrates that the implementation of lenient referral criteria of visual acuity and stereoacuity tests had the desired effect of increasing specificity. Also, despite the

reduction of these criteria, sensitivity and NPV were not sacrificed as they were exactly the same for both variations.

Table 32. Validity of five vision test combinations. Numbers in bold represent those estimates surpassing the cut-off criteria for an effective program.

Visual Functions	N*	Sensitivity	Specificity	PPV	NPV
Post-CS Children					
Autorefraction + CS + Visual Acuity + Stereoacuity + Alignment/ Motility	62	**97**	87	**89**	**96**
Lenient Criteria/All Children Who Received the Gold Standard Exam					
Autorefraction + CS + Visual Acuity + Stereoacuity + Alignment/ Motility	110	**98**	75	**82**	**97**
Original Criteria/All Children Who Received the Gold Standard Exam					
Autorefraction + CS + Visual Acuity + Stereoacuity + Alignment/ Motility	110	**98**	65	**76**	**97**

* Note, to be included in validity calculations, children were required to complete at least one of the tests in the combination.

Importantly, the Table also reveals that the inclusion of an extra test to a four-test combination may introduce redundancy to the battery. Specifically, the second variation possessed the exact same validity estimates as the four-test combination of autorefraction, visual acuity, stereoacuity, and alignment/motility tests (Table 31), implying that the addition of CS testing to this combination is redundant. Note however, the addition of autorefraction or ocular alignment/motility to a four-test combination was not redundant

as both greatly improved the sensitivity and NPV of the battery. For instance, adding autorefraction to the four-test combination of CS, visual acuity, stereoacuity, and alignment/motility (see Table 31) raised sensitivity and NPV by 14% and 12%, respectively. Likewise, the addition of alignment/motility tests to the four-test combination of autorefraction, visual acuity, CS, and stereoacuity increased sensitivity and NPV by 13% and 12%, respectively. This reiterates the importance of both autorefraction and ocular alignment/motility to a screening program that desires to detect even subtle cases of vision disorders.

Thus, three general conclusions may be reached in light of the data provided from five-test combinations. First, the addition of CS to a four-test combination that already includes two indices of spatial vision (i.e., autorefraction + visual acuity) was redundant as it did not improve any single measure of validity. Second, the addition of autorefraction and/or ocular alignment/motility tests to a four-test combination was *not* redundant as sensitivity and NPV were substantially improved. Finally, when implementing combinations of many tests, it was essential that the referral criteria of visual acuity and stereoacuity tests were relatively lenient in order to maximize specificity. Importantly, this reduction of referral criteria was not detrimental to overall sensitivity or NPV.

Discussion

This ambitious study was successful on a number of practical and clinical levels. First, we assessed 946 toddlers and preschoolers on up to five separate visual functions. This comprehensive assessment included several innovative, experimental screening

tests/tools such as the Welch-Allyn SureSight autorefractor, the contrast sensitivity (CS) cards, the Lea Symbols, and the Patti Pics symbols. Thus, the present study is the first, or among the first to implement these techniques in a preschool vision screening program. Furthermore, thorough analyses regarding the effectiveness of these tests *and* combinations of tests are provided. Second, the overall program was very accurate as the vast majority of children referred did indeed possess vision disorders which required immediate treatment. This is also of particular importance from a clinical standpoint as these children are in a sensitive period of brain plasticity during which treatment may still prevent permanent visual dysfunction. Third, based on the present study, I have provided the first detailed data regarding estimated prevalence of vision disorders in toddlers and preschoolers in Newfoundland and Labrador, Canada. The implications of the results of the present study are discussed further below.

Comparison of Prevalence Rates to Those of Other Studies

The prevalence estimates of the present study are re-presented below in Table 33. As mentioned above, these estimates are of particular interest as few exist for Canada or the province of Newfoundland and Labrador. Moreover, relative to other parts of North America, the population of Newfoundland and Labrador is genetically isolated and has been shown to possess relatively high prevalence of several disorders such as hemophilia, colorectal cancer, and Bardet-Beidl syndrome (Rahman, Jones, Curtis, Bartlett, Peddle, Fernandez, & Freimer, 2003; Woods, Hyde, Curtis, Stuckless, Green, Pollett, Robb, Green, Croitoru, Careen, Chaulk, Jegathesan, McLaughlin, Galllinger, Younghusband, Bapat, & Parfrey, 2005; Woods, Young, Parfrey, Hefferton, Green, & Davidson, 1999;

Xie, Zheng, Leggo, Scullyum, & Lillicrap, 2002). As certain vision disorders such as strabismus and astigmatism have a strong genetic component (Abrahamsson, Magnusson, & Sjostrand, 1999; Grosvenor, 1978; Lorenz, 2002), it is quite likely that the population of Newfoundland and Labrador also possesses a high prevalence of these disorders. However, because the sample only included children from the St. John's metropolitan area, these estimates should not be considered representative of the entire province. For comparison, Table 33 shows the prevalence estimates for other areas of Canada (Ontario, Saskatchewan, New Brunswick, and British Columbia) and other industrialized countries (Britain, the United States, Germany, Sweden, and Australia). A couple of caveats must be mentioned regarding the estimates in the Table. First, prevalence estimates from different areas/studies are often based on different disease criteria. Therefore in some cases, discrepancies between estimates may not reflect true differences in disease prevalence, but the use of different criteria. Second, some studies based prevalence estimates on a single, older age group such as children in grade 1 (Kornder, Nursey, Pratt-Johnson, & Beattie, 1974; Robaei, Rose, Ojaimi, Kifley, Huynh, Mitchell, 2005; Robinson, 1999; Ross et al., 1977; Woodruff, 1986). This is important as preschoolers and toddlers tend to be hyperopic but shift toward emmetropization and in some cases, may begin to become myopic (Flitcroft, 1998). In light of this, studies of older children may underestimate prevalence of hyperopia and overestimate the prevalence of myopia.

Table 33. Prevalence estimates of vision disorders in young children from the present study compared to other areas of Canada and other industrialized countries.

Vision Disorder	Estimated Prevalence Present Study (percent)	Estimates From Other Canadian Studies (percent)	Estimates From Other Countries (percent)
Strabismus, Motility/ Fixation Disorders	3.1%	1.2 - 4.5%[1,2]	2 - 5.3%[10-17]
Significant Refractive Error	7.1%	*No estimate	5.8 - 7.7%[13,15,16]
Astigmatism	3.1%	5 - 7.2%[3,4]	3 - 4.9 [15, 16,18]
Hyperopia	4.2%	†4.6%[4]	2.5 - 6.4[19,20]
Myopia	1.0%	1.2 - 6.0%[4,5]	1.4 - 5.2,[16, 18-20]
Anisometropia	1.4%	†3.6%[4]	1 - 3.5[15, 16, 18]
Suspected Amblyopia	4.3%	0.83 - 5.6%[1,6,7]	0.2 - 5.5%[12-14, 16, 17, 20]
All Disorders	14.4%	10 - 15%[1,7-9]	5 - 15%[15, 21-24]

[1]Feightner (1994); [2]Kornder et al. (1974); [3]Cowen & Bobier (2003); [4]Woodruff (1986); [5]Robinson (1999); [6]Ross et al. (1977); [7]Robinson et al. (2000); [8]Robinson et al. (1999); [9]Ontario Association of Optometrists (1997);[10]Fathy & Elston (1993); [11]Graham (1974); [12]US Public Health Service (1994); [13]Kvarnström et al. (2001); [14]Barry & König, (2003); [15]Donnelly, Stewart, & Hollinger (2005); [16]Preslan & Novak (1998); [17]Williams et al. (2001); [18]Mayer, Hansen, Moore, Kim, & Fulton (2001); [19]Robaei et al. (2005); [20]Junghans & Crewther (2003); Simons (1996); [22]Brown (1975); [23]Snowdon & Stewart-Brown (1997); Feldman, Milner, Sackett, & Gilbert, (1980).

* No estimate could be obtained from Canadian studies.

† A range of estimates could not be provided as prevalence estimates could only be found for a single study.

Despite these caveats, the Table demonstrates that the estimates of the present study are consistent with those from other areas of Canada and other industrialized nations. Note however, the estimated prevalence of myopia is somewhat low. Yet, this discrepancy may be due to the relatively high disease criterion for myopia in the present study (i.e., sphere < -1 D). Thus, although Newfoundland and Labrador is genetically

homogeneous to some degree, the children in the present study do not appear to possess an abnormally high prevalence of vision disorders.

Comparison of Completion Times and Completion Rates

Despite the obvious importance of completion rates and completion times to the development and implementation of a program to assess young children, few screening studies provide clear data on these measures. This is particularly true for completion times, for which data are virtually nonexistent (but see Shallo-Hoffman et al., 2004). Generally, only those studies that assess children with a single, often experimental technique (e.g., photoscreening or autorefraction) provide completion rates. In light of this, it is difficult to compare the total completion rates and completion times of the present study to those of other studies. Nevertheless, a comparison to similar screening studies, i.e., those that implement two or more tests (Enzenauer et al, 2000; Köhler & Stigmar, 1973), reveals that the overall completion rate of the present study is relatively low (79% vs. 96 – 99%). Note however, that the present study placed great attentional demands on each child as a total of eight tests were often attempted. Furthermore as mentioned above, the completion rates of the present study *do not* reflect the true testability of children because in many cases, certain tests were not even attempted to avoid distressing the child.

A summary of completion times and completion rates for the tests in the present study is provided in Table 34 below. Mean completion times and completions rates are calculated across all age groups and all tests for each visual function. Note that the tests are ranked on completion time and completion rate and an average rank is provided. The

tests are listed in ascending order of their average rank. The Table indicates that as noted above, autorefraction was the easiest procedure for children to complete. Importantly, this procedure possesses additional merits as it is remarkably simple and limited training is required. Stereoacuity and visual acuity tests were also generally easy to complete (completion rates of 92% and 87% respectively) even though the completion time for acuity tests was slightly high (completion time = 2.8 min). Interestingly however, completion rates and completion times of the visual acuity and stereoacuity tests were poorer than those reported by Shallo-Hoffman et al. (2004) and the VIP Study Group (2004; 2004b)[39] who screened children with similar tests (i.e., Lea linear optotype chart and Randot E Stereotest). The reason for this discrepancy is unclear. However, the testers in the present study were extremely cautious to avoid distressing the child. Thus, it is likely that in many cases the child could have been coaxed to complete these tests.

Table 34 indicates that CS and ocular alignment/motility tests were relatively difficult for children to complete (completion rates of 77% and 81%, respectively). For the most part, this limited the completion rate of the entire battery. In particular, children had difficulty completing the cover-uncover tests as they were often unable to fixate straight ahead during the procedure. Yet it should be pointed out that the battery also included the Hirschberg corneal reflex, a simple, albeit crude test of alignment that could be completed by almost all children (96%). Similarly, the CS booklet was difficult for children to complete (63%). Note however, that the booklet was only recently added to

[39] The VIP Study Group (2004; 2005b) reported completion rates of > 98% for the Lea Symbols and ~ 90% for the Randot E stereotest. Similarly, Shallo-Hoffman et al. (2004) reported completion rates of ~ 90% and ~ 93% for Lea Symbols and Randot E stereotest, respectively. Also, Shallo-Hoffman et al. (2004) obtained a mean completion time of ~ 2.1 minutes for the Lea linear optotype test.

the screening program and the testers were still adapting to the testing procedure. In addition, the Table demonstrates that CS was also hampered by a long mean completion time of 3.4 minutes (CS cards = 3.3 minutes; CS book = 5.6 minutes). This suggests that the procedure may require modification before it can gain widespread acceptance into preschool vision screening programs. One potential modification is to test children with fewer spatial frequencies. This change is justified as no single case of visual dysfunction was detected due to failure at the lowest spatial frequency (0.75 c/deg).

Table 34. Mean completion times and completion rates across all tests and age groups for each visual function.

Visual Function/Test (Overall Rank)	Completion Time (Rank)	Completion Rate (Rank)	Average Rank
(1) Autorefraction	1.4 min (1)	95% (1)	1
(2) Stereoacuity	1.8 min (2)	87% (3)	2.5
(3) Visual Acuity	2.7 min (4)	92% (2)	3
(4) Ocular Alignment/ Motility	2.3 min (3)	81% (4)	3.5
(5) Contrast Sensitivity	3.4 min (5)	78% (5)	5
Overall	**11.6**	**78%**	

Comparison of the Validity of Tests of Each Visual Function to Other Studies

As the first step of the progressive validity analyses above (see Progressive Validity Analyses in the Results section) validity was estimated for tests of each visual function. In essence, these data represent a simulation of a vision screening program that used a single test. As a result, it is possible to compare these data to those of other studies that utilized a single test to allow us to determine whether the most appropriate tests were chosen for the present study. Thus, this comparison is conducted for each aspect of functional vision assessed in the present study in the subsections below. Yet, it should be noted that only those studies that provide complete validity data on tests of a *single* aspect are included for the comparison. This measure was taken to ensure the data reflect the validity *solely* of the tests being compared, and not tests of additional visual functions. Also, though the Vision in Preschoolers (VIP) Study Group (2004; 2005a; 2005b) studies provide data only on sensitivity and specificity, these studies have been included as they implement tests that are identical or similar to those used in the present study. Furthermore, the validity estimates of these studies are based on an extremely large sample of children (N = 1452 to 2588), and are therefore, likely to be accurate. Finally, for each visual function, the validity estimates provided represent each study's effectiveness across *all* vision disorders.

Visual Acuity

The validity of visual acuity tests from the present study and other screening studies (N = 20) are presented in Table 35 below. All studies (including ours) are ranked on all four measures of validity (except for the VIP studies which are ranked only on

sensitivity and specificity), and a mean rank on the four measures is provided. The studies are listed in ascending order based on this mean rank. Note that although the validity estimates of the present study have been combined for all acuity tests, the vast majority of children who received the optometric gold standard exam were tested with the Lea Symbols or the highly similar Patti Pics.

The Table reveals some interesting findings regarding the results of the present study. First, despite their poor sensitivity (54%), the visual acuity tests under the lenient referral criteria still ranked highly at the 5th position overall. On the other hand, under the original, stricter referral criteria, the acuity tests of the present study ranked poorly at 11th, due in part to its low specificity (84%; ranked 17th). Second, the NPV under both sets of criteria was fairly low (lenient criteria = 66%; original criteria = 69%), ranking 12th (original criteria) and 13th (lenient criteria) overall. Conversely, the PPV ranked highly under both referral criteria (lenient criteria = 97%; original criteria = 82%) at the 2nd and 5th positions. Collectively, these data suggest that although most of the children who failed the acuity tests possessed visual deficits, it was difficult to identify children with vision disorders. In fact, at least one-third of the children who possessed vision disorders were able to pass these tests, resulting in poor sensitivity and NPV. Moreover, these results do not appear to be due to poor choice of referral criteria as they were found for both lenient and strict conditions. Thus, it appears more likely that the poor sensitivity and NPV of the present study reflect the limits of the tests themselves. This notion is supported by the findings of the VIP Study Group (2004; 2005a; 2005b) who implemented a similar test of acuity and also reported low sensitivity.

Table 35. Summary of the validity of visual acuity tests from vision screening studies. Note, the results of the present study are provided in bold. Numbers in parentheses represent each study's rank on that particular measure of validity. Studies in the Table are listed in ascending order based on mean rank on all validity measures.

