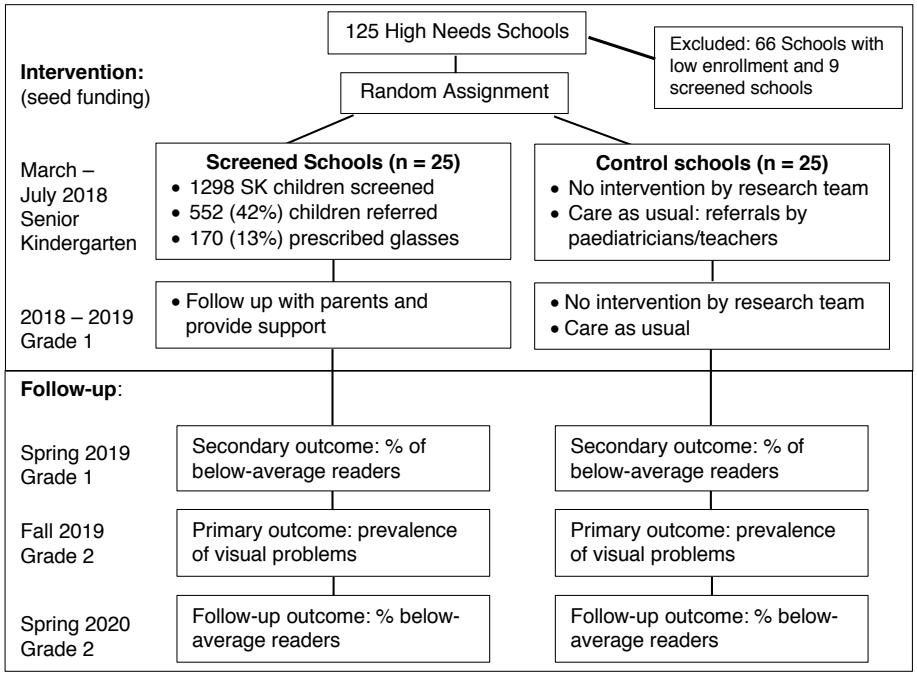
The efficacy of a visual screening program to reduce later amblyopia and untreated refractive errors.

Design. The proposed study is a single-masked cluster randomized clinical trial, with randomization and analyses occurring at the level of "schools" (i.e., we are not randomly assigning *individuals* to groups). In order to assess the efficacy of a visual screening *program,* we will compare outcomes in schools randomly chosen to receive the program and schools allocated to "care as usual". The screening tools match those to be rolled out by Ontario public health under a new standard in 2018-19 and ones recommended in the literature (see below). The outcome measures will assess later differences between schools receiving the visual screening program and control schools for prevalence of (1) visual and (2) reading problems. A lower prevalence of either or both problems in schools where the visual screening program was offered would provide evidence for the efficacy of vision screening.



The trial is being conducted with students in senior kindergarten in the Toronto District School Board's model schools for inner cities, where high needs can be predicted based on neighbourhood income and education [1]. We chose to conduct the trial with children in senior kindergarten, when children are typically 5 years old, because that is the grade during which vision screening is mandated by the new Ontario public health standard [2], because our previous research indicates that the screening tools are more accurate in senior than in junior kindergarten [3], and because age 5 is sufficiently early to effectively treat or prevent amblyopia and appears to be early enough to offset the adverse effects of untreated refractive error on learning [see Background]. Phase 1—which includes sample size calculations, randomization of schools, and vision screening in the 25 selected schools —has already occurred with seed funding (details below). This was necessary in order to

conduct the first phase of the study before all Ontario children in senior kindergarten are mandated to receive visual screening in the 2018-19 school year.

The visual outcomes will measure amblyopia, reduced stereopsis (an index of abnormal binocular vision that is linked to amblyopia), and untreated clinically significant refractive errors. They will be measured in Grade 2, after the children in the screened schools have had more than 1 year for treatment. This is sufficient time for treatment for amblyopia and its risk factors to be effective (see Background), so that, if the screening program is effective, there should be fewer cases of visual problems in the children attending schools that received the screening program than in the control schools. There should also be fewer cases of children with clinically significant refractive errors that are not being treated with glasses. Note that with this design the predictions are at the school level, not the level of individual children, and hence the analyses are not restricted to children who actually received screening in senior kindergarten or follow-up in Grade 2. Analyses at this level obviate against finding an apparent benefit of an intervention that is actually the result of which children (or parents) chose to participate in the intervention. That said, <1% of parents opted out of the screening offered in senior kindergarten.

