Evaluation of the Light Sensitivity Project in Pomona Unified School District

John Michael Duffy, Ph.D.

Department of Educational Research and Evaluation
Pomona Unified School District
2/12/98

This is an evaluation of the Light Sensitivity Project in the Pomona Unified School District. The findings of the evaluation support the argument that participation in the program leads to higher achievement scores. The evaluation design, however, needed better controls for selection effects, history effects, Hawthorne effects, and instrumentation effects. A better designed study is planned for the near future.

Introduction: General education students have varying levels of visual perceptual skills, and some of these conditions can be "corrected" by the use of differently colored acetate sheets. (See Spafford, 1995; Tyrrell, 1995; Craig, 1991; Padack, 1992; Otto, 1994) The condition, called Scotopic Sensitivity Syndrome, causes the student to see the letters distorted in some way. For example, the letters may be seen as blurry, moving, double, reversed, inverted or not on the page at all. Some students see letters that switch around in the words or words that change places in the sentences. It also affects numbers for some students. Other characteristics of the printed page can also be affected. Dorothy Henson-Parker, a school psychologist at Pomona Unified School District, has studied this problem, collecting research data on students that have this condition.[1] (See Henson-Parker, 1997, for a description of the assessment.) This paper evaluates this program at two of Pomona Unified School District's elementary schools, Kingsley and Montvue. While Ms. Henson-Parker has provided diagnostic and interview data for this evaluation, the achievement scores are from the records of the school district. One acid test that demonstrates the effectiveness of the program is whether it is successful in increasing the achievement scores of participating students. This evaluation uses the program participants, the diagnosed students that did not have the condition, and the general population of the elementary schools' students as comparison groups.

Literature: Generally, the literature on this subject finds that visual acuity is improved for students using the Irlen filters. (See Spafford, 1995; Tyrrell, 1995; Craig, 1991; Padack, 1992; Otto, 1994.) Strong research designs, however, have found little or no effect on changing reading scores. O'Connor (1990), Blaskey (1990), Solan (1990), Fletcher (1994), and Cardinal (1993) found that using the lenses did not lead to increased reading scores. In contrast, Robinson (1990) found that reading comprehension and accuracy increased but reading rate did not increase. Solan (1990) reviewed the literature and concluded that carefully designed and controlled studies do not currently lend support to the Irlen hypothesis. Robinson(1994) reviewed the literature on Irlen filters and concluded that "not all studies of

[1] Ms. Henson-Parker wishes to thank her co-investigator, School Psychologist Jami Lawson, and the teachers who participated in the study: Phyllis Eschleman, Carole Glass, and Karen Griffin, Resource Specialist Cheryl Burch, and their principals, Mary Binch and Constance Struve.

reading achievement were positive, which emphasizes the fact that reducing print distortions, whether by optometric prescription or color, is likely to be only one factor in reading development." (Robinson, 1994, p.10)

Model, assumptions, and variables: The program calls for screening an entire classroom of students in general education classrooms, interviewing all of the students, selecting students that meet the criteria of this visual perceptual condition, identifying the corrective color needed to use with all print-based media, and collecting achievement score data on the students in two time periods (once previous to the use of the color screen and once after the introduction of this intervention).

Hypotheses: It is hypothesized that the screening for the visual perceptual skills debility and using the colored sheets for reading and for test taking will lead to higher scores relative to the higher scores experienced by the general population of students and the students that were screened as not having the debilitating condition. Background variables that are normally collected in the schools will be held constant. Thus, it is not enough that the scores rise in the treatment group. The scores of this group must rise significantly higher than the scores of the other groups. Tests will be made of the achievement scores of reading, language, and math components of the CTBS/4 and SABE II achievement tests.

Moreover, it is also hypothesized that the increase in scores will be greater for the reading and language achievement tests components than for the math computation sub-test components. This differential effect is due the different text spacing on the tests. The reading text is generally in single-spaced lines across the page. The math test questions are shorter in total word count and shorter across the page, usually providing a visual separation between problems. Thus, a student using a colored plastic transparency to provide accurate visual input will gain relatively more clarity, stability, and accuracy in the language and writing portions of the test.

Sample of students: A convenience sample of students was selected in the second and third grades at Kingsley and Montvue elementary Schools as part of the Light Sensitivity Project in the Pomona Unified School district. The notes that are regularly taken during the screening exams were used to collect some of the student data, notably, which students were screened with the condition and which wore glasses. Normal curve equivalent achievement scores, other demographic variables, and program participation variables were then selected, matching the students that had both the previous year's achievement scores from the second grade and the present year's score. Thus, the numbers in the samples reflect the continuation of students in the schools over this two year period. approximately 15% of the students had scores in the first year but not the second. There was no attempt to randomly select students for screening; there was also no attempt to randomly assign students to the treatment group.