Study (Rank)	Test	Sensitivity	Specificity	PPV	NPV	Mean Rank
(1) Nordlöw & Joachimson (1962)	Boström Hooks (5m; single)	95 (3)	99 (2)	86 (3)	100 (1)	2.25
(2) Raina (1998)	Wright Cards	100 (1)	95 (7)	77 (7)	100 (1)	4
(3) Ladenwall (1988)	HOTV (distance and optotype not reported)	87 (7)	97 (4)	86 (3)	98 (5)	4.75
(4) Kvarnström et al. (1998)	HOTV (3m; linear)	92 (4)	97 (4)	69 (10)	99 (3)	5.25
(5) Spowart et al. (1998)	Stycar (6m; single), Glasgow Acuity Cards (3m; linear)	83 (8)	95 (7)	40 (13)	99 (3)	7.75
(5) Present Study (Lenient Criteria)	**Lea Symbols Chart & Patti Pics Chart (3m; linear), Lea/Patti Pics Isolated Symbols (3m; single); Patti Pics Card (3m; single), Teller Acuity Cards**	**54 (13)**	**98 (3)**	**97 (2)**	**66 (13)**	**7.75**
(7) Raina (1998)	Teller Acuity Cards	15 (20)	100 (1)	100 (1)	87 (10)	8
(8) Simon et al. (2004)	Enfant II VEP Headband	97 (2)	81 (18)	71 (9)	98 (5)	8.5
(9) Arnold et al. (2004)	HOTV (distance and optotype not reported)	91 (6)	86 (16)	79 (6)	94 (7)	8.75
(10) Kennedy et al. (1995)	Snellen Test (distance and optotype not reported), Stycar Graded Balls	33 (19)	97 (4)	54 (12)	94 (7)	10.5
(11) Present Study (Original Criteria)	**Lea Symbols Chart & Patti Pics Chart (3m; linear), Lea/Patti Pics Isolated Symbols (3m; single); Patti Pics Card (3m; single), Teller Acuity Cards**	**67 (9)**	**84 (17)**	**82 (5)**	**69 (12)**	**10.75**
(12) Schmidt (1994)	Broken Wheel Test	92 (4)	53 (20)	57 (11)	89 (9)	11

(12) VIP Study Group (2004)	Lea Symbols (3m; linear)	61 (11)	90 (11)			11
*(12) VIP Study Group: Lay Screeners (2005b)	Lea Symbols (1.5m; single)	61 (11)	90 (11)			11
(15) Schmidt (1994)	Teller Acuity Cards	64 (10)	81 (18)	75 (8)	71 (11)	11.75
(16) VIP Study Group (2005a)	Lea Symbols (3m; linear)	49 (15)	94 (9)			12
(17) VIP Study Group: Nurses (2005b)	Lea Symbols (3m; single)	49 (15)	90 (11)			13
(18) VIP Study Group (2004)	HOTV (3m; linear)	54 (13)	89 (15)			14
(18) VIP Study Group (2005a)	HOTV (3m; linear)	36 (18)	93 (10)			14
*(18) VIP Study Group: Lay Screeners (2005b)	Lea Symbols (3m; single)	37 (17)	90 (11)			14

* Testers were trained lay screeners.

Despite the shortcomings of the tests of the present study, it is difficult to determine which acuity tests are superior as the overall ranks of some of the most commonly used tests vary greatly. For instance, studies that used the popular Stycar-HOTV letter tests (Arnold et al., 2004; Kvarnström et al., 1988; Ladenwall, 1988; Spowart et al., 1998; VIP Study Group, 2004, 2005a) attained overall ranks ranging from 3rd to 18th. Similarly, two studies tested children with the Teller Acuity Cards (Raina, 1998; Schmidt, 1994) and attained ranks of 7th and 15th, respectively. Interestingly, the highest ranking study in Table 35 tested children with the Boström Hooks, a test which is highly similar to the Landolt C but is no longer manufactured. However, Schmidt (1994) tested 2-3 year-olds using the Broken Wheel Test, a variant of the Landolt C Test, and

achieved a far lower ranking of 12th overall. Yet it should be pointed out that the sample size of the Schmidt screening study was relatively small (n = 30) and the children were 1 to 2 years younger than those screened by Nordlöw et al. (1962). Also, despite the success of the Boström Hooks, the Landolt C test is rarely used to assess preschoolers, perhaps due to the difficulty in completing the procedure (Becker et al., 2000; 2002; Simons, 1996). Indeed, the present study supports this conclusion. Although at the onset of the study, only two children were tested with the Landolt C, the procedure was difficult and it was immediately apparent that the picture optotype tests (i.e., Lea Symbols and Patti Pics) were far easier for young children to complete.

Ocular Alignment/Motility

The validity of the tests of ocular alignment/motility from the present study is provided in Table 36 below. It should be mentioned that the comparison to other studies is limited to the VIP Study Group (2004) as they are the only other researchers to date, to provide validity data solely for a test of ocular alignment (namely, the cover-uncover test). The Table indicates that the sensitivity of ocular alignment/motility tests of both the present study and the VIP study is relatively poor at 28% and 16%, respectively. However this is expected as these values represent sensitivity to all vision disorders, the vast majority of which primarily affected spatial vision (e.g., reduced visual acuity, significant refractive error). As mentioned above, alignment/motility tests are relatively insensitive to these disorders. On the other hand, the sensitivity of the tests of the present study and the VIP study to disorders of alignment and motility (e.g., strabismus) is much higher at 64% and 60%, respectively. Importantly, the higher sensitivity of the present

study to all disorders and alignment/motility disorders is likely because it utilized four different tests of ocular alignment/motility whereas the VIP study utilized just one. Therefore, if a child in the present study possessed a subtle ocular alignment/motility disorder that was not detected by the cover-uncover test, it was possible that it would be detected by one of the three remaining tests. Conversely, the Table shows that while the specificity of both studies is relatively high (88% to 98%), the specificity of the present study is slightly lower. Again, this may be explained by the implementation of more alignment/motility tests in the present study. Specifically, as each child was assessed with four tests, there was greater likelihood that a child with normal vision would fail a test during screening and be incorrectly classified as positive, thereby leading to poorer specificity. In light of these findings, it can be speculated that if one wishes to increase sensitivity, more ocular alignment/motility tests could be added. On the other hand, if one wishes to increase specificity, one or more tests could be removed from the program. This latter suggestion is important as it could perhaps increase the relatively low specificity of the test combinations that include alignment/motility tests.

Table 36. Summary of the validity of tests of alignment/motility from the present study and the VIP Study Group (2004).

Study (Rank)	Test	Sensitivity	Specificity	PPV	NPV
Present Study	**Cover-Uncover Test, Hirschberg Corneal Reflex, Ocular Motility, Convergence/ Divergence**	**28**	**88**	**73**	**52**
VIP Study (2004)	Cover-Uncover Test	16	98		

Stereoacuity

The validity of the stereoacuity tests of the present study under both the lenient and original referral criteria is provided in Table 37. Note that due to the limited number of children assessed with each test, the data for both have been combined.[40] For comparison, the validity data from other stereoacuity screening studies (N = 12) are also presented in the Table. As with the subsection above on visual acuity, the studies here are ranked on each of the four measures of validity (except for the VIP Studies which are ranked only on sensitivity and specificity) and a mean rank for all four measures is provided. Also, the studies are listed in ascending order of their mean rank.

As shown in the Table, the stereoacuity tests of the present study ranked moderately (4^{th}) under the original, strict referral criteria, but more poorly under the lenient referral criteria (11^{th}). It is important to point out two of the top three studies summarized in the Table assessed school-aged children and teenagers (Hope & Maslin, 1990; Walraven & Janzen, 1993), a factor which may have contributed to their relatively high validity. That notwithstanding, an inspection of the Table reveals that the specificity and NPV of the present study were notably low for both sets of referral criteria (original criteria: specificity = 84%, NPV = 65%; lenient criteria: specificity = 88%; NPV = 58%), whereas PPV was relatively high (original criteria = 78%, ranked 3^{rd}; lenient

[40] Another reason for the combination of these tests is that a hybrid version of the two tests was used. Specifically, in most cases when a child failed only the Randot Preschool Stereoacuity Test, he/she was immediately tested with the Randot E Stereotest. If the child then passed the Randot E Stereotest, he/she was not retested/referred. Thus, the true disease status of these children is not known and as a result, the true validity of the Preschool Stereoacuity Test can not be calculated.

criteria = 77%, ranked 4th). Moreover, it is doubtful that further adjustment of the referral criteria would have improved the relative standing of the present study as an attempt to increase the relatively low specificity would likely decrease the moderate to poor sensitivity (original criteria = 56%; lenient criteria = 38%). Thus, it is likely that the overall stereoacuity rankings of the present study reflect the limitations of the combination of the Randot E Stereotest and the Randot Preschool Stereoacuity Test.

As with the results from acuity tests above, it is difficult to determine which specific test should be recommended to measure stereoacuity. Although the TNO test attained the highest mean validity rank (Walraven & Janzen, 1993), the children tested with this procedure were far older (4 – 18 years of age) than those tested in the present study. Therefore, it is not clear whether the TNO can yield similarly high validity estimates in a preschool population. Furthermore, it is uncertain whether the procedure is easy for toddlers and preschoolers to complete. The Table demonstrates that the majority of stereoacuity screening studies tested children with the Randot E Stereotest, yet there is a great deal of variability in their rankings, ranging from the 2nd to the 13th position. Conversely, studies that implemented the Randot Stereo Smile II ranked relatively consistently (8th to 13th), but yielded poor sensitivity.

Table 37. Summary of the validity of stereoacuity tests from various screening studies. The results from the present study are shown in bold. Numbers in parentheses represent each study's rank on that particular measure of validity. Studies are listed in ascending order based on mean rank on all validity measures.

Study (Rank)	Test	Sensitivity	Specificity	PPV	NPV	Mean Rank
*(1) Walraven & Janzen (1993)	TNO	100 (1)	93 (2)	44 (7)	100 (1)	2.75
(2) Schmidt (1994)	Randot E Stereotest	77 (3)	88 (9)	83 (2)	83 (4)	4.5
*(3) Hope & Maslin (1990; School-aged)	Randot E Stereotest	53 (7)	92 (3)	92 (1)	55 (8)	4.75
(4) Hope & Maslin (1990; Preschoolers)	Randot E Stereotest	89 (2)	75 (14)	17 (8)	99 (2)	6.5
(4) Ruttum & Nelson (1991)	Randot E Stereotest	54 (6)	87 (11)	54 (6)	87 (3)	6.5
(4) Present Study (Original Criteria)	**Randot E Stereotest; Randot Preschool Stereoacuity Test**	**56 (5)**	**84 (12)**	**78 (3)**	**65 (6)**	**6.5**
*(7) Hope & Maslin (1990; School-aged; strict criteria)	Randot E Stereotest	64 (4)	81 (13)	75 (5)	71 (5)	6.75
(8) VIP Study Group (2004)	Randot Stereo Smile II	33 (13)	94 (1)			7
(8) VIP Study Group (2004)	Randot Stereo Smile II	44 (9)	91 (5)			7
(8) VIP Study Group: Nurses (2005a)	Randot Stereo Smile II	45 (8)	90 (6)			7
(11) Present Study (Lenient Criteria)	**Randot E Stereotest; Randot Preschool Stereoacuity Test**	**38 (12)**	**88 (9)**	**77 (4)**	**58 (7)**	**8**
(12) VIP Study Group (2004)	Randot E Stereotest	42 (10)	90 (6)			8
† (13) VIP Study Group: Lay Screeners (2005a)	Randot Stereo Smile II	40 (11)	90 (6)			8.5
(13) VIP Study Group (2004)	Randot E Stereotest	22 (14)	92 (3)			8.5

* These studies assessed mainly school-aged children.
† Testers were trained lay screeners.

Contrast Sensitivity

Unfortunately, the validity of CS tests can not be compared to that of other studies as it has not yet been implemented as part of any previous vision screening program. However, it is possible to compare the CS tests to another index of spatial vision, namely visual acuity. This comparison is of particular importance as researchers have argued that CS provides a more comprehensive index of spatial vision compared to visual acuity testing alone (Adams et al., 1992; Banks & Dannemiller, 1987; Banks & Salapatek, 1981; Drover et. al., 2002). This therefore raises the possibility that if valid, CS tests could eventually replace visual acuity as the measure of spatial vision within screening programs. Table 38 shows the validity of CS presented with the validity of visual acuity studies summarized in Table 35 above. The Table reveals that CS obtained an overall rank of 10th out of 21 studies. Interestingly, CS actually obtained a higher mean validity rank than visual acuity testing in the present study under the original referral criteria. Moreover in some cases, CS ranked higher than commonly used acuity tests such as the HOTV test (VIP Study Group, 2004; 2005ab), the Teller Acuity Cards (Schmidt, 1994), and the Snellen Test (Kennedy et al., 1995). As with the visual acuity and stereoacuity tests within the present study, CS measurement yielded high PPV (88%, ranked 3rd overall) but low NPV (70%; ranked 12th overall). In addition, the Table indicates that CS possessed low sensitivity (56%). These latter two findings are likely the result of a methodological problem inherent within the current version of the CS cards, namely a ceiling effect. Specifically, children with vision disorders could sometimes detect even the lowest contrast gratings at low to mid SFs. However, this problem has been remedied

Table 38. A comparison of the validity of CS in the present study to the validity of visual acuity tests. The results of the present study are provided in bold. Numbers in parentheses represent each study's rank on that particular measure of validity. Studies are listed in ascending order based on mean rank on all validity measures.

Study (Rank)	Test	Sensitivity	Specificity	PPV	NPV	Mean Rank
(1) Nordlöw & Joachimson (1962)	Boström Hooks (5m; single)	95 (3)	99 (2)	86 (4)	100 (1)	2.5
(2) Raina (1998)	Wright Cards	100 (1)	95 (8)	77 (8)	100 (1)	4.25
(3) Ladenwall (1988)	HOTV (distance and optotype not reported)	87 (7)	97 (4)	86 (4)	98 (5)	5
(4) Kvarnström et al. (1998)	HOTV (3m; linear)	92 (4)	97 (4)	69 (11)	99 (3)	5.5
(5) Spowart et al. (1998)	Stycar (6m; single), Glasgow Acuity Cards (3m; linear)	83 (8)	95 (7)	40 (14)	99 (3)	8
(6) Present Study (Lenient Criteria)	**Lea Symbols Chart & Patti Pics Chart (3m; linear) Lea/Patti Pics Isolated Symbols (3m; single); Patti Pics Card (3m; single), Teller Acuity Cards**	**54 (14)**	**98 (3)**	**97 (2)**	**66 (14)**	**8.25**
(6) Raina (1998)	Teller Acuity Cards	15 (21)	100 (1)	100 (1)	87 (10)	8.25
(8) Simon et al. (2004)	Enfant II VEP Headband	97 (2)	81 (19)	71 (10)	98 (5)	9
(9) Arnold et al. (2004)	HOTV (distance and optotype not reported)	91 (6)	86 (17)	79 (7)	94 (7)	9.25
(10) Contrast Sensitivity (Present Study)	**CS Cards, CS Booklet**	**56 (13)**	**93 (10)**	**88 (3)**	**70 (12)**	**9.5**
(11) Kennedy et al. (1995)	Snellen Test (distance and optotype not reported), Stycar Graded Balls	33 (20)	97 (4)	54 (13)	94 (7)	11
(12) Schmidt (1994)	Broken Wheel Test	92 (4)	53 (21)	57 (12)	89 (9)	11.5

(12) VIP Study Group (2004)	Lea Symbols (3m; linear)	61 (11)	90 (12)			11.5
(12) Present Study (Original Criteria)	**Lea Symbols Chart & Patti Pics Chart (3m; linear) Lea/Patti Pics Isolated Symbols (3m; single); Patti Pics Card (3m; single), Teller Acuity Cards**	**67 (9)**	**84 (18)**	**82 (6)**	**69 (13)**	**11.5**
*(12) VIP Study Group: Lay Screeners (2005b)	Lea Symbols (1.5m; single)	61 (11)	90 (12)			11.5
(16) Schmidt (1994)	Teller Acuity Cards	64 (10)	81 (19)	75 (9)	71 (11)	12.25
(17) VIP Study Group (2005a)	Lea Symbols (3m; linear)	49 (16)	94 (9)			12.5
(18) VIP Study Group: Nurses (2005b)	Lea Symbols (3m; single)	49 (16)	90 (12)			14
(19) VIP Study Group (2005a)	HOTV (3m; linear)	36 (19)	93 (10)			14.5
(20) VIP Study Group (2004)	HOTV (3m; linear)	54 (14)	89 (16)			15
*(20) VIP Study Group: Lay Screeners (2005b)	Lea Symbols (3m; single)	37 (18)	90 (12)			15

* Testers were trained lay screeners.

with the recent development of the CS booklet, which possesses much lower contrast levels at all SFs. As a result, the CS booklet should be able to detect substantially more cases of vision disorders and therefore, yield higher sensitivity and NPV.

