Article

Learning-Related Visual Problems in Baltimore City: A Long-Term Program

Paul Harris, OD

A longitudinal, single-masked, random sample study of children at a Baltimore City Public Elementary school documents the prevalence of learning-related visual problems in the inner city of Baltimore and tests the effectiveness of vision therapy. Vision therapy was provided to one of the randomly selected groups and data were collected on optometric tests, visual performance tests, and standardized achievement tests before and after treatment was provided. Data presented show that the vision therapy program has made a significant difference in the demand level of reading that could be read for understanding, in math achievement on standardized testing, and in reading scores on standardized testing, as well as on infrared eye-movement Visagraph recordings, which show significant changes on nearly all mechanical aspects of the reading process. This article makes a strong case that untreated learning-related vision problems are a significant public health concern and that the profession of optometry has a treatment modality that can address these problems in a significant way. The article presents many of the difficult questions that had to be addressed during the early formation stage of the study and during the execution of the study. The rationale behind the key decisions that had to be made during each step of the program is provided so that future researchers may be able to replicate this study with full knowledge of what to expect. **Key Words:** *learning-related vision problems, vision development, vision therapy, visual training, perception, performance, school-based vision therapy, Visagraph, inner-city, ocular motor dysfunction, binocular dysfunction, visual attention problems, convergence insufficiency.*

INTRODUCTION

In 1997 representatives of the Abell Foundation of Baltimore came across an article in the *Boston Globe* chronicling the work of Dr. Antonia Orfield with inner-city students in Boston. The article started the foundation on a search for information, which led them to investigate to what extent a problem similar to that which Dr. Orfield was addressing in Boston existed in the city of Baltimore. After suitably convincing themselves that a public health problem did potentially exist in the youth of the city of Baltimore, representatives of the Abell Foundation asked me to be the chief investigator for a program that would

Correspondence regarding this article should be addressed to Paul Harris, OD, 110 Old Padonia Road Suite 300, Cockeysville, MD 21030; e-mail: HarriOpto3rd@compuserve.com.

first identify the prevalence of learning related vision problems in the Baltimore City Public Schools and then test the efficacy of treatment for those conditions identified. At the time, I was living in Denmark on a 1-year sabbatical to teach behavioral vision care in Europe. The following article chronicles step by step how I worked toward achieving these goals. I describe how the formal research protocols were established, report the initial data collected on the children, the preliminary data following the first year of the study, and the data at the end of the second year of the study, and then comments on each of the findings in detail.

Many decisions were made at critical steps in the development of the protocols. It is my intent to share the factors that were weighed in each of these decisions so that 1) the reader may fully understand the rationale that went into the choices that were made in the development of the protocol and 2) those who want to may use the information presented here to guide them in making the key decisions necessary to establish similar programs in other communities.

STATEMENT OF THE PROBLEM

Baltimore, Maryland, "The City That Reads," has had trouble living up to its motto. Like most large cities, its public schools have lagged behind in basic academic achievement as well as scores on standardized tests compared with suburban or rural school districts. During academic year 1997–1998, Baltimore City had 17.6% of its students in special education. This figure is well above the national average of 12%. During the same time, Baltimore City spent an average of $9,700 per pupil enrolled in special education. The city spent only $3,100 per pupil per year for those students not receiving any extra help in school. With almost 110,000 students in school, Baltimore would have nearly 1,936 students in special education, costing the city nearly $12,777,000 extra each year.* The Abell Foundation, a Baltimore-based foundation, has dedicated itself to working to improve education in Baltimore and to identify interventions that show promise in helping to make Baltimore's motto a reality.

WHAT TO TEST FOR? LEARNING-RELATED VISUAL PROBLEMS

Learning-related visual problems are defined as those problems or dysfunctions of the visual process that affect the child's ability to learn. These problem areas can be broken down into three types, which may be labeled nontechnically as tracking, teaming, and focusing.

The category of "tracking" problems in this article is defined as the inability to accurately and/or efficiently move through space and time the area of space the person selects from which the person is deriving meaning and directing action in. For example, in a sport such as tennis, the better players generally have a better ability to keep their "spotlight of attention" extremely close or nearly exactly centered on the ball as it travels in flight on a continuous basis. In the act of reading, the person must sequentially move his or her fixation point from place to place across the line of text. Different patterns of fixations, regressions, and return sweeps when looked at on a statistical basis or on a global basis reveal various scan patterns, each of which is characteristic of different types of reading and/or different levels of development in reading ability. These scan patterns are indicators of the degree to which a person has developed the ability to sequentially move his or her spotlight of attention across the page. Excessive amounts of head movement vs eye movement free of head and upper-body movement are measures of visual development level. Excessive numbers of fixations or regressions or an excessive number of regressions in proportion to the number of forward fixations are all indicators of a lack of development of the ability to extract information from the printed page. All of these would be considered as "tracking" problems. The most common diagnostic codes used clinically would be the group of codes labeled ocular motor dysfunction (OMD), which includes the diagnoses OMD-Deficiencies of Saccadic Eye Movements (ICD-9-CM 379.57),

**Figures based on Maryland School Performance Reports compiled by LEA:30 release 1.5, July 1999.*

OMD-Deficiencies of Pursuit Eye Movements (ICD-9-CM 379.58), and OMD-Abnormal Oculomotor Studies (ICD-9-CM 794.14).

The category of "teaming" problems in this article is defined as any difficulty or interference in using the bidirectional flow of data through each of the neurological channels (eyes to the brain and brain to the eyes) in a unified manner. This has been termed "binocularity." However, this term puts too much emphasis on the eyeballs, as in: "The two eyes are not teamed together." What is being hinted at here is a broader understanding of how the person builds his or her internal representation of reality. Hoffman[1] states that, "*Vision is construction. Vision is not merely a matter of passive perception; it is an intelligent process of active construction. What you see is, invariably, what your visual intelligence constructs.*" When the person is effectively using the flow of data through both channels to and from the outside world in a seamless manner, then their understanding, insights, and use of the space-time continuum, which is the real world, are greatly enhanced. (A single channel can be thought of as the connections from one eye to the brain and back again. Because we have two eyes and each is connected reciprocally with the brain, there are two channels of flow. Alternately, one could think functionally of the connections the brain has to each area of visual space. The right side of visual space goes to the left visual cortex through connections from the left side of the retina from each eye. In this way of organizing the view of channels, the channels would be functionally organized as linkages between the brain and the outside world, with each channel involving parts of each eye.) Interruptions in this seamless use of both channels of flow are clinically diagnosed as the following binocular dysfunctions: General Binocular Vision Dysfunction (ICD-9-CM 368.30), Suppression of Binocular Vision (ICD-9-CM 368.31), Fusion with Defective Stereopsis (ICD-9-CM 368.33), Convergence Insufficiency (ICD-9-CM 378.83), and Convergence Excess (ICD-9-CM 378.84), *etc.*

The category of "focusing" problems in this article is defined as any difficulty in either (1) the selection of an area of space from which to derive meaning and direct action or (2) the inability to maintain visual attention at that location for as long as the person needs to sustain deriving meaning and direction of action at that place in space. These types of problems manifest clinically in one form as an inability to lock on to a target. Rather than fixate on an object directly, several attempts may be made in or around the location of the object. A person with this difficulty in achieving a lock-on of the spotlight of attention and of fixation displays behavior that is observed as high levels of distractibility and may be classified as having an attention problem such as attention deficit disorder with or without the hyperactivity.

The peripheral or magnocellular visual system is highly specialized as an alarm system and is excellent at detecting change in the environment and acts as an alarm system or threat indicator. If the person is unable to get a good central lock-on to the object, these threat indicators may trigger forms of the primitive startle reflex, resulting in the person's attention being drawn off task, which may thus appear to be a fundamental attention problem. Unfortunately, the formal diagnostic categories available to the profession of optometry do not fully address the manner in which visual attention is currently understood. It has become very difficult to tease out of attention models the exact degree to which visual process difficulties appear or are revealed as general problems of attention. In many instances the problems become absolutely intertwined so that it becomes impossible to delineate where the visual attention problem begins and ends and where the general attention problem begins and ends. It is not within the scope of license of optometry to make the diagnosis of attention deficit disorder (ADD) either with or without the hyperactive component. The formal diagnosis of Accommodative Dysfunction (ICD-9-CM 367.5) applies only to one very small subset of the overall issue of visual attention, which is being addressed by using the term "focusing." Focusing in this context refers to both the actual appreciation of clarity or lack thereof in the perception of our lighted world and to the focusing and channeling of overall visual attention to perform the tasks necessary to derive meaning and direct action accurately and efficiently.

FIRST STEPS: WHAT IS THE PREVALENCE OF "THESE PROBLEMS" IN BALTIMORE?

Different research protocols may require vastly different resources, may have very different price tags to conduct, as well as expose the program to various types of critiques. In each design, there will be trade-offs between what is affordable and what will give the best information on which to extrapolate to larger populations. Questions that had to be answered all kept coming back to a core question: what is the prevalence of learning-related visual problems (LRVPs) in the population to be studied? If the prevalence was lower, more schools would need to be included to get the number of subjects and the number of matched controls in the study for significance to emerge from the data. If the prevalence of LRVPs was higher, fewer schools would be needed for significance to emerge.

In general, the minimum number of subjects needed in a study to provide data that can readily be extrapolated to larger populations has been established by those involved in statistics as 30. Estimates were needed to determine how many children should be included on the front end of the study to guarantee that at least 30 completed the program.

There were other circumstances that had to be considered in working to achieve N = 30. The treatment protocols would mirror that done in my private practice. The curriculum of vision therapy done in my practice is exactly that which is taught by the Baltimore Academy for Behavioral Optometry (BABO) for the treatment of LRVPs. The average length of treatment for LRVPs in private practice is from six to eight months. Abell Foundation representatives with work experience in the Baltimore City Public Schools warned me to expect major problems with attendance as well as with children moving in and out of specific schools. These problems could easily diminish the numbers in both the control and treatment groups once a study protocol was established and later as the study moved well into the execution phase. Thus, any protocol design would have to be robust enough to take into account these factors leading to potentially high dropout rates.

Visual Screening To Establish Prevalence

During academic year 1997–1998, the Baltimore City Public School system had 112 elementary schools. Representatives of the Abell Foundation helped by selecting two schools that they had experience with in which visual screenings would be performed to collect the prevalence data needed to develop the research protocol that would be used in the full treatment program. The two schools selected, Hampstead Hill Elementary School and Westport Elementary School, were from two neighborhoods in Baltimore City that represented a cross section of the city. Both schools were in economically depressed areas of the city, one on the west side of the downtown area and one on the east side of the downtown area. One area had predominantly Caucasian children; the other had children who were predominantly African American. Testing was performed on all first- and third-graders present in these schools on the day of their respective screenings.

I was given the school scores for the previous 3 years on the Maryland School Performance Assessment Program (MSPAP) testing. This is a test given in the state of Maryland to assess the level of education at each school and is used as a measure of how well that school is performing its duties. No individual scores are obtained; only grouped scores are used to compare one school with another and each school against a minimum standard level. The lower the number the worse the children have assimilated what is being taught and/or can apply what they have learned to the test. Westport Elementary School had scores in the lowest rankings in the state at 2.8 in 1994, 3.5 in 1995, and 3.3 in 1996. Hampstead Hill ranked in the top 10% of the Baltimore City Public Schools but still had ratings of only 13.8 in 1994, 11.4 in 1995, and 15.9 in 1996.† (These numbers are out of a possible 100.)

The following is the screening battery, performed by optometric office staff including vi-

†Figures extracted from an article in the Baltimore Sun newspaper by Abell Foundation Staff in the fall of 1997.

sion therapists. No optometrists were directly involved in the visual screening. Tests performed included the New York State Optometric Association (NYSOA), King-Devic Saccadic Test, convergence near point (CNP) or near point of convergence (NPC), Randot Stereo Acuity, Keystone Visual Skills 4-Ball Test at both distance and near, 30-second +/−2.00 flipper cycle testing, +1.50 diopter test at far point, and distance visual acuity. Table 1 shows the criteria used for each test to determine the presence or absence of a visual problem. The pass/fail criteria were the same for both age groups tested on all tests except for the NYSOA King Devic Saccadic Test, for which a longer time was allowed for the first-graders than the third-graders.

Screening Results

A total of 129 students were tested at Hampstead Hill Elementary School, divided into 60 first-graders and 69 third-graders. A total of 136 students were tested at Westport Elementary School, divided into 79 first-graders and 57 third-graders. Table 2 shows the number of children who failed each of the screening areas.

How Many Were OK?

Of all children tested at Westport Elementary School, only 20 of 136 (14.7%) did not fail any area of the screening. Of all children tested at Hampstead Hill Elementary School, only 34 of the 129 (26.47%) did not fail any area of the screening. Across both schools this means that 54 of the 265 children (only 20.4%) passed the entire screening battery.

Discussion of the Screening Data

Several things are interesting to note. In both schools, the failure rates on the NYSOA King-Devic Saccadic test dropped from the first to the third grades. The score for failure in both instances was 1 SD slower than the expected value for the age group being tested. This is a slightly looser standard than might be more formally used, which would often be 2 SDs below age expected. However, 1 SD below was used here as a general indicator of problems in the overall population being tested. In addition, the higher the N values are above 30, the more significant smaller differences become, allowing tighter standards. In both schools, there were some exceptionally long times for a few of the first-graders on this test, because some of these children didn't know their numbers well enough or reliably enough to get through the test. Some did not even know the names of all 10 single-digit numbers.

The percentage of children who demonstrated a lack of ability to converge up to the 4-inch standard on the CNP test increased in both schools as the children got older. The percentages were very similar in both schools, increasing from the low of 15% and 17.7% with receded CNPs in the first grade to 21.7% and 21.0% in the third grades. A longitudinal study following the same group of children would be needed to detect significance in the emergence of these problems during the first few years of school. Lieberman *et al.*[2] used the same criteria in the validation work on the NYSOA Screening Battery. With an N of 1963 students tested, they found a failure rate of only 4.6%. Cooper and Duckman[3] found a median rate of 7% of the population with convergence insufficiency across a number of studies.