Time frame of data collection, analyzes, and reports: Achievement data collection occurred in April, 1996, and April, 1997. Data collection from the district database occurred in October, 1997. Students were provided the plastic transparencies in late February, 1997. thus, the "treatment" group used the plastic transparencies for a maximum of two months. Statistical analysis and preliminary reports were conducted and written in December, 1997.

Statistical tests: A simple t-test is considered inappropriate in this case, as is an analysis of variance procedure. Generally, both of these methods rely exclusively on random assignment of students to the treatment and control groups, which is usually a part of the design of the evaluation. In this case, students were chosen for the treatment group because they were diagnosed as having the condition. A regression analysis is needed to control statistically for all of the possible "blocking conditions" that could not be controlled in the design of the "experiment." Thus, in the presence of all of the group differences among the students, the addition of the diagnosis and treatment using the colored acetates is expected to be significant and positive. This test can be done using the 1997 achievement test scores as the relevant dependent test score, using the group categories (treatment, diagnosis but no treatment, and all other regular education students at that school in that grade level) as independent dummy variables, and controlling for the level of the achievement scores by including the previous year's score as an independent variable. The test for the first hypothesis is that the coefficient for the treatment dummy variable is positive, significant and significantly greater than the coefficient for the other dummy variables of interest. The test for the second hypothesis is that the coefficients for the treatment dummy variables for the reading and writing regression models are positive, significant, and greater than the coefficient for the math computation regression model.

Results: Overall, this study found supporting evidence for the use of plastic transparencies to aid reading learning. As can be seen from the second and third grade means of students' math, reading, and language scores (see Tables 1 and 2), students in the treatment classes had higher mean NCE scores before and after the use of color transparencies. Moreover, these students gained relatively more during the year in which the use of color transparencies occurred. It is unclear whether the treatment "caused" this increased learning in these subjects, as both the students with plastic transparencies and those without plastic transparencies increased their NCE scores. A better designed evaluation study could clear up this question.

The models used identical sets of explanatory (but non-experimental) variables, and the set of models as a whole provided robust results for these explanatory variables. Reviewing Tables 3 and 4, we find that the lagged NCE values were always positive and strong explainers of the dependent variables. Being female had a negative effect for math scores, and were only significant for the computation subtest. LEP categorization had a negative effect for language scores and a positive effect or math scores, principally the computation subtest. Third graders scored higher for language, math, and reading, as expected. GATE identified children scored higher only on reading. Finally, Title I students scores significantly lower only on the math sections, although the signs were negative for the other two other subject areas. Thus, the tables show expected results for all the subject areas, making the models robust in their explanatory results, aside from the experimental variables.

As to the first hypothesis, this study found that there was an independent effect for the students who were assigned the plastic transparencies. consider Tables 3 and 4. Using regression analysis, which controls for the independent effects of previous score levels, program participation, and student demographics, we find that participation in the treatment increases scores on the achievement tests for every subject area. This is *prima facie* evidence that the treatment increases student achievement scores.

Unfortunately, the control group of students, who were in the same classroom as the treatment group but who did not receive the plastic transparencies, also increased their achievement scores. This suggests that the teacher or type of student in the classroom was at least partly responsible for the increased scores. In the table below, all of the relative contributions (beta values) of participating with transparencies were positive, but so are the relative contributions of being in the same class as the "treatment" students but not receiving the transparencies. Participation in the treatment was about two-thirds as important as prior math scores and as important (but in the opposite direction) as participation in the Title I program. A better designed and implemented evaluation study could tease out the effects of these confounding variables.

Model Quality and the Relative Contributions of the Transparencies in Each Model
Adjusted R-squares and Beta Values in Each Model

	Total Language	Total Reading	Concepts & Applications Math	Total Math	Computation Math
Adjusted R-squared	.809	.437	.453	.429	.242
Relative Contribution of Transparencies, Beta Value	.164	.162	.183	.231	.222
Relative Contribution of only being in the class with experimental group, Beta Value	.088	.153	.281	.223	.122

As to the second hypothesis, this study found that betas, which are the regression equivalents for effect sizes (standardized coefficients), for the total math battery were nearly 50 % larger than those of either reading or language. (See Tables 3 and 4). The standardized coefficients for reading and language were approximately the same; the standardized coefficient for the Math Computation subtest was higher. This finding was not expected, as we hypothesized that the effects for reading and language would be higher. The standardized coefficient for the Math Concepts and Applications subtest was more like the reading and language coefficients. Moreover, the effect for the "classroom control" students (those students not assigned a transparency but residing in the same classroom as the treatment students) was as high for math and reading but only half as much for language. Again, this points out that something else is happening to the students in the class besides participation in the Light Sensitivity Project.