The learning outcomes will be the prevalence of reading one standard deviation below grade level in schools offered the visual screening program versus control schools. Reading performance is assessed routinely at the end of each year by the Toronto District School Board in the 50 schools that are part of this study. The TDSB research department has agreed to share the distribution of scores with us, as they did in the pilot study described in the Background. If the screening program is effective in identifying children with amblyopia, its risk factors, and significant refractive errors, and getting them needed follow-up to receive treatment, then the distributions should differ between screened and control schools: there should be fewer children reading below grade level in the screened schools. We will compare the distributions at the ends of Grade 1 and Grade 2 in order to detect both immediate and any delayed effects.

Intervention: Vision screening and facilitated follow-up

Sample size. Although we will be measuring a number of outcomes, we chose the sample size based on the prevalence of amblyopia, because of its well-documented lifetime health consequences. Specifically, we chose a sample size that would allow us to detect a 50% reduction in the prevalence of amblyopia. A pilot study involving 1200 children in junior and senior kindergarten in 6 of Toronto's model schools for inner cities revealed a prevalence of amblyopia of 6%, a prevalence consistent with that reported for other multi-ethnic communities (see Background). Our experimental hypothesis is that a successful school-based screening program will reduce the prevalence of amblyopia by half to 3%. For sample size, standard calculations for cluster random clinical trials were performed. Sample size calculations assumed an ICC of 0.01 and used a two-sided alpha of 0.05 with 80% power. Our final sample size is based on two estimates of cluster size: when senior kindergarten enrollment was set at 50 children per school, the estimated sample size was 46 schools (N = 2,300 SK children), and when senior kindergarten enrollment was set at 41 children per

school, the estimated sample size was 52 schools (n = 2,050 children). Because senior kindergarten enrollment is not uniform across schools, we chose a sample size that fell between these two estimates — 50 schools —with actual kindergarten enrollment ranging from 38-93 children in the 25 "screened" schools and 39-109 children in the control "care as usual" schools (Total N = approximately 3,000 senior kindergarten children).

Selection of schools. The Learning Opportunities Index (LOI), a needs-based criterion of the TDSB [1], was utilized to select schools. LOI ranks schools based on external challenges affecting student success, such as median neighborhood income. The 125 schools with the most challenges are part of the Model Schools for Inner Cities program, and, as part of that program, receive annual reading assessments. We chose the 50 "model schools" with the highest senior kindergarten enrollment and randomly allocated 25 schools each to the "screened" versus control "care as usual" groups, excluding nine schools that had received vision screening (3 schools in our pilot study and 6 schools with ad hoc volunteer screening). Our biostastician, Prof. Kevin Thorpe, conducted this randomization utilizing a simple computer-generated allocation and was not involved in the actual screening.

All 25 schools selected for the vision screening program participated. We expect similar cooperation from all 50 schools for the Grade 2 visual assessments because of the support of the TDSB (see letter from collaborator Maria Yau) and because all cooperate with the Gift of Sight and Sound, which provides a similar program for children in Grade 5.

Vision screening. We conducted vision screening in the 25 selected schools between March and May 2018. There were three screening tools, the same tools to-be-used by Ontario Public Health in the newly mandated screening:

(1) HOTV with crowding bars. This is one of the most sensitive tests of acuity for vision screening of young children [4]. The child is asked to recognize a letter (H, O, T, or V) that is surrounded by a bar on each side to induce the crowding effects typical of amblyopia. The child can name the letter or match it to the same letter on a card he/she is holding. Although the VIP studies recommend testing at 5 feet for children 3-5 years old, our previous studies indicated that the standard distance of 10 feet is not problematic for children 5-6 years old, i.e., senior kindergarten. Testing occurs in a well-lit room and begins with large letters and progresses to smaller letters until the child makes too many mistakes. The threshold is the smallest size for which the child correctly identified (or matched) 3 of the letters (in a maximum of 5 presentations). Testing is monocular with the other eye patched with a Fresnel eye patch that prevents peaking. In order to pass acuity screening, the child must achieve at least 20/32 in each eye [5]. Otherwise, he/she is referred for an eye exam.