The Random Dot Stereo test used is the special Randot test (Stereo Optical, Inc. Chicago, Ill), with both object and background done completely with random dot patterns. This test often shows lower measures of stere-

TABLE 1. Visual Screening Pass/Fail Criteria

Test	First-Grade Criteria	Third-Grade Criteria
NYSOA King-Devic Saccadic Test	> 150 sec	> 100 sec
CNP/NPC	> = 4 inches	> = 4 inches
Randot Stereo Acuity	> 50 sec of arc	> 50 sec of arc
Keystone 4-Ball Test Distance & Near	2 or 4 balls	2 or 4 balls
30-second +/− 2.00 Flipper Test	< 8 flips	< 8 flips
+1.50 Distance Test	20/20 or better VA	20/20 or better VA
Visual Acuity	Worse than 20/40	Worse than 20/40

TABLE 2. Visual Screening Data

Test	Failures (% of total) Hampstead Hill First Grade (N = 60) (%)	Hampstead Hill Third Grade (N = 69) (%)	Westport First Grade (N = 79) (%)	Westport Third Grade (N = 57) (%)
NYSOA King-Devic Saccadic Test	34 (56.6)	12 (17.4)	55 (69.6)	19 (33.3)
CNP/NPC	9 (15)	15 (21.7)	14 (17.7)	12 (21.0)
Randot Stereo Acuity	23 (38.3)	10 (14.5)	29 (36.7)	21 (36.8)
Keystone 4-Ball Test Distance	23 (38.3)	29 (42.0)	30 (37.9)	24 (42.1)
Keystone 4-Ball Test Near	18 (30)	17 (24.6)	20 (25.3)	22 (38.6)
30-second +/– 2.00 Flipper Test	25 (41.6)	14 (20.3)	33 (41.8)	22 (38.6)
+1.50 Distance Test	8 (13.3)	8 (11.6)	8 (10.1)	3 (5.3)
Visual Acuity	0	0	0	0

opsis than those tests with solid objects on random dot backgrounds or those stereo tests with solid objects in the absence of random dot backgrounds, such as Wirt Circles, which appear on the opposite side of the Stereo Fly. These more traditionally used tests have monocular clues to depth that may be used by the subject that are not present in the fully random dot test used in this screening. This may account for the higher than expected fail rate on the Randot Stereo Acuity test and may actually be more indicative of the teaming problems described above. Lieberman *et al.*[2] used the Wirt Circles stereo acuity test as part of the NYSOA Screening Battery, and their failure criteria was less than 60 seconds of arc. The test they used did not have finer samples below 40 seconds of arc. Despite this, they still had a 16.6% failure rate on an N of 1,979 subjects. Our rates, which were in the middle 30th percentile for three of the four groups tested, were significantly higher.

On review of the data, the number of failing responses on the 4-Ball Keystone Skills Test distance and near cards seems higher than expected. Some types of stereoscopes require different degrees of separation on the cards to measure orthophoria. Using a card designed for one stereoscope in one with a different separation may generate false-positive results. However, the stereoscope used for the screening was the Keystone Ophthalmic Telebinocular model 46C for which the cards were developed. Therefore, those responses that had many children seeing either two or four balls—which indicates either suppression of binocular vision or a misalignment of the two channels—are not to be discounted and may indeed be an indicator of general teaming or binocularity problems in this population. Lieberman *et al.*[2] using the same criteria we did, found a 7.8% failure rate at distance with an N of 1,973 and a failure rate of 10.6% at near with an N of 1,971. A rather high percentage of the children tested could not complete the minimum eight flips of the +2.00/–2.00 diopters binocular flippers in the 30 seconds. It is felt that this relates to their inability to shift attention efficiently from one location in space to another, particularly along the z-axis, which is toward and away from the self. Lieberman *et al.*[2] used six flips in 30 seconds and he found a 37.0% failure rate in an N of 1,969 subjects. The +1.50 diopter distance test was used to screen for adverse hypertopia. Adverse hyperopia is defined as an amount of hyperopia, or farsightedness, which when left fully uncompensated for at near, often is a major factor in an inability to sustain visual attention at near for long periods of time. The number of children with adverse hyperopia ranged from 5.3% to 13.3% in each of the four groups tested. Lieberman *et al.*[2] used the same criteria and found a much higher failure rate of 30.6% based on an N of 1,963 subjects.

The final finding was not fully unexpected. However, one would have expected that of a total of 265 children or 530 eyes that at least one channel in one child would be found to be incapable of reading at least 20/40 visual acuity. This was not so. Every child with each eye was able to read at least 20/40 in each condition. Therefore, there were no high degrees of refractive conditions beyond the adverse hyperopia already mentioned above, nor was any amblyopia detected in this particular sampling.

Looked at as a whole, the data from the visual screenings demonstrated that there was a very high level of prevalence of tracking, teaming, and focusing difficulties in the populations tested. This was very significant in moving to phase 2 of the research program—the formal design of the testing and treatment protocols.

HOW MANY SUBJECTS?

The data collected from the screening showed that a very high percentage (approximately 80%) of the children in the Baltimore City Public Schools displayed evidence of LRVPs. The target minimum sample size of 30 in both the control and treatment groups would need to be higher to account for the expected dropout rates. To strengthen the conclusion of the study, the target number of subjects in both the control and treatment groups was bumped to 50 children. Therefore, 75 students were to be selected for and given vision therapy. Over the length of the study, if one third of each group were lost to follow-up given then there would still be 50 in each of the groups. If the attrition in each group were greater, then the ability to extrapolate to the broader base of students in Baltimore City would be compromised. However, the study could have sustained as much as a 50% loss of subjects in each group and still have had more than the required 30 subjects in each to merit publication.

HOW MANY SCHOOLS?

With the target sample data size set at 150 with 75 to be randomly assigned in each group, the next question to address was how many schools would be needed to satisfy these numbers. Six candidate schools were invited to a meeting at which the main concepts of the study were presented, and what was expected of each school in terms of space and access to children was explained. Of the six schools, four indicated a strong desire to have their school selected as a site for the program. Site visits were arranged to each of these four schools to determine suitability of space for the testing and treatment phases of the program. The total number of fourth-grade students enrolled at each school for the academic year of 1997–1998 was obtained. Additional data included the percent of the children in the third and the fifth grades at each school who had scored satisfactory on the MSPAP testing performed in 1997. The final piece of data supplied was the percentage of children on the Free/Reduced Lunch Program, which could be considered a measure of the general economic status of the area. Neither of the two schools at which the screening had been performed was included on the school candidate list. Table 3 shows the candidate school profile.

School #254 and school #27 decided after the meeting to withdraw their school from consideration. School #36 and school #98 had the required number of students to perform the entire program in a single school. If either of the other schools were chosen, then at least two schools would have to be involved in the study to achieve the full size of the two groups as determined above.

There were rather strong pros and cons to working with one, two, or four schools. There were several benefits of doing the entire program in one school. By working with one school, only one set of equipment would need to be purchased and only one team of vision

TABLE 3. Candidate School Profile

School	Fourth-Grade Enrollment	MSPAP Reading, % Satisfactory, Third Grade	MSPAP Reading, % Satisfactory, Fifth Grade	Free/Reduced Lunch Program, % of Enrollment
#254	192	5.2	6.9	84
#27	125	6	1.6	90
#215	139	10	15.6	61
#36	203	6.5	11.7	94
#95	120	12.3	6.2	91
#98	156	6.7	8	100

therapists would be required to conduct training and be monitored for quality control. Because the controls and treatment groups would be in the exact same environment, same school, same teachers, same curriculum, same neighborhoods, same socioeconomic profiles, *etc.,* the controls would truly be as matched as possible with the treatment group. Thus, working all within one school had great appeal.

However, by working in a single school, the potential existed for some problems as well. One such problem relates to the ability to make generalizations to other schools or other parts of the city regardless of how successful things looked at the one school chosen. Another great concern was whether an event that took place either at that school or in that community during the period of the study could greatly influence the measured effects of the treatment.

Another consideration was to use a single school as a treatment school and a second school as a matched control. The benefits of this would be that it would be very easy to keep the two groups apart to make sure that no transfer occurred during off time between the two groups. Potential problems of this would include the constant nagging suspicion, no matter how the numbers looked, of whether the two schools were truly matched in every respect. Even if the front-end data looked like the samples were the same, could it be said with any degree of certainty that every aspect that would have been controlled in a single school would be controlled to the same degree? I did not believe that this could be guaranteed. In addition, it would be hard to keep the testing staff fully masked when they did the post-testing. On the basis of which school they went to, to perform the post-testing they would know which children were treatment children and which were control children. The staff could never be allowed to know that the protocols specified an entire school as a treatment group and the other as a control. It wouldn't take a tester too long to figure out which school they were testing, which could potentially affect their post-testing results. I did not want such an easy "give-away" to those doing the post-testing.

An additional protocol could have been developed, which would work in up to four different schools at the same time. Could this ensure more homogeneity by using two schools for treatment and two for control? Better yet, could half of the children in each of the four schools be randomly assigned to the treatment group and half assigned to the control group so that statistics could eventually be measured between groups in a single school and then among groups in different schools? Clearly this was the best choice. However, the implications were that four schools would have to be equipped with staff and equipment. Logistics of training four treatment teams and arranging for the testing at four separate sites would also have to be worked out. Quality control and supervision issues appeared overwhelming with this type of protocol, although it was recognized that it was potentially the best overall. Lastly, there were some challenges to think about in terms of the added levels of complexity that would be associated with the statistics of such a complex design.

The strongest plea for our services was coming from the largest of the four schools still in the running, school #36. When space for treatment at each of the other three schools was deemed to be inadequate and the rental of a temporary trailer for the vision therapy facilities at each of the other schools was factored in, school #36 was chosen as the single school in which to conduct phase 2 of the research, phase 1 having been the screening programs.

WHAT GRADE/AGE GROUPS TO WORK WITH?

At this point, another critical decision had to be made. What age group of students should be worked with? There were quite a few things that went into the selection that was made. I was given the impression by the foundation that within approximately 2 years they wanted to see some preliminary results that indicated that moneys were being spent on a worthwhile program. I did not know to what degree the problems were developmental in nature vs being influenced by other factors beyond the scope of what is seen in a typical cross section of private practices around the country that provide vision therapy services. In other words, were there factors that had not yet been encountered in the private practice setting that would alter the potential for improvement? If so, to what degree would these factors militate against the work showing benefits? If these other factors were present, to what degree would picking a specific age group maximize the potential to show an impact in 2 years? If the study showed benefits on this age group, then additional funding might be granted later to investigate other age groups.

It was anticipated that the curriculum of treatment would mirror what I had been using in my private practice for many years and was being taught in the core courses by the Baltimore Academy for Behavioral Optometry (BABO) in its series of postgraduate clinical education for optometry. This same curriculum was responsible for a 73% improvement in reading speed with improved comprehension on higher grade-level demand reading passages in a study of 44 consecutive cases in my private optometric practice. This study used the OBER2 Eye Movement Recording Device before and after vision therapy to record eye movements and provided objective proof of the effectiveness of the treatment.[4] To what degree could the same levels of change be expected in an inner-city population and what age groups would show similar types of change?

If the problem in this population is purely a vision development problem, then a graph of performance levels vs age would look similar to that shown in Figure 1.

Figure 1 shows the developmental age plotted against the chronological age of two different groups. At the origin, they are assumed to be equal. The normal group progresses 1 year of development for 1 year of life on earth. The abnormal group progresses a 0.5 year for each year of life on earth. The later in life one looks at the two groups, the further apart the two groups are from each other and the easier it is on the front end of a study to know for certain that the two groups are indeed different. For statistical significance to emerge, one needs base numbers on performance testing to be at a level that expected differences, and expected change as a result of an intervention could demonstrate such significance.

One could easily make the case that the problems would be relatively easy to identify and much easier to treat in the very young child. However, to demonstrate significance of a treatment modality it might take a very long time until such significance were to emerge.

Figure 2 simulates what one might expect if treatment was given at age 5, the average age of a kindergartner, and the treatment group is plotted with an accelerated growth rate of 1.2 years' performance change for each year of life until they level out with the normal group at age 15. The untreated abnormal development group continues to make 0.5 year of improvement for each year of life. In this scenario, it can be seen that a very significant change has occurred within the treatment group, which will show its impact well down the line. However, at the end of year 1, there is only a 0.6-year difference between the treatment and nontreatment groups and only a 1.2-year difference between the two groups at the

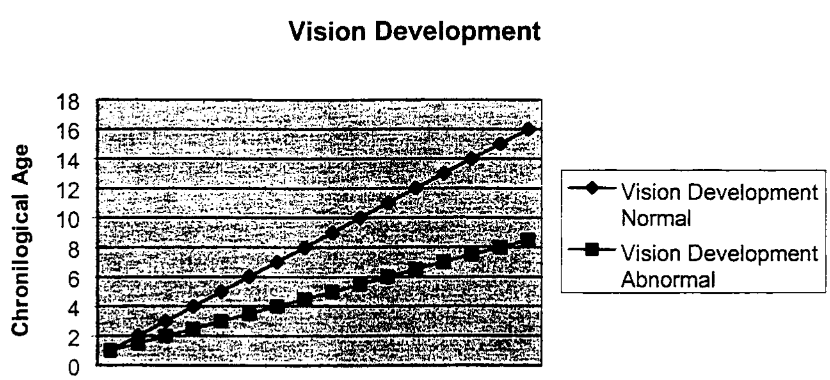

Fig. 1. Normal vs abnormal vision development.

Vision Development – Treatment at Age 5

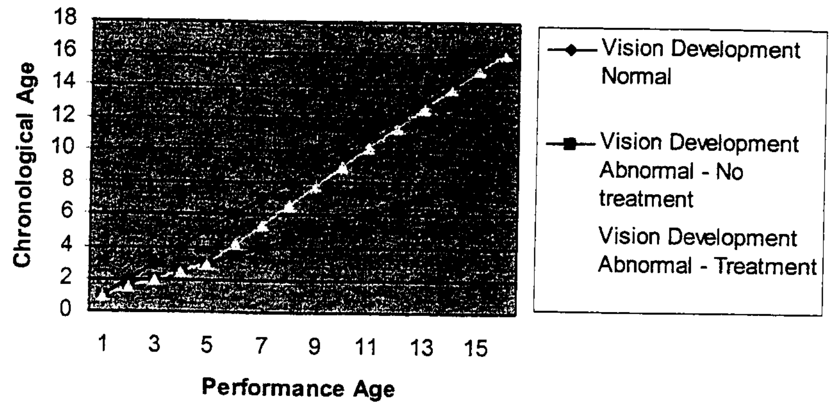

Fig. 2. Vision development: treatment at age 5.

end of 2 years. Unless the groups are observed for a much longer period, the evidence of significance, particularly if the performance numbers on the tests given have very low absolute values, may be difficult to establish.

In Figure 3, the hypothetical treatment was begun at age 9 or in fourth grade. The nontreatment group is again assumed to continue to develop at the rate of 0.5 year per 1 year of life. As of the time of treatment, the treatment group is now assumed to develop at the rate of 1.5 years per year of life. Working with older children whose developmental lags could be more readily identified and to whom specific remediation offered, rather than giving mostly an across the board general treatment as would have to be designed for the 5-year-old group, allows for the assumption that the benefits and the growth rate coming out of the treatment would be greater in the 9-year-old group than in the 5-year-old group. Differences at the end of year 1 would now be 1.0 year of performance difference and would grow to a 2.0-year difference at the end of the second year after the conclusion of the treatment. This larger difference on top of a higher base of performance in absolute terms on the tests should reveal significance over a shorter period of time. In this case, it takes 7 years after treatment for nearly complete normalization, whereas it took 10 years for the same to occur with the treatment begun at age 5. This all assumes a high degree of linearity, which most likely could not be safely assumed. These graphs offer some insight into the thought process that led up to the selection of an older rather than a younger group for identification and treatment. If the study been established with a 5- to 7- to 10-year follow-up from its inception, then younger age groups would have been selected because the intervention would have consisted of a simpler pro-

Vision Development – Treatment at age 9

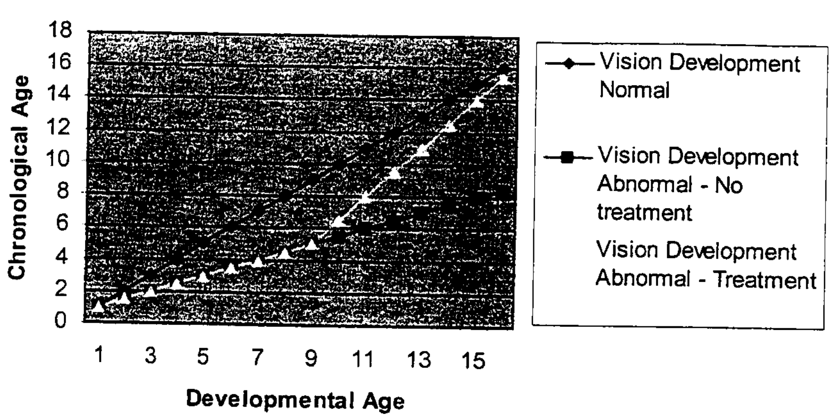

Fig. 3. Vision development: treatment at age 9.

gram that could have served more children for the same investment of time, effort, and energy.

