Test-immediate retest; The reliability of the Developmental Eye Movement (DEM) test

Steven L. Wilson
Pacific University

Craig L. Kuntz
Pacific University

Recommended Citation
https://commons.pacificu.edu/opt/1216
Test-immediate retest; The reliability of the Developmental Eye Movement (DEM) test

Abstract
Thirty-one 4th and 5th grade subjects were tested using the Developmental Eye Movement Test (DEM). A test-retest schedule of four trials, with two sessions of two back to back trials each, separated by one week was performed. An Analysis of Variance (ANOVA) of times for each of the four trials showed significant differences in performance times between trials during session one when compared to trials during session two (one week later). No significant difference in time was found between trials 1 and 2 or 3 and 4 (the back to back trials). There was no significant difference in errors between any trials. A plateau in performance was reached after trial 3 (first trial of second session). A learning effect is indicated requiring a retest with a one week interval between trials to determine true baseline performance.

Degree Type
Thesis

Rights
Terms of use for work posted in CommonKnowledge.

This thesis is available at CommonKnowledge: https://commons.pacificu.edu/opt/1216
Copyright and terms of use

If you have downloaded this document directly from the web or from CommonKnowledge, see the “Rights” section on the previous page for the terms of use.

If you have received this document through an interlibrary loan/document delivery service, the following terms of use apply:

Copyright in this work is held by the author(s). You may download or print any portion of this document for personal use only, or for any use that is allowed by fair use (Title 17, §107 U.S.C.). Except for personal or fair use, you or your borrowing library may not reproduce, remix, republish, post, transmit, or distribute this document, or any portion thereof, without the permission of the copyright owner. [Note: If this document is licensed under a Creative Commons license (see “Rights” on the previous page) which allows broader usage rights, your use is governed by the terms of that license.]

Inquiries regarding further use of these materials should be addressed to: CommonKnowledge Rights, Pacific University Library, 2043 College Way, Forest Grove, OR 97116, (503) 352-7209. Email inquiries may be directed to: copyright@pacificu.edu

This thesis is available at CommonKnowledge: https://commons.pacificu.edu/opt/1216
TEST-IMMEDIATE RETEST; THE RELIABILITY OF THE DEVELOPMENTAL EYE MOVEMENT (DEM) TEST

By

Steven L. Wilson
Craig L. Kuntz

A thesis submitted to the faculty of the
College of Optometry
Pacific University
Forest Grove, Oregon
for the degree of
Doctor of Optometry
February, 1997

Advisor:

Dr. Paul Kohl
Signatures:

Authors

Steven L. Wilson

Craig L. Kuntz

Advisor

Paul Kohl, OD
Biographies

Steven L. Wilson attended Arizona State University in Tempe, Arizona where he majored in Biology. He received his Bachelor of Science degree in Visual Science from Pacific University in December 1994. He is a student member of the American Optometric Association, Arizona Optometric Association, and Colorado Optometric Association. Upon graduation he plans to enter private practice with an emphasis in primary care and sports vision.

Craig L. Kuntz attended Brigham Young University in Provo Utat and The University of Nevada, Las Vegas where he majored in Chemistry. He received his Bachelor of Science degree in Visual Science from Pacific University in December 1994. He is a student member of American Optometric Association and Nevada Optometric Association. Upon graduation he plans to enter a private or group practice with emphasis in primary care, pediatrics, contact lens, and sports vision.
Acknowledgments

We would like to thank West Union Elementary School Principal, Mr. Roy Nickerson, the office staff, Mrs. Heagh-Avritt, and Mr. Beals for their generosity in allowing us to utilize the school facilities for this project. We would also like to thank the third and fourth grade students that participated. Without the cooperation of all involved, this thesis project would not have been successfully completed.

We would also like to thank Paul Kohl, OD for his wisdom and patience as our advisor. His assistance was sincerely appreciated.
Abstract

Thirty-one 4th and 5th grade subjects were tested using the Developmental Eye Movement Test (DEM). A test-retest schedule of four trials, with two sessions of two back to back trials each, separated by one week was performed. An Analysis of Variance (ANOVA) of times for each of the four trials showed significant differences in performance times between trials during session one when compared to trials during session two (one week later). No significant difference in time was found between trials 1 and 2 or 3 and 4 (the back to back trials). There was no significant difference in errors between any trials. A plateau in performance was reached after trial 3 (first trial of second session). A learning effect is indicated requiring a retest with a one week interval between trials to determine true baseline performance.