Another important point of interest regarding CS was whether it could provide valuable information on functional vision that was not provided by visual acuity tests. In particular, we wished to determine whether the CS tests could detect cases of visual dysfunction that escaped detection from the visual acuity tests. An inspection of the data

revealed that the CS cards identified two children who possessed a vision disorder (one with reduced stereoacuity, a second with hyperopia) that were not detected by visual acuity tests. Thus, data are scant at this time but it is expected that given its greater precision, such results may become more common with the implementation of the CS booklet.

Autorefraction

The autorefraction validity data from the present study, and those from other studies (N = 12) are presented in Table 39 below. Note that autorefraction studies that assessed children *only* on cylindrical refractive error (i.e., astigmatism) or spherical refractive error (i.e., hyperopia, myopia, or anisometropia) were not included in the Table (Cordonnier & Dramaix, 1998; Cordonnier & Dramaix, 1999). These studies were excluded as noncycloplegic autorefraction tends to provide more accurate measures of cylindrical refractive error as opposed to spherical refractive error (Iurno, Grant, Noël, 2004; Steele, Ireland, & Block, 2003; Suryakumar & Bobier, 2003; Zhao, Mao, Luo, Li, Pokharel, & Ellwein, 2004). Thus, relative to studies that screen for all amblyogenic factors, studies that measure only cylindrical refractive error only would likely provide artificially high validity estimates. Conversely, studies that assess only spherical refractive error would likely yield artificially low validity estimates.

Table 39. Summary of the validity of autorefraction screening studies. Note that the results of the present study are provided in bold. Numbers in parentheses represent each study's rank on that particular measure of validity. Studies are listed in ascending order based on mean rank on all validity measures.

Study (Rank)	Test	Sensitivity	Specificity	PPV	NPV	Mean Rank
(1) Present Study	**SureSight Autorefractor**	**64 (6)**	**94 (1)**	**93 (1)**	**70 (5)**	**3.25**
(2) Büchner et al. (2005)	SureSight Autorefractor	94 (1)	70 (11)	82 (2)	89 (4)	4.5
*(2) Williams et al. (2000)	Topcon PR-2000	88 (2)	82 (10)	51 (3)	97 (3)	4.5
(2) VIP Study Group: Nurses (2005b)	Nikon Retinomax Autorefractor	68 (5)	90 (4)			4.5
(5) VIP Study Group (2004)	Nikon Retinomax Autorefractor	64 (6)	90 (4)			5
(5) VIP Study Group: Nurses (2005b)	SureSight Autorefractor	64 (6)	90 (4)			5
(7) Barry & König (2001)	Nikon Retinomax Autorefractor	80 (3)	58 (13)	5 (4)	99 (1)	5.25
(8) Barry & König (2001) lenient criteria	Nikon Retinomax Autorefractor	70 (4)	60 (12)	4 (5)	99 (1)	5.5
(9) VIP Study Group (2004)	Nikon Retinomax Autorefractor	52 (12)	94 (1)			6.5
(9) VIP Study Group (2004)	SureSight Autorefractor	63 (9)	90 (4)			6.5
(11) VIP Study Group (2004)	SureSight Autorefractor	51 (13)	94 (1)			7
† (11) VIP Study Group: Lay Screener (2005b)	Nikon Retinomax Autorefractor	62 (10)	90 (4)			7
† (13) VIP Study Group: Lay Screeners (2005b)	SureSight Autorefractor	61 (11)	90 (4)			7.5

* Data represent weighted means of validity measures obtained when screening for anisometropia, astigmatism, and hyperopia/myopia (see the three Williams et al. studies in Table 3).
† Children were screened by trained lay screeners.

The Table indicates that the results of the present study attained the highest mean validity, ranking 1st overall out of 13 studies, and yielded the highest specificity and PPV estimates. However, sensitivity ranked moderately (6th overall) whereas NPV ranked last overall. It is likely that slightly stricter criteria such as those of Büchner et al (2005)[41] would have increased the sensitivity of the present study, and perhaps NPV. Interestingly, the Table demonstrates that two of the three highest ranking studies utilized the SureSight autorefractor to measure refractive error (Büchner et al, 2005; the present study). Yet, it is still not clear whether the SureSight autorefractor is the best tool to assess refractive error, as two other studies that used this test ranked last and second last overall (VIP Study Group, 2004; 2005b). The major competing autorefractor, the Nikon Retinomax, was also variable, ranking from 2nd and 11th overall.

A secondary interest of the present study was to compare the accuracy of the SureSight autorefractor to the accepted gold standard estimate of refractive error, namely cycloplegic retinoscopy as conducted by an eyecare specialist. Thus, Pearson r was calculated to determine the relationship between two procedures. Specifically, estimates obtained from the right eye of children (N = 31) with the autorefractor during screening were compared to those obtained from the same eye with cycloplegic retinoscopy during the optometric gold standard exam. This analysis revealed a significant correlation for both spherical ($r = 0.64$, $p < 0.001$) and cylindrical refractive error ($r = 0.94$, $p < 0.001$). Note that the stronger relationship between estimates of cylindrical refractive error as opposed to estimates of spherical refractive error is in agreement with the studies cited

[41] In comparison to the present study, the criteria of the Büchner et al. (2005) are stricter for hyperopia (> 3D vs. > 3.5D), astigmatism (> 1.25D vs. > 1.5D) and anisometropia (>1D vs. > 1.75D).

above (Iurno, Grant, Noël, 2004; Steele, Ireland, & Block, 2003; Suryakumar & Bobier, 2003; Zhao, Mao, Luo, Li, Pokharel, & Ellwein, 2004).

Comparison of Combinations of Tests of Visual Functions to Other Studies

As part of the progressive analyses above, the validity of 29 different combinations of tests of functional vision was analyzed. In this subsection, these results are compared to those of other screening studies to determine the relative effectiveness of the combinations of the present study, and to consider which visual functions should be assessed as part of an effective preschool vision screening program. The validity of the top 10 combinations within the present study is compared to that of the vision screening studies summarized in Table 3 (see Introduction: Validity of Vision Screening Programs subsection above). To determine the top 10 combinations of the present study, an average validity estimate (i.e., average of sensitivity, specificity, PPV, and NPV) was calculated for each combination. Those combinations possessing the 10 highest validity averages were chosen and are shown in Table 40 below. Note that two of the three variations of the five-test combination (i.e., see Table 32) were in the top 10 validity averages. However, only the variation that exclusively considered those children who received the gold standard exam *after* CS was added to the program (i.e., the first variation in Table 32) truly represents a five test combination.[42] Thus, it is the lone version represented in Table 40. The validity of each combination was then compared separately to the studies in Table 3 and for each measure of validity, given a rank out of 71 (i.e., the 70 studies in Table 3 plus the combination of the present study). The mean

[42] This is because almost half of the subjects included in the other two versions of the five test combination were screened when the program included only *four* tests.

rank on the four measures is also provided in the rightmost column. Finally, the leftmost column of the Table contains each combination's overall rank out of 71 studies, and the combinations are listed in ascending order of overall rank.

Table 40. A comparison of the validity of test combinations from the present study to previous vision screening studies from Table 3. Numbers in parentheses represent that combination's rank (out of 71) on that measure of validity. Combinations are listed in descending order of mean validity rank.

Visual Functions (Overall Rank)	N	Sensitivity	Specificity	PPV	NPV	Mean Rank
(10) Autorefraction + CS + Visual Acuity + Stereoacuity + Alignment/Motility	63	**97 (5)**	87 (40)	**89 (14)**	**96 (24)**	20.75
(12) Autorefraction + Visual Acuity + CS + Alignment/Motility	62	**94 (10)**	**90 (30)**	**91 (13)**	93 (35)	22
(16) Autorefraction + CS + Stereoacuity + Alignment/Motility	62	**94 (10)**	87 (40)	**88 (15)**	93 (35)	25
(18) Autorefraction + Visual Acuity + Stereoacuity + Alignment/Motility	110	**98 (2)**	75 (59)	**82 (24)**	**97 (22)**	26.75
(21) Autorefraction + CS + Alignment/Motility	62	**90 (23)**	**90 (30)**	**90 (14)**	90 (45)	28
(27) Autorefraction + Stereoacuity +Alignment/Motility	110	**95 (8)**	76 (57)	**82 (24)**	93 (35)	31
(36) Autorefraction + Visual Acuity +Alignment/Motility	110	**92 (15)**	78 (57)	**83 (22)**	89 (47)	35.25
(39) Autorefraction + Visual Acuity + CS + Stereoacuity	62	84 (34)	87 (40)	**87 (15)**	84 (56)	36.25
(41) CS + Visual Acuity + Stereoacuity + Alignment/Motility	60	83 (36)	87 (40)	**86 (15)**	84 (56)	36.75
(41) Autorefraction + Visual Acuity + Stereoacuity	110	88 (27)	80 (52)	**84 (20)**	85 (55)	38.5

The Table confirms the trends of the Results section. Specifically, autorefraction is essential to vision screening as it was included in 9 of the top 10 combinations. This finding contradicts the literature review above (see Tables 3 and 4) which indicated that visual acuity was generally the most critical component of an effective vision screening program. It should be pointed out however, that tests of visual acuity were relatively important in the present study as they were included in 7 of the top 10 combinations. The superior performance of autorefraction is even more surprising given its relatively poor rankings in vision screening studies summarized in Table 3. Yet these poor rankings may reflect the fact that all previous autorefraction studies utilized the autorefractor as the lone screening tool to assess children. This is unfortunate as our results suggest that the addition of a single test to a screening program that includes autorefraction can substantially increase sensitivity and NPV. In addition, the Table indicates that tests of alignment/motility were included in 8 of the top 10 combinations, confirming their importance as outlined above (see Results section). Finally, stereoacuity tests were included in 7 of the top 10 combinations, whereas CS was included in 6 of the top 10 combinations.

Further inspection of the Table 40 reveals important findings regarding the estimates on each measure of validity. For instance, the sensitivity and PPV of the combinations in the Table are relatively high and range in rank from 2nd to 36th, and 13th to 24th, respectively. Collectively, this suggests that most combinations detected over 90% of children with vision disorders and that the majority of children who tested positive did indeed possess a vision disorder. In contrast, the combinations yielded

relatively low specificity, ranging in rank from 30th to 59th. This implies that compared to other studies, children without disorders were often incorrectly classified as positive during screening. Furthermore, the NPV of the combinations in Table 40 are also low, ranging in rank from 22nd to 56th, suggesting that children who truly possessed vision disorders were often incorrectly classified as negative during screening. The low NPV of test combinations reflects similar results that were reported above for tests of individual visual functions. Moreover, as above, these results may reflect the limitations of the tests rather than referral criteria as adjustment of referral criteria to increase specificity could potentially reduce NPV. In fact, this was the case when stereoacuity and visual acuity criteria were reduced in the progressive validity analyses. Therefore, in order to maximize the validity of our screening program, it is perhaps necessary to remove certain tests and replace them with those that provide both high specificity and NPV. Yet, this is problematic because it is not clear which tests are superior as their validity estimates vary widely across studies (e.g., Randot E Stereotest, Nikon Retinomax, Stycar-HOTV tests). Also, because different studies often assess different age groups, it is difficult to determine whether tests that yield high validity (e.g., TNO stereoacuity test) can be completed by young children who often become fussy or distressed during testing.

Cost-Effectiveness of Government Funded Vision Screening

Given that the results of the present study suggest that preschool vision screening is useful in detecting cases of treatable vision disorders, it is important to determine whether permanent, wide-scale, vision screening is feasible. Therefore, this subsection provides an economic evaluation of a population-based, preschool vision screening

program based on the results of the present study. Specifically, I evaluate a hypothetical program funded by government health care, and conducted entirely by a dedicated health care professional (a public health nurse) with a strong background in community health, assessment, and preventive care. In order to determine which visual functions should be assessed in the program, the cost-effectiveness per annum of several of the most effective combinations summarized in Table 40 is calculated. Cost-effectiveness is an index of the relationship between the cost of a program and its beneficial effects and is defined as the cost of vision screening per the number of newly detected cases of vision disorders (König, Barry, Leidl, & Zrenner, 2000). It should be pointed out that calculation of cost-effectiveness of the present study is modeled closely on the procedure of König et al. (2000). Note that only those combinations that yielded a sensitivity of at least 90% were considered for this evaluation, as this was the criterion for an effective and valid program for this measure of validity.

Calculation of Costs

Firstly, all costs were based on the assumption that testing could be conducted over 156 work days per year, with 2.5 hour sessions per day. These numbers were chosen as they reflect our experiences with local daycares over the past 2 years. The costs considered for cost-effectiveness calculations include the cost of labour, transportation, materials, and optometric examinations. Cost of labour was based on a wage of $28.28, the average salary of a registered nurse in the province of Newfoundland and Labrador. Estimates of hours worked included time spent on actual assessments (including time required to set up and dismantle equipment, and time spent testing),

consultation with daycare directors and parents, as well as travel time, and reflect hours worked in the present study. Cost of transportation was determined based on an average driving distance of 8 km per testing session (round trip) at a price of 8.6¢ per km. The cost of materials included all testing equipment along with consent forms and data sheets. Finally, the cost of an optometric exam was $45.00, the standard fee, and that charged by the optometrist who was part of our research team. Note that the cost of unnecessary optometric exams (i.e., exams of children with normal vision) was also included based on the proportion of false positives for each combination. It is important to point out that the costs calculated in these analyses only included those incurred up to the diagnosis and do not include treatment as these costs are borne by the patient's family or private health care insurance.

Effectiveness

The index of effectiveness for a screening combination was the number of previously undetected cases of vision disorders as identified by that combination. The calculation of this number involved several steps. First, the number of children that could be assessed with each combination was calculated based on the number of testing sessions conducted during the year and the mean completion times from Table 21. Note that mean "completion times" were also determined for children who could not complete one or more tests based on our experiences from the present study. Second, the number of cases of undetected vision disorders was calculated based on the prevalence rate of undetected disorders from the present study, i.e., 12.5%. Note that all conditions in Table 18 were considered target conditions as the vast majority of them require

treatment and/or monitoring from an eyecare specialist. Finally, the number of cases of successfully detected disorders was calculated by determining the number of existing cases and multiplying that value by the sensitivity of that combination.

Cost-Effectiveness Ratios

Cost-effectiveness ratios (CERs) were calculated as the total cost required for the combination of tests divided by the number of cases of newly detected disorders. Thus, the lower the CER, the greater the cost-effectiveness of the combination. The CERs for seven different testing combinations are provided in Table 41 below in order of decreasing cost-effectiveness. Note that for combinations that include visual acuity testing, CERs were calculated both including and excluding the cost of the Teller Acuity Cards. This measure was taken as the Teller Acuity Cards is a relatively expensive test of acuity ($3474.20) and accounts for an extremely high proportion (90%) of the total cost of visual acuity tests. However, it was rarely used to assess children in the present study. Moreover, we have recently replaced this test with the cheaper Patti Pics cards ($46.40; Precision Vision, La Salle Illinois, USA). Also, cost-effectiveness calculations of combinations that include CS are based on the CS cards as the CS booklet has only recently been added to the screening program.