Lastly, it was important to work with scales of performance, which were linear in nature and could allow the use of higher-order statistical analysis. Standardized testing was already being performed regularly on these children. At a young age, many children might achieve a score of zero on a particular subtest or test. This score, although at an absolute floor, may not be linearly related to very low positive scores. This would introduce a degree of nonlinearity into the overall scheme. For example, on the screenings reported above, some children did not know their numbers and, therefore, could not perform on the NYSOA King-Devic Saccadic Test. It is one thing to say that they failed the screening. However, if one wanted to see reports on actual means and standard deviations for each of the groups, one would have to decide how to handle a child who failed because they didn't know their numbers. Would an artificially long time be assigned to that child to show clearly that they had failed? And after they learned their numbers and subsequently they did the test well and faster than the chosen slow time for them, would this demonstrate an actual improvement attributable to any specific treatment other than just learning their numbers? This conclusion could not be made in that instance. If these long times were used in the calculation of group means and group statistics, then the entire endeavor could be called into question.

For all the reasons cited above, fourth-graders were selected for this study. They would be available for two years in the elementary school. On most tests that were selected for them to take all children should be able to perform at or above the minimum scores on those tests so that either (1) no children would have to be eliminated from the study because their performance was too low to register on our measuring scales or (2) we would not have to manufacture specific scores for these low performers, which could affect the overall study.

TESTING BEGINS

Funding for the research phase was approved mid-September of 1998 and by mid-October 1998, the testing phase of all fourth-graders at Harford Heights Elementary School (school #36) commenced. During that month, the testing protocols were established and a statistician was consulted on the design of the entire program. During this phase, three staff were hired as vision therapists for the research and their formal training began. Their training will be reported on later.

The following testing was administered to all 178 children in the fourth grade. Some of these children were involved in special education for a portion of the day. None of the children in the full-day dedicated special education classroom nor in the one classroom dedicated for emotionally and physically handicapped fourth-graders were included in the study. The 178 children were spread over nine separate classrooms. Fortunately, all nine classrooms were on a single floor around a central core, which included a library and a multipurpose room. A standard optometric testing lane with chair, stand, phoropter, projector, lensometer, trial lens set, and all hand instruments was set up in the health suite in another part of the school. All other testing was performed in the centrally located library.

Testing performed in the health suite included visual acuities distance and near; cover test distance and near; motility testing; convergence near point; color vision testing with Ishihara plates; the Special Randot Test; stress point retinoscopy; lensometry, if the child wore glasses; Worth 4-dot testing at distance, near, and at near through +/–2.00 diopter flippers; and a complete analytical examination, which followed the guidelines published as Appendix A in Harris.[5]

Three optometrists were trained to perform the testing in a uniform manner. All had been active members of a Baltimore area study group for many years or had worked with me in my office and were fully familiar with the testing routines to be followed. None of these optometrists were involved in the training, the training of the vision therapists, or the analysis of the data in any way.

Testing performed in the library area was broken down into three groups. One group was the Visagraph testing. Visagraph testing protocol will be explained in detail in the section that reports on the findings on the Visagraph. The second group of tests included the NYSOA

King-Devic Saccadic Test, the Groffman Visual Tracing Test, the Wold Sentence Copy Test, parts I and II of the Jordan Left-Right Test, the Eye-Hand Coordination Sub-Test of the Developmental Test of Visual Perception II, the Gates Oral Reading Survey from the Handbook of Diagnostic Tests, and an oral history including a symptoms checklist. This section was called the visual performance testing area.

Dr. Ronald Berger, an optometrist from Columbia, Maryland, who was also a member of the Baltimore area study group, put together the third area of testing performed in the library area. This drew heavily on the work of Swiss psychologist Jean Piaget and the work of Piagetian scholar Harry Wachs, a Maryland optometrist currently practicing in Washington, DC. Dr. Berger included tests of visual recall, visual transposition, auditory matching, auditory manipulation, visual-auditory integration, and learning motivation. These tests will be described in a separate article; they are mentioned here only for completeness.

The children were given each part of the test on a different day so they would not become fatigued. No one section lasted more than 30 minutes; the Visagraph station was the shortest test, and the visual performance testing, including the oral symptoms checklist, took the longest time to complete. The testers in the library area were different staff than would be involved in the vision therapy.

The testing phase lasted approximately 6 weeks and was completed before Thanksgiving of 1998. A meeting was held with the statistician to decide the best way to separate the 178 children into treatment and control groups. In addition, test scores for the children on their national standardized testing were provided to us for the fall of 1998. These were the California Diagnostic Math Test and the California Diagnostic Reading Test by McGraw Hill.

TESTING RESULTS

The following are the results obtained on the 178 children as a whole. These data are reported here to give an idea of the total performance level of the entire fourth grade before the children were broken down into their respective groups.

Symptoms Checklist

Complete data were obtained on 168 children on the symptoms checklist (Fig. 4; Appendix A). The checklist was filled out by one of the testers; each question was asked of the child verbally. Only one child did not respond to any of the items of the checklist with a "Yes" response. The average response was "Yes" to 6.53 items (SD = 3.53). Table 4 shows the top five complaints and the percentage of children responding to each.

The following findings are selected findings from the initial optometric analytical and chair tests (Table 5).

Of the 171 subjects on which this finding was recorded, four subjects were found to have constant esotropia, and one subject was found to have an intermittent exotropia, which was more than 10 prism diopters. Four had an esophoria estimated to be more than 10 prism

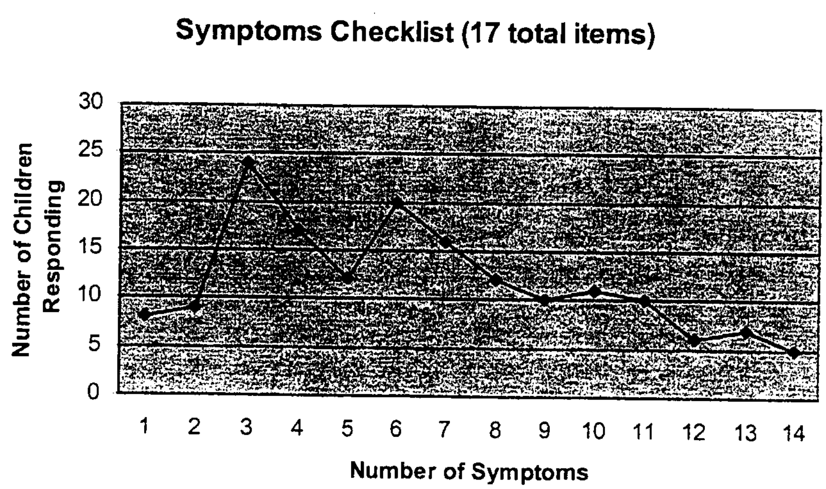

Fig. 4. Symptoms checklist.

TABLE 4. Top Five Complaints on the Symptoms Checklist

Headaches	84.3%
Uses finger to keep place with reading	78.1%
Eyes tired with reading/school work	70.8%
Loses place when reading	70.2%
Rubs eyes during the day	52.8%

TABLE 6. Near Cover Test Results

Esotropia	4
Exotropia	1
Low esophoria	4
Low exophoria	42
Moderate exophoria (>10 pd)	6
Orthophoria	106

diopters, and 13 had exophoria estimated to be more than 10 prism diopters. Phorias were not neutralized with prisms but were estimated by the examiner on the basis of the amount of movement noted during the alternate cover test.

At near, the number of strabismics remained the same (Table 6). Those with esophorias at distance that were estimated to be more than 10 prism diopters still showed esophoria at near, but the amount was now estimated, in all four cases, to be lower than 10 prism diopters. At distance 13, children showed exophorias greater than 10 prism diopters, but at near this number was reduced to 6. Forty-two children showed a low exophoria on near cover testing.

Convergence Near Point

Convergence Near Point (CNP) or Near Point of Convergence (NPC), as some call this test, was performed by using a Wolff Wand that was moved slowly from the patient's Harmon distance toward the bridge of his or her nose. The patient was asked to report the first instance when he or she noticed that the Wolff Wand no longer appeared to be single. The examiner also watched for any sign of either of the two eyes losing fixation on the Wolff Wand. The break point was recorded as the distance from the subject where either of these events occurred first. The average distance of the break point was found to be 1.82 inches (SD = 1.4 inches). Of the 171 students on whom data on this test were recorded, only 4 students had CNPs of greater than 4 inches. An additional 7 subjects had CNPs between 3 and 4 inches, and 24 subjects had CNPs between 2 and 3 inches. Depending on the criteria used for the diagnosis of convergence insufficiency, different levels of this condition would be diagnosed. Using a more conservative measure of 4 inches or greater would yield a very low percentage of only 2%. Using 3 inches or more would yield 11 subjects or 6% with convergence insufficiency, and the strictest definition of 2 inches or more yields 35 subjects, or 20% of the population. Convergence insufficiency has a reported prevalence of from 1% to 25% in clinic population.[6–9] The median prevalence of convergence insufficiency in the general population is 7%, and it is similar for adults and children.[16]

TABLE 5. Distance Cover Test Results

Esotropia	4
Intermittent exotropia (>10 pd)	1
Significant esophoria (>10 pd)	4
Significant exophoria (>10 pd)	13
Orthophoria, low exophoria, low esophoria	149

Ocular Motilities

Ocular motility testing was performed with a Wolff Wand. Separate pursuit and saccadic testing were not performed. Rather than look at each of these components of ocular motility, the ability to keep the eyes on a moving target was tested as a whole. The Wolff Wand was moved in a pattern, which begins with a horizontal sweep, followed by a vertical sweep, and then several diagonal sweeps. The movements then become circular and include z-axis movements and variable speeds, and then start and stop as they cross the mid-line. Three things are rated, each on a scale of 1–4: the amount of head movement, the amount of supportive body movement, and the quality of the tracking (rated in terms of the percentage of time that the person is fixating the target directly).

The head movement scale is as follows:

1. No movement
2. Small
3. Moderate
4. All head movement

Of the 171 children for whom data are available on this test, the average score for the amount of head movement was 1.59 (SD =

0.82). Nine children moved their heads exclusively to track the Wolff Wand as it moved through space. When a target is tracked this way, the person is essentially keeping the eyes steady in the head and using the supportive muscles of the neck and upper back to perform the tracking movement. Ten children were rated as a 3, which meant that a moderate amount of head movement was noted during the tracking of the target. This yields an incidence of 11% of the children (19/171) showing significant evidence of an ocular motor dysfunction based solely on the degree of head movement during a very simple tracking activity.

The amount of movement in the torso was very small. A rating for the entire group of 1.00 would indicate that all subjects were rock-steady with their upper bodies. The actual rating for all 171 subjects was 1.11 (SD = 0.37). No children were given a rating of 4, which would have indicated that the tracking was being performed exclusively by using movements of the upper torso to track the moving target. Three children were given a rating of 3, which means that a significant amount of torso movement was noted during the testing. An additional 13 children were rated as having some movement in the upper body.

The accuracy of ocular motility was rated according to the following scale:

1. >95% accurate
2. <95% >75% of the time on target
3. <75% >50% of the time on target
4. <50% of the time on target

The total tracking score was rated for the 171 subjects at 1.71 (SD = 0.93). Eleven of the subjects were rated as a 4, and an additional 19 were rated as a 3 on the above scale. Thus, 30 of the 171 subjects (18%) had accuracy of eye-tracking skills well below expected and well below that necessary to perform well in the educational environment.

Random Dot Stereo Acuity

The Special Random Dot Stereo Acuity test was administered. This test has both the object and the background done entirely with computer-generated random dot patterns. The test was given at a working distance of 16 inches with the patient's habitual lenses on or no lenses if the patient did not wear glasses or did not have their glasses with them.

Forty-three of the 177 children (24%) for whom data were collected on this test had stereo acuity that measured worse than 40 seconds of arc (Table 7). This number is much higher than has been seen in other testing of random samples of school-age children. Williams, in a study of children aged 7–11 found that stereo acuity levels improved as the children got older. In the study, which used the TNO stereo test, they found that 52.7% of the 7 year olds got 60 seconds of arc or better. By the age of 9, 83.52% were now at this level and by age 11, 84.8% had reached this level.

Ishihara color plates were used to screen for any color deficiencies. Of the 170 children for whom data were collected, only 3 children demonstrated any color vision problems.

Refractive Findings

The following information shows the average refractive findings in the total group along with standard deviation and minimum and maximum refractive powers on the basic distance subjective. The finding as reported here is a binocular finding performed after cylinder testing and binocular balance by using a dissociation technique with a low amount of plus fog used for balancing. After balance is achieved, the examiner decreases plus binocularly or increases minus binocularly. The end point reported for the subjective refraction (OEP #7) is most plus or least minus to 20/20 visual acuity (Table 8).

The average values for the right and left eye refractive findings at just over one-half a

TABLE 7. Randot Stereo Scores

Score (seconds of arc)	Number of Subjects
20	112
30	14
40	8
50	3
70	5
100	6
200	3
400	7
600	6
None seen*	13

* The child was not able to appreciate depth on even the grossest targets, which had 600 seconds of arc of stereo acuity.

TABLE 8. Values and Variation of the Subjective (OEP #7) Finding

Subjective Sphere	Value	Standard Deviation	Most Plus	Most Minus
OD Sphere	+0.53	±1.36	+7.75	−7.75
OS Sphere	+0.58	±1.36	+8.50	−8.50

diopter of plus is what would be expected in a random sample of the general population of young children who had yet to encounter lots of sustained near-point demands. The ranges for each eye are similar, with the left eye having a slightly wider overall range (17 diopters vs 15.50 diopters), and the standard deviations for each eye are the same at 1.36 diopters.

Prism Testing

An integral part of the analytical examination is the use of prisms at both distance and near. The patient is asked to look at a target, usually an eye chart, and prisms are symmetrically increased in front of both eyes simultaneously. The testing performed by all optometrists in this study followed the same protocol. Base-out prisms were performed first followed by base-in prisms. The prisms are increased until the patient reports seeing diplopia and then decreased slowly until they report seeing the images come back together. Table 9 gives the average responses to these tests on all subjects from whom data were obtainable. Some patients did not report seeing diplopia or did not report when they saw the images reconverge, even on a repeat of the test probe. This is why the number of subjects is not the same for all findings.