Introduction

Over the years clinicians have found that children with poor eye movement skills have increased difficulty in learning to read and in later years these problems are intensified as these skills are used when reading to learn. Studies have shown that the evaluation of eye movements can provide considerable information on how well a child assimilates visual information during the reading process. Clinicians indicate that the typical signs of poor reading skills are, regressions, omitting words, skipping whole lines of text, losing place while reading, and needing to reread material to fully understand. Poor eye movements can lead to the use of a finger for visual support while reading, excessive head movement, and/or the avoidance of reading entirely.
In optometry there are a number of methods utilized to test eye movements including bead skills, norm scored tests, and eye movement recording devices. Three frequently used tests include the Developmental Eye Movement Test (DEM), the King-Devick (K-D), and the OBER/Visagraph eye position monitoring devices.

Out of the above tests, the DEM is probably the most frequently used for several reasons. The test itself is inexpensive to purchase, with no expensive equipment required. It is also quick and easy to administer. In addition, readily available norms have been established, making it possible to quantify the level of performance of each patient when compared to the general population.

The DEM includes three sub tests in which letters are visually presented and verbally called off by the patient. Sub tests A and B present numbers in vertical columns which test automaticity of number-naming (requiring very easy eye movements). In sub test C, numbers are presented in a horizontal fashion. This requires saccadic eye movements which are identical to those utilized for reading.

The DEM is not only used to diagnose eye movement deficiencies, but is also used to assess the efficiency of optometric vision therapy by comparing pre and post therapy scores.

A previous study entitled, "Test - Retest; The Reliability Of The DEM, King Devick, and Visagraph" was completed in May 1994 by Dawn M. Clary and Joseph R. Peters. Their study was designed to determine if three different eye movement tests would produce repeatable results when administered on three separate occasions one week apart. Their study indicated a statistically significant improvement in performance
between the first and second as well as between the first and third administrations with no intervening vision therapy administered. This indicated that test familiarity or a learning effect may play a role in the improvement of test performance. Since significant difference in performance was found between the 1st and 2nd test administrations, when separated by one week interval, with no difference between 2nd and 3rd administrations, also separated by a week, it seems that the practitioners are obliged to retest patients at least twice to get a “true” result. If this is so, do we also have to wait the intervening week to do the retest, to reach a plateaued performance, or can the test-retest be done back to back during the same test session?

To answer this, we administered two DEM tests in a test-immediate retest fashion on the first day with a second administration of the two tests given one week later (4 administrations in total). In doing so, we will be able to reevaluate the test-retest reliability of the DEM and whether or not a learning effect is present regardless of the testing schedule.

Methods

Our study involved 31 subjects between the ages of nine and eleven years. Permission was obtained from each subject’s parent or guardian for admission into our study. We screened each student to assess near visual acuity and ocular alignment via a cover test at both near and far. The Developmental Eye Movement Test was administered to each subject a total of four times with back to back administrations one week apart.
Testing was performed at the subject’s school in a small multi-purpose room. The room had standard over-head fluorescent lighting and was free from significant distractions. Each subject was given a pretest of ten single-digit numbers presented in a horizontal fashion placed at standard reading distance. The subject was asked, “Do you see this row of numbers?” (Motioning with finger from left to right). “Please read these numbers out loud for me.” If the numbers were correctly called within twelve seconds, the subject was given the DEM series of tests.

The DEM was administered in the same order for each subject, starting with the vertical sub tests A and B. The student was seated comfortably at a table and the following instructions were given, “I want you to carefully read the numbers down the two columns like this as quickly as you can.” We pointed to the top of the first column and motioned with our finger the direction in which the numbers were to be read. We instructed the subject to use their eyes only and to not use their fingers to keep their place. We then asked them if they understood the instructions. If they did, we started the test. After test A, without any significant delay, we proceeded with test B. An identical instruction set was given. The instruction set for test C was similar except the subject was told, “I want you to carefully read the numbers across the row like this as quickly as you can.” We told each subject to call out each number as quickly and as carefully as possible. When the subject understood these instructions, the test was administered. A second administration of the DEM was given in the same fashion as the first with a 1 to 3 minute interval between tests. The times and errors for each sub test were recorded immediately following each administration.