We are reluctant to give up on the second hypothesis. It may be that the higher error in the math models is leading to the higher estimates of the relative contributions of the participation in the experimental classrooms. In the table above, as the quality of the model decreases, i.e., as the R-squared decreases, the relative contributions of being in the experimental classrooms increases. The absence of better explanatory variables for the math models increases the variability of the experimental variables. While the relative contributions of being in an experimental classroom are greater for the math scores, the models have a great deal more variability. This reasoning is only suggested as one possible explanation of the contrary results.

Generally, the transparency treatment group gained higher achievement scores than the control group of students. This leads to the conclusion that the filter treatment did contribute to increased learning, including increased language achievement in the treated students, although many other things, including design effects and model effects, confound this finding.

Threats to internal and external validity: This is an ex post facto research design with two treatment groups (screened students, ones issued plastic transparencies and ones not issued plastic transparencies) and a control group (all other non-screened students at these grade levels). Threats to internal validity may lead to finding the same results that were found using the present methodology. Better studies will include responses to these threats, taking account of any additional effects before the treatment tests are made.

Selection threats: Selecting the students using a diagnostic technique precludes the use of random assignment of students to the treatment and control groups. Educational research often has this threat to internal validity. We do not know how much bias was introduced by allowing the screening staff to include children from the class into one group or another. Our results suggest that the control classes and the other students had different pre-treatment achievement scores. We used the entire non-screened classes as the control group, but this is unsatisfactory for the selection of the control. One way to attempt a better control of this threat is to randomly select classes for the program and then to randomly assign treatment and control status. First, we could give the diagnostic to each of the classes of students. In one of the classes, we assign them the plastic transparencies, giving a placebo to the students not needing a transparency. In the control class, we keep track of the students diagnosed as needing the filter and those not needing it. The test of increased reading learning could then be between the condition-identified treatment group, the placebo treatment group, the filter-needing control group, and the filter-not-needing control group. Without the random assignment, which is the design presented in this study, the evaluation could have inadvertently included the biases of the screening staff.

Instrumentation Threats: The CTBS/4 or SABE II are regularly given once a year to these grade level students, and the pre- and post-testing does not seem to affect the results. However, did the filter treatment students actually use the sheet during the time between the prescription and the second testing? No formal attempt was made in this study to follow-up on the students to determine this fact, although calls to teachers were made to check on what was happening. A better design would be to observe the students using these sheets, recording the frequency of the use of the acetates during class work and the test. While every student considered has, indeed, taken both the “pre” and “post” CTBS/4 or SABE II exams, we have no formal evidence that the students have used the sheets of acetate throughout the prescription period and during the second testing. This stronger design, one including the collecting of filter use data for students during their classroom activities, would be able to show the direct effect of student use of plastic transparencies and increased achievement test scores.

Hawthorne Effect Threats: These effects confound the results by supplying the treatment classroom students with a “special” feeling that they are more important than other students. Consider the students in the treatment classes. Most are using plastic transparencies, and they may be getting “special” treatment from teachers. The students who were not using plastic transparencies could have felt the need to “compensate” or “compete” with these students. This could have accounted for the elevated score gains in these classes. A possible solution could be to assign a “real” or placebo filter to every student in the class, making the filter a part of normal classroom operation and not placing relatively any more emphasis on the types of filter users. A double blind study could even include both types of plastic transparencies without the teacher’s knowledge of the differences between them.

Implementation Threats: Gottfredson and Gottfredson (1990) comment that the original study at the General Electric plant included special organizational training, special incentives for increasing production, and the developing of a team spirit among the Hawthorne lighting-treatment subjects. These are implementation effects that were not controlled for in the Hawthorne study. Similar problems could have occurred in this case. The use of the plastic transparencies by the treatment subjects could have had similar implementation effects: special training by the screening staff besides the initial training, team or group identity building by the treatment subjects that was not recorded, or internalized positive feedback incentives (higher scores leading to more parental praise or school incentives leading to higher incentives). there appears to be some teacher effects in these results. We could tease out these effects if we collect teacher specific data. This would allow us to glean a “true” estimate of the effect of the filter treatment.