The Table highlights several important findings. For instance, the Table indicates that the greater the number of children that can be assessed per year, the greater the cost-effectiveness (see the N column). Related to this point, the fewer tests included in the combination, the greater the cost-effectiveness. In fact, the top three combinations that possess the lowest CERs, consist of three tests, whereas the remaining combinations

possess the highest CERs and consist of four or five tests. This suggests that combinations consisting of fewer tests require less equipment and thus, reduce cost (see Materials column). Furthermore, combinations that include fewer tests generally require less time and therefore, allow more children to be tested per year (see N column once again). Finally, the Cases Detected column indicates that generally, the greater the number of cases of newly detected vision disorders, the higher the cost-effectiveness of a combination. Collectively, these results imply that in order to optimize cost effectiveness, a sensitive three-test combination should be implemented to detect a large number of cases of undetected disorders while minimizing cost and time spent per screening.

Despite these conclusions, it should be noted that some of the results in the Table may be artificial. In particular, combinations that include tests which are sometimes difficult to complete, such as CS and alignment/motility tests, have relatively deflated CERs. This is because as mentioned above, completion times used in the calculations not only accounted for children who could complete each test, but also those who could not complete one or more tests. Specifically, in many cases during the study, no attempt was made to assess fussy or distressed children with these procedures and thus, "completion times" calculated for these children are notably low (i.e., 15-30 seconds). As a result, completion times for combinations that include these difficult tests are perhaps artificially low and the number of children that can be assessed is relatively high, leading to low and perhaps, inaccurate CERs. This caveat notwithstanding, the general conclusions stated above still apply, i.e., cost-effective vision screening should limit the number of tests,

time required for screening, and cost of materials, while at the same time, maximizing sensitivity in order to detect a large number of cases of undetected vision disorders.

Table 41. Simulated cost-effectiveness ratios of different combinations of tests. Note that the numbers in parentheses represent calculations that exclude the Teller Acuity Cards. N column denotes the number of children that can be assessed per year.

Visual Functions (CER Rank)	N	Wages + Transportation	Materials	Optometric Exams	Total Cost	Cases Detected	Cost-Effectiveness Ratio
(1) Autorefraction + Stereoacuity +Alignment/Motility	1919	$20 636.24	$10 211.65	$12 285.00	$43 132.89	228	$189.18 per detected case
(2) Autorefraction + CS + Alignment/Motility	1663	$20 415.75	$10 069.20	$9 405.00	$39 899.95	187	$213.31 per detected case
(3) Autorefraction + Visual Acuity +Alignment/Motility	1663	$20 409.35	$13 508.58 ($10 034.38)	$10 260.00	$44 177.93 ($40 703.73)	191	$231.30 per detected case ($213.11)
(4) Autorefraction + CS + Stereoacuity + Alignment/Motility	1279	$20 071.41	$10 408.10	$7 650.00	$38 129.51	150	$254.20 per detected case
(5) Autorefraction + Visual Acuity + Stereoacuity + Alignment/Motility	1279	$20 071.41	$13 847.50 ($10 373.30)	$8 415.00	$42 333.91 ($38 859.71)	157	$269.64 per detected case ($247.51)
(6) Autorefraction + Visual Acuity + CS + Alignment/Motility	1151	$19 960.36	$13 752.80 ($10 278.60)	$6,705.00	$40 418.16 ($36 943.96)	135	$299.39 per detected case ($273.66)
(7) Autorefraction + CS + Visual Acuity + Stereoacuity + Alignment/Motility	1023	$19 844.42	$14 199.95 ($10 725.75)	$6 255.00	$40 299.37 ($36 825.17)	124	$324.99 per detected case (296.98)

Other Major Issues and Directions for Preschool Vision Screening

As the preceding subsections suggest, two major concerns of the present study were cost effective vision screening and the evaluation of screening tests/tools. Therefore, it is relevant to discuss two recent developments in these areas which may impact the future of preschool vision screening. For instance in a recent preschool vision screening study, Lim, Yu, Park, Ahn, Kim, Lee, Jeong, Shin, and Koo (2004) conducted a stepwise vision screening program in which the first step required parents to prescreen their children at home. Screening kits containing cards that consisted of five picture optotypes (fish, butterfly, airplane, duck, and car) and instructions explaining how to measure monocular acuity using the cards were delivered to preschoolers' kindergarten classrooms. Also included in the kits were questionnaires concerning the visual health of the children (existence of strabismus, ptosis, sensitivity to light, frequent blinking, etc.). Preschoolers who failed the visual acuity test or possessed poor visual health as determined by the questionnaire, were screened by nurses at a public healthcare centre. Children who failed this screening were sent to eye clinics for an ophthalmological gold standard exam. Although complete validity data were not provided, the program reported a relatively high PPV of 77%. Perhaps even more importantly, 35 226/36 973 (95%) preschoolers were screened on a limited budget from the Korean healthcare system. The study demonstrated that prescreening could be conducted by parents at home for a fraction of the cost required for screening conducted by health care professionals. Thus, the addition of home prescreening to a vision screening study would greatly improve

cost-effectiveness by limiting labour costs. Yet, one must be cautious regarding the implications of the Lim et al. (2004) study as they do not provide data regarding the true disease status of children who were prescreened at home and who tested negative. This is an important consideration as parents were given no formal training in vision testing and thus, it is possible that a relatively high percentage of children who tested negative during prescreening may actually possess a vision disorder. If this was indeed the case, prescreening may not represent a viable first step in preschool vision screening.

The second development discussed here, receiver operator characteristics (ROC) curves, is relevant to the future of the present study as it deals with precise, systematic evaluation of a screening tool. ROC curves arise from signal detection theory, which attempts to quantify reasoning and decision-making while under conditions of uncertainty (Egan, 1975). In terms of health screening, this refers to the attempt to detect diseases/disorders which may or may not be present. ROC curves have been used over the past two decades to evaluate the validity of screening tests/tools in detecting a variety of diseases/disorders including cancer, iron deficiency, and pneumonia, (Baker, Bowton, & Haponik, 1995; Carter, Lau, Fowler, Carlson, Carson, & Twiggs, 1995; Kim, Pollitt, Leibel, Viteri, & Alvarez, 1984; Kodoi, Yoshishara, Sumii, Haruma, & Kajiyama, 1995). Importantly, ROC curves have recently been applied to vision screening (Miller, Dobson, Harvey, & Sherill, 2001). As some tests of functional vision allow only one of two possible outcomes, i.e., pass or fail (e.g., tests of alignment/motility), the performance of

these tests can be summarized with a single measure of sensitivity and specificity (Park, Goo, & Jo, 2004).[43] However, most tests of functional vision allow a number of possible outcomes (i.e., scores), either of which may be used as the criterion for referral. For example, a researcher/clinician may choose one of several scores on a test of visual acuity (20/25, 20/32, 20/40, 20/50, 20/64, etc.) as the referral criterion. Thus, the *true* performance of these tests cannot be represented by a single sensitivity and specificity estimate, but must be represented by several pairs of sensitivity and specificity estimates, namely one pair for each referral criterion. In such a case, the performance of the tool/test can be represented in the form of a ROC curve, a graphical representation of the validity of a screening tool which consists of a plot of sensitivity vs. 1 – specificity for all possible referral criteria (see Figure 17 below; also see Park et al., 2004 for an excellent description). Thus, the curve provides an index of both sensitivity and specificity across all referral criteria. Although a ROC curve may be constructed by plotting sensitivity vs. 1 – specificity for all criteria, a less time consuming option is to plot sensitivity vs. 1 - specificity for fewer referral criteria and create a fitted ROC curve from these criteria based on mathematical assumptions (Park et al., 2004). Two examples of fitted ROC curves are shown in Figure 17 below.

An important aspect of the ROC curve is the area under the curve which represents the average sensitivity at all specificity levels. Therefore, the greater the area under the curve, the higher the average sensitivity across all specificity values and thus, the better the validity of the screening test/tool. This concept has important implications

[43] Note, the inclusion of both PPV and NPV allow a more complete measure of validity.

as it suggests that different screening tests/tools can be compared simply by inspecting the area under their respective ROC curves. Moreover, this comparison can be made across all possible referral criteria at once. This point is illustrated nicely in Figure 17. Note that the ROC curve in Figure 17 B possesses greater area under the curve than the ROC curve in Figure 17 A, particularly at high specificity levels. This indicates that the test/tool represented in Figure 17 B possesses higher sensitivity, especially at high levels of specificity and is therefore, the superior screening tool. Despite the potential of ROC curves as an evaluative technique, they have only recently been applied to vision screening (see Miller et al., 2001). Unfortunately, due to the flexible, evolving nature of our screening battery, a number of alternative (eg., TAC, Patti Pics cards) and replacement tests (Patti Pics chart, Randot Preschool Stereoacuity Test) were implemented, and as a result, relatively few children were assessed with each single test.[44] This precluded evaluation of individual tests using ROC curves. However, given the advantages of this method of evaluation, ROC curves will likely be used to compare screening tests in the future of our program when more children are assessed with each test. This will allow definitive recommendations to be made regarding which tests should be included in preschool vision screening programs.

---

[44] This problem is discussed further in the next subsection. Note also that ROC curves can not be used to evaluate alignment/motility tests as they allow only two outcomes, nor can they be used to evaluate autorefraction as the referral criteria are multi-dimensional (i.e., it possesses separate referral criteria for hyperopia, myopia, anisometropia, and astigmatism; VIP Study Group, 2004).

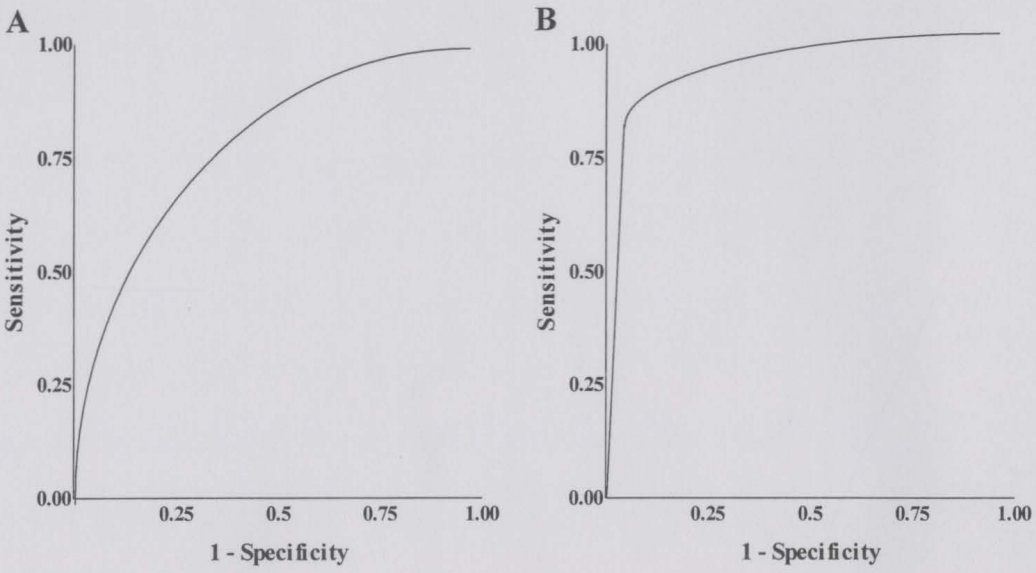

Figure 17. Hypothetical, fitted receiver operator characteristics (ROC) curves. Note that the area under curve B is greater than under curve A. Thus, curve B represents the superior screening tool.

Methodological Limitations of the Present Study

Despite the merits of our vision screening program, there are important methodological limitations that must be addressed. For instance, although 946 children were tested within the vision screening program, only 110 of them actually received the optometric gold standard exam. Ideally, all children would have received the exam as this would ensure that all validity calculations were precise. However, this is impractical as it is very expensive and time-consuming (but see the VIP Study Group 2004; 2005a; 2005b for rare exceptions). Importantly, only 34 of the 110 children who received gold standard exams were among those who tested negative during the screening (i.e., negative children). This suggests that children who passed all screening tests were

underrepresented, a factor which may have affected the accuracy of validity estimates. However, the disproportionately low number of negative children was unavoidable for two reasons. First, although parents of 145 of these negative children were invited for *free* screening exams, very few parents accepted the invitation. This is not surprising however, as parents often consider the time and effort required to attend an eye exam as unnecessary as their child already passed all of the tests in our comprehensive screening battery. Second, because the gold standard exams of these children were paid for with research funds designated primarily for other purposes, relatively few negative cases could be invited for the follow-up.

A second methodological limitation is that because children who tested negative were not randomly chosen to attend the optometric gold standard exam, they may represent a biased sample. For instance, one might expect that even though a child tested negative, many of those parents who accepted the invitation for the gold standard exam may have suspected that their child possessed a vision disorder. Thus, an invitation for a free exam provided parents with an opportunity to confirm their suspicions. In such a case, there would be a high number of false negatives in the sample of children who received the optometric gold standard exam. As a result, sensitivity and NPV (both of which require the number of false negatives as part of the denominator) may be artificially low. On the other hand, those who accepted the invitation may represent a sample of parents who are particularly knowledgeable, vigilant, or conservative with respect to their child's health. Therefore, one might expect that due to their persistent vigilance, few of their children would possess a vision disorder. If this was indeed the

case, our validity estimates would be artificially high. However, despite these assumptions, because the direction of bias is unknown, so are the effects on the results of the present study. In light of this, it would be wise in the future to contact the parents of these children and inquire about their reasons for accepting our invitation for an eye exam.

A third methodological limitation of the present study is whether the team optometrist was truly masked to the screening status of each child during the optometric gold standard exam. Although parents were instructed not to inform the optometrist of their child's screening result, it is likely that in some cases that this information was revealed during the exam or even the scheduling of the exam. Note however, that the optometrist was completely masked to the nature of the data analysis of the present thesis until the study was completed. Furthermore, the optometrist was not concerned with the screening status of the child or the data analyses as his mandate is simply to detect visual dysfunction.

The present study was also limited by the flexible nature of the screening battery. Specifically, alternative tests (e.g., Patti Pics cards, Teller Acuity Cards) or replacement tests (e.g., CS booklet, Randot Preschool Stereoacuity Test, Patti Pics Chart) were added for three of the five aspects of functional vision that were assessed (visual acuity, stereoacuity, and CS). As a result, relatively few children completed each single test within these aspects of functional vision, thus precluding the calculation of validity on a test by test basis. Therefore, recommendations can not be made regarding which particular tests should be employed to screen preschoolers. However, the addition of

these tests is justified as they are often easier for young children to complete or possess merits that may allow them to provide more accurate estimates of visual functioning. Furthermore, the inclusion of these additional tests does not prevent us from making recommendations regarding which visual functions should be assessed as part of a vision screening program.

Another important methodological limitation of the present study was the limited experience of the primary examiner (JRD). Although this examiner conducted a thorough literature review in order learn the testing procedures and was given instructions from the program supervisor, Dr. Adams, no formal training procedure was provided. As a result screening results may be affected by the examiner's limited experience, particularly at the beginning of the program. Indeed, the lower specificity in the first half of the screening program (i.e., before CS testing was added to the program) relative to the second half of the program (see Results subsection entitled "Calculation of Validity: Five-Test Combinations") suggests that screening followed a learning curve. That is, the accuracy of the program increased as the primary examiner gained more experience.