One week later, the same procedure was repeated with a third and fourth administration of the DEM test. All instructions and conditions remained the same as in the previous week.
The results of all four trials were compiled on an Excel spreadsheet and imported into the Statview 512+ program. The data was then analyzed using repeated measures analysis of variance (ANOVA), comparing all possible trial combinations. A Scheffe-F test with .90 level of significance was used for post hoc analysis. Only data from subjects present at all four trials was included in the analysis.

Results

An Analysis of Variance (ANOVA) was performed for the horizontal time, number of errors, and percentile ranking for the group as a whole. The horizontal time showed a significant difference between trials 1 and 3, 1 and 4, 2 and 3, and 2 and 4. No significant difference was found between 1 and 2 or 3 and 4. There was no significant difference in the number of errors between any of the trials. The percentile ranking of the subjects reveal a significant difference between trials 1 and 2, 1 and 3, 1 and 4, 2 and 3, and 2 and 4. No significance difference was found between 3 and 4. (See figures 1 and 2 and tables 1 and 2 for descriptive statistics and Scheffe F-Test results.)

After reviewing the statistics of the whole group, it appeared there may be a difference in test performance from one trial to another when comparing those who scored above the 50th percentile with those below the 50th percentile. Therefore, an ANOVA for each group was performed.

The ANOVA for the 14 subjects that scored below the 50th percentile showed a significant difference in horizontal times between trials 1 and 3, 1 and 4, and 2 and 3. No significance was found between 1 and 2, 2 and 4, and 3 and 4. There was no significant difference in the number of errors between any of the trials. Percentile
ranking indicated a significant difference between 1 and 3, 1 and 4, and 2 and 3, but no significant difference between 1 and 2, 2 and 4, and 3 and 4.

The ANOVA for the 17 subjects that scored at or above the 50th percentile showed a significant difference in horizontal times between trial 1 and 2, 1 and 3, 1 and 4, 2 and 3, and 2 and 4. No significant difference was found between trials 3 and 4. Again, no significant difference was found in the number of errors between any of the trials. Percentile rankings displayed a significant difference between trials 1 and 3, and 1 and 4, with no significant difference between 1 and 2, 2 and 3, 2 and 4, and 3 and 4.

Conclusion

In analyzing the data for the group as a whole, horizontal times showed no significant difference on tests given back to back on the same day (trials one and two or three and four). However, all tests that were given on separate days did show a significant difference when compared to each other (trials one to three, one to four, two to three, and two to four). This indicates to the practitioner that there is no benefit to administering the test in the test-immediate retest fashion. After the third test, we saw a performance plateau effect indicated by horizontal mean test times of 44.6, 42.9, 39.5, and 39.6 seconds for each of the four sequenced trials. This shows that when giving the tests on separate days, a learning effect apparently occurs. This could indicate that the assimilation of information on how the test is done occurs over a period of time and is not apparent immediately. The numbers of errors did not change between any of the trials. Thus test-retest does not improve accuracy of performance. Baseline for accuracy is achieved at the first test. The percentile ranking data indicates that there is improvement in performance level when comparing all trials.
except 3 and 4 which are approximately equal. The mean percentiles for the four trials are; 54.2, 62.0, 71.2, and 70.0. The greatest change was between trials 2 and 3 (see figure 2). Again, a plateau effect was present after the third trial. It is apparent that a test-retest, with a one week interval, is required to arrive at results which closely approximate the patient's peak performance capabilities without any intervening therapy. Knowing this level can help practitioners more accurately determine the patient's baseline prior to implementing a program of vision training.

