Spotlight on the Developmental Eye Movement (DEM) Test

Abstract: The developmental eye movement (DEM) test is a practical and simple method for assessing and quantifying ocular motor skills in children. In this review, a summary of the literature relevant to the DEM test has been made, its psychometric properties and its pros and cons have also been considered. The DEM test provides clinicians with a simple method of measuring eye movement using a psychometric test. Over the years, many studies have enabled the identification of the strengths and weaknesses of this test apart from outlining the psychometric properties. The validity of the test has been checked and expanded over time and studies have shown that the DEM test measures an aspect of eye movement related to reading, rather than purely parameters associated with eye-movement. Some reservations have emerged regarding the repeatability of the test because a degree of learning effect emerges over multiple sessions. Being aware of this point allows correct clinical application and interpretation of the test. Normative data in children were available for nine languages and countries. So far, DEM test could be applied clinically in each case when a rapid test of eye movement was required, such as in testing for vision-related visual problems.

Keywords: DEM test, eye movement, psychometric test, normative data, reading

Introduction
During the process of reading, the eyes change their position quickly in a series of saccades. These alternate with fixations in which the information is acquired. Several studies have demonstrated that learning disabled (LD) and generally poor readers may perform inadequate ocular movements, producing more regression and fixation, and additional, although short, left to right saccades.

With these indications, clinicians require a practical and basic instrument to assess ocular movement skills during reading in a simple way as an alternative to eye-tracking instruments. The DEM test (Figure 1) is a practical paper-based psychometric test designed for the assessment of ocular movement in a reading-like condition. This review outlines a summary of the recent literature, together with giving consideration to the classic studies that are the key to understanding its application, the psychometric properties, and pros and cons of the DEM test.

Eye Movement Testing
The examination of eye movements can be performed by using different techniques, adapted according to the age, the aims of evaluation and the time required. Each method has its pros and cons. A simple summary of all the possible methods for ocular movement testing is presented in Table 1.
The history of the psychometric measurement of eye movements originates from the Pierce test, which was composed of four cards, one pre-test and three test cards. In this test, a subject was required to perform saccades from a column of numbers on the left to one on the right, while naming the numbers. The pre-test and the first cards were used as a training phase. The execution time was measured, and it was compared to the norms supplied with the test.

The King Devick (KD) and the NYSOA K-D tests, are tests similar to that of Pierce, with subjects required to name the number quickly. The KD differs from the Pierce test, in that the numbers were arranged in a reading-like condition. The evolution of Pierce to KD is in the arrangement of the numbers, from left-to-right equal-space saccades to irregular and unpredictable horizontal saccades for reading.

Table 1 An Outline of the Available Methods for Measuring Eye Movements

<table>
<thead>
<tr>
<th>Method</th>
<th>Pros</th>
<th>Cons</th>
<th>Test or instrumentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct clinical observation</td>
<td>Low cost</td>
<td>Subjective evaluation</td>
<td>NSUCO Ophthalmoscopy</td>
</tr>
<tr>
<td></td>
<td>Speed</td>
<td>Interexaminer repeatability</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Easy method to assess fixations</td>
<td>Difficulty of scoring</td>
<td></td>
</tr>
<tr>
<td>Indirect or psychometric test</td>
<td>Evaluation of performance in a</td>
<td>Verbal response (not all)</td>
<td>King Devick</td>
</tr>
<tr>
<td></td>
<td>standard task</td>
<td>A high degree of collaboration</td>
<td>DEM</td>
</tr>
<tr>
<td></td>
<td>Speed</td>
<td>Interference of other cognitive function</td>
<td>Groffman Visual Tracing</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Visual search tasks</td>
</tr>
<tr>
<td>Eye-tracking systems</td>
<td>Objective measurement</td>
<td>Expensive</td>
<td>Limbus reflection eye tracker</td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>Long-time of execution</td>
<td>Purkinje image tracker</td>
</tr>
<tr>
<td></td>
<td>Accuracy</td>
<td>Calibration</td>
<td>Pupil or reflected corneal image eye tracker</td>
</tr>
<tr>
<td></td>
<td>Non-invasive</td>
<td>Head artefacts</td>
<td></td>
</tr>
<tr>
<td>Electrophysiological systems</td>
<td>Non-invasive, Cost</td>
<td>Noise, artefact, Uncomfortable and technically difficult (EMG)</td>
<td>Electro Oculography (EOG)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Electro Ocular Myography (EMG)</td>
</tr>
</tbody>
</table>

Note: Modified with data from Leigh and Zee. Abbreviations: NSUCO, Northeastern State University College of Optometry oculomotor test; DEM, developmental eye movement test.
However, in both of these tests, there is the problem of the influence of baseline naming in the total time of execution. The overall performance could be slowed both by eye movements and/or the naming process. Consequently, the DEM test was developed because it was realized that naming deficit represents an important problem in dyslexic and learning disabled children. The vertical time and the calculation of ratio, together with the four clinical response types, could differentiate naming problems from oculomotor disorders.