A final methodological limitation of the present thesis is that we were unable to obtain medical records of children (N = 23) who received gold standard exams from other ophthalmologists/optometrists. Thus, it was necessary to base true/false positive classifications on whether these children received treatment (e.g., spectacles, patching, orthoptic therapy, etc.). As a result, it is possible that this crude criterion of disease classification may have affected the accuracy of validity measures. In fact, this notion is supported by the finding that children who failed screening and received gold standard

exams from other ophthalmologists/optometrists were more likely to be classified as having a disease/disorder (83% were classified as true positives) compared to those who failed screening and received the gold standard exam from the team optometrist (74% were classified as false positives). Therefore, it is possible that this higher rate of disease/disorder classification may have inflated the overall sensitivity and PPV of the screening program.

Recommendations for Preschool Vision Screening

In spite of the methodological limitations of the present study, the results suggest that the screening program was highly successful and that several conclusions can be drawn regarding the validity and cost-effectiveness of preschool vision screening. For instance, the validity data of the present study indicate that unlike many previous studies, a single test (or tests of a single visual function) should *not* be utilized as the sole procedure in a vision screening program. Specifically, sensitivity and NPV for each test is quite low (sensitivity = 28% to 67%; NPV = 52% to 70%), indicating that many children who possess vision disorders are not correctly identified during screening. Moreover, low sensitivity and NPV are particularly detrimental for vision screening programs because as mentioned above, parents are unlikely to bring their children in for an optometric exam following a negative screening (i.e., the child passed all screening tests), even if the exam is free.

In light of this conclusion, it is apparent that preschoolers should be assessed with a combination of tests of several visual functions. However, when testing children with a combination, we recommend the implementation of relatively lenient referral criteria to

avoid low specificity and therefore, overreferrals. Furthermore, because vision tests are specialized in terms of the category of the vision disorder that they can detect (i.e., spatial vision disorders vs. alignment/motility disorders), it is important that the combination includes at least one test that assesses spatial vision, and at least one that assesses alignment/motility. The best option for a test that assesses spatial vision is the Welch-Allyn SureSight autorefractor. This screening tool possesses three important merits that warrant its use as part of any screening program. First, the autorefractor can obtain fairly accurate estimates of refractive error from almost all toddlers and preschoolers in less than 1.5 minutes; almost half the time required by most other procedures. Second, the procedure is very simple and therefore, the tester requires limited training. Third, the autorefractor yields the highest estimates on almost all measures of validity (see Table 25 from the Results section).

Although autorefraction is the obvious choice to assess spatial vision in preschoolers, it is difficult to determine which test should be used to assess alignment/motility. The alignment/motility tests (i.e., cover tests, Hirschberg, motility, and convergence/divergence) of the present study were the most sensitive to alignment/motility disorders (sensitivity = 64%; see Table 26 from the Results section). However, it should be pointed out that stereoacuity tests are generally easier for children to complete and also require less tester training. Thus, we feel that either would be appropriate for preschool vision screening. Furthermore, a hybrid of the two procedures may also be considered. For example, if a child cannot complete the most difficult of the

alignment/motility tests, i.e., the cover-uncover tests, perhaps he/she can be assessed with a stereoacuity test.

Finally, the data from the present study suggest that children should be assessed with combinations of three to four tests (i.e., tests of three to four functions) such as those presented in Table 41. These combinations possess relatively high validity estimates, implying that most cases of disorders will be detected and few children without disorders will be classified as positive. Furthermore, the cost-effectiveness calculations of the present study imply that perhaps three-test combinations, such as the top three combinations presented in Table 41 (i.e., autorefraction, stereoacuity, and alignment/motility; autorefraction, CS, and alignment/motility; autorefraction, visual acuity, and alignment/motility) should be considered for preschool vision screening as they possess a number of advantages over four- and five-test combinations. First, screening combinations that include three tests limit the cost of materials, the cost of labour (per child), and time spent per screening (i.e., completion time). Second, due to the relatively short time required per screening, a greater number of children can be tested with these combinations. Third, because more children with vision disorders can be detected with three-test combinations, they are more cost-effective (see Table 41 above).

The Future of Our Vision Screening Program

The impetus for preschool vision screening is the improved prognosis due to early detection and treatment of visual dysfunction. However, some researchers posit that treatment outcomes may be further improved by screening younger children such as

infants and toddlers (Eibschitz-Tsimhoni et al., 2000). Indeed, there is empirical evidence that treatment earlier in the period of plasticity (i.e., before 2 years of age) reduces the duration of pre-existing visual dysfunction, thereby providing greater benefits than treatment during the preschool years (Birch, Fawcett, & Stager, 2000; 2002; Fawcett, Leffler, & Birch, 2000). In light of this evidence, we are currently considering screening younger children as part of our vision screening program. Specifically, we propose to screen children as young as 12 months of age with screening techniques that require limited attentional demand such as photorefraction and autorefraction. Importantly, both techniques are relatively objective, and if autorefraction is combined with the Hirschberg test, both can potentially detect spatial vision *and* alignment/motility deficits. Thus, these techniques may allow for effective, early treatment of visual dysfunction, and as a result, improve visual outcomes and substantially reduce the prevalence of amblyopia.

Conclusions

In the present study, we implemented a comprehensive yet flexible vision screening program designed to detect even subtle cases of vision disorders in toddlers and preschoolers. This thesis represents a critical first step in the implementation of a long-term screening program to be conducted by our laboratory in the coming years. This is an important undertaking as children of this age are undergoing a period of brain plasticity during which diseases/disorders must be detected and treated promptly to prevent permanent visual dysfunction. Our results indicate that the program was highly successful as to date, a total of 58 children have been identified who require treatment for

a variety of undetected vision disorders. Furthermore, the program was remarkably accurate as the majority of children referred based on screening displayed some form of visual dysfunction. The present study is also particularly relevant to current vision screening research for several reasons. For instance, it is first to provide estimates of prevalence of vision disorders in toddlers and preschoolers in Newfoundland and Labrador, and is one of few studies to do so in all of Canada. Importantly, these rates were shown to agree with estimates from other areas of Canada, and other industrialized nations obtained by licensed eyecare professionals. Also, this study is the first to evaluate contrast sensitivity as part of a vision screening program, and one of the first to evaluate the Lea Symbols, the Patti Pics symbols, the Randot Preschool Stereoacuity Test, and the Welch-Allyn SureSight autorefractor. Finally, this study provides the most detailed evaluation of vision screening tests/tools of any screening study conducted to date and is the first to make clear recommendations regarding which visual functions should be assessed as part of a preschool vision screening program.

References

Abrahamsson, M., Magnusson, G., & Sjostrand, J. (1999). Inheritance of strabismus and the gain of using heredity to determine populations at risk of developing strabismus. *Acta Ophthalmologica Scandinavica, 77,* 653-657.

Adams, R. J. & Courage, M. L. (1996). Monocular contrast sensitivity in 24- and 36-month-olds as assessed with the contrast sensitivity card procedure. *Optometry and Vision Science, 70,* 97-101.

Adams, R. J., Dalton, S. M., Murphy, A. M., Hall, H. L., & Courage, M. L. (2002). Testing young infants with the Welch Allyn Suresight non-cycloplegic autorefractor. *Ophthalmic and Physiological Optics, 22,* 546-551.

Adams, R. J., Hall, H. L., Drover, J. R., Dalton, S. M., Vernescu, R., & Courage, M. L. (2001). Noncycloplegic autorefraction in infants and young children. *Investigative Ophthalmology and Visual Science, 42/4,* S391.

Adams, R. J., Mercer, M. E., Courage, M. L., & van Hof-van Duin, J. (1992). A new technique to measure contrast sensitivity in human infants. *Optometry and Vision Science, 69,* 440-446.

Adobe Systems Incorporated (1986). *Postscript Language Tutorial and Cookbook.* Adobe Systems Incorporated: Author

Adobe Systems Incorporated (1990). *Postscript Language Reference Manual/Adobe Systems Incorporated.* (2nd Edition). Adobe Systems Incorporated: Author

Adoh, T. O., Woodhouse, J. M., & Oduwaiye, K. A. (1992). The Cardiff test: a new visual acuity test for toddlers and children with intellectual impairment. A preliminary report. *Optometry and Vision Science, 69,* 427-432.

Allen, J. W. & Bose, B. (1992). An audit of preschool vision screening. *Archives of Disease in Childhood, 67,* 1292-1293.

Appelboom, T. M. (1985). A history of vision screening. *Journal of School Health, 55,* 138-141.

Arnold, R. W., Arnold, A. W., Stark, L., Arnold, K. K., Leman, R., Armitage, M. D. (2004). Amblyopia detection by camera: Gateway to portable inexpensive vision screening. *Alaska Medicine, 46,* 63-72.

Atkinson, J. (1984). Human visual development over the first 6 months of life. A review and a hypothesis. *Human Neurophysiology, 3,* 61-74.

Atkinson, J., Anker, S., Evans, C., & Pimm-Smith, E. (1988). Visual acuity testing of young children with the Cambridge crowding cards at 3 and 6 m. *Acta Ophthalmologica, 66,* 505-508.

Atkinson, J. & Braddick, O. (1983). Vision screening and photorefraction-the relation of refractive errors to strabismus and amblyopia. *Behavioral Brain Research, 10,* 71-80.

Atkinson, J. & Braddick, O. (1983). The use of isotropic photorefraction for vision screening in infants. *Acta Ophthalmologica Supplementum, 157,* 36-45.

Bailey, I. L. & Lovie, J. E. (1976). New design principles for visual acuity letter charts. *American Journal of Optometry and Physiological Optics, 53,* 740-745.

Baker, A. M., Bowton, D. L., & Haponik, E. F. (1995). Decision making in nosocomial pneumonia. An analytic approach to the interpretation of quantitative bronchoscopic cultures. *Chest, 107,* 85-95.

Banks, M. S. & Dannemiller, J. L. (1987). Infant visual psychophysics. In P. Salapatek & L. Cohen (Eds.). *Handbook of Infant Perception* (pp. 115-177). New York: Academic Press.

Banks, M. S., Geisler, W. S., & Bennett, P. J. (1987). The physical limits of grating visibility. *Vision Research, 27,* 1915-1924.

Banks, M. S. & Salapatek, P. (1981). Infant pattern vision: a new approach based on the contrast sensitivity function. *Journal of Experimental Psychology, 31,* 1-45.

Barry, J. C. & König, H. H. (2001). Non-cycloplegic screening for with the Nikon Retinomax hand held autorefractor in 3-year-old kindergarten children. *British Journal of Ophthalmology, 85,* 1179-1182.

Barry, J. C. & König, H. H. (2003). Test characteristics of orthoptic screening examination in 3-year-old kindergarten children. *British Journal of Ophthalmology, 87,* 909-916.

Barry, J. C. & Loewen, N. (2001). Zykloplegie tropferfahrungen in deutschsprachigan zentren für Kinderophthalmologie und strabologie-Umfrageergebnisse 1999. *Kliniusche Monotsblatter für Augenheilkunde, 218,* 26-30.

Becker, R., Hübsch, S., Gräf, & Kaufmann, H. (2000). Preliminary report: examination of young children with Lea symbols. *Strabismus, 8,* 209-213.

Becker, R., Hübsch, S., Gräf, & Kaufmann, H. (2002). Examination of young children with Lea symbols. *British Journal of Ophthalmology, 86,* 513-516.

Birch, E. E. (1993). Stereopsis in infants and its developmental relation to visual acuity. In K. Simons (Ed.). *Early Visual Development, Normal and Abnormal* (pp. 224-234). New York: Oxford University Press.

Birch, E. E., Fawcett, S., & Stager, D. R. (2000). Why does early surgical alignment improve stereoacuity outcomes in infantile esotropia. *Journal of the American Association of Pediatric Ophthalmology and Strabismus, 4,* 10-14.

Birch, E. E., Fawcett, S., & Stager, D. R. (2002). Risk factors for the development of accommodative esotropia following treatment for infantile esotropia. *Journal of the American Association of Pediatric Ophthalmology and Strabismus, 6,* 174-181.

Birch, E. E. & Salamão, S. (1998). Infant random dot stereoacuity cards. *Journal of Pediatric Ophthalmology and Strabismus, 35,* 86-90.

Birch, E. E., Stager, D. R., & Wright, W. W. (1986). Grating acuity development after early surgery for congenital cataracts. *Archives of Ophthalmology, 104,* 1783-1787.

Birch, E. E., Williams, C., Hunter, J., Lapa, M. C., & ALSPAC. (1997). Random dot stereoacuity of preschool children. *Journal of Pediatric Ophthalmology and Strabismus, 34,* 217-222.

Bolger, P., Stewart-Brown, S., Newcombe, E., & Starbuck, A. (1991). Vision screening in preschool children: comparison of orthoptists and clinical medical officers as primary screeners. *British Medical Journal, 303,* 1291-1294.

Boothe, R. G., Dobson, V. & Teller, D. Y. (1985). Postnatal development of vision in human and nonhuman primates. *Annual Review of Neuroscience, 8,* 495-595.

Bray, L. C., Clarke, M. P., Jarvis, S. N., Francis, P. M., & Colver, A. (1996). Preschool vision screening: a prospective comparative evaluation. *Eye, 10,* 714-718.

Brown, M. S. (1975). Vision screening of preschool children: how to check on visual acuity and heterophoria as part of a routine physical examination. *Clinical Pediatrics, 14,* 968-973.

Büchner, T. F., Schnorbus, U., Grenzebach, U. H., & Busse, H. (2005). Examination of preschool children for ametropia: first experiences using a new hand-held autorefractor. *Strabismus, 12,* 111-117.

Carter, J. R., Lau, M., Fowler, J. M., Carlson, J. W., & Twiggs, L. B. (1995). Blood flow characteristics of ovarian tumors: Implications for ovarian cancer screening. *American Journal of Obstetric Gynecology, 172,* 901-907.

Cashell, G. T. W. & Durran, I. M. (1980). *Handbook of Orthoptic Principles.* Churchill Livingstone: London.

Cheng, K. P., Hiles, D. A., Biglan, A. W., & Pettapiece, M. C. (1991). Visual results after early surgical treatment for unilateral congenital cataracts. *Ophthalmology, 98,* 903-910.

Chui, L., Fraser, T., Hoar, K., LaRoche, G. R. (2004). Negative predictive value of a vision screening program aimed at children aged 3 to 4 years old. *Journal of AAPOS, 8,* 566-570.

Ciner, E. B., Dobson, V., Schmidt, P. P., Allen, D., Cyert, L., Maguire, M., Moore, Orel-Bixler, E., & Schultz, J. (1999). A survey of vision screening policy of preschool children in the United States. *Survey of Ophthalmology, 43,* 445-457.

Ciner, E. B., Schanel-Klitsch, E, & Herzberg, C. (1996). Stereoacuity development: 6 months to 5 years. A new tool for testing and screening. *Optometry and Vision Science, 73,* 43-48

Ciner, E. B, Schmidt, P.P., Orel-Bixler, D., Dobson, V., Maguire, M., Cyert, L., Moore, B., Schultz, J. (1998). Vision screening of preschool children: evaluating the past, looking toward the future. *Optometry and Vision Science, 75,* 571-584.

Ciuffreda, K. J., Levi, D. M., & Selenow, A. (1991). *Amblyopia, Basic and Clinical.* Butterworth-Heinemann: Stoneham.

Cool, S. J. (1979). The amblyopic eye: Is it downtrodden or just lazy. *Journal of the American Optometric Association, 6,* 694-698.

Cooper, J. & Feldman, J. (1978). Random-dot-stereogram performance by strabismic, amblyopic, and ocular-pathology patients in an operant-discrimination task. *American Journal of Optometry and Physiological Optics, 55,* 599-609.

Cooper, C. D., Bowling, F. G., Hall, J. E., Colville, D. J., Dortmans, R. J., Munch, J., & Gole, G. A. (1996). Evaluation of photoscreener instruments in a childhood population. *Australian and New Zealand Journal of Ophthalmology, 24,* 347-355.