The data for the group of 14 that scored below the 50th percentile showed a significant difference in horizontal mean times between trials 1 and 3, 1 and 4, and 2 and 3, with no significant change for trials 1 and 2, 2 and 4, and 3 and 4. The mean times for the four trials were, 51.0, 49.9, 45.2, and 47.1 seconds. It appears that the changes in times are greatest between tests that were performed on separate days. The mean percentile rankings were; 28.2, 38.2, 51.4, and 44.3, which mirror the time changes shown for this group. The lack of significant percentile and time changes between trials 2 and 4 and 3 and 4 may indicate a possible fatigue factor or a lack of interest on the subject's part. In addition, when looking at individual subject data, there appears to be a larger variability in performance between trials within this group when compared to the group as a whole. This group went from the bottom quartile on the first trial to almost the 50th percentile on the third and fourth trials. These findings support the assumption that children who have learning or attention problems may also have more variable results on the DEM, and thus retests for this population may be even more important. Without any intervening therapy, this group experienced an increase in performance with a plateau effect by trial three, the same as seen in the group as a whole. This further supports the premise that the administration of the DEM on a separate occasions may give more accurate baseline information for those subjects who do poorly on the DEM.
In the group of 17 that scored above the 50th percentile, the horizontal times changed significantly between all trials except 3 and 4. The mean horizontal times for trials one through four were; 39.3, 37.1, 34.9, and 33.3 seconds. This group shows a steady improvement with each subsequent trial with the same plateau effect by trial 3 as with previous groups. However, this group exhibited less variability between trials and as a whole experienced a more consistent improvement in performance as indicated by a lower average standard error score than the below 50 percentile group (see table 1). There was no significant difference in percentile ranking among any of the groups except groups 1 and 3, and 1 and 4. The mean times for the four trials were; 39.3, 37.1, 34.9, and 33.3 seconds. The lack of significant difference between the majority of the trials may be due to the higher initial level of performance of this group indicating that they were already performing close to their peak ability. The significant improvement that did occur between trials 1 and 3, and 1 and 4 was probably due to initial test jitters, test familiarity over time, and the repetition of test taking at a later date. This finding is consistent with previous findings that by the third trial the performance level of the group begins to plateau near their peak performance level.

From our results and those of the Clary and Peters study, we recommend in all cases in which the DEM is used to assess eye movements, that it be performed at least twice with a one week interval between trials. It appears that this interval allows the subject to perform closer to their peak capabilities. Back to back administrations, however, do not appear to factor out initial test jitters or allow time for learning to occur. Future studies should investigate whether or not this one week interval could be shortened to possibly one day and still allow the learning process necessary for peak performance to occur. Obtaining true baseline information more quickly will allow the clinician to implement a program of vision training (when necessary) in a more expeditious manner.
Figure 1  Mean Horizontal Times

![Graph showing Mean Horizontal Times across different trials and groups.](image)
Figure 2  Mean Percentile Rankings

- Trial 1
  - Percentile Rank All Groups: 28.2
  - Percentile Rank < 50% Group: 54.2
  - Percentile Rank > 50% Group: 75.6

- Trial 2
  - Percentile Rank All Groups: 38.2
  - Percentile Rank < 50% Group: 62
  - Percentile Rank > 50% Group: 81.6

- Trial 3
  - Percentile Rank All Groups: 51.4
  - Percentile Rank < 50% Group: 71.2
  - Percentile Rank > 50% Group: 87.4

- Trial 4
  - Percentile Rank All Groups: 44.3
  - Percentile Rank < 50% Group: 69.9
  - Percentile Rank > 50% Group: 90.9
Table 1  Summary of Results