These findings are noteworthy in that all averages on the break points are higher than the OEP expecteds, and all the recoveries are

TABLE 9. Prism Duction Findings (Equilibrium)

Finding	Data	OEP Expecteds[19]	# of Subjects
DV Base-out break	25.35	19	151
DV Base-out recovery	6.73	9	147
DV Base-in break	15.87	9	152
DV Base-in recovery	2.73	5	151
NV Base-out break	28.74	21	147
NV Base-out recovery	8.68	15	144
NV Base-in break	23.04	22	150
NV Base-in recovery	7.75	18	149

lower than the OEP expecteds. When large groups of findings are systematically higher on one side and lower on the other side, one possible explanation is that some of the members of the group are not good observers. By this is meant that there is a longer than normal delay between the observance of an event and the reporting of an event. This could be due to systematic slowness of verbal response or overly fast moving of the prisms by the testers. It is also possible that the just noticeable difference in terms of detecting the diplopia or the refusion is wider than normal in these subjects. For example, the images may have doubled much earlier, but they were not yet wide enough apart to trigger the person's report of diplopia. The two images had to be quite far apart before the person would report seeing diplopia. In this instance then the high number would be a false indicator for that finding and could not be construed to mean that the aspect of vision being measured by this test was indeed "normal." No attempt was made during the testing to tease out the exact reasons for these systematic differences in these findings vs the expected findings. The only way to do this would be to perform eye movement recordings at the same time as the base-in and base-out testing is being administered and to analyze these recordings along with other brain wave recordings to assess when cognitive awareness of the events in question took place. This may yield interesting information but is beyond the scope of this article.

Additional Test Results

One more set of findings from the analytical examination are the findings classically called Positive Relative Accommodation (PRA/OEP# 20) and Negative Relative Accommodation (NRA/OEP # 21). Table 10 summarizes this data and shows values here to be within the expected range again.

TABLE 10. Positive/Negative Relative Accommodation Results

Finding	Value in Diopters	Standard Deviation	OEP Expected
PRA #20	−2.57	±1.47	−2.25 to −2.50
NRA #21	+2.65	±0.67	+1.75 to +2.00

The data in Table 11 is from some selected performance tests.

The NYSOA King-Devic Saccadic Test[10–12] consists of three paragraphs of widely spaced numbers that must be called off as quickly as possible. Performance on this test relates directly to the eye movements associated with reading. The total time to complete all three paragraphs is reported in Table 11. The average for the entire group was 82.91 seconds (SD = 19.51 seconds). The norm for the average fourth grader is 73.4 seconds. Twenty-five children (14.0%) took longer to complete the test than the norm for the average 7 year old, which is 100.89 seconds.

The Groffman Visual Tracing Test[13] consists of five intertwined lines that must be followed from beginning to end. Two factors are measured—speed and accuracy. No points are awarded if the child does not get to the correct end point. If they do get to the correct end point, they are awarded more points for going faster and fewer points for going slower. The total points for all five lines are added together, and this score is reported in Table 11. The maximum number of points that can be received on each trial is 10, with 50 points being the maximum for the entire test. This test relates to the eye-monitoring skills necessary in those tasks where eye-hand coordination is essential, such as handwriting, and also involves sustained visual attention. The full group averaged 13.3 points (SD = 10.1 points). The norm for the average fourth grader is 22 points. On the Groffman Visual Tracing test, 70 subjects (39.3%) scored below the lowest norm on the test, which is 10 points for the average 7 year old.

The Eye-Hand Coordination Sub-Test of the Developmental Test of Visual Perception II requires the child to draw a line with a pencil between progressively narrower corridors, some straight and wide, others narrow, some with turns, and others with curves. The more accurate the lines are drawn, the more points are awarded on each line segment of each trial. Each line segment is graded, and the total points for all parts of the test are added up to derive a raw score. The raw score is then converted to an age-equivalent score by using the tables in the back of the testing booklet. The entire group averaged 8.41 years (SD = 1.82 years). Fifty-seven children (32.0%) performed below the age of 7.5 years. The range of performance on this test was from 4.2 to 11.4 years of age.

The Wold Sentence Copy Test[14] consists of a sentence that must be copied as quickly as possible. The test was originally created as a vehicle to afford the tester an opportunity to observe postural and mechanical problems with copying tasks. Later, Wold applied normative data from other copying tasks to quantify how fast or slow the individual performed the copying task. The test group completed the test in an average time of 127.24 seconds (SD = 32.35 seconds). The norm for the average fourth-grader is 144 seconds. Twenty students took longer than the 166 seconds, which is the norm for the average second-grader.

The Oral Reading Diagnostic Test is from the Handbook of Diagnostic Tests for the Developmental Optometrist and is interpolated from the Gates-Mckillop Reading Diagnostic Test.[15] This oral reading test consists of seven paragraphs, which when taken together are a single story about a boy, a dog, and a rat. Each paragraph is written at a progressively higher

TABLE 11. Additional Performance Test Results

Test Name	Test Emphasis	Average Score	Fourth-Grade Norm
NYSOA-King Devic Saccadic Test	Eye movements for reading	82.91 ± 19.51 sec	73.4 sec
Groffman Visual Tracing	Sustained visual attention and tracking for writing	13.3 ± 10.1 points	22 points
EH Subtest of DTVP-II	Eye hand coordination	8.41 ± 1.82 yr	9 yr
Wold Sentence Copy Test	Eye hand coordination and tracking	127.24 ± 32.35 sec	144 sec
Gates Oral Reading Test	Oral reading without time pressure and without comprehension testing	Grade 4.62 ± 1.06 years	Grade 4

level of demand in terms of the decoding skills and/or the size of the sight word vocabulary required to read the passage successfully. The examiner listens carefully and marks on a recording sheet any errors made by the subject. For each paragraph, a maximum of six points can be awarded. Based on the number of errors made for each paragraph, a chart is consulted to determine how much of the maximum of six points per paragraph is awarded to the subject. The total scores for all seven paragraphs are totaled together to get the raw score for the test. (As soon as the subjects scores only one point or below on any given paragraph, the test is terminated. Thus, a child may only read two or three paragraphs and a raw score of zero is assumed for all of the more difficult paragraphs. The total raw score then is converted by using yet another table to a grade level or to an age score.)

This test does not investigate comprehension as part of the test and, therefore, is thought of as a decoding test only. Typically, a child can decode words at a reading level higher than their comprehension level. The average child in the fourth grade at Harford Heights Elementary School was able to decode at the grade 4.62 level (SD = 1.06 grades). This is significant because it shows that these children are essentially at grade level in their ability to decode text. Only seven children had decoding abilities below grade 3.0.

Visagraph

The 178 fourth-grade students were tested with the Visagraph, an infrared eye-movement recording device. A pair of goggles is placed on the child. The goggles are large enough to fit over any eyeglasses that a person might be wearing. The goggles are connected to an interface box, which in turn is connected to the serial port of a computer. In this case, a Dell Inspiron 3500 running Windows 98 was used as the recording machine. Version 4 of the Visagraph program was used to collect and analyze the data.

The children were given a story to read silently. Testing is performed with silent reading only. If a child starts to read the passage out loud, the testing on that sample was discontinued and the directions to read silently were repeated and a new reading passage was chosen. There are 7–10 different passages at each of the different reading levels available for use during the testing. The child reads from a printed test booklet. Some early versions of the eye-movement recording devices allowed for the reading passage to be read directly off a computer screen. All testing with the Visagraph for this study was performed with the child reading from the printed texts in the booklet supplied with the device. Reading from a video display terminal (VDT) screen induces many additional variables into the recordings, which other researchers have documented. Because most of the reading done by the children in this school and in most inner-city public schools across the country is not performed on a VDT, the standard printed reading samples supplied with the device by Taylor Associates, Inc. (Huntington, NY) were used.

All first sample recordings were used to acclimate the child to the testing situation and were not used in the data reported here. After reading the passage, the recording device was turned off and the child was asked 10 "Yes/No" questions to determine the level of comprehension. These questions are supplied in the test booklet. The same questions also come up on the computer screen after the recording has been completed. The questions are asked verbally of the patient and the examiner enters the subject's responses into the computer.

The mechanics of reading vary significantly as the demand level of passage reading changes. When the reading passage is more than 1 year below the person's instructional level, there are characteristic changes in the scan pattern used to move the eyes over the text. When the demand of the reading passage is more than 1 year above the person's instructional reading level and the child begins to get frustrated, there are often significant differences in the visual scan patterns. These variations can be very significant and could affect the data reported if the demand level is not controlled accurately in this type of testing.

Therefore, it is critical to have the reading passage be at the appropriate demand level for each subject. This is achieved by using comprehension level as the determining factor. The child was required to get at least 70% correct on the passage being read. If a passage was read and the child got less than 70% correct on the reading passage, then a new passage was read with the demand level lowered by one grade. The process was repeated until a demand level was found where the child got at least 70% correct on the reading passage. Fortunately, with this sample of fourth-graders, the lowest-level readers were able to read with a minimum of 70% comprehension at the first-grade level. There were no children who could not read at the first-grade level. Stated positively, all children in the fourth grade at Harford Heights Elementary School were able to read with understanding at least at the first-grade level.

On the higher performance side, when a child got either 90% or 100% correct on the reading passage, they were given a chance to try a harder reading passage. As long as they kept getting 90% or 100% correct on each passage, the child was given a new reading passage at a higher demand level. This was repeated until the highest level of demand was reached on which the child got a minimum of 70% comprehension. In this sample of fourth-grade children, only seven children were able to reach the fifth-grade level stories and still get the minimum of 70% comprehension. No children were able to reach the 70% minimum on a sixth-grade or higher level reading passage.

(First- through third-grade level demand stories are nine lines long. The middle seven lines on these passages have a total 50 words on them, and the print is of the size normally seen in most early reading books [approximately Times Roman 16]. Fourth-grade cards and higher have 12 lines of text with a total of 100 words on the middle 10 lines. The print on these cards is smaller than for the first- through third-grade cards and is similar in size to the fonts used in most elementary and middle-school text books [approximately Times Roman 12].)

On the computerized reading testing, the average reading level based on comprehension scores was grade 3.0 (SD = 0.96 years). Table 12 shows the distribution of children and the maximum grade level with at least 70% comprehension.

Analysis of the Visagraph data also yields significant data relative to the mechanics of reading. By mechanics I am referring to the scan patterns, which are determined by the number of fixations, regressions, and return sweeps, as well as the average amount of time the person stops at each fixation during the reading act. A fixation is defined as occurring whenever the eyes remain at a fixed location on the text or the eyes are not in motion. A forward fixation is one that occurs as the person moves their eyes from left to right (assuming they are reading a left-to-right-read language such as English) within a single line of text. A regression or backward fixation occurs after the person has already had one or more fixation(s) on the same line of text at a position further to the right than where the eyes have just landed. Most of the time a person is not consciously aware of having made a regression. At times a reader may be aware of not understanding something and consciously going back to reread a section, a sentence, a passage, or a paragraph. This is entirely different from most regressions, which occur below conscious awareness in most readers.

TABLE 12. Highest Grade Level Demand on Visagraph

Highest Level Attained with at Least 70% Comprehension	Number of Children	Percentage of Total
1	15	8.47
2	33	18.64
3	79	44.63
4	43	24.29
5	7	3.95

Return sweeps are made at the end of each line of text. In theory, there should be a single right-to-left movement from the location of the last fixation on the line to the starting point for reading on the next line. Often the return sweep in a beginning reader or a poor reader consists of a complex of eye movements in which most often there is a major movement from right to left, followed by a small correcting saccadic eye movement to get the eyes to the actual starting place for reading. Return sweep complexes are not normally analyzed separately. However, both the number of fixations and the number of regressions would tend to be higher in readers whose return sweep complexes have these small corrections on a routine basis.

Table 13 summarizes the mechanics of the full group of 178 fourth-graders tested.

The full group averaged 193.6 fixations per 100 words, which is equivalent to a reader who

TABLE 13. Visagraph Mechanics (178 subjects)

Mechanical Ability	Average Score	Grade Equivalent	Fourth-Grade Norm
Fixations/100 words	193.6	1.5	139
Regressions/100 words	43.5	1.8	31
Average Duration of Fixation (seconds)	0.33	1.0	0.27
Reading Speed (words per minute)	105.8	1.5	158

is halfway through the first grade. These are many more fixations than are expected from the average fourth-grader, which is 139. This demonstrates that the children are stopping much more often with their eyes than the average fourth-grader to read the passages. This can be thought of as adding to the demand of the reading task for comprehension as more separate pieces need to be assembled together from which the meaning is to be extracted.

The full group averaged 43.5 regressions per 100 words read, which is equivalent to a reader at grade 1.8 or nearing the end of first grade. Return sweeps are not counted in the number of regressions. This number includes only those backward fixations that occur within each line of text being read. The average fourth-grade reader is expected to have only 31 regressions per 100 words of text. The ratio of 43.5 regressions to 193.6 fixations shows that in this sample 22.4% of all fixations are regressions, which is exactly at the ratio of the regressions to fixations expected for the fourth-grader; 31 regressions to 139 fixations is 22.3%.

The average duration of fixation for the full group is 0.33 second, which is equal to that expected of a beginning first-grader. The norm expected for the average fourth-grader is 0.27 second. Although the difference of 0.06 second seems small, it is helpful to keep in mind that a college-level reader has an average duration of fixation of 0.24 second, which is only 0.03 second faster than the average fourth-grader. Thus, the 0.06-second slowness in the average duration of fixation should be interpreted as significant.

Many things must occur during a fixation. The person must decode the word they are looking at. This acts as a key, which is used through associative memory to access the meaning of that word. The meaning or multiple meanings, depending on context, are integrated with what has come before in the reading and, typically, guesses based on knowledge of the syntax of the language being read and on contextual clues are made as to what is coming up. In addition, and often as a parallel process in good readers, a process is functioning to preprocess to the right of fixation the spatial layout of the next area to be fixated so that the exact correct size saccadic eye movement can be programmed to be executed when needed. Good readers generally fixate one-third of the way into the next word. To do this accurately, because words are of varying length, there must be a preprocessing of the areas to the right of fixation to figure out where one-third the way into the next word is. Once the preprocessing is done, the target area for landing with the next fixation can be plotted. All of these separate tasks may require very different amounts of time and can vary significantly during the reading. There is a huge variability in each fixation in terms of its length. The Visagraph data table only reports the average duration of fixation and does not give any sense of the degree of variability from one subject to another or from one passage to another within a single subject. However, viewing the actual raw recording data shows that some fixations may be as short as 175 milliseconds, whereas others may be 750–1,000 milliseconds long. Often one sees these very long, single fixations occur at a time when the person stops taking in data from the printed text and goes internal to think about the data or to access other stored information they may have about the data being collected. A reinterpretation of the information already taken in may have to occur because something unexpected was encountered. When this occurs, it is as if some readers go off-line for a second or more before resuming parsing of the text. Thus, in some subjects the average duration of fixation could be longer because of either small increments of each fixation in time or the inclusion of several very long fixations.

The Visagraph currently does not offer an automatic way to gain insights into this data.

Combining all factors on the children tested yields an average reading speed of 105.8 words per minute for the entire group, which is equivalent to the average child who is halfway through the first grade. The average reading speed for fourth-graders is 158 words per minute. Thus, the group of fourth-graders at Harford Heights Elementary School is reading 49.3% slower than average fourth-graders across the country.