Conclusions: This was an *ex post facto* evaluation study of the Light Sensitivity Project. Evidence was found supporting the value of the treatment. The results, however, were not completely satisfying, as the design could not completely rule out alternative explanations for the findings. Specifically, while strong effect sizes were found for the training in the presence of past achievement score levels, of student demographic variables, and of student program participation, the data also strongly suggest other mechanisms may be at work.

These include a possible biased choice of classrooms for the treatment, a possible Hawthorne effect among the filter users and non-users in the classroom, possible teacher differences, and no formal evidence of a direct link between the use of the plastic transparencies and increased scores. While the results of this evaluation were generally positive for the value of the treatment, a future study with a stronger design would yield better information.

Bibliography:

Blaskey, P. et al, (1990), “The Effectiveness of Irlen Filters for Improving Reading Performance, A Pilot study,” *Journal of Learning Disabilities*, 23.

Cardinal, D. H. et al, (1993), “Do Tinted Lenses Really Help Students with Reading Disabilities?”, *Intervention in School and Clinic*, 28: 275-279.

Fletcher, J. and G. Martinez (1994), “An eye Movement Analysis of the Effects of Scotopic Sensitivity Correction on Parsing and Comprehension,” *Journal of Learning Disabilities*, 27: 67-70.

Gottfredson, G. D. (1996), “The Hawthorne Misunderstanding (and How to get the Hawthorne Effect in Action Research),” *Journal of Research in Crime and Delinquency*, 33: 28-48.

Henk, W. A. (1992), “Scotopic Sensitivity Syndrome and It’s Treatment by Colored Overlays and Lens Filters: an Update,” in Padack, N. D. et al, eds., *Literacy research and Practice: Foundations for the Year 2000: 14th Yearbook of the College Reading Association*.

Henson-Parker, D. (1997), *See it Right!*, Assessment Manual for Scotopic Sensitivity Syndrome, Dorothy Henson-Parker, Rancho Cucamonga, CA.

O’Connor, P. D. et al, (1990), *“Reading Disabilities and the Effects of Colored Filters,”* Journal of Learning Disabilities, 23: 597-603.

Otto, W. (1994), “Tinted Polyester Overlays and Other Stories (Views and Reviews),” *Journal of Reading*. 37: 694-697.

Parker, R. M. (1990), “Power, Control, and Validity in Research,” *Journal of Learning Disabilities,* 23: 613-620.

Robinson, G. L., and R. N. Conway (1990), “The Effects of Irlen Colored Lenses on Students’ Specific Reading Skills and Their Perception of Ability: a 12 Month Validity Study,” *Journal of Learning Disabilities,* 23: 589-596.

Robinson, G. L. (1994), “Coloured Lenses and Reading: A Review of Research into Reading Achievement, Reading Strategies, and Causal Mechanisms,” *Australasian Journal of Special Education,* 18: 3-14.

Spafford, C. S., et al (1995), “Contrast Sensitivity Differences between Proficient and Disabled Readers Using Colored Lenses,” *Journal of Learning Disabilities,* 28: 240-252.

Solan, H. A. (1990), “An Appraisal of the Irlen Technique of Correcting Reading Disorders Using Tinted Overlays and Tinted Lenses,” *Journal of Learning Disabilities,* 23: 621-623.

Tyrrell, R. et al (1995), “Coloured Overlays, Visual Discomfort, Visual Search, and Classroom Reading, *Journal of Research and Reading,* 18: 10-23.

Statistical Appendix:

Table 1: Normal Curve Equivalent Values for
CTBS/4 Survey and SABE II
Students in grade 2, 1997

Total Language Battery

Mean; St. Dev.; N	1996	1997
Transparencies, Yes	**54.20;** 24.14; 10	**66.00;** 15.90;10
Transparencies, No	**60.50;** 13.65; 8	**63.75;** 18.46; 8
Not Screened	**24.72;** 25.31;140	**23.48;** 24.99; 140
Total	**24.80;** 26.75; 158	**28.21;** 27.53; 158

Total Math Battery

Mean; St. Dev.; N	1996	1997
Transparencies, Yes	**58.50;** 25.15; 10	**73.90;** 18.13;10
Transparencies, No	**56.88;** 19.98; 8	**82.75;** 15.92; 8
Not Screened	**39.10;** 22.06;140	**36.62;** 17.58; 140
Total	**41.23;** 22.82; 158	**41.32;** 21.87; 158

Total Reading Battery

Mean; St. Dev.; N	1996	1997
Transparencies, Yes	**53.00;** 23.82; 10	**66.90;** 19.87;10
Transparencies, No	**53.00;** 19.01; 8	**60.00;** 18.07; 8
Not Screened	**41.31;** 18.30;140	**36.48;** 15.62; 140
Total	**42.64;** 18.95; 158	**39.59;** 18.18; 158