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Age</th>
<th>Trial Number 1</th>
<th>Trial Number 2</th>
<th>Trial Number 3</th>
<th>Trial Number 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9.10</td>
<td>52.83</td>
<td>35</td>
<td>1</td>
<td>68.03</td>
</tr>
<tr>
<td>2</td>
<td>11.3</td>
<td>46.44</td>
<td>30</td>
<td>0</td>
<td>45.75</td>
</tr>
<tr>
<td>3</td>
<td>9.9</td>
<td>58.13</td>
<td>26</td>
<td>5</td>
<td>65.75</td>
</tr>
<tr>
<td>4</td>
<td>10.1</td>
<td>50.13</td>
<td>35</td>
<td>6</td>
<td>45.75</td>
</tr>
<tr>
<td>5</td>
<td>10.6</td>
<td>46.97</td>
<td>45</td>
<td>0</td>
<td>40.35</td>
</tr>
<tr>
<td>6</td>
<td>11.1</td>
<td>50.93</td>
<td>15</td>
<td>1</td>
<td>45.88</td>
</tr>
<tr>
<td>7</td>
<td>10.3</td>
<td>49.74</td>
<td>35</td>
<td>1</td>
<td>44.75</td>
</tr>
<tr>
<td>8</td>
<td>11.0</td>
<td>47.96</td>
<td>25</td>
<td>2</td>
<td>39.06</td>
</tr>
<tr>
<td>9</td>
<td>10.3</td>
<td>62.35</td>
<td>10</td>
<td>1</td>
<td>63.72</td>
</tr>
<tr>
<td>10</td>
<td>11.0</td>
<td>45.99</td>
<td>20</td>
<td>0</td>
<td>52.82</td>
</tr>
<tr>
<td>11</td>
<td>9.8</td>
<td>48.62</td>
<td>45</td>
<td>0</td>
<td>45.63</td>
</tr>
<tr>
<td>12</td>
<td>10.2</td>
<td>53.79</td>
<td>30</td>
<td>1</td>
<td>51.94</td>
</tr>
<tr>
<td>13</td>
<td>11.6</td>
<td>47.00</td>
<td>25</td>
<td>0</td>
<td>47.5</td>
</tr>
<tr>
<td>14</td>
<td>11.6</td>
<td>50.96</td>
<td>20</td>
<td>7</td>
<td>45.2</td>
</tr>
</tbody>
</table>

Mean | 51.02 | 28.21 | 1.57 | 49.34 | 38.21 | 1.43 | 45.20 | 51.43 | 1.36 | 47.13 | 44.29 | 2.71 |
Std. Dev. | 4.50 | 10.30 | 2.17 | 9.63 | 23.42 | 1.34 | 8.36 | 24.53 | 1.78 | 8.84 | 27.09 | 3.56 |

Table 2 - ANOVA (Scheffe F-Test)

<table>
<thead>
<tr>
<th>Subjects</th>
<th>Horizontal Times</th>
<th>Percentile Ranking</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 50% Yr. Mo.</td>
<td>Trial 1 vs. Trial 2</td>
<td>0.262</td>
<td>1.585</td>
</tr>
<tr>
<td>Trial 1 vs. Trial 3</td>
<td>7.593*</td>
<td>8.598*</td>
<td>0.034</td>
</tr>
<tr>
<td>Trial 1 vs. Trial 4</td>
<td>3.352*</td>
<td>4.121*</td>
<td>0.953</td>
</tr>
<tr>
<td>Trial 2 vs. Trial 3</td>
<td>5.035*</td>
<td>2.766*</td>
<td>0.004</td>
</tr>
<tr>
<td>Trial 2 vs. Trial 4</td>
<td>1.762</td>
<td>0.568</td>
<td>1.205</td>
</tr>
<tr>
<td>Trial 3 vs. Trial 4</td>
<td>0.84</td>
<td>0.841</td>
<td>1.344</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subjects &gt; 50%:</th>
<th>Horizontal Times</th>
<th>Percentile Ranking</th>
<th>Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 1 vs. Trial 2</td>
<td>2.711*</td>
<td>1.876</td>
<td>0.134</td>
</tr>
<tr>
<td>Trial 1 vs. Trial 3</td>
<td>11.03*</td>
<td>7.15*</td>
<td>0.953</td>
</tr>
<tr>
<td>Trial 1 vs. Trial 4</td>
<td>20.042*</td>
<td>12.056*</td>
<td>0.73</td>
</tr>
<tr>
<td>Trial 2 vs. Trial 3</td>
<td>2.808*</td>
<td>1.7</td>
<td>0.372</td>
</tr>
<tr>
<td>Trial 2 vs. Trial 4</td>
<td>8.011*</td>
<td>4.418*</td>
<td>0.230</td>
</tr>
<tr>
<td>Trial 3 vs. Trial 4</td>
<td>1.334</td>
<td>0.637</td>
<td>0.015</td>
</tr>
</tbody>
</table>

* Significant at 90%
References