Table 14 compares the actual reading speed for the group, broken down by reading level, against norms published by Taylor *et al.*[16] In general, the higher the level material that can be read and understood, the faster the reading speed. Taylor's norms show that the average first-grader reads at about 80 words per minute and that by fifth grade the average reader is reading at 173 words per minute. I wanted to determine whether this type of relationship held in this sample.

Table 14 shows some very interesting information in that as the level of reading ability improved, there were no commensurate increases in the average reading speed, as was expected. The 15 fourth-grade children who could only handle first-grade material as the highest level on which they could attain at least 70% of the comprehension test averaged 100.5 words per minute. The seven fourth-grade children who were able to handle fifth-grade material with at least 70% comprehension but who could not reach this same level of comprehension on sixth-grade material averaged just 106.0 words per minute. The average fifth-grade-level reader in Taylor's study read at 173 words per minute.

Speculating on why this might be the case leads to the following hypothesis, which is based on some observation. It appears that many teachers at this school may have concluded that the primary channel through which these children learn is not vision but audition. If these teachers are presenting material that they believe is important for the child to learn, they present the information verbally, usually in a group discussion format as opposed to the child being given a passage of text to read about the topic, followed by a group discussion. The end result is that there is actually a decrease in the amount of total time spent on reading for gathering information. A separate study would need to be conducted to address this concern and would involve charting the amount of time inner-city children spend at different tasks vs how suburban children spend their time during their academic day. My hypothesis is that significant variations exist that may explain some of the differences seen in these populations and may lead to recommended changes in the structure of the school day and in the manner of instruction.

Along the same lines, I wanted to look at the average duration of fixation in each of these groups to see how this finding varied or did not vary with reading demand level differences. These results are shown in Table 15.

As in the reading speed breakdown, the average duration of fixation does not follow the expected shortening with increased reading abilities in this population. The average duration of fixation is exactly the same for the 15 children whose highest level of reading on which they could get at least 70% comprehension was 0.33 second as it was for the seven children who could read and understand the fifth-grade-level stories.

One issue this raises is the impact that this may be having on the overall education of these children. It has been said that until

TABLE 14. Reading Speed by Comprehension Level

Highest Level Attained with at Least 70% Comprehension	Number of Children	Taylor Reading Speed Norms	Reading Speed (words/min)
1	15	80	100.5 ± 49.5
2	33	115	111.9 ± 49.5
3	79	138	104.4 ± 39.6
4	43	158	105.9 ± 35.2
5	7	173	106.0 ± 21.0

TABLE 15. Average Duration of Fixation by Comprehension Level

Highest Level Attained with at Least 70% Comprehension	Number of Children	Taylor Average Duration of Fixation Norms	Average Duration of Fixation
1	15	0.33	0.33
2	33	0.30	0.31
3	79	0.28	0.34
4	43	0.27	0.34
5	7	0.27	0.33

fourth grade one learns to read and that from fourth grade on one uses reading to learn. If we assume that a significant portion of the child's school day is involved in using their own reading ability to learn from the printed page, then the children at Harford Heights Elementary School—and these children look representative of many other inner-city populations—are taking 50% longer on average to complete tasks than children with average skills in suburban areas. To compensate for this, the school day would have to be longer for the inner-city population simply to make up for the difference in the reading speeds between them and suburban schools. If, for example, the school day was 6 hours long and 2 hours of that day involved children reading for information, then it would take the inner-city population 3 hours to complete what the suburban children were doing in 2 hours. Thus, to have the same intensity of educational experience would require that the inner-city schools not only run a 7-hour day, but also be certain that the entire extra time was devoted exclusively to reading for information tasks.

Note also that this does not at all compensate for the fact that the average reader in this school is 1 year behind in the demand level at which they can read and get information with good comprehension. This means that if the inner-city schools use fourth-grade texts that are written at the level that requires fourth-grade reading skills, then the textbooks are too difficult for the children to learn from. (I am unaware whether the issue of differing textbooks for inner-city vs suburban schools has been addressed.) Stated more plainly, in the subject of social studies, for example, is the fourth-grade child in the inner-city school given a social studies textbook that covers the fourth-grade curriculum on a fourth-grade intellectual basis but written for a third-grade-level reader? And is the child given 50% more time to read this special type of textbook than the average fourth-grader in a suburban school reading the material written at the fourth-grade level? To my knowledge, this is not the case and this aspect of matching the needs of the population being served in both levels of demand and time required to complete the task has not been addressed. If I am correct, the inner-city child is given textbooks that match the curriculum needs, not the reading abilities of the child. Because of budgetary constraints, many of the textbooks used in the inner-city are out of date. Over time, these discrepancies will lead to a wider and wider separation between the average inner-city child and the average child in suburbia, separate from all other influences such as socioeconomic, family structure, or cultural, which have been studied.

In the fall of 1998, the entire fourth grade was administered both the reading and math portions of the California Diagnostic Test. Table 16 gives the scores for these two tests on the entire fourth-grade sample and the scores on the Visagraph grade level and the Gates Oral Reading Test.

HOW MANY GROUPS?

Another question to be answered was whether to have a two-group design (one control and one treatment) or a three-group design (one control, one treatment, and one placebo treatment). There were many factors to consider. In many of the early meetings with schools, school administrators, and people within the Abell Foundation who had experience working in the schools, I learned that one of the most difficult things we would have to overcome would be simply getting access to the children. We would not be allowed to take children at any time that conflicted with their

TABLE 16. Standardized Tests Scores Compared with Visagraph and Gates

All Scores Reported as Grade Level	California Diagnostic Test Reading	California Diagnostic Test Math	Visagraph-Highest Grade Level with Comprehension	Gates Oral Reading Test
Score	3.23	3.44	2.95	4.62
SD	1.42	1.09	0.98	1.06
Variance	2.02	1.19	0.95	1.13
Minimum	0.00	0.00	1.0	2.2
Maximum	10.7	7.7	5.0	7.6

"special" classes: physical education, library time, music, art, etc. We would not be allowed to take children during any time when tests were being administered. Initially, it was stated that the morning time was a critical time, when reading instruction and actual reading assignments were assigned and that no pullouts, even by special education teachers, were to be allowed during this time. This later turned out not to be the case, or we were able to get a special dispensation allowing us to work with children during this morning reading time. There were also space limitations for our program. In fact, the vision therapy area was moved after the therapy part of the program had been going on after just 2 months. If a three-group protocol would have been selected, a second area for therapy as well as additional staff and a placebo form of therapy would have had to be implemented.

In addition, during the year before starting the research, Sylvan Learning Centers had been subcontracted to provide certain tutoring services directly to students at this elementary school. Nearly 25% of the children in the fourth grade were already being taken out of the regular classroom for extra help through the Sylvan Learning Center. Information on which students were receiving this tutoring service was recorded in the children's records with the idea that on *post hoc* evaluations the Sylvan Learning Center work might be factored in as a potential placebo treatment group. By this I mean that certain subgroups of children were being brought out and given special attention on a regular basis. The intensity of the tutoring did not match the level of attention that the students were given in their vision therapy program because the vision therapy was done on a one-on-one basis, whereas the Sylvan tutoring was done closer to a one-on-four staff-to-student ratio.

For all these reasons it was decided that two groups—one treatment group and one control group—would be randomly selected from the entire pool of fourth-graders at Harford Heights Elementary School.

RANDOMIZATION

After consulting with Daniel Agley, who served as statistician for this program, nine performance variables were chosen for analysis of variance. These included the scores on the Wold Sentence Copy Test, the Groffman Visual Tracing Test, the NYSOA King-Devic Saccadic Test, the Eye-Hand Coordination subtest of the Developmental Test of Visual Perception II, the Gates Oral Reading Test, the number of symptoms checked off on the symptoms checklist, the Berger Perception Tests, the Visagraph highest grade level with the minimum of 70% comprehension, and the reading subsection of the California Diagnostic Test.

A computer program was written to go through the entire database and to randomly assign 75 students to the treatment group; everyone else was assigned to the control group. The program was run six times and the values of "C" for control or "T" for treatment were stored for each randomization. Variances were then calculated for each of the nine interest areas for each of the randomization runs. The calculated F statistic was determined for each pair of independent samples by dividing the larger variance by the smaller variance. The null hypothesis is that the means of each group are the same at the beginning of the program. For most of the areas tested the final size of each group was 69 in the control group and 47 in the treatment group. For other areas tested the final "N" values for each independent group did not vary more than ±3. This yields degrees of freedom of 68 by 46. For example, the comparison of the variances of the Visagraph Grade scores between the control and treatment groups yielded a calculated F statistic at 1.0256 at 68 and 46 df. This finding demonstrates the independent sample groups are similar in variance, because they did not exceed either value at the critical ends of the two-tailed analysis (upper critical value F = 1.79; lower critical value F = 0.558). Table 17

TABLE 17. Control vs Treatment Group Scores

Interest Area	Calculated F Statistic
Visagraph Grade	1.0256
CDT Reading	1.0101
Wold Sentence Copy Test	1.3271
Groffman Visual Tracing Test	1.2167
Symptoms Checklist	1.2361
DTVP II Eye-Hand Subtest	1.0676
Berger Perception	1.2567
Gates Oral Reading	1.0026
NYSOA King-Devic Saccadic Test	1.0270

gives the calculated F statistic for each of the interest areas of the independent samples.

TRAINING THERAPY STAFF

At the time the grant was approved in September of 1998, in parallel with the testing phase reported on above, three staff were hired and trained as vision therapy staff. They were put through an 8 week program to train them to perform the vision therapy. All three had teaching experience, and one is the parent of a patient who had undergone vision therapy in my practice.

The chief vision therapist from my practice, Dennis Hoover, a Certified Optometric Vision Therapist, devoted most of his time to training the three therapists. Their training included some didactic work given by myself as well as hands-on work in each of the activities with each other under the supervision of Mr. Hoover. At about 3-week point, each of the therapists began working with patients in my private practice under the direct supervision of my vision therapy staff.

THERAPY PROTOCOL

Over my 20 years of private practice and as a result of founding the Baltimore Academy for Behavioral Optometry (BABO), I had developed a curriculum of vision therapy for children with learning-related visual problems. This curriculum is based very much on the concepts of Robert Kraskin, as espoused in his series "VT in Action"[17] in which he recommends a broad-based treatment program that addresses providing the patient with the opportunity to have the necessary meaningful experiences to develop the skills and abilities necessary to meet the demands of the things that they want to do in life. This philosophy does not assign specific therapy procedures or activities based on specific findings but rather provides a well-rounded curriculum to improve the person's vision. As a secondary effect one sees changes in the patient's findings.

The BABO Curriculum of Vision Therapy for Learning Related Vision Problems was adopted as the treatment protocol for this program. Because some of the more specialized equipment in the full BABO protocol was not available, the grid of activities was modified slightly to accommodate this. (See Appendix B for the grid of activities established for this treatment program. The grid shows only the names of the activities. Each activity procedure has been written up in detail and will be made available to anyone wishing to replicate the research protocol.)

This program is essentially the same program that in my private practice demonstrated a 73% improvement in reading speed with a 12% improvement in comprehension in a study done on 44 consecutive subjects as tested by the Visagraph.[1]

Homework?

An integral part of vision therapy as conducted in my practice as well as in most private practices around the country is to have the patient do homework on a regular basis. The protocol followed by most patients in my practice has them coming to the office for a weekly 50-minute visit. They are asked to perform 15–20 minutes of home practice on some of the activities they worked on in the office that week. They do not do homework on the day of their scheduled office appointment, and they are given 1 day off during the week. Thus, I am looking for 5 days of practice between office visits. The typical office-centered vision therapy session has the patient working on five different activities for 8 minutes each. This leaves 10 minutes at the end of the session for the parent to be taught how to do the homework with their child. A good bit of time is also spent during the session having the child demonstrate how they did an activity at home to be sure that it was done correctly before adding on the next level of an activity or before moving on to a new activity.

Avi Karni, an Israeli neurologist, makes the case very strongly for the need for regular practice sessions over an extended period of time to facilitate permanent changes in brain secondary to the learning and development that is taking place.[18] Karni states:

> Our observations suggest two stages in the acquisition of improved perception. A fast improvement, occurring early in training, can be induced by a limited number of trials, on a time scale of a few minutes or less. . . and then saturates, with performance remaining stable within the session and for up to 8 hours afterward. After this latent period, large and long-

> lasting improvements in performance were found. Performance continued to improve over days and was maximal after 5 to 10 consecutive training sessions spaced 1 to 3 days apart. Once a maximal level of performance was reached, most of the gain was retained over months and even years.

This finding supports the methodology followed by most successful therapy interventions in many fields. Short practice on a new procedure or activity in a control setting where performance changes rapidly during the session. This then must be followed by regular practice on the new activity at least 5–10 times with no more than a 2-day break between practice sessions.

To be successful, I wanted to be certain to include the appropriate amount of regular practice. Involving the parents and having the child be responsible for regular vision therapy homework would be one way to make this happen. Our symptoms checklist included asking about the home life of the child. We found that many were living with single parents, many were living with a single grandparent, and some were living with people to whom they were not related. A meeting for parents and guardians was set up and announced in fliers sent home with the children on two different days about a week apart. The meeting was held after school in the library, but only one parent attended the meeting. Because I doubted that a fourth-grader without parental support could be counted on to do the vision therapy homework on the basis needed to guarantee success, we used a different approach. The vision therapy was set up so that the children involved in active treatment would be seen on a one-on-one basis each school day for 30 minutes a day. During this time, they were given four different activities, each of which was done for an average of 7.5 minutes. Because the therapists would not have to check on homework notes or to check whether the activity had been done correctly at hóme, nor would they have to assign new homework, nearly the entire 30-minute session was productive vision therapy. In a way, the children were receiving just as much actual working therapy in a 30-minute session in school as a private patient during a 50-minute session in a private practice setting.

Length of Treatment

The average LRVP patient following the BABO VT-II therapy curriculum will average nearly 32 treatment sessions when done in private practice. The in-school version of the therapy curriculum calls for a minimum of 70 treatment sessions. Because of holidays, school closings for teacher training and advancement, and absences for illness, it was hoped that the children would average 4 days of training per week. Thus, the target of 70 treatment sessions could be completed in 17.5 weeks. Depending on when the extended breaks in the school calendar fell, this could stretch to 6 months for the fastest children to complete the tasks. The therapists were given latitude to extend the amount of time on any particular activity if they believed the child would benefit from additional time on that particular activity. It was expected that the average child would need 15% longer than the minimum to complete the treatment protocol, some would need only the bare minimum, and some would require as much as 50% longer to complete the entire curriculum. Thus, it was expected that some would finish in 70 sessions, and some might require as many as 105, with an average of approximately 80 sessions. Each child was given the same exact activities in the same exact sequence. The demand levels of the activities were adjusted by the therapist to the level necessitated by the patient's level of development. By making this adjustment, each therapy activity could become a meaningful developmental experience for the child. (See Kraskin[17] for more information on how to adjust difficulty levels of specific activities to meet the needs of specific patients.)