Cooper, C. D., Gole, G. A., Hall, J. E., Colville, D. J., Carden, S. M., & Bowling, F. G. (1999). Evaluating photoscreeners II: MTI and Fortune videorefractor. *Australian and New Zealand Journal of Ophthalmology, 27,* 387-398.

Courage, M. L., Drover, J. R., Vernescu, R., Keough, T. M., Adams, R. J. (2001). Efficiency and accuracy of the Welch-Allyn SureSight autorefractor for assessing school children [CD-ROM]. *Investigative Ophthalmology and Visual Science, 44*

Cordonnier, M. & Dramaix, M. (1998). Screening for abnormal levels of hyperopia in children: a non-cycloplegic method with a hand held refractor. *British Journal of Ophthalmology, 82,* 1260-1264.

Cordonnier, M. & Dramaix, M. (1999). Screening for refractive errors in children: accuracy of the hand held refractor Retinomax to screen for astigmatism. *British Journal of Ophthalmology, 83,* 157-161.

Cordonnier, M. & Kallay, O. (2001). Non-cycloplegic screening for refractive errors in children with the hand-held autorefractor Retinomax: final results and comparison with non-cycloplegic photoscreening. *Strabismus, 9,* 59-70.

Cornsweet, T. N. & Crane, H. D. (1970). Servo-controlled infrared optometer. *Journal of the Optical Society of America, 60,* 548-554.

Cowen, L. & Bobier, W. R. (2003). The pattern of astigmatism in a Canadian preschool population. *Investigative Ophthalmology and Visual Science, 44,* 4593-4600.

DeBecker, I., MacPherson, H. J., LaRoche, G. R., Braunstein, J., Cottle, R., Mcintyre, L. & Kozousek, V. (1992). Negative predictive value of a population-based preschool vision screening program. *Ophthalmology, 99,* 998-1003.

Donnelly, U. M., Stewart, N. M., & Hollinger, M. (2005). Prevalence and outcome of childhood disorders. *Ophthalmic Epidemiology, 12,* 243-250.

Drover, J. R., Earle, A. E., Courage, M. L., & Adams, R. J. (2002). Improving the effectiveness of the contrast sensitivity card procedure. *Optometry and Vision Science, 79,* 52-59.

Drummond, G. T., Scott, W. E., & Keech, R. (1989). Management of monocular congenital cataracts. *Archives of Ophthalmology, 107,* 45-51.

Egan, J. P. (1975). *Signal detection theory and ROC analysis.* NewYork:Academic Press.

Eibschitz-Tsimhoni, M., Friedman, T., Naor, J., Eibschitz, N., & Friedman, Z. (2000). Early screening for amblyogenic risk factors lowers the prevalence and severity of amblyopia. *Journal of AAPOS, 4,* 194-198.

Enzenauer, R. W., Freeman, H. L., Larson, M. R., Williams, T. L. (2000). Photoscreening for amblyogenic factors by public health personal: the Eyecor Camera System. *Ophthalmic Epidemiology, 7,* 1-12.

Erlich, M. I., Reinecke, R. D., & Simons, K. (1983). Preschool vision screening for amblyopia and strabismus. Programs, methods, guidelines, 1983. *Survey of Ophthalmology, 28,* 145-163.

Fantz, R. L. (1965). Visual perception from birth as shown by pattern selectivity. *Annals of the New York Academy of Sciences, 118,* 793-814.

Fathy, V. C. & Elston, P. J. (1993). Orthoptic screening of three- and four-year-olds. *Public Health, 107, 19-23.*

Fawcett, S., Leffler, J., & Birch, E. E. (2000). Factors influencing stereoacuity in accommodative esotropia. *Journal of the American Association of Pediatric Ophthalmology and Strabismus, 4,* 15-20.

Feightner, J. W. (1994). Routine preschool screening for visual and hearing problems. In Canadian Task Force on the Periodic Health Examination. *Canadian Guide to Clinical Preventive Health Care* (pp.297-304). Ottawa: Health Canada.

Feldman, W., Milner, R. A., Sackett, B., & Gilbert, S. (1980). Effects of preschool vision and hearing on prevalence of vision and hearing problems 6-12 months later. *Lancet, 82,* 1014-1016.

Fern, K. D. & Manny, R. E. (1986). Visual acuity of the preschool child: a review. *American Journal of Optometry and Physiological Optics, 63,* 319-345.

Flitcroft, D. I. (1998). A model of the contribution of oculomotor and optical factors to emmetropization and myopia. *Vision Research, 38,* 2869-2879.

Freedman, H. L. & Preston, K. L. (1992). Polaroid photoscreening for amblyogenic factors. *Ophthalmology, 99,* 1785-1795.

Friendly, D. S. (1993). Development of vision in infants and young children. *Pediatric Clinics of North America, 40,* 693-703.

Friendly, D. S., Jaafar, M. S., & Morillo (1990). A comparative study of grating and recognition visual acuity testing in children with anisometropic amblyopia without strabismus. *American Journal of Ophthalmology, 110,* 293-299.

Gottlob, I., Weghaupt, H., Vass, C., & Auff, E. (1989). Effect of levodopa on the human pattern electroretinogram and pattern visual evoked potential. *Graefe's Archives of Clinical and Experimental Ophthalmology, 277*, 421-427.

Graham, P. A. (1974). Epidemiology of strabismus. *British Journal of Ophthalmology, 58,* 224-231.

Granet, D. B., Hoover, A., Smith, A. R., Brown, S. I., Bartsch, D., & Brody, B. (1999). A new objective digital computerized vision screening system. *Journal of Pediatric Ophthalmology and Strabismus, 36,* 251-256.

Grosvenor, T. (1978). Etiology of astigmatism. *American Journal of Optometry and Physiological Optics, 55,* 214.218.

Guo, X., Jia, X., Guo, L., Xiao, X., Shen, H., Li, S., & Zhang, Q. (2000). Comparison of computer-photoscreening with non-cycloplegic retinoscopy for amblyopiogenic risk factors in children. *Chinese Medical Journal, 113,* 1007-1010.

Haegerstrom-Portnoy, G., Brabyn, J., Schneck, M. E., & Jampolsky, A. (1997). The SKILL card: an acuity test of reduced luminance and contrast. *Investigative Ophthalmology and Visual Science, 38,* 207-216.

Hall, D. & Elliman, D. (2002). Screening for vision defects. *Health for All Children* (pp. 230-244). London: Oxford University Press.

Hamer, R. D., Norcia, A. M., Day, S. H., Haegerstrom-Portnoy, G, Lewis, D., & Hsu-Winges, C. (1992). Comparison of on-and off-axis photorefraction with cycloplegic retinoscopy in infants. *Journal of Pediatric Ophthalmology and Strabismus, 29,* 232-239.

Hatch, S. W., Tibbles, C. D., Mestito, I. R., Read, R., Traveis, L., & Richman, J. (1997). Validity and reliability of the MTI photoscreener. *Optometry and Vision Science, 74,* 859-864.

Hazel, C. A., Cox, M. J., & Strang, N. C. (2003). Wavefront aberration and its relationship to the accommodative stimulus-response function in myopic subjects. *Optometry and Vision Science, 80,* 151-158.

Hered, R. W., Murphy, S., & Clancy, M. (1997). Comparisons of the HOTV and Lea Symbols charts for preschool vision screening. *Journal of Ophthalmological Nursing Technology, 16,* 68-73.

Hess, R. F. & Holliday, I. E. (1992). The spatial localization deficit in amblyopia. *Vision Research, 32,* 1319-1339.

Hickey, T. L. & Peduzzi, J. D. (1987). Structure and development of the visual system. In P. Salapatek & Cohen (Eds.). *Handbook of Infant Perception* (pp. 1-52). New York: Academic Press.

Holopigian, K., Blake, R., & Greenwald, M. J. (1986). Selective losses in binocular vision in anisometropic amblyopes. *Vision Research, 26,* 621-630.

Hope, C. & Maslin, K. (1990). Random dot stereogram E in vision screening in children. *Australian and New Zealand Journal of Ophthalmology, 18,* 319-324.

Howland, H. C., Atkinson, J., Braddick, O., & French, J. (1978). Infant astigmatism measured by photorefraction. *Science, 202,* 331-333.

Howland, H. C., Braddick, O., Atkinson, J., & Howland, B. (1983). Optics of photorefraction: orthogonal and isotropic methods. *Journal of the Optical Society of America, 73,* 1701-1708.

Hubel, D. (Ed.) (1988). *Eye, Brain, and Vision.* New York: Scientific American Library.

Imamura, K., Richter, H., Fischer, H., Lennerstrand, G., Franzen, O., Rydberg, A., Andersson, J., Schneider, J., Onoe, H., Watanabe, Y., & Langstrom, B. (1997). Reduced activity in the extrastriate cortex of individuals with strabismic amblyopia. *Neuroscience Letters, 225,*173-176.

Ingram, R. M. (1989). Review of children referred from the school vision screening programme in Kettering during 1976-1978. *British Medical Journal, 298,* 935-936.

Iurno, J. D., Grant, W. D., & Noël, L. P. (2004). Clinical comparison of the Welch-Allyn SureSight handheld autorefractor versus cycloplegic autorefraction and retinoscopic refraction. *JAAPOS, 8,* 123-127.

Jarvis, S. N., Tamhe, R. C., Thompson, L., Francis, P. M., Anderson, J. & Colver, A. F. (1991). Preschool vision testing. *Archives of Disease in Childhood, 66,* 288-294.

Junghans, B. M. & Crewther, S. G. (2003). Prevalence of myopia among primary school children in eastern Sydney. *Clinical and Experimental Optometry, 86,* 339-345.

Kani, W. (1978). Stereopsis and spatial perception in amblyopes and uncorrected amblyopes. *British Journal of Ophthalmology, 62,*756-762.

Kennedy, R. A. & Sheps, S. B. (1989). A comparison of photoscreening techniques for amblyogenic factors in children. *Canadian Journal of Ophthalmology, 24,* 269-264.

Kennedy, R. A., Sheps, S. B., & Bagaric, D. (1995). Field trial of the Otago photoscreener. *Canadian Journal of Ophthalmology, 30,* 193-197.

Kennedy, R. A. & Thomas, D. E. (2000). Evaluation of the iScreen digital screening system for amblyogenic factors. *Canadian Journal of Ophthalmology, 35,* 258-262.

Kim, I., Pollitt, E., Leibel, R. L., Viteri, F. E., & Alvarez, E. (1984). Application of receiver-operator analysis to diagnostic tests of iron deficiency in man. *Pediatric Research, 18,* 916-920.

Kodoi, A., Yoshihara, M., Sumii, K., Haruma, K., & Kajiyama, G. (1995). Serum pepsinogen in screening for gastric cancer. *Journal of Gastroenterology, 30,* 452-460.

Kohl, P. & Samek, B. M. (1988). Refractive error and preferential looking visual acuity in infants 12-24 months of age: year 2 of a longitudinal study. *Journal of the American Optometric Association, 59,* 686-690.

Köhler, L. & Stigmar, G. (1973). Vision screening of four-year-old children. *Acta Pediatrica Scandinavica, 62,* 17-27.

Köhler L. & Stigmar, F. (1978). Visual disorders in 7-year-old children with and without previous screening. *Acta Ophthalmologica Scandinavica, 67,* 373-377.

König, H. H., Barry, J. C., Leidl, R., & Zrenner, E. (2000). Cost-effectiveness of orthoptic screening in kindergarten: a decision analytic model. *Strabismus, 8,* 79-90.

Kornder, L. D., Nursey, J. N., Pratt-Johnson, J. A., & Beattie, A. (1974). Detection of manifest strabismus in young children. I. A Prospective Study. *American Journal of Ophthalmology, 77,* 207-210.

Kushner, B. J. (1998). Amblyopia. In L. B. Nelson (Ed.). *Hartley's Pediatric Ophthalmology* (pp. 125-139). Philadelphia: Hartley.

Kushner, B. J., Lucchese, N. J., & Morton, G. V. (1995). Grating visual acuity with Teller cards compared with Snellen visual acuity in literate patients. *Archives of Ophthalmology, 113,* 485-493.

Kvarnström, G., Jakobsson, P. & Lennerstrand, G. (1998). Screening for visual and ocular disorders in children, evaluation of the system in Sweden. *Acta Pediatrica, 87,* 1173-1179.

Kvarnström, G., Jakobsson, P. & Lennerstrand, G. (2001). Visual screening of Swedish children. *Acta Ophthalmologica Scandinavica, 79,* 240-244.

Ladenvall, G. (1988). Screening for ocular and vision defects among Swedish small children. In J. Marube (Ed.). *Transactions of the European Strabismological Association Meeting, 17,* 27-33.

Last, J. M. (1988). *A Dictionary of Epidemiology,* 2nd Edition. New York: Oxford University Press.

Leguire, L. E., Rogers, G. L., Bremer, D. L., Walson, P. D., & McGregor, M. L. (1993). Levodopa/carbidopa for childhood amblyopia. *Investigative Ophthalmology and Visual Science, 34,* 3090-3095.

Leguire, L. E., Walson, P. D., Rogers, G. L., Bremer, D. L., & McGregor, M. L. (1993). Longitudinal study of levodopa/carbidopa for childhood amblyopia. *Journal of Pediatric Ophthalmology and Strabismus, 30,* 354-360.

Leguire, L. E., Walson, P. D., Rogers, G. L., Bremer, D. L., & McGregor, M. L. (1995). Levodopa/carbidopa treatment for amblyopia in older children. *Journal of Pediatric Ophthalmology and Strabismus, 32,* 143-151.

Lennerstrand, G., Jakobsson, P. & Kvarnström, G. (1995). Screening for ocular dysfunction in children: approaching a common program. *Acta Ophthalmologica Scandinavica, (Suppl), 214,* 26-38.

Lennie, P. & van Hemel, S. B. (2002). Assessment of vision in infants and children. *Visual Impairments: Determining Eligibility for Social Security Benefits* (pp. 140-159). Washington, D.C.: National Academy Press.

Lim, H. T., Yu, Y. S., Park, S.-H., Ahn, H., Kim, S., Lee, M., Jeong, J.-Y., Shin, K. H., & Koo, B. S. (2004). The Seoul metropolitan preschool vision screening programme: results from South Korea. *British Journal of Ophthalmology, 88,* 929-933.

Lloyd, I. C., Dowler, J. G. F., Kriss, A., Speedwell, L., Thompson, D. A., Russell-Eggitt, I., & Taylor, D. (1995). Modulation of amblyopia following early surgery for unilateral congenital cataracts. *British Journal of Ophthalmology, 79,* 802-806.

Lorenz, B. (2002). Genetics of isolated and syndromic strabismus: facts and perspectives. *Strabismus, 10,* 147-156.

MacPherson, H., Braunstein, J., & La Roche (1991). Utilizing basic screening principles in the design and evaluation of vision screening programs. *American Orthoptic Journal, 41,* 110-121.

Manny, R. E., Martinez, A. T., & Fern, K. D. (1991). Testing stereopsis in the preschool child: is it clinically useful? *Journal of Pediatric Ophthalmology and Strabismus, 28,* 223-231

Maternal and Child Health Bureau and National Eye Institute Task Force on Vision Screening in the Preschool Child (2001). Preschool vision screening. *Optometry Vision Science, 78,* 122-126.

Maurer, D. & Lewis, T. L. (1993). Visual outcomes after infantile cataract. In K. Simons (ed.), *Early Visual Development: Normal and Abnormal* (pp. 454-484). New York: Oxford University Press.

Maurer, D., Lewis, T. L. & Brent, H. P. (1989). The effects of deprivation on human development: studies of children treated for cataracts. In F. J. Morrision, C. Lord, & D. P. Keating (eds.). *Applied Developmental Psychology. Vol 3. Psychological Development in Infancy.* San Deigo: Academic Press. pp. 39-227.