(2) Preschool randot stereoacuity. The Randot Preschool Stereoacuity test is a reliable screening test for stereo depth perception that can be performed by children as young as 2.5 years old [6-7]. When tested against results from a gold standard eye exam, the Randot yielded better results than the Stereo Smile (originally used in the VIP study) for detecting strabismus, amblyopia, and anisometropia, although neither was very good at detecting anisometropia [7].

We used this test in our previous study of the sensitivity and specificity of 5 tools against a full cycloplegic exam. For the 267 children in senior kindergarten, and a cutoff of 100 arcsecs (consistent with the VIP study), almost all children passed, yielding a sensitivity for the Randot on its own of only .27 (95% CI .14, .4) but a specificity of .93 (95%CI: .89, .97) [3]. However, 100 arcsecs may have been too lenient a criterion. The original normative data with 4355 children aged 3-18 years found that the mean stereoacuity for 3-year-olds (n=138) and 4-year-olds (n=217) was 100 arcsec while for 5-year-olds it was 60 arcsec (n = 104) [6]. Later normative data [7] indicated that >80% of children between 60 and 72 months (5-6 years, i.e., the age of children in senior kindergarten) achieve a score of 60 arcsecs or better. Therefore, to avoid missing children with binocular eye problems, we used 60 arcsecs as the criterion for passing in Phase 1 of the current study.

Testing occurs in a well-lit area, with a testing distance of 40 cm. The child is given a booklet in which the left page shows test shapes, such as a tree, and the right page shows random dot patterns in four quadrants. The child wears polarizing glasses that divide the random patterns into different images to the left and right eye, which, if fused, can be resolved into a three-dimensional stereoscopic image of the test shapes in 3 of the 4 quadrants. Otherwise, the child sees only random dot patterns. The child's task is to indicate which shape, if any, is present in each quadrant of the dot patterns by naming it or matching it to a shape on the left page. Children are tested beginning with easily resolved patterns (disparity level of 800 arcsec) and then the task is gradually made more difficult with successive disparity levels of 400, 200, 100, and 60 arcsec levels. Children unable to get 2/3 shapes correct at the 60 arcsec level are referred for an eye exam.

(3) Plusoptix autorefractor. Our earlier studies used both the Spot (WelshAllyn) and the Plusoptis. Both measure refractive errors by focusing a light onto the child's eye and recording how its reflection from the retina returns to the camera—as an appropriately focused spot, a blurred spot indicating myopia or hyperopia, or as a slit indicating astigmatism. It can also detect whether there is anisometropia (unequal refractive errors in the two eyes), a condition that can quickly leads amblyopia. The results, like previous studies [8-10}, indicated that they have reasonable sensitivity and specificity and that the results are very highly correlated, with the PlusoptiX having slightly higher sensitivity (.72 vs. .67) and the Spot having slightly higher specificity (.94 vs. .90). Therefore, we decided to use only one autorefractor for this study, namely the PluxoptiX, the one with higher sensitivity to detect eye problems.

The child is tested in a dimly lit space, either a separate room or a large pop-up tent within a larger room. The child sits 100 cm away from the handheld device and is asked to look at the lighted smiley face in the middle. The device automatically takes measurements when the child is at the right distance and fixating. Testing in a dim space facilitates capturing the child's attention on the lighted smiley face and obtaining good measurements because of the increased pupil size.

In our original study of 712 kindergarten children, all of whom received full eye exams, we used the autorefractor screening cutoffs recommended by the American Association of Pediatric Ophthalmology and Strabismus [5]: hyperopia >+3.5D; astigmatism >1.5D; myopia >-1.5 D; anisometropia 1.5D SE. ROC analysis of the results suggested that a cutoff of +2 for hyperopia would improve sensitivity without lowering specificity significantly. Therefore, we used the following referral criteria in Phase 1: hyperopia >+2.0D; astigmatism >1.5D; myopia >-1.5 D; anisometropia 1.5D SE [3].