Schedules

The teachers, administration, and the master teacher for the fourth grade requested that we make up a schedule that had each child coming to their therapy session at a different time each day so that they didn't miss a full 6 months of a particular subject. A good bit of work went into setting up rotating schedules that still allowed every child to attend every one of their special classes (physical education, art, music, library, Sylvan Learning Center, reading tutor, speech, *etc.*). The vision therapy program began the week after

Thanksgiving in the fall of 1998. A schedule that took into account the amount of time the three vision therapists could work and the physical limitations of the vision therapy room was made up. The schedule accommodated 25 children. Three additional children were selected as alternates. If any of the group of 25 were absent on a particular day, then the alternates would be called. This would decrease the amount of down time for the vision therapists and possibly allow several additional children to move through the process beyond the calculated maximum capacity of 25. An ambitious schedule was worked out over the academic years of fall 1998 to spring 1999 and fall 1999 to spring 2000 to attempt to complete the three groups of 25 students chosen to be treated. This was not the ideal way to do this. The ideal way would be to have had all 75 in the treatment group getting their treatment beginning at the same time. However, space and time limitations set the maximum number of children that could be seen daily by three vision therapists working one-on-one to 25.

Stress-Relieving Lenses

Each member of the treatment group was given a pair of stress-relieving lenses to be kept at school and to be used during all sustained close work and desk work in the classroom. The analytical evaluations along with the results of the stress point retinoscopy were used to determine the appropriate level of plus that was prescribed for near point work. Each prescription was customized for that particular child. Prescribing principles as taught as part of the BABO Behavioral Vision Care course were followed. It is beyond the scope of this article to go into the exact steps whereby the prescription was derived. However, it should be noted that a standard prescribing system was used, which resulted in each child getting the lens that was appropriate for them. The lenses were all supplied in single-vision form. Generally speaking, no plus was given that was higher than the modified stress point retinoscopy lens, and no lenses were given that were less plus than the #14B (fused cross cylinder at near). Embeddedness was used as a guide as to which cases would most benefit from lenses closer to the #14B or closer to the maximum modified stress point retinoscopy lens. Other factors taken into account include the equilibrium findings, the #20 (Positive Relative Accommodation), the #21 (Negative Relative Accommodation), the #7/7A distance subjective, and all phorias. Marchon and Marcolin (Garden City, NY) donated frames for use in the study, and New City Optical, a division of Southern Optical, (Baltimore, Maryland) supplied the lenses and edging services free of charge.

There were many problems with the glasses. They were often broken, necessitating regular repair and replacement. Despite directions to keep the glasses in school, many children took the glasses home and then proceeded to lose them. Some teachers did not, on a regular basis, begin the day by giving the children their glasses. Often when they did give out the glasses, their continued use during the day was not supported and they were not collected at the end of the day. The end result was that the effectiveness of the glasses as a potential aid was greatly reduced. Typically, I see a difference in those children who use their stress-relieving lenses regularly during therapy. This usually results in therapy being more effective, reaching higher levels of cure in shorter periods of time than when the lenses were not an integral part of the treatment program.

THE FIRST COHORT FINISHES

Toward the end of the 1998–1999 school year, 16 children of the first group of 25 had completed the therapy program. Three of the 25 had left the school, and the alternates had taken their place in the schedule but had not yet completed the protocol. Six of the 25 were still working through the protocol but because of either absences or slow progression through the protocol, they were not yet far enough along to be considered as having completed the program. For comparison, 19 controls were selected on the basis of gender and class. Children from the same classes as the 16 children in the treatment group were chosen at random and with the same ratio of gender mix to get a glimpse into how the study was progressing. This was done to see what changes were taking place and as a quality control measure. The summer would provide an opportunity to continue the training of the therapists to bring them to a new level of effectiveness and to improve the quality of the program.

Repeat testing with the Visagraph showed the following changes in treatment group 1 compared with the matched subset of controls (Table 18).

Approximately 7 months had passed between the initial testing and this interim point. The control group had made 0.48 year improvement in the grade level material that they could read and understand at the 70% minimum level of comprehension. The 16 treatment subjects improved by 1.22 years in the grade level material that they could read over the same 7-month period. The control subjects had essentially no change in their reading speed. (On the basis of the above discussion of reading speeds, which showed that across the board the fourth-graders, regardless of the grade level read with comprehension, averaged around 100 words per minute, one would not have expected any significant change in the control group, and none was noted.) The treatment subjects increased their reading speed by 13.7%. Although this difference seems significant compared with a 1.7% decrease in speed in the control group, it fell far short of the 73% improvement in reading speed as was demonstrated when the vision therapy was done with private practice patients in my office.[4] The change in reading speed was achieved by a 20.8% decrease in the number of fixations, compared with a 3.9% increase in number of fixations by the control group, and a 32.4% decrease in the number of regressions, compared with a 15.0% decrease in the control group.

Table 19 shows changes in the same groups of children in some of the selected performance tests.

On the Wold Sentence Copy Test, the control group had an insignificant increase in the time to complete the test of 1.6%, which would mean that based on norms the control group regressed by 2 months in their speed of copying over the 7 months since they had been tested in the fall of 1998. The treatment group decreased the amount of time to copy the sentence by 9.1%, which translates into a 10.7-month improvement over the same 7-month period.

On the Groffman Visual Tracing test, both groups improved significantly. The control group made a 71.0% improvement in their raw score, which translated into a 15-month improvement in performance. The treatment group improved their raw score by 157%, which translates into a 28-month improvement in performance, nearly double the improvement made by the control group.

On the NYSOA King-Devic Saccadic Test, both groups improved their time by nearly exactly the same amount.

MID-TERM CRITIQUE

Although several significant gains had been achieved, as can be seen, I thought that the research subjects should have improved more significantly. The possible reasons for the shortfall of improvement may include but are not limited to:

1. The alteration of my in-office program, which normally takes an average of 8 months with weekly visits and daily homework being compressed to a daily in-school 70- to 105-treatment session program.
2. The use of new trainees rather than seasoned vision therapists.
3. The lack of understanding by the children of what the program was about, why they had been chosen to be involved, and what benefits they personally would derive.
4. The fact that specifics of the optometric and performance data of the children in the treatment program were withheld from the vision therapists.
5. Poor compliance with the use of the stress-relieving lenses.
6. The fact that, because of using a random-

TABLE 18. Visagraph Data (First Cohort-Spring 1999)

	Control (N = 19)	Treatment (N = 16)
Grade level change	+0.48 yr.	+1.22 yr.
Reading speed change (%)	−1.7 slower	+13.7 faster
Number of fixations (%)	+3.9 increased	−20.8 decreased
Number of regressions (%)	−15.0 decreased	−32.4 decreased

TABLE 19. Selected Performance Tests (First Cohort-Spring 1999)

	Control (N = 19)	Treatment (N = 16)
Wold Sentence Copy Test (time)	+1.6% increased	-9.1% decreased
Wold Grade Level Change	-2 months	+10.7 months
Groffman Visual Tracing (score)	+71.0%	+157.0%
Groffman Grade Level Change	+15 months	+28 months
NYSOA King-Devic Saccadic Test (time)	-14.6%	-14.1%

ization process of all children, some children in the research group may not have had a significant need for the vision therapy.

7. Failure to integrate or connect vision therapy with classroom experiences.

CORRECTIVE ACTIONS INSTITUTED

During the summer between the completion of the first treatment group and the beginning of the second, the following actions were taken to improve the quality of care that was being delivered.

1. Additional comprehensive education for the therapists was given over the summer with additional hands-on work in my office under my direct supervision.
2. Before beginning treatment, the research subjects would meet with a vision therapist one-on-one so that the child could understand why they had been chosen to receive this service and how it would help them. The vision therapist would share with them what vision therapy is and what changes they may expect as a result of the work being done.
3. The comprehensive testing record of each child was reviewed with the therapists before beginning treatment so that they knew exactly what each child's visual condition was before beginning vision therapy. Goals for each of the students on the therapy procedures were set, in the same manner as is done in my private practice.
4. An orientation meeting was to be set up with all fifth-grade teachers before the beginning of the next school year. The purpose of this meeting was to fully inform the teachers of the purpose and value of the program.
5. A positive reward for positive use of the glasses system of behavioral modification will be instituted. In unannounced classroom visits by therapists or Dr. Harris, children would be awarded pencils and other educational items for demonstrating use of the glasses.
6. Some children needed extra time to master some of the foundation activities. The therapists were mistakenly under the impression that for the sake of the research and uniformity they had to move the child along anyway to keep to the time schedule. The fact that this was going on despite directions to the contrary did not surface until after a detailed quality control review of the records showed that every child was perfectly on schedule. This misinformation was cleared up so that the therapists fully understood that a particular child could stay with an activity until mastery of that activity was achieved to the level of the therapist deemed necessary.

WE ARE THROWN A CURVE!

A major concern about a single-school design vs a four-school design was that an event might occur in that school or in that neighborhood that would affect in a major way the outcome of the study. During the summer of 1999, several major changes took place at Harford Heights Elementary School. The school was deemed by the Baltimore City Board of Education to be too large to be administered properly by a single administration. During the summer, the school-board divided the school into two separate schools. The physical structure of the school was not altered; rather, the administration of the school and all accounting and business aspects of the school were cut in two: the kindergarten, first, and second grades were grouped into a single entity, and the third through the fifth grades were grouped into a second entity. A new principal was hired, and administrative staff was brought in for the third through fifth grades. The prior administration was also changed somewhat and stayed with the kindergarten through the second grade.

In addition, the contract that the school had with the Sylvan Learning Center was not renewed. Half of the fifth-grade teaching corps transferred to a different school, retired from teaching, or left the area. Two of the teaching positions were filled with just 1 week to go before school started, one was filled the first week of school, and the final position was not filled until 2 weeks into the new school term. Because of all these changes, a meeting with the new principal was scheduled. The meeting originally scheduled for early August 1999 was rescheduled several times because other pressing business, which required urgent attention of the new school principal, took priority. Two weeks before the start of academic year 1999–2000, I met with the new principal. A representative of the Abell Foundation was present at this meeting. Our concern was that without the support of the new administration the study would have to be terminated. The Abell representative was present to help salvage the study if such objections came up. Fortunately, we learned that the new principal had heard something about what we had done and that what she had heard was good. She was aware of the previous administration's commitment to the program, and she stated that she would honor the agreement and that we could continue our work.

As a result of these changes and the late restaffing at the school, the meeting with the teachers, which had been planned to be a half-day meeting, took place for 45 minutes one afternoon the week before the beginning of school

THE RACE TO THE FINISH

The few children from the first treatment group who had not finished by the end of the prior year completed their treatment in September or early October. A new group of 30 more students from the original treatment group selection was started. Of the first cohort of 25 who had been started, 20 completed the treatment and were still enrolled in the school in the early spring of 2000 when it was time to perform the post-treatment testing. Of the 30 who began the treatment in September and October of 1999, 27 completed the treatment and were still enrolled at the school in the spring of 2000. This brought to 47 the total number of subjects of the original 75 chosen by the randomization process, just three under the target of 50 students through the program.

In mid-March of 2000 we were notified that because the out-going fifth-graders would be taking the CTBS tests and the MSPAP in May, we would not have access to any of the children in the month of April as they prepared for the testing. It would take a full month for the testing team to complete the full testing regimen on the full control and treatment groups, so therapy had to end the last day of March 2000. The administration granted permission for the testing phase to continue during the month of April.

Of the original 178 children tested, 75 had been randomly selected for treatment. Of the 60 or so children to receive some vision therapy, 47 children completed the program and were still enrolled at Harford Heights Elementary School in the spring of 2000. Of the 103 other children who were in the control group, 73 children were present for the perceptual/performance testing, 69 took the CTBS standardized tests, and 71 were present for the final Visagraph testing and served as controls. Some children were present on the days that the Visagraph was tested but were absent every one of the days that the perceptual/performance testing was administered. This accounts for the differences in the numbers of subjects in the control group reported in some of the tables that follow.

POST-TREATMENT DATA

Table 20 compares the children's reading level scores. They took the California Diagnostic Test Reading and Math tests in the fall of 1998 as fourth-graders. In the spring of 2000, as fifth-graders, they took the CTBS diagnostic tests. Approximately 17 months elapsed between these two tests. It is unfortunate that the children were given two different standardized tests because it would have been best to compare their scores before and after on the same exact tests. Both tests are produced by the same testing service, but they are not the same test.

When the scores for the CTBS were received during the summer of 2000, two additional statisticians were consulted on how to compare the CDT Reading and CDT Math scores from the fall of 1998 to the CTBS scores on tests given in the spring of 2000. The best

TABLE 20. Standardized Reading Scores*

	Reading Score in Grade Level		
	Control (N = 69)	Treatment 1 (N = 20)	Treatment 2 (N = 27)
CDT Reading, pretest	3.25	2.87	3.08
CTBS Reading post-test	4.54	4.69	4.50
NET change (years)	+1.29	1.82 (P = 0.137)	+1.42 (P = 0.701)

* For all statistics reported in this article changes in the means from pre- to post-testing conditions were used. Analysis of variance (ANOVA) was performed on the net change in means between the groups. In tables with separate treatment groups, a comparison was first performed between treatment group 1 and the control to derive the significance of group 1, and then a separate analysis was performed to compare treatment group 2 to the control group. No direct comparisons between treatment group 1 and treatment group 2 were performed. The statistics package SPSS 10.0 was used for all ANOVA calculations.

way to make the analysis on these would have been to compare either the scaled scores for each subtest with each other or to use the normalized curve equivalent scores. Both of these types of scores were supplied for the CTBS test scores given in the spring of 2000. However, the only scores that were supplied for the CDT Math and the CDT Reading tests given in the fall of 1998 were the grade equivalent scores for each test for each child.

The Abell Foundation has funded a central database for all students in the Baltimore City Public Schools so that research programs such as this one can access information for studies from one central source. The CTBS scores are in this central database, but the CDT Math and CDT Reading test scores are not. The responsible parties in control of the test scores for the Baltimore City Public Schools are not cooperating at the present time with the central statistical database project run by the Center for Social Organization of Schools at the Johns Hopkins University. The representatives of the Baltimore City Public Schools have stated that their concern for the privacy of the individual students scores overrides the potential benefits that might derive from studies such as this. Therefore, the scaled scores or the normalized curve equivalent scores for the subtests on the CDT Math and CDT Reading are not currently available to me. The statisticians noted that a comparison of the means could be done on the grade equivalent scores and that if the scaled scores or normalized curve equivalent scores were made available in the future, more powerful statistics could be done.

Testing in this inner-city school over the past 5 years had shown that, on average, the children were making only a 0.50-year increase in their reading scores for each year they were in school beginning after the second grade. While in second grade, testing had shown that these children scored near the national norms for all second graders. From this point on, for every year in school, the children fell back a 0.05-year behind average students across the country. Thus, despite the efforts to concentrate on reading, this school and many in Baltimore City were, from the second grade on, making only about 6 month's progress in reading achievement for each child in school.

At the time of the pretesting, the students in the control group were at the beginning of the fourth grade and showed a 0.75-year deficit from where they should have scored. At post-testing, the control group was at the end of the fifth grade ready to go into sixth grade. Now the discrepancy was just under 1.5 years, which reinforces the earlier observation of slower than normal growth on standardized tests for reading.

The control group made 1.29 years improvement in the 17 months. The first treatment group made 1.82 years improvement, and the second treatment group made 1.42 years improvement in the same time period. The scores for treatment group 1 are approaching significance, whereas those for group 2 show no significant change over the control group. The mean starting reading levels for each of these groups is not the same. Treatment group 1 started below each of the other two groups and finished with the highest level. The variances of the control and treatment group were matched before the testing. However, only 47 of the 75 treatment subjects were left, and 69 of the 103 control subjects were left. The variances in the beginning scores and the means in these scores may now not be matched as they were before the dropouts over the 17-month period. There was no way to control for these losses and it was hoped that the random loss of some controls and some treatment subjects would not skew the beginning sets of data too much. The various measures of reading will be contrasted and unified into a continuum of reading later in this article.