Table 2: Normal Curve Equivalent Values for
CTBS/4 Survey and SABE II
Students in grade 3, 1997

Total Language Battery

Mean; St. Dev.; N	**1996**	**1997**
Transparencies, Yes	**35.24;** 21.74; 29	**52.17;** 18.78; 29
Transparencies, No	**43.55;** 16.61; 11	**54.09;** 21.46; 11
Not Screened	**27.66;** 27.20; 108	**29.34;** 26.30; 108
Total	**30.32;** 25.89; 148	**35.66;** 26.66; 148

Total Math Battery

Mean; St. Dev.; N	**1996**	**1997**
Transparencies, Yes	**45.76;** 16.64; 29	**54.83;** 20.97; 29
Transparencies, No	**52.64;** 17.44; 11	**55.09;** 23.20; 11
Not Screened	**48.61;** 20.42; 108	**50.94;** 21.72; 108
Total	**48.35;** 19.49; 148	**52.01;** 21.61; 148

Total Reading Battery

Mean; St. Dev.; N	**1996**	**1997**
Transparencies, Yes	**42.52;** 15.31; 29	**50.10;** 19.85; 29
Transparencies, No	**46.45;** 18.74; 11	**55.73;** 17.58; 11
Not Screened	**40.30;** 18.51; 108	**46.90;** 15.96; 108
Total	**41.19;** 17.92; 148	**48.18;** 16.96; 148

Independent Variables	Total Language Battery, 1997		Total Math Battery 1997		Total Reading Battery, 1997	
Lagged NCE Values	Language NCE 1996 **0.651***	Language NCE 1996 **0.645***	Math NCE 1996 **0.385***	Math NCE 1996 **0.383***	Reading NCE 1996 **0.547***	Reading NCE 1996 **0.530***
Dummy for Special Education	0.038	0.018	-0.049	**-0.091***	0.012	-0.017
Dummy for Female	0.007	0.002	-0.074	-0.082	0.037	0.032
Dummy for LEP	**-0.261***	**-0.207***	0.078	**0.176***	0.034	**0.101***
Dummy for FEP	-0.039	-0.011	0.006	0.056	0.003	0.037
Dummy for Third Grade	**0.055***	0.028	**0.158***	**0.118***	**0.244***	**0.214***
Dummy for GATE	0.057	0.038	0.063	0.037	**0.131***	**0.113***
Dummy for Title I	-0.014	-0.029	**-0.223***	**-0.241***	-0.030	-0.055
Dummy for Transparency, Yes	----	**0.164***	----	**0.231***	----	**0.162***
Dummy for Transparency, No	----	**0.088***	----	**0.223***	----	**0.153***
R^2	0.789	0.815	0.371	0.448	0.419	0.456
Adjusted R^2	0.784	0.809	0.354	0.429	0.403	0.437
F-Ratio for model	**139.24***	**130.31***	**21.92***	**23.91***	**26.75***	**24.68***
Degrees of freedom for model	297	295	297	295	297	295

Notes: **Bold*** -ed valued indicate significance at the 5% level. NCE is Normal Curve Equivalent.
JD, 12/9/97

Independent Variables	Computation Math Subtest, 1997		Concepts and Applications Math Subtest, 1997	
Lagged NCE Values	Comp NCE 1996 **0.272***	Comp NCE 1996 **0.282***	C & A NCE 1996 **0.295***	C & A NCE 1996 **0.276***
Dummy for Special Education	-0.029	-0.057l	-0.034	-0.081
Dummy for Female	**-0.135***	**-0.141***	-0.006	-0.013
Dummy for LEP	**0.124***	**0.203***	-0.038	0.053
Dummy for FEP	0.055	0.094	-0.038	0.013
Dummy for Third Grade	0.057	0.021	**0.199***	**0.168***
Dummy for GATE	0.079	0.053	0.048	0.030
Dummy for Title I	**-0.221***	**-0.233***	**-0.272***	**-0.295***
Dummy for Transparency, Yes	----	**0.222***	----	**0.183***
Dummy for Transparency, No	----	**0.122***	----	**0.281***
R^2	0.220	0.267	0.384	0.4471
Adjusted R^2	0.199	0.242	0.368	0.453
F-Ratio for model	**21.85***	**10.77***	**23.16***	**26.26***
Degrees of freedom for model	297	295	297	295

Notes: **Bold*** -ed valued indicate significance at the 5% level. NCE is Normal Curve Equivalent.