Maurer, D., Lewis, T. L., Brent, H. P., & Levin, A.V. (1999). Rapid improvement in the acuity of infants after visual input. *Science, 286,* 108-110.

Mayer, D. L. (1986). Acuity of amblyopic children for small field gratings and recognition stimuli. *Investigative Ophthalmology and Visual Science, 27,* 1148-1153.

Mayer, D. L., Beiser, A. S., Warner, A. F., Pratt, E. M., Raye, K. N., & Lange, J. M. (1995). Monocular acuity norms for the Teller acuity cards between ages one month and four years. *Investigative Ophthalmology and Visual Science, 36,* 671-685.

Mayer, D. L., Hansen, R., Moore, B. D., Kim, S., & Fulton, A. B. (2001). Cycloplegic refraction in healthy children aged 1 through 48 months. *Archives of Ophthalmology, 119,* 1625-1628.

McDevitt, H. I. (1977). Automatic retinoscopy-the 6600 auto-refractor. *Optician, 173,* 33-42.

McDonald, M., Ankrum, C., Preston, K., Sebris, S. L., & Dobson, V. (1986). Monocular and binocular acuity estimation in 18- to 36-month-olds: acuity card results. *American Academy of Optometry, 63,* 181-186.

McDonald, M. A., Dobson, V., Sebris, L., Baitch, L., Varner, D., & Teller, D. Y. (1985). The acuity card procedure: a rapid test of infant acuity. *Investigative Ophthalmology and Visual Science, 26,* 1158-1162.

Mills, M. D. (1999). The eye in childhood. *American Family Physician, 60,* 907-916.

Miller, J. M., Dobson, V., Harvey, E. M., & Sherrill, D. L. (2001). Comparison of preschool vision screening methods in a population with a high prevalence of astigmatism. *Investigative Ophthalmology and Visual Science, 42,* 917-924.

Miller, J. M., Dobson, V., Harvey, E. M., & Sherrill, D. L. (2003). Cost-efficient vision screening for astigmatism in Native American preschool children. *Investigative Ophthalmology and Visual Science, 44,* 3756-3763.

Millodot, M. (1986). *Dictionary of Optometry.* London: Butterworths

Mohan, K., Dhankar, V., & Sharma, A. (2001). Visual acuities after levodopa administration in amblyopia. *Journal of Pediatric Ophthalmology and Strabismus, 38,* 62-67.

Mohindra, I. (1977). Early treatment of anisometropic astigmatism and strabismus. *American Journal of Optometry and Physiological Optics, 54,* 479-484.

Mohn, G. & van Hof-van Duin, J. (1991). Development of spatial vision. In D. Regan (Ed.). *Vision and Visual Dysfunction. Vol. 10. Spatial Vision* (pp. 179-192). London: MacMillan

Molgaard, I. L., Biering-Sorenson, K, Michelson, N., Elmer, J., & Rydberg, A. (1984). Amblyopia screening in kindergarten with TNO stereotest. *Acta Ophthalmologica, 62,* 156-162.

Morad, Y., Werker, E., & Nemet, P. (1999). Visual acuity tests using chart, line, and single optotype in healthy and amblyopic children. *Journal of the American Academy of Pediatrics Ophthalmology and Strabismus, 3,* 94-97.

Morgan, K. S. & Johnson, W. D. (1987). Clinical evaluation of a commercial photoscreener. *Archives of Ophthalmology, 105,* 1528-1531.

Moseley, M. J. (1998). Preschool vision screening: a recent report calls for a halt. *British Journal of Ophthalmology, 82,* 722-723.

Newman, D. K. & East, M. M. (1999). Preschool vision screening: negative predictive value for amblyopia. *British Journal of Ophthalmology, 83,* 676-679.

Norcia, A. M., Tyler, C. W., & Hamer, R. D. (1990). Development of contrast sensitivity in the human infant. *Vision Research, 30,* 1475-1486.

Nordlöw, W. & Joachimsson, S. (1962). A screening test for visual acuity in four year old children. *Acta Ophthalmologica, 40,*453-462.

Odom, J. V., Hoyt, C. & Marg, E. (1981). Effect of natural deprivation and unilateral eye patching on visual acuity of infants and children. *Archives of Ophthalmology, 99,* 1412-1416.

Olitsky, S. E. & Nelson, L. B. (1998). Strabismus disorders. In L. B. Nelson (Ed.). *Hartley's Pediatric Ophthalmology* (pp. 125-139). Philadelphia: Hartley.

Olver, J. M. (1988). Simple retinoscopic screening. *Eye, 2,* 309-313.

Ontario Association of Optometrists. (1997). Guidelines for preschool vision screening. Toronto: Publication of the Association of Optometrists.

Ottar, W. L. (1997). The ABCs of visual acuity assessment. *Insight, 22,* 81-87.

Ottar, W. L., Scott, W. E., & Holgado, S. I. (1995). Photoscreening for amblyogenic factors. *Journal of Pediatric Ophthalmology and Strabismus, 32,* 289-295.

Park, S. H., Goo, J. M., & Jo, C. H., (2004). Receiver operating characteristic (ROC) curve. Practical review for radiologists. *Korean Journal of Radiology, 5,* 11-18.

Pediatric Eye Disease Investigator Group (2002). A randomized trial of atropine vs. patching for treatment of moderate amblyopia in children. *Archives of Ophthalmology, 120,* 268-278.

Pediatric Eye Disease Investigator Group (2003). A comparison of atropine and patching treatments for moderate amblyopia by patient age, cause of amblyopia, depth of amblyopia, and other factors. *Ophthalmology, 110,* 1632-1638.

Pickert, S. M. & Wachs, H. (1980). Stimulus and communication demands of visual acuity tests. *American Journal of Optometry and Physiological Optics, 57,* 875-880.

Pinel, J. P. (1997). Human brain damage and animal models. *Biopsychology, (3rd Edition)* (pp. 130-149). Boston: Allyn and Bacon.

Pirchio, M., Spinelli, D., Fiorentini, A., & Maffei, L. (1978). Infant contrast sensitivity evaluated by evoked potentials. *Brain Research, 141,* 179-184.

Preslan, M. W. & Novak, A. (1998). Baltimore vision screening project: Phase 2. *Ophthalmology, 105,* 150-153.

Rahman, P., Jones, A., Curtis, J., Bartlett, S., Peddle, L., Fernandez, B. A., & Freimer, N. B., (2003). The Newfoundland population: a unique resource for genetic investigation of complex diseases. *Human Molecular Genetics, 12,* R167-R172.

Raina, J. (1998). A comparison of Wright card with Teller card for detection of amblyopia. *Journal of Pediatric Ophthalmology and Strabismus, 35,* 17-21.

Reinecke, R. D. & Simons, K. (1974). A new stereoscopic test for amblyopia screening. *American Journal of Ophthalmology, 78,* 714-721.

Repka, M. X. (1998). Refraction in infants and children. In L. B. Nelson (Ed.). *Hartley's Pediatric Ophthalmology* (pp. 112-124). Philadelphia: Hartley.

Repka, M. X. & Ray, J. M. (1993). The efficacy of pharmacological penalization. *Ophthalmology, 100,* 769-774.

Ricci, F., Cedrone, C., & Cerulli, L. (1998). Standardized measurement of visual acuity. *Ophthalmic Epidemiology, 5,* 41-53.

Richman, J. E. & Lyons, S. (1994). A forced choice procedure for evaluation of contrast sensitivity in preschool children. *Journal of the American Optometric Association, 65,* 859-864.

Robaie, D., Rose, K., Ojaimi, E., Kifley, A., Huynh, S., & Mitchell, P. (2005). Visual acuity and the causes of visual loss in a population of 6-year-old Australian children. *Ophthalmology, 112,* 1275-1282.

Robinson, B. E. (1999). Factors associated with the prevalence of myopia in 6-year-olds. *Optometry and Vision Science, 76,* 266-271.

Robinson, B., Bobier, W. R., & Martin, E. (2000). Oxford County preschool vision screening: prevalence of vision problems. *Canadian Journal of Optometry, 62,* 19-24.

Robinson, B., Bobier, W. R., Martin, E., & Bryant, L. (1999). Measurement of the validity of a preschool vision screening program. *American Journal of Public Health, 89,*193-198.

Rosner, J. (1978). The effectiveness of the random dot E stereotest as a preschool vision screening instrument. *Journal of the American Optometric Association, 49,* 1121-1124.

Ross, E., Murray, A. L., & Stead, S. (1977). Prevalence of amblyopia in grade 1 schoolchildren in Saskatoon. *Canadian Journal of Public Health, 68,* 491-493.

Rosser, D. A., Laidlaw, D. A. H., & Murdoch, I. E. (2001). The development of a "reduced logMAR" visual acuity chart for use in routine clinical practice. *British Journal of Ophthalmology, 85,* 432-436.

Rovamo, J., Mustonen, J., & Näsänen, R. (1994). Modeling contrast sensitivity as a function of retinal illuminance and grating area. *Vision Research, 34,* 1301-1314.

Rubin, S. E. & Nelson, L. B. (1993). Amblyopia. Diagnosis and management. *Pediatric Clinics of North America, 40,* 727-735.

Ruttum, M. S. (1988). Visual screening with random dot stereograms. *Seminars in Ophthalmology, 3,* 175-180.

Ruttum, M. S. & Nelson, D. B. (1991). Stereopsis testing to reduce overreferral in preschool vision screening. *Journal of Pediatric Ophthalmology and Strabismus, 28,* 131-133.

Salamão, S. R. & Ventura, D. F. (1995). Large sample population age norms for visual acuities obtained with Vistech-Teller acuity cards. *Investigative Ophthalmology and Visual Science, 36,* 657-670.

Scharre, J. E., Cotter, S. A., Stein-Block, S. S., & Kelly, S. A. (1990). Normative contrast sensitivity data for young children. *Optometry and Vision Science, 67,* 826-832.

Schmidt, P. P. (1994). Vision screening with the RDE stereotest in pediatric populations. *Optometry and Vision Science, 71,* 273-281.

Schwartz, S. H. (Ed.). (1999). Spatial Vision. *Visual Perception* (pp. 175-204). Stamford: Prentice Hall.

Sekuler, R. & Blake, R. (1994). Spatial vision and pattern recognition. In J. Vaicunas & J. Akers (Eds.), *Perception* pp. 141-179. New York: McGraw Hill, Inc.

Shallo-Hoffmann, J., Coulter, R., Oliver, P., Hardigan, P., & Blavo, C. (2004). A study of pre-school vision screening tests' testability, validity, and duration: do group differences matter? *Strabismus, 12,* 65-73.

Sheedy, J. E., Bailey, I. L., & Raasch, T. W. (1984). Visual acuity and chart luminance. *American Journal of Optometry and Physiological Optics, 61,* 595-600.

Simmers, A. J., Gray, L. S., & Spowart, K. (1997). Screening for amblyopia, a comparison of paediatric letter tests. *British Journal of Ophthalmology, 81,* 465-469.

Simon, J. W. & Kaw, P. (2001). Commonly missed diagnoses in the childhood eye examination. *American Family Physician, 64,* 623-628.

Simon J. W., Siegfried, J. B., Mills, M. D., Calhoun, J. H., & Gurland, J. E. (2004). A new visual evoked potential system for vision screening in infants and young children. *Journal of AAPOS, 8,* 549-554.

Simons, K. (1981). A comparison of the Frisby, Random-Dot E, TNO, and Randot Circles stereotests in screening and office use. *Archives of Ophthalmology, 99,* 446-452.

Simons, K. (1983). Visual acuity norms in young children. *Surveys of Ophthalmology, 8,* 84-92.

Simons, K. (1996). Preschool vision screening: rationale, methodology and outcome. *Survey of Ophthalmology, 41,* 3-30.

Simpson, T. (1991). The suppression effect of simulated anisometropia. *Ophthalmic and Physiological Optics, 11,* 350-357.

Snowdon, S. K. & Stewart-Brown, S. L. (1997). Preschool vision screening. *Health Technology Assessment, 1,* 1-83.

Spowart, K. M., Simmers, A. & Tappin, D. M. (1998). Vision testing in schools: an evaluation of personnel, tests, and premises. *Journal of Medical Screening, 5,* 131-132.

Steele, G., Ireland, D., & Block, S. (2003). Cycloplegic autorefraction results in pre-school children using the Nikon Retinomax Plus and the Welch-Allyn SureSight. *Optometry and Vision Science, 80,* 573-577.

Stewart-Brown, S. L. & Haslum, M. (1988). Screening of vision in school: could we be doing better by doing less? *British Medical Journal, 297,* 1111-1113.

Stewart-Brown, S. L., Haslum, M. N., & Howlett, B. (1988). Preschool vision screening: a service in need of rationalization. *Archives of Disease in Childhood, 63,* 356-359.

Sturr, J. F., Kline, G. E., & Taub, H. A. (1990). Performance of young and older drivers on a static test under photopic and mesopic luminance conditions. *Human Factors, 32,* 1-8.

Suryakumar, R. & Bobier, W. R. (2003). The manifestation of noncycloplegic refractive state in preschool children is dependent on autorefractor design. *Optometry and Vision Science, 80,* 578-586.

Taylor, D. (1987). Screening for squint and poor vision. *Archives of Disease in Childhood, 62,* 982-983.

Tong, P. Y., Bassin, R. E., Enke-Miyazaki, E., Macke, J. P., Tielsch, J. M., Stager, D. R., Beauchamp, G. R., Parks, M. M., & The National Children's Eye Care Foundation Vision Screening Study Group. (2000). Screening for amblyopia in preverbal children with photoscreening photographs. *Ophthalmology, 107,* 1623-1629.

Tong, P. Y., Macke, J. P., Bassin, R. E., Everett, M., Enke-Miyazaki, E., Tielsch, J. M., Stager, D. R., Parks, M. M., Beauchamp, G., R., & The National Children's Eye Care Foundation Vision Screening Study Group. (2000). Screening for amblyopia in preverbal children with photoscreening photographs. *Ophthalmology, 107,* 1630-1636.

U. S. Public Health Service (1994). Vision screening in children. *American Family Physician, 50,* 587-590.

Vaughan, D., Asbury, T., & Riordan-Eva, P. (Eds.) (1992). *General Ophthalmology.* Stamford, CT: Appleton & Lange.

The Vision in Preschoolers Study Group (2004). Comparison of preschool vision screening tests as administered by licensed eye care professionals in the Vision in Preschoolers study. *Ophthalmology, 111,* 637-650.

The Vision in Preschoolers Study Group (2005a). Sensitivity of screening tests of detecting Vision in Preschoolers-targeted vision disorders when specificity is 94%. *Optometry and Vision Science, 82,* 432-438.

The Vision in Preschoolers Study Group (2005b). Preschool vision screening tests administered by nurse screeners compared with lay screeners in the Vision in Preschoolers study. *Investigative Ophthalmology and Visual Science, 46,* 2639-2648.

von Noorden, G. K. (Ed.) (1990). *Binocular Vision and Ocular Motility: Theory and Management of Strabismus.* St. Louis: Mosby.

Wagner, R. S. (1998). Pediatric eye examination. In L. B. Nelson, (Ed.). *Hartley's Pediatric Ophthalmology* (pp. 82-95). Philadelphia: Hartley.

Wali, N., Leguire, L. E., Rogers, G. L., & Bremer, D. L. (1991). CSF interocular interaction in childhood amblyopia. *Optometry and Vision Science, 68*, 81-87.

Walraven, J. (1975). Amblyopia screening with random dot stereograms. *American Journal of Ophthalmology, 80,* 893-900.

Walraven, J. & Janzen, P. (1993). TNO stereopsis test as an aid to the prevention of amblyopia. *Ophthalmic and Physiological Optics, 13,* 350-356.