The control group made 0.84-year improvement in the 17 months in the area of mathematical abilities as opposed to 1.41 years in treatment group 1 and 1.12 years in treatment group 2 (Table 21). Treatment group 1 scores show highly significant changes beyond the 0.01 level. The scores for treatment group 2 approach significance. Many of the areas that vision therapy addresses directly affect spatial thinking, visual thinking, and visual problem-solving involving time and space relationships, which are the foundation on which mathematics is based. What begins to emerge from looking at the two treatment groups over time is that the first treatment group, having completed their therapy in the spring of 1999, had nearly 1 year of using their new skills in the classroom to facilitate the generalization of their new-found skills into higher-level testing such as the CTBS standardized tests. The groups will be tracked over time by using the central database system. The hypothesis for scores over the next year in both math and reading would be that the control groups would continue to make slow progress—less than 1 year of progress for each year in school. Treatment group 2 should make a significant jump in this next year in all areas. Over the next few years both treatment groups should then continue to progress at or slightly better than the rate of 1 year growth in scores for each year remaining in the school system. If this trend continues, then the true merits of the intervention will be demonstrated and thus leave no doubt of the potential benefits of such a program.

Eye Movements and Reading

Table 22 shows the average in each group of the highest reading comprehension levels attainable with a minimum of 70% on the comprehension test, as tested with use of the infrared eye movement-recording device, the Visagraph.

The control group made a 1.25-year improvement in the 17 months of the study. Treatment group 1 made a 2.40-year improvement in the same time, and treatment group 2 made a 2.11-year improvement. Both groups showed highly significant changes. Treatment group 1 would have closed almost entirely the gap in reading level scores between their performance and the level that at which they should be reading by the time they enter the sixth grade. It is anticipated that treatment group 2 only needs some additional time and additional practice using their new-found tools and skills in reading to fully close the gap as well over the next 6–12 months.

Table 23 compares reading speeds before and after treatment, based on the highest level reading passage on which students could achieve at least 70% comprehension both before and after treatment. This means that, on average, the children in the control group were being tested on 1.25 year's more difficult reading passages and those in the treatment groups on a more than 2-year increase in the difficulty of the passage read.

TABLE 21. Standardized Math Scores

	Math Score in Grade Equivalent		
	Control (N = 69)	Treatment 1 (N = 20)	Treat 2 (N = 27)
CDT Math	3.64	3.30	3.24
CTBS Math	4.48	4.71	4.36
Net change (years)	+0.84	+1.41 (P = 0.007)	+1.12 (P = 0.118)

TABLE 22. Visagraph Reading Grade Level Improvement

	Highest Grade Level Attainable with Minimum 70% Comprehension		
	Control (N = 71)	Treatment 1 (N = 20)	Treatment 2 (N = 27)
Visagraph grade before	2.83	3.30	3.00
Visagraph grade after	4.08	5.70	5.11
Net change (years)	+1.25	+2.40 (P = 0.004)	+2.11 (P = 0.015)

The control group improved their reading speed by 11.74 words per minute over the 17-month period. Based on the Taylor norms, the average fourth-grade reader should be reading at 158 words per minute, at 173 words per minute by fifth grade, and at 185 words per minute by sixth grade, which these children were about to enter. Extrapolating from the norms, the control group moved from grade equivalent (GE) score of grade 1.91 to grade 2.39, or 0.48 year change in 17 months. Treatment group 1, which finished its vision therapy in May 1999, showed an actual slowing of just less than one word per minute. Treatment group 2, which had just finished their vision therapy, had made 31.82 words per minute improvement. Their GE scores improved from a grade of 1.48 to 2.65, or 1.17-year change in 17 months.

The lack of speed change at this time in treatment group 1 on their 2.40-years more difficult reading passage is a bit puzzling. The final phase of the vision therapy uses a computer-based reading scan drill program, ReadFast, to train better eye movements. It does this by using a moving window, which parses the text to be read from left to right, showing about two to two-and-a-half words at a time. The speed of the window movement is controlled and is generally set from 10 to 20 words per minute faster than the person is currently comfortable reading. This helps to pace the person during reading and to stop them from making regressions. Because the window only shows the current fixation point and to the right of that, regressions are futile because they would be made to a blank zone on the computer screen.

The differences between the two treatment groups could be explained in several different ways. The first would be that the speed improvement seen directly after intense work with the ReadFast program is when reading speeds reach their maximum and that over time there is a rolling back of the reading speed to the former habitual speed.

A second explanation might be that many of the individual members of treatment group 1 had just recently jumped another level in the demand reading level at which they were able to read. For example, one of the subjects in this group began reading at 167 words per minute on a second-grade-level card. At the end of the treatment, this child was now able to handle a junior high school-level reading card. However, this child's reading speed had now dropped to 116 words per minute. I am certain that most teachers and parents would rather have the child working at a much higher grade level even if they had to slow down their reading speed to achieve the increase in difficulty level.

Of the 20 children in treatment group 1, 8 read slower. Only 2 of the 20 children were not able to handle passages with any greater difficulty. All of the other children were able to

TABLE 23. Visagraph Reading Speed on Highest Demand Level

	Tested at Maximum Reading, Post-Treatment (words per minute)		
	Control (N = 71)	Treatment 1 (N = 20)	Treatment 2 (N = 27)
Reading speed before	112.37	107.75	97.81
Reading speed after	124.11	106.85	129.96
Net change (grade equivalent)	+11.74 (+0.48)	−0.9 (P = 0.378) (no change)	+31.82 (P = 0.106) (+1.17)

read more difficult stories. (Above the sixth-grade level of difficulty the reading samples for the Visagraph have only two reading levels: junior high school and high school/college. Six of the eight children in this group were now able to read at the junior high school level of difficulty. These cards actually cover the spread from the traditional seventh-grade to ninth-grade levels of difficulty and most likely should have been counted in the statistics for maximum reading level as an average demand of eighth grade. However, they were counted as seventh-grade-level stories to keep the scales linear. This resulted in an underestimation or under-reporting of the increase in grade level difficulty handled, as reported above. Not only were six of the eight making this jump into the junior high school-level stories, but also all but two were making a 3-year or more jump in demand level as well, even counting the junior high school-level stories at a grade value of 7.0. Six of 20 subjects (30%) in treatment group 1 and 5 of 27 subjects (18.5%) in treatment group 2 had advanced to the junior high school level. Only 3 of 71 control students (4.2%) advanced to this level.

Table 24 shows the grade level improvements now with the junior high school-level cards assigned a value of 8.0 for all subjects who successfully read this level story with a minimum level of 70% comprehension. As a result of the change, the spread between the control and treatment groups has increased, and the measures of significance on a one-way analysis of variance has also increased for each treatment group.

A combination of both of these effects may also be in play. The only way to know for sure would be to watch groups of children closely as they improve in reading ability to see how comprehension levels and reading speeds interplay. The hypothesis is that as the child jumps to a new level of demand competency, first there is a loss in speed, which rebounds over time as the child automates the skill at the new higher level. This would plateau for a period of time and then the cycle would continue.

A third explanation might be that the therapists put more emphasis on those activities that would foster greater speed changes with treatment group 2 than with treatment group 1. The therapy staff was extensively retrained during the summer between the times that treatment group 1 finished their treatment and treatment group 2 began treatment. However, this explanation would tend to be refuted by the fact that in the spring of 1999, 16 members of treatment group 1 showed a 13.7% increase in speed vs 19 members of the control group, who showed a 1.7% slowing of their reading speed. Thus, at one point, soon after the completion of the treatment phase, there was a speed increase. However, this was measured on the same level reading passage. This will be discussed more in the Discussion section.

I wanted to remove the grade level differential as a factor in comparing reading speeds before and after testing. Table 25 represents this reading-level-free comparison. To compile these data, the children were given a reading passage on the post-test that matched the highest reading level they attained at the pretest session. Thus, if the highest level a child could read at and still get at least 70% correct on the comprehension in the pretest was fourth grade, then they were given a fourth-grade text at the post-test, even if they could now read on the junior high school grade level with comprehension greater than or equal to 70%.

With the reading level kept constant the control group made a GE change of from grade

TABLE 24. Visagraph Reading Grade Level Improvement (with Junior High School = 8.0 Modification)

	Highest Grade Level Attainable with Minimum 70% Comprehension		
	Control (N = 71)	Treatment 1 (N = 20)	Treatment 2 (N = 27)
Visagraph grade before	2.83	3.30	3.00
Visagraph grade after	4.13	6.00	5.30
Net change (years)	+1.30	+2.70 (P = 0.001)	+2.30 (P = 0.009)

TABLE 25. Visagraph Reading Speed (Demand-Controlled)

	Words per Minute		
	Control (N = 71)	Treatment 1 (N = 20)	Treatment 2 (N = 27)
Reading speed before	112.37 (1.91)	107.75 (1.79)	97.81 (1.48)
Reading speed after	128.17 (2.53)	131.30 (2.71)	139.41 (3.05)
Net change (grade equivalent)	+15.80 (+0.62)	+23.55 (P = .577) (+0.92)	+41.60 (P = .039) (+1.57)

Students were tested at the same reading level post-test as pretest.

1.91 to grade 2.52 or a 0.61-year improvement. The first treatment group made a GE change of from grade 1.79 to grade 2.71 or a 0.92-year improvement. The second treatment group made a GE change of from grade 1.48 to grade 3.05 or a 1.57-year improvement. This showed significance at the 0.05 level (P = 0.039).

Selected Performance Tests

Table 26 reports on the differences in the NYSOA King-Devic Saccadic Test before and after the treatment. Four subjects from treatment group 2 were not present in school when this portion of the testing was administered.

In each group, the average times got faster and a net decrease in time to call off the numbers was noted (Table 27). The control group pared 19.49 seconds from their time to shift from a score equivalent to that of the average 8.0 year old to that of the average 10.6 year old. The ending score for all three groups is within a second of each other. Treatment group 1 made a bigger net change, but that may have been because they started at a lower level of performance. These differences were not significant.

Both the control group and treatment group 2 made very similar changes of 5.81 and 6.00 points, respectively, over the 17-month time of the project, which correspond to 0.95 and 1.00 year change, respectively. Their scores moved them from approximately a 7.5-year-old performance level to an 8.5-year-old performance level. Treatment group 1, however, moved from a 6.94-year-old performance level to a 9.15-year-old performance level. This was significant to the 0.05 level (P = 0.024).

Changes in eye-hand coordination are summarized in Table 28.

The net change in the ability in both treatment groups showed an improvement in the actual score, but these changes did not approach significance. The control group made 1.04 years of improvement in the 17 months of the program, whereas the treatment groups made 1.41 and 1.45 years of improvement. The results from the Gates Oral Reading Survey are shown in Table 29.

The reading survey results for the three groups are virtually indistinguishable from one another. There appears to be no difference in the children's ability to decode what they read. On average, the entire group has the ability to decode at or slightly above their current grade placement. Certainly, mastery of reading at the same level for comprehension is not at the same level, and all measures of the mechanics of reading measured by the Visagraph are much below these levels.

At the time of this writing, much additional raw data remain that have not been given the attention it merits. Much of this are the data from the analytical examination as

TABLE 26. NYSOA King-Devic Saccadic Test

	Seconds to Complete All 3 Samples (Age in Years)		
	Control (N = 73)	Treatment 1 (N = 20)	Treatment 2 (N = 23)
Before	80.60 (7.94)	84.68 (7.74)	78.48 (8.11)
After	61.11 (10.63)	59.29 (10.79)	59.52 (10.77)
Net change	−19.49 (+2.69)	−25.39 (P = 0.209) (+3.05)	−18.96 (P = 0.963) (+2.66)

TABLE 27. Groffman Visual Tracing

	Points Scored (Grade Equivalent)		
	Control (N = 73)	Treatment 1 (N = 20)	Treatment 2 (N = 23)
Before	13.23 (7.46)	9.40 (6.94)	13.57 (7.51)
After	19.04 (8.41)	22.60 (9.15)	19.57 (8.51)
Net change	+5.81 (+0.95)	+13.20 (P = 0.024) (+2.21)	+6.00 (P = 0.949) (+1.00)

well as the other more "pure" optometric tests such as visual acuity, ocular motility, stereo acuity, near point retinoscopy, accommodative ranges, equilibrium findings, phoria measures, plus acceptance findings, cover testing, and more. Attention in reporting data has been slanted toward "performance" testing and/or standardized testing, because this is the area that the public is most concerned with. This represents the unmet needs of the population that the profession of optometry serves through the treatment modality vision therapy or visual training. Regardless of what changes occurred with the specific optometric findings, those concerned (the Abell Foundation, the parents of the children with LRVPs everywhere, teachers of children with LRVPs, and administrators of schools with children with LRVPs) were looking for changes in performance in the classroom and on standardized tests. Making a difference in these areas of performance will lead to continuance and future expansion of these types of programs.

DISCUSSION

Reading performance on a standardized test requires the integration of all aspects of reading skills into the ability to get information from written material and to use that information to answer questions and to do so over a sustained period of time. When "reading" is investigated, few researchers report on the spectrum or continuum of reading abilities and scores, which correlate to the specific aspects of the reading process the tests are investigating. Attempting to quantify reading ability into a single number or a single score cannot begin to describe a child's strengths or weaknesses or to communicate the degree to which that same child can efficiently and effectively derive meaning from the printed page and put that information to use. The following is an attempt to give the reader a view of how "reading" has been investigated in this research.

The Gates Oral Reading Survey probes specifically into to what degree the child has broken the code of being able to say the words from the printed page. It is a decoding test only. This test typically yields the highest "reading level" of any reading test because it does not involve comprehension and is not a test that asks for performance to be sustained over any great period of time. The groups on the Gates Oral Reading Survey are all essentially the same, being somewhere in the middle of fourth grade at the beginning of the testing period and at the beginning to middle of sixth grade at the end of the program. There were no significant differences between the degree of change or the beginning or ending points. The test measures reading at the basic ability level of decoding only.

As stated before, when Visagraph recordings are made, comprehension is an integral part of the testing. A minimum of 70% on the comprehension test is required, and reading demand levels are adjusted up and down until

TABLE 28. Eye-Hand Coordination Sub-Test of the Developmental Test of Visual Perception II

	Age Score		
	Control (N = 73)	Treatment 1 (N = 20)	Treatment 2 (N = 23)
Before	8.17	8.14	8.47
After	9.21	9.55	9.92
Net change	+1.04	+1.41 (P = 0.518)	+1.45 (P = 0.480)

TABLE 29. Gates Oral Reading Survey

	Grade		
	Control (N = 73)	Treatment 1 (N = 20)	Treatment 2 (N = 23)
Before	4.49	4.72	4.47
After	6.28	6.56	6.22
Net change	+1.79	+1.84 (P = 0.813)	+1.75 (P = 0.862)

the appropriate demand level has been found. It is important to note that on reading passages below the fourth grade the passages are nine lines long and analysis is done on the middle seven lines, which average 50 words, and that all passages of fourth-grade-level difficulty and higher are 12 lines long and analysis is done on the middle 10 lines, which average 100 words. This test is more involved than the Gates Oral Reading Survey because it incorporates comprehension while measuring mechanical aspects of reading at the same time. When the Visagraph is recorded, the subject is aware that their reading speed is a part of the test. They are told to read for comprehension and they have to answer questions on what they read. They are not told directly that we are measuring their reading speed but they are told that their eye movements are being recorded into the computer. This causes some to be a bit more stressed during a formal recording than if they had been asked to read the story to themselves by just holding the card without the goggles on and without the computer being nearby.