Children who did not pass all three screening tests were referred for full cycloplegic optometry exams, scheduled at school for the convenience of the parents. Children already wearing glasses were tested on acuity and stereoacuity while wearing their glasses and referred if they did not pass, because they might need a prescription change. Vision screening was completed by internationally trained medical graduates working with the Toronto Gift of Sight and Sound. They were trained and monitored by our research team. Screening time averaged 15 minutes/child (range 11-23 minutes). Overall, 552 (42%) of the screened children were referred for an optometry exam, although the referral rate ranged from 25 to 63% across the 25 schools.

The follow-up optometry exams were standardized and occurred in the school, unless it was more convenient for the parent to go to the optometrist's office. They were full exams, with cycloplegia, with the parent present, and followed a standard protocol specified by the research team and recorded on a standardized assessment sheet for each child (Appendix **1).** Specifically, they were asked to assess visual history, monocular visual acuity at near and far, strabismus, binocular function, abnormalities of the anterior segment, and cycloplegic refraction. A member of the research team was present at each in-school exam and assured that each child's optometry assessment sheet was completed fully. If the child needed glasses, the frames were picked out at the time of the exam and dispensed either at the school or the optometrist's office. Overall, 364 (66%) of the parents of referred children arranged for an in-school optometry exam, a percentage similar to that we found when we offered a similar program in 42 schools in **15 communities** across Ontario [3]. Glasses were dispensed to 170 (13%) of the screened children.

To facilitate follow-up, optometrists were chosen with offices near the schools. For 3 of the 25 schools where this was not possible, an out-of-town optometrist conducted the in-school exams, but a community optometrist with an office near the school accepted the patients for any needed follow-up. All parents agreed to be contacted by the community optometrist and our program when it is time for follow-up. With the parent's continuing consent, we will monitor children in "screened" schools who have a visual problem for two years, by reminding parents about follow-up appointments and replacing broken and lost glasses. The frequency of monitoring will vary depending on the child's condition and decisions by the treating optometrist. Because the goal of our research is to evaluate the efficacy of a *screening program,* our primary concern is that screening correctly identifies children with visual problems and facilitates their care for amblyopia and/or refractive errors. Whether the recommended treatments are optimal or whether children comply with the prescribed treatments will not be addressed in this study. Rather, the study will assess

whether a program of vision screening and facilitated follow-up is effective, when it is implemented in the real world, complete with real-world limitations like student absences on the day of screening, lost glasses that are not replaced, parents who do not show up for optometry exams, etc. The control "care as usual" schools will not receive any intervention by our research team, but children may receive care through regular referral channels (e.g., a teacher may recommend that a child see an optometrist; a parent may take the child for a routine exam).

Follow-up: Visual outcomes in Grade 2

We will employ different testers for visual assessments in Grade 2 than those involved in Senior Kindergarten, in order to prevent experimenter bias; specifically, Grade 2 testers will not know whether a particular school was part of the "screened" or control "care as usual" group. Each screener will also be blind to the results of the other tests.

All assessment will occur during a 2-month period in the late fall of 2019, after the school routine has settled, and be randomly scheduled among the "screened" and control "care as usual" schools. We anticipate that principals of all 50 schools will agree to participate because no Toronto principal has declined our vision screening program in the past and because principals of these high-needs schools welcome extra services for their children.

Parents will be sent a letter describing the vision checking and will be given the option to opt out. We anticipate that few parents will opt out, because less than 1% did so when the children were in senior kindergarten. Parents will receive a letter indicating whether or not the child passed the vision checks. When the child did not, the letter will indicate the need for a follow-up optometry exam and provide advice on how to book an OHIP-covered eye exam.

Primary Clinical Outcomes: Amblyopia and Untreated Refractive Error

Amblyopia. One primary outcome measure is the prevalence of amblyopia in "screened" vs. control "care as usual" schools in Grade 2 (i.e. after allowing at least one year for treatment). We will use the HOTV Paediatric wall chart with proportional spacing to measure the acuity of each eye, while the untested eye is patched. By Grade 2, children can be tested with linear letter charts, which are more sensitive to the crowding effects of amblyopia than charts with crowding bars or other arrangements [11]. If any child is unsure of the names of the letters (e.g., because they are in the process of learning English), the test can be based on matching to a card held by the child with the same letters. Amblyopia will be defined as a 2-line or greater difference in acuity between eyes [12], or, in the rarer cases of bilateral amblyopia, acuity worse than 20/40 in the eye with better acuity [13]. Children wearing glasses will be tested while wearing the glasses and, should they fail with glasses, the parents will be informed that the child may have outgrown the glasses and to check with the optometrist to see if a follow-up appointment is warranted.