Watts, P., Walker, K., & Betts, L. (1999). Photoscreening for refractive errors in children and young children with severe learning disabilities using the MTI photoscreener. *Eye, 13,* 363-368.

Waugh, S. J. & Levi, D. M. (1993). Visibility, luminance and vernier acuity. *Vision Research, 33,* 527-538.

Weinand, F., Gräf, M., & Demming, K. (1998). Sensitivity of the MTI photoscreener for amblyogenic factors in infancy and early childhood. *Graefe's Archives of Clinical Experimental Ophthalmology, 236,* 801-805.

Wesemann, S. & Rassow, B. (1987). Automatic infrared refractors-A comparative study. *American Journal of Optometry and Physiological Optics, 64,* 627-638.

Williams, C., Harrad, R. A., Harvey, I., Sparrow, J. M., & the ALSPAC Study Team. (2001). Screening for amblyopia in preschool children: results of a population-based, randomized controlled trial. *Ophthalmic Epidemiology, 8,* 279-295.

Williams, C., Lumb, R., Harvey, I., & Sparrow, J. M. (2000). Screening for refractive errors with the Topcon PR 2000 pediatric refractometer. *Investigative Ophthalmology and Visual Science, 41,* 1031-1037.

Wilson, H. R. (1988). Development of spatiotemporal mechanisms in infant vision. *Vision Research, 28,* 611-628.

Wood, I. C. J., Fox, J. A., & Stephenson, M. G. (1978). Contrast threshold of random dot stereograms in anisometropic amblyopia: a clinical investigation. *British Journal of Ophthalmology, 62,* 34-38.

Wood, I. C. J., & Stephenson, M. G. (1981). Contrast thresholds in anisometropia, Mein, J. & Moore J., (Eds.): *Orthoptics Research and Practice.* London, Henry Kimpleton.

Woods, M. O., Hyde, A. J., Curtis, F. K., Stuckless, S., Green, J. S, Pollett, A. F., Robb, J. D., Green, R. C., Croitoru, M. E., Careen, A., Chaulk, J. A. W., Jegathesan, J., Younghusband, H. B., Bapat, B. V., & Parfrey, P. S. (2005). High frequency of hereditary colorectal cancer in Newfoundland likely involves novel susceptibility genes. *Clinical Cancer Research, 11,* 6853-6861.

Woods, M. O., Young, T. L., Parfrey, P. S., Hefferton, D., Green, J. S., & Davidson, W. S. (1999). Genetic heterogeneity of Bardet-Beidl syndrome in a distinct Canadian population: evidence for a fifth locus. *Genomics, 55,* 2-9.

Woodruff, M. E. (1986). Vision and refractive status among Grade 1 children in the province of New Brunswick. *American Journal of Optometry and Physiological Optics, 63,* 545-552.

Wormald, R. P. L. (1991). Preschool vision screening in Cornwall: performance indicators of community orthoptists. *Archives of Disease in Childhood, 66,*917-920.

Xie, Y. G., Zheng, H., Leggo, J., Scullym, M. F., & Lillicrap, D. (2002). A founder factor VIII mutation, valine 2016, in a population with an extraordinarily high prevalence of mild hemophilia A. *Thrombosis and haemostasis, 87,* 178-179

Yuodelis, C. & Hendrickson, A. (1986). A qualitative and quantitative analysis of the human fovea during development. *Vision Research, 26,* 847-855.

Zhao, J., Mao, J., Luo, R., Li, F., Pokharel, G., & Ellwein, L. B. (2004). Accuracy of noncycloplegic autorefraction in school-age children in China. *Optometry and Vision Science, 81,* 49-55.

Appendix A

St. John's Regional Preschool Vision Screening Program

Study Information and Request to Participate

Dear parents,

We are a team of researchers at Memorial University who are currently initiating a vision screening study at all daycare centers in the St. John's area. The purpose of the study is to detect children with early, subtle visual disorders such as a turned eye, poorly developing visual acuity or focusing problems. At the same time, we wish to evaluate the effectiveness of the tests used in the screening process. There is a critical need for screening research, because if undetected and untreated, early disorders may lead to permanent visual deficits which in later years, are very difficult to correct surgically or with medication or therapy. Thus, it is important to detect and treat any existing visual disorders well before the school years in order to allow the child to perform to the best of his/her abilities both academically and socially. Furthermore, our research team here in Newfoundland is at the forefront as there are currently no effective screening programs within any Canadian province. We hope that the results of this study will provide the basis for effective early vision screening in Newfoundland and across the country.

In this study, each child's vision will be assessed with a battery of tests that are not typically used until the elementary school years and also go well beyond the typical public health pre-kindergarten vision check. The tests include: (1) the Landolt C visual acuity test, (2) the cover-uncover test, (3) the Randot stereo test, (4) the contrast sensitivity cards, and (5) autorefraction. All of these tests are simple, non-threatening and most children enjoying doing them as they are designed for preschoolers. Specific details of each test are provided at the end of this letter.

The entire screening procedure will be conducted at your child's daycare center and should be completed in approximately 20 minutes. The tests will be performed by very experienced examiners who have tested thousands of infants and children in the past. Although most children usually find the testing enjoyable, we will be careful not to proceed if the child gives any indication that s/he is uncomfortable, or becomes uninterested.

We expect that most children will show normal levels of vision. However if a child scores below the norms for other children of the same age, he/she will be retested at a later date, likely within 2 weeks. If after the second test, his/her scores are still below the norm, you will be offered the opportunity to bring him/her to the optometrist or ophthalmologist who is part or our team, to receive a follow-up eye exam. Each child's results are confidential, will be safeguarded, and will not be released without parental permission. Note however, that your child's results can be made available to you any time upon your request. You also have the right to withdraw from the study at any point

(even *after* your child has been tested) and all of the results from your child will be discarded.

In our opinion, there are no apparent harms to participation and the benefits may be substantial, especially if we determine that your child has a vision problem and may benefit from treatment. Finally, participation in this study (or not) will in no way affect your child's regular medical evaluations, including the preschool vision check which is usually conducted by a Public Health Nurse prior to Kindergarten.

DETAILS OF THE TESTS TO BE ADMINISTERED:

(1) The **Landolt C test** is a chart (like the adult BIG E chart) containing rows of Cs of different sizes. The child must locate the position of the Cs opening or gap. Children who cannot complete the Landolt C test (usually 2- and 3-year-olds) will be assessed with the Lea Symbols test which is a chart (or plastic booklet) with symbols (houses, hearts, squares, and circles) of different sizes. The smallest Landolt C or Lea Symbol that the child can see gives us an indication of his/her visual acuity, traditionally the most important clinical aspect of one's vision.

Children who can not complete either of these tests will be assessed with the Teller Acuity Cards. These are a set of rectangular cards that contain black and white stripes of different sizes. Children are shown cards containing stripes of progressively smaller size and asked to point to them. The smallest size of the stripes detected provides an estimate of visual acuity.

(2) The **cover-uncover test** is used to detect strabismus (an eye turn). During this test, the child looks at a stuffed toy while one eye is covered very briefly with a small plastic paddle. The eyes are observed after the cover is removed to see if they move and function normally. The test is then quickly repeated with the other eye covered. Children will also be assessed with the Hirschberg corneal reflex in which a penlight is briefly shone into his/her eyes. If the reflection of the light is asymmetrical, the child may possess and eye turn. Also, each child's eye movements will be examined as he/she will be asked to follow the penlight as it is moved in several directions (with the light off).

(3) The **Randot E Stereotest** measures depth perception. The test consists of two cards: one contains an "E" that can be seen only with special polarized "stereo" glasses that the child wears, whereas the other plate is a blank. A child with normal stereo (3-D) vision will be able to correctly identify the "E". Children who are too young to complete this test will be tested with the simpler Randot Stereosmile Cards which consists of a series of large rectangular cards, each containing a 3-D smiling face.

(4) The **contrast sensitivity card procedure** consists of a series of rectangular cards and is similar to the Teller Acuity Cards. Each card contains black and white stripes of a specific size and contrast. Children will be instructed to point at the stripes if they can see them.

(5) Finally, each child will be assessed with the Welch-Allyn SureSight autorefractor, a hand-held camera-like device that uses a light to obtain a rapid measurement of the eye's optics (the eye's ability to focus an image). This instrument measures the degree of myopia (nearsightedness), hyperopia (farsightedness), or astigmatism in each eye.

This study has been approved by the Interdisciplinary Committee on Ethics in Human Research (ICEHR) at MUN. The results of this study will likely be published in well-established medical, neuroscience and psychology journals. If you wish to have your child participate, please complete the portion of the form below as soon as possible and return it to your child's daycare teacher. If you have additional questions or concerns, please contact the study supervisor, Dr. Russell J. Adams (737-8496), James Drover (737-4786) or the secretary of the ICEHR at 737-8368. Please keep this sheet as a reference. There is also a copy of this letter on file at your child's daycare Centre if you happen to misplace this information. Thank-you.

Very sincerely,

Russell J. Adams, PhD.
Department of Psychology
Department of Pediatrics

Mary L. Courage, PhD.
Department of Psychology
Department of Pediatrics

James R. Drover, .
Ph.D. Candidate
Department of Psychology

Please return this portion as soon as possible to your child's daycare Centre or teacher. We hope to begin testing in the upcoming week.

I have read and understand all of the information pertaining to the **St. John's Regional Preschool Vision Screening Program** and wish to have my child participate.

Child's Name_____
Child's Daycare _____
Child's Birth Date_____
Days and sessions (AM/PM) that your child attends daycare
_____

Parent's Name_____
Parent's Signature_____
Today's Date_____
(Optional) : Your phone # _____and/or email
_____

Have we tested your child before?? _____ If yes, when (approximately) _____ and where_____.

Is there anything that you would like to communicate to the researchers about your child or any question that you may have?

_____
_____
_____

Appendix B

St. John's Regional Preschool Vision Screening Program

Study Information and Request to Participate

Dear parents,

We are a team of researchers at Memorial University who are currently initiating a vision screening study at all daycare centers in the St. John's area. The purpose of the study is to detect children with early, subtle visual disorders such as a turned eye, poorly developing visual acuity or focusing problems. At the same time, we wish to evaluate the effectiveness of the tests used in the screening process. There is a critical need for screening research, because if undetected and untreated, early disorders may lead to permanent visual deficits which in later years, are very difficult to correct surgically or with medication or therapy. Thus, it is important to detect and treat any existing visual disorders well before the school years in order to allow the child to perform to the best of his/her abilities both academically and socially. Furthermore, our research team here in Newfoundland is at the forefront as there are currently no effective screening programs within any Canadian province. We hope that the results of this study will provide the basis for effective early vision screening in Newfoundland and across the country.

In this study, each child's vision will be assessed with a battery of tests that are not typically used until the elementary school years and also go well beyond the typical public health pre-kindergarten vision check. The tests include: (1) the Patti Pics visual acuity test, (2) the cover-uncover test, (3) the Randot stereo test, (4) the contrast sensitivity booklet, and (5) autorefraction. All of these tests are simple, non-threatening and most children enjoying doing them as they are designed for preschoolers. Specific details of each test are provided at the end of this letter.

The entire screening procedure will be conducted at your child's daycare center and should be completed in approximately 20 minutes. The tests will be performed by very experienced examiners who have tested thousands of infants and children in the past. Although most children usually find the testing enjoyable, we will be careful not to proceed if the child gives any indication that s/he is uncomfortable, or becomes uninterested.

We expect that most children will show normal levels of vision. However if a child scores below the norms for other children of the same age, he/she will be retested at a later date, likely within 2 weeks. If after the second test, his/her scores are still below the norm, you will be offered the opportunity to bring him/her to the optometrist or ophthalmologist who is part or our team, to receive a follow-up eye exam. Each child's results are confidential, will be safeguarded, and will not be released without parental permission. Note however, that your child's results can be made available to you any time upon your request. You also have the right to withdraw from the study at any point

(even *after* your child has been tested) and all of the results from your child will be discarded.

In our opinion, there are no apparent harms to participation and the benefits may be substantial, especially if we determine that your child has a vision problem and may benefit from treatment. Finally, participation in this study (or not) will in no way affect your child's regular medical evaluations, including the preschool vision check which is usually conducted by a Public Health Nurse prior to Kindergarten.

DETAILS OF THE TESTS TO BE ADMINISTERED:

(1) The **Patti Pics Visual Acuity test** (like the adult BIG E chart) containing rows of simple symbols (a house, an apple, a circle) of different sizes which the child either names or points to. Younger children (usually 2-and-3 year-olds) who cannot complete the full version of the Patti Pics are tested with a simpler version which presents the symbols one at a time. The smallest symbol that the child can see gives us an indication of his/her visual acuity, traditionally the most important clinical aspect of one's vision.

(2) The **cover-uncover test** is used to detect strabismus (an eye turn). During this test, the child looks at a stuffed toy while one eye is covered very briefly with a small plastic paddle. The eyes are observed after the cover is removed to see if they move and function normally. The test is then quickly repeated with the other eye covered. Children will also be assessed with the Hirschberg corneal reflex in which a penlight is briefly shone into his/her eyes. If the reflection of the light is asymmetrical, the child may possess and eye turn. Also, each child's eye movements will be examined as he/she will be asked to follow the penlight as it is moved in several directions (with the light off).

(3) The **Randot Preschool stereo test** measures depth perception. This test consists of three booklets that contain simple objects (a heart, a car, a hand) that can be seen only with special polarized "stereo" glasses that the child wears. A child with normal stereo (3-D) vision will see the figures. Children who cannot complete this test will be tested with the **Randot E Stereotest**. This test consists of two cards: one contains an "E" that can be seen only with special polarized "stereo" glasses that the child wears, whereas the other plate is a blank. A child with normal stereo (3-D) vision will be able to correctly identify the "E".

(4) The **contrast sensitivity booklet** consists of a series of ring binder with black and white stripes of a specific size and contrast. Children will be instructed to point at the stripes if they can see them.

(5) Finally, each child will be assessed with the **Welch-Allyn SureSight autorefractor**, a hand-held, camera-like device which uses a light to obtain a rapid measurement of the eye's optics (the eye's ability to focus an image). This instrument measures the degree of myopia (nearsightedness), hyperopia (farsightedness), or astigmatism in each eye.

This study has been approved by the Interdisciplinary Committee on Ethics in Human Research (ICEHR at MUN.. The results of this study will likely be published in well-established medical, neuroscience and psychology journals. If you wish to have your child participate, please complete the portion of the form below as soon as possible and

return it to your child's daycare teacher. If you have additional questions or concerns, please contact the study supervisor, Dr. Russell J. Adams (737-8496), the project coordinator, Christina Dove (737-7684), James Drover (737-4786) or the secretary of the ICEHR at 737-8368. Please keep this sheet as a reference. There is also a copy of this letter on file at your child's daycare Centre if you happen to misplace this information. Thank-you.

Very sincerely,

Russell J. Adams, PhD.
Department of Psychology
Department of Pediatrics

Mary L. Courage, PhD.
Department of Psychology
Department of Pediatrics

James R. Drover, MSc.
PhD. Candidate
Department of Psychology

Christina Dove,
MSc. Candidate (Neuroscience)
Department of Psychology

Please return this portion as soon as possible to your child's daycare Centre or teacher. We hope to begin testing in the upcoming week.

I have read and understand all of the information pertaining to the **St. John's Regional Preschool Vision Screening Program** and wish to have my child participate.

Child's Name_____
Child's Daycare _____
Child's Birth Date_____
Days and sessions (AM/PM) that your child attends daycare
_____

Parent's Name_____
Parent's Signature_____
Today's Date_____
(Optional) : Your phone # _____and/or email
_____

Have we tested your child before?? _____ If yes, when (approximately) _____ and where_____.

Is there anything that you would like to communicate to the researchers about your child or any question that you may have?

_____
_____
_____