It is important to note that the Visagraph test is very short in duration. For a reader doing a third-grade passage at 100 words per minute, the total time to read the passage may be just 35–45 seconds. For the readers of the longer stories, the total time to read the entire 12 lines, which contain more than 100 words, may run from 70 to 85 seconds. Thus, those who have not yet fully integrated these new skills may perform well on a short-burst sprint like the Visagraph but have not yet generalized or integrated these skills and abilities such that they can perform at that level for the sustained periods of time, which is precisely what standardized testing requires.

Retesting at the end of the program was done at two different demand levels. In one instance, testing was done at the same demand level that the children had read at during the pretreatment testing. The second was to test the child at the highest reading level they could understand at the 70% minimum comprehension level. It is known that as the demand level of the reading passage changes for a single reader, very large changes in reading speeds can be made. For example, if a person is reading a passage 2 years below their instructional level, they can breeze through it very quickly with little effort. Their speed drops a bit when they are at their instructional level. When given something that is too hard, most people at first drop their speed significantly. Once in a while, when a person perceives a passage to be just too hard, they may give up or move into "flight." They don't actually quit or leave the situation physically; they continue to work on the task but their commitment in effort, energy, and resources drops significantly. At that time, one may observe that any outward appearance that the person was concentrating is now gone. They back away from the text; tension is gone from the face and the posture, *etc.* At that point, the reading speed usually increases dramatically. They have switched from "reading" to "skimming." When asked if they read the passage they will say, "Yes," but in fact about all they have done is pass their eyes over the text with little effort and even less comprehension. Therefore, a 1- or 2-year level change on the demand can greatly affect the speed measures. So, one of the post-test reading passages was done at the previous demand level so that speeds could be directly compared. We believed this would maximize the speed differences and demand would be factored out.

The second set of data looked at the new maximum level of difficulty that the child could handle as a result of the treatment. What emerged from our data here and what has been seen in my general practice is the following. A vision therapy patient is able to shift up to harder difficulty levels rather early in the vision therapy process. The mechanics of reading may appear to regress during the initial phase of working on the higher demand material. Most teachers and parents are very happy that the child is handling more difficult material, and few parents and teacher use stop watches to determine how quickly their children are reading. One would have to spend a good deal of time with a child to be able to detect 30% or 40% differences in reading speeds when a child is reading silently.

Many parents and teachers do notice changes in reading performance early in therapy. The primary way they become aware of this is through oral reading. What they typically detect are cadence differences in the oral reading. The child who is developing and learning better tracking, eye teaming (binocular), and focusing of visual attention skills typically can more quickly and more smoothly decode the words in front of him and recite them in a more connected manner. The Gates Oral Reading Survey helps us to see the reality, which is that the actual decoding ability is not really enhanced; rather, the system is working more efficiently and smoothly and, therefore, those who work with the child perceive an improvement in reading when in reality the decoding skills are roughly unchanged.

Over time the child does begin to understand more difficult reading material. This does not happen because the vision therapy has increased his intellect nor because there has been a significant change in his reading ability. Rather, the change in the child's ability to derive meaning from the printed page has been enhanced very significantly because the process of vision has been improved by working on the basic tracking, teaming, and focusing problems mentioned in the introduction. The end result is that now the child can use more of his intellect. Rather than using so much effort and energy to perform the mechanical aspects of performing reading, he can now channel his energies into learning from reading.

At first this will occur only in short bursts and will be picked up by the Visagraph well before the same changes are revealed by testing that takes longer than just a few minutes to complete. Only after 6–12 months using newly acquired skills and abilities may the gains that the child gets from having been involved in an optometric vision therapy program become consolidated or embedded enough that the child can use these new skills and abilities for long periods of time. Only then will the changes from the treatment reveal themselves on standardized testing.

1. The resultant hypothesis would rank the expected order of seeing changes in the tests, from earliest to latest, as followings: Gates Oral Reading Test—decoding only
2. Visagraph—difficulty change, no mechanics changes yet
3. Visagraph—mechanics changes at lower level demands
4. Visagraph—mechanics changes at higher level demands
5. Standardized reading test scores

Where Are We Now?

During the academic year of fall 2000 to spring 2001 we have begun working with a new group of fourth-graders. Some of the testing has been streamlined. Rather than working with randomly selected children from a pool of all children, the classroom teacher and the master fourth-grade teacher were asked to submit names of children they believed could benefit from the testing and treatment program. This single change has helped us gain significant support from the teachers and the administrators. From the pool of children identified with visual difficulties, children have been randomly chosen to participate in the treatment program. Those not chosen constitute the control group for those receiving treatment.

Standardized testing on the initial control and treatment groups will be administered during the spring of 2001 and every spring for the next few years even though the children will have moved on to middle school. Results on these tests will be made available for analysis over the years. An attempt will be made to locate as many of the original groups as possible to repeat the Visagraph testing in the spring of 2001, 2002, 2003, and 2004.

Randomization Difficulties

When the randomization was done at the beginning of the study, the group of children chosen to receive vision therapy was taken from the group of all children. The early screening showed that the prevalence of LRVPs was between 80% and 85%. Because the numbers were so high, the samples were randomly taken from the entire student body. However, it is possible that some of the results did not show a great difference between the two groups because some of the children in the treatment group did not have primary vision problems. A *post hoc* analysis will be performed to determine whether this might have been significant. Future designs should first cull from the general population of students those who have significant primary visual problems and then perform the randomization on that subgroup.

What Does the Future Hold?

The program will move in one of three directions. Plans are being made to expand initially to one additional school and then to expand over a 10-year period to all schools within the Baltimore City Public School system. Expansion would depend on the continued strengthening of the performance differences between the treatment and control groups. Launching a program from scratch at a second school with the benefit of our 3 years of experience at Harford Heights Elementary School should help to make the new program far more effective. If such a program is even more effective in another location, then the evidence that vision therapy for LRVPs in inner-city school populations would be strengthened. At some point, the evidence that the problem exists and that there is a safe and effective treatment for these problems may become overwhelming. Municipalities and school boards around the country and around the globe will need rational, cost-conscious plans and programs to address this major public health concern.

Where would funding for such a program come from? School-based optometric vision therapy may save significantly on the amount of money being spent on special education. Five to 7 years from now we should be able to perform a retrospective study of the children in this study, comparing the treatment groups with the control group to determine the level of special education services required by these children and calculate the actual costs to educate the different groups in our study. The Baltimore City Public Schools spend an average of $6800 per year on children who, under Public Law 94-142 and its regulations, require additional educational services over and above the base of $3100 per student per year for those not receiving extra services.

A detailed business plan of a program to bring this treatment to a major school district has been created, and projected gross cost per student served, including equipment, hiring, training, supervision, and all administrative costs, is between $1600 and $1750. This is a one-time expense per child with a learning-related visual problem. The $6800 is a recurring cost each year for as long as the child remains in school and needs extra educational and supportive services. If a child is identified and treated in the fourth grade and if the optometric vision therapy reduces, on a conservative basis, by one-third the needs of the child served, then the savings would be $13,586.40 per student served—more than an 8-to-1 return on investment. This assumes that, on average, these children stay in school only 6 years after this time. If either the treatment is more effective or the children on average stay in school longer, then the potential savings could be significantly higher.

ACKNOWLEDGMENTS

Assistance has been given by Drs. Valerie Whittaker, Ted Sober, Ron Berger, Dana Taylor, Jason Sober, and Thelma Lombardi, Katherine McKearn, Emily Harris, Betty Francis, Cindy Hubbard, Carol Zimmerman, Sidney Chernick, Theresa Krejci, Dennis Hoover, Liz St. Ours, Jennifer Kungle, and Phyllis Lloyd. Essilor and Marchon/Marcolin as well as New City Optical have given additional assistance. Daniel Agley, PhD in statistics and faculty at Towson University, served as the statistician and consultant for the program. Additional thanks to Molly Mullen who edited the first version of the article and to Irwin Suchoff, Selwyn Super, and Lenny Press for their input.

REFERENCES

1. Hoffman DD. *Visual Intelligence—How We Create What We See.* WW Norton and Company, 1988;0-393-04669-9:8,XII.
2. Lieberman S, Cohen A. Validation study of the New York State Optometric Association (NYSOA) vision screening battery. *Am J Optom Physiol Opt.* 62(3):165–168.
3. Cooper J, Duckman R. Convergence insufficiency: Incidence, diagnosis, and treatment. *JAOA.* 1978;49:673–680.
4. Hoover D. The effects of using the ReadFast computer program on eye movement abilities as measured by the OBER2 Eye Movement Device. *JOVD.* 1997;28(Winter):227–234.
5. Harris PA. *The Prevalence of Visual Conditions in a Population of Juvenile Delinquents.* OEPF, 1989.
6. Duane A. A new classification of the motor anomalies of the eye, based upon physiologic principles. Part 2. Pathology. *Ann Ophthalmol.* 1897;6:247–260.
7. Cooper J, Duckman R. Convergence insufficiency incidence, diagnosis and treatment. *J Am Optom Assoc.* 1978;49:673–680.
8. Kent PR, Steeve JH. Convergence insufficiency incidence among military personnel and relief by orthoptic methods. *Mil Surg.* 1953;114:202–205.
9. Norn MD. Convergence insufficiency: incidence in ophthalmic practice results of orthoptic treatment. *Acta Ophthalmol.* 1966;44:132–138.
10. Cohen A, Lieberman S, Stolzberg, Ritty MJ. The NYSOA vision screening battery—A total approach. *JAOA.* November 1983;7:979–984.
11. Kulp PT, Schmidt P. Reliability of the NYSOA King-Devic Saccadic Eye Movement Test in kindergartners and first graders. *JAOA.* September 1997;68(9):589–595.
12. Lieberman S, Cohen A, Rubin. NYSOA K-D Test. *JAOA.* July 1983;7:631–637.
13. Groffman S. Visual Tracing. *JAOA.* 1966;37(2):139–141.
14. Wold R. *Screening Tests To Be Used by the Classroom Teacher.* Academic Therapy Publications, 1970.
15. Lowry R. *Handbook of Diagnostic Tests for the Developmental Optometrist.* OEPF, 1970.
16. Taylor SE, Frackenpohl H, Pettee JL. *Grade Level Norms for the Components of the Fundamental Reading Task,* Research and Information Bulletin, No. 3, Huntington, NY: Educational Developmental Laboratories, Inc., 1960.
17. Kraskin R. *VT in Action.* OEPF. 1965–1968.
18. Karni A. Adult cortical plasticity and reorganization. *Sci Med.* January/February 1997.
19. Birnbaum M. *Optometric Management of Nearpoint Vision Disorders.* Butterworth-Heinemann, 1993;129.
20. Williams S, Simpson A, Silva PA. Stereo acuity levels and vision problems from 7–11 years. *Ophthal Physiol Opt.* 1988;8.

APPENDIX A Visual History

Name: ______________________________ (M/F) Teacher: ______________________________

Room: __________

Parent's Names: ____

Birth Date: ______________________________ Has a grade been repeated (**Y/N**)?

Has glasses (**Y/N**)?
If yes, when are they worn (**full time/school/home/close work/not worn**)?
If no, did they ever have glasses? (**Y/N**)?

Besides screenings have you ever had a professional vision examination? (**Y/N**)?
If yes, how long since your last exame? (**<1 year/>1 and <2 years/>2 years**)

Ask each of the following: Do you get or do you ever experience. . . .

Headaches	Yes	No	Cover one eye	Yes	No
Blurred distance sight	Yes	No	Rub eyes excessively	Yes	No
Blurred near sight	Yes	No	Eyes red a lot of the time	Yes	No
Double vision	Yes	No	Get lost when reading	Yes	No
Eyes feel tired	Yes	No	Use your finger to keep your place	Yes	No
Eye hurt	Yes	No	Bump into objects	Yes	No
Hold reading close	Yes	No	Bothered by bright lights	Yes	No
Hold reading far away	Yes	No	Dislike reading	Yes	No
Close one eye	Yes	No			

How much TV do you watch (**<1 hour/1> <3 hours/>3 hours**)?

How well do you do in school (**average/better than average/lower than average**)?

Do you read (**faster/slower/the same speed**) as everyone else in your grade?

Do you understand (**more/less/the same**) as everyone else in your grade?

APPENDIX B

Baltimore Academy for Behavioral Optometry Learning-Related Visual Problems Curriculum (VT-II) for School-Based Vision Therapy

Session	Activity 1	Activity 2	Activity 3	Activity 4
1	Eye Control	Square Balance Board	Handball	Hart/Chart Near/Far Rock
2				
3		Walking Rail	Punch Ball	
4				
5			Thumb Ball	MAR
6	Chalk Board Circle Trace			
7			3 ways to catch	
8				
9	Chalk Board Racetracks			
10			Look and catch	
11	Chalkboard Double O's	Line Tracing		
13			3 O's and catch	
15			O's around head	
16	Comp. Video Saccades			
17				
18			Bunting	BAR
19	Coin Circles			
20				
21			"V" Bunting	
22		Mazes		
23			Half circles	
24	Visicare Disassociated Pointing			
25			Column Jumping	
26	Visicare Associated Pointing			
27				Keystone Fusion Games
28	Visicare Jump Ductions			
29		Ann Arbor Letter Tracking		
30	Circle Jumping			Near Vision Vectograms

APPENDIX B Continued

Session	Activity 1	Activity 2	Activity 3	Activity 4
31			Physiological Diplopia	
32	○Square on table			
33				Vectogram Localization
34	○▶ Single-Stage 1		Brock String	
35				
36	○▶ Single-Stage 2		Bead Jumping	Vectogram Jump Ductions
37				
38	Slap Tap	Rotator Golf Tees	Bead Sliding	Vectogram BOP/BIM
39				
40			Bug on the string	Rotator O's
41		Rotator Tees with Ping Pong Balls		
42				
43		C-P Saccades	Bug on the wall	Comp Video Pursuits
44				
45				
46		Flashlight Pointing-1 Light	ReadFast Comprehension Testing	Comp Vergence
47			Visualization-Arrow Jumping	
48	Cheiroscopic Tracing			Comp BI/BO JD
49			Visualization-Draw on Back	
50				Comp Directionality
51	Van Orden Stars	Flashlight Pointing 2 Lights		
52			Visualization-Mystery Bag	
53	Comp. Tachistoscope			Comp Visual/Motor Integration
54		Read Fast Moving Windows		
55			Visualization-Rotating Pictures	

APPENDIX B Continued

Session	Activity 1	Activity 2	Activity 3	Activity 4
56				Comp Auditory/Visual Integration
57	Harry's Blocks			
58			Spelling Method	Comp Visual Scan
59				
60				Read Fast Tachistoscope
61				
62				
63				
64				
65				
66				
67				
68				
69				
70				