A secondary measure will be stereoacuity, which is reduced in patients with some types of amblyopia [14], often secondary to strabismus. We will use the Randot test of stereopsis because it includes values down to 20 arcsecs (unlike the preschool version that stops at 40 arcsecs) and because by Grade 2, virtually all children can understand the instructions. The cut-off for normal stereoacuity will be 30 arcsecs, the mean value found in normative data for children 7-8 years old [6]. Although monocular cues can lead to spurious results at the larger disparities of the Randot, they are not an issue for disparities less than 140 arcsecs [15]. As with acuity, the measure of interest is best corrected vision and hence children with glasses will be tested while wearing them. Should the child fail with glasses, the parent will receive a note that the child's glasses may have been outgrown and to check with the optometrist about follow-up.

Untreated refractive errors. The second primary outcome measure is the prevalence of clinically significant refractive errors that are not being treated by prescription glasses. We will measure refractive error using the PlusoptiX autorefractor, as we did for the senior kindergarten children, but with the cut-offs adjusted to be appropriate for Grade 2. There are variable opinions about what level of refractive error is clinically significant at this age, but a reviews of the evidence and of current practice [16-17] suggest: hyperopia ≥+1.5D, myopia ≥ -1.0D, astigmatism ≥0.75D, and anisometropia ≥1.0D.

As a clinical service, parents of children who do not pass all three tests will receive a letter recommending that they book a full eye exam. Children already wearing glasses will not be tested with the autorefractor because any refractive error is already under treatment and because autorefractors do not work accurately through glasses. However, if they do not pass the visual acuity and/or stereoacuity tests, their parents will receive a letter indicating that the child may have outgrown the glasses and that they should check with their optometrists to determine if it is time for follow-up.

Secondary outcomes: Reading scores

Educational assessments are collected annually by TDSB at all high needs schools. using the Canadian Test Centre's CAT4 [18]. These are standardized achievement tests, with normative data based on the testing of over 60000 students in kindergarten through Grade 12 from over 300 schools across Canada, stratified to capture a representative sample. TDSB has agreed to share with us the distribution of scores from each of the 50 schools for the Reading subtest. We will analyze the distribution of scores in "screened" versus control "care as usual" schools in Grade 1 (shortly after the intervention) and Grade 2 (after over 1 year for treatment).

Analyses of data. The statistical analyses have been planned with our co-investigator, Dr. Thorpe. Demographic data describing the students at each school will be summarized with descriptive measures (age, sex). The primary outcome (prevalence of amblyopia) will be analyzed using the method of Generalized Estimating Equations [19] for binary outcomes with a logit link and school as the unit of clustering. The detection effect will be expressed as an odds ratio with 95% confidence interval. Similar analyses will be performed separately for the proportion of children with reduced stereoacuity and untreated refractive errors.

Children already wearing glasses will be included in the analyses of amblyopia and stereoacuity but not in the analysis of untreated refractive errrors.

For the reading assessments. the data will consist of the distribution of scores across 9 standardized categories, centred on the normative value of 5. Because ~75% of children will not have a visual problem, we do not expect vision screening to increase the mean reading level; instead, we hypothesize that screening will lead to fewer children with untreated eye problems and thereby reduce the proportion of children experiencing reading difficulties (i.e., 1 standard deviation below the norm; children farther below the norm are likely to have additional problems). Therefore, as with amblyopia and other visual disorders, we will compare the proportion of children in screened versus control schools with a score below the norm. Data for Grade 1 will reveal if the vision screening program results in fewer initial reading problems. Data for Grade 2 will reveal if any difference emerges after more than 1 year of treatment for children in the "screened" schools. We will use percentage turn-over (i.e., number of children leaving and joining a school between SK and Grade 2) for each school as a factor in the regression analyses because such turnover will work against finding an effect of screening at the school level.

A positive effect of the visual screening program will manifest as significantly fewer children in Grade 2 with amblyopia, reduced stereopsis, untreated refractive errors, and/or below average reading scores in screened schools compared to control "care as usual" schools. If so, subsequent studies can assess which parts of the visual screening program are essential in the Ontario health care context in order to identify children with eye problems (e.g., is screening alone effective without in-school follow-up exams?) and assure they receive treatment, thereby preventing or ameliorating amblyopia and eye-related learning difficulties.

References

1. Toronto District School Board. The 2017 Learning Opportunities Index: Questions and Answers. 2017; Toronto.

2. Ministry of Health and Long-Term Care Ontario. Ontario Public Health Standards: Requirements for programs, services, and accountability. Effective Date January 1, 2018: p.50-52.

3. Maurer, D., Wong, A., & Nishimura, M. Kindergarten Vision-Testing Programme 2015-2018 Report. (Appendix 2).

4. Cotter, S. A., Cyert, L. A., Miller, J. M., & Quinn, G. (2015). Vision screening for children 36 to <72 months: recommende practices. Optom Vis Sci, 92(1), 6-16.

5. Donahue, S. P., Arthur, B., Neely, D. E., Arnold, R. W., Silbert, D., Ruben, J. B. et al. (2013). Guidelines for automated preschool vision screening: a 10-year, evidence-based update. J AAPOS, 17(1), 4-8.

6. Birch, E., Williams, C., Drover, J., Fu, V., Cheng, C., Northstone, K. et al. (2008). Randot Preschool Stereoacuity Test: normative data and validity. J AAPOS, 12(1), 23-26.

7. Afsari, S., Rose, K. A., Pai, A. S.-I., Gole, G. A., Leone, J. F., Burlutsky, G. et al. (2013). Diagnostic reliability and normative values of stereoacuity tests in preschool-aged children. Br J Ophthalmol, 97(3), 308-313.

8. Panda, L., Barik, U., Nayak, S., Barik, B., Behera, G., Kekunnaya, R. et al. (2018). Performance of Photoscreener in Detection of Refractive Error in All Age Groups and Amblyopia Risk Factors in Children in a Tribal District of Odisha: The Tribal Odisha Eye Disease Study (TOES) # 3. Transl Vis Sci Technol, 7(3), 12.

9. Matta NS, Singman EL, Silbert DI. Performance of the plusoptiX S04 photoscreener for the detection of amblyopia risk factors in children aged 3 to 5. J AAPOS. 2010;14:147-9.

10. Moghaddam AA, Kargozar A, Zarei-Ghanavati M, Najjaran M, Nozari V, Shakeri MT. Screening for amblyopia risk factors in pre-verbal children using the Plusoptix photoscreener: a cross-sectional population-based study. Br J Ophthalmol. 2012;96:83-86.

11. Norgett, Y., & Siderov, J. (2017). Effect of stimulus configuration on crowding in strabismic amblyopia. J Vis, 17(13), 1-14.

12. Maurer, D., & McKee, S. P. (2018). Classification and diversity of amblyopia. Vis Neurosci, 35, E012.

13. Hunter, D., & Cotter, S. (2018). Early diagnosis of amblyopia. Vis Neurosci, 35, E013.

14. McKee, S. P., Levi, D. M., & Movshon, J. A. (2003). The pattern of visual deficits in amblyopia. Journal of Vision, 3(5), 380-405.

15. Hahn, E., Cornstock, D., Durling, S., MacCarron, J., Mulla, S., James, P. et al. (2010). Monocular clues in seven stereotests. DMJ, 37(1), 4-13.

16. Leat, S. J., Mittelstaedt, A., McIntosh, S., Machan, C. M., Hrynchak, P. K., & Irving, E. L. (2011). Prescribing for hyperopia in childhood and teenage by academic optometrists. Optom Vis Sci, 88(11), 1333-1342.

17. Leat, S. J. (2011). To prescribe or not to prescribe? Guidelines for spectacle prescribing in infants and children. Clin Exp Optom, 94(6), 514-527.

18. https://www.canadiantestcentre.com/CAT4/CAT4-About.php

19. Donner A & Klar N. Design and analysis of cluster randomized trials in health research. Arnold, London: 2000.