The DEM Test

The Developmental Eye Movement (DEM) test is a psychometric paper test for the assessment of eye movement performance in children. The quantification of eye movement is performed by naming numbers in a reading-like condition without sophisticated instruments. The task consists of naming numbers in a reading-like condition without using a complex instrument.

The DEM test is composed of a pre-test and three test cards: A, B, and C. The pre-test card is composed of 12 horizontal numbers intended to assess the basic skills for completing the test. The first two cards (A and B) comprise two columns of 20 numbers for each card. Card C is composed of the same 80 numbers from the A and B cards but arranged in a horizontal pattern, similar to a reading text. The subjects tested were usually seated in a comfortable position with an inclined lectern positioned at 40 cm, and they were subsequently presented with the different cards. The subjects were required to read the numbers on the different cards as fast as possible. For the vertical cards A and B, the participants had to read aloud the two columns of numbers vertically, while for card C, the participants had to read aloud the same 80 numbers horizontally in a condition analogous to reading. The time of execution for all cards and the errors made for card C were recorded. Subsequently, two scores were calculated. The Vertical Time (VT) is calculated by adding the times for the two cards A and B. The Adjusted Horizontal Time (AHT) is the time for card C, adjusted for addition or omission errors. The Ratio score is calculated by dividing the adjusted horizontal time by the vertical time. The total number of errors is the sum of the different errors made for card C. The VT represents the time spent naming 80 numbers, while the AHT reflects the total time taken to name numbers plus the time to perform saccadic eye movements. The Ratio is the main measure used to evaluate ocular movement performance. Each score was successively compared with the specific language normative data for the detection of number naming or eye movements problems, or a combination of the two.

Application of DEM Test

Who Needs to Use the DEM Test and Why?

As mentioned earlier, compared to Pierce and KD, only the DEM test takes into account the naming skills and, for this reason, the target population of this test is children without, with, or with suspected learning disabilities aged between 6 and 13 years. However, the age at which the test could be applied depends primarily on the existing norms. The different norms that are available in different languages have been limited in higher ages from 11 to 18 years old. Application outside these norms has been possible, but while being aware of the limitations of the evaluation. Taking into account the higher age available and the percentile scoring, it is also almost possible to detect the severe cases of naming and oculomotor problems in adults. Ratio score was demonstrated as a predictor of an oculomotor dysfunction through the complete life span. In fact, a modified version of the DEM test shows a mean ratio of about 1.04 (SD 0.10), for participants aged between 14 and 68 years, without differences by year groups.

Studies that investigate visual problems have shown that 62% of developmental dyslexia or learning disabled (DD/LD) children present oculomotor problems on application of the DEM test, compared to 15% of those whose development was typical. In the neurological population with acquired brain injury (ABI) or traumatic brain injury (TBI), the DEM test has proven to be useful in the assessment of baseline behavior and or specific oculomotor treatment, other than the screening capacity in a baseline clinical practice. Children with developmental coordination disorder (DCD) showed mild atypical performance in the DEM test.

The DEM test has no formal restriction to its use, and it could be performed by different professionals involved in the care of learning-disabled children. The latest manual reports that the test could also be administered by classroom and special education teachers, reading specialists, and school psychologists rather than only by vision specialists. However, the personnel applying the test need to be aware of its standardized administration, and
detailed interpretation of the results could only be performed by professionals who have a complete understanding of the vision and cognitive status of the child.\textsuperscript{21} In these cases, a multidisciplinary team is preferable for addressing the different clinical manifestations of LD problems.\textsuperscript{25} However, the different legislation and professionals involved in various countries make a unique statement very difficult to perform.

**What We Can and Cannot Do with the DEM Test?**

Some procedures, such as gross observations of eye movements, have found a relationship with reading disability.\textsuperscript{26,27} Compared to the DEM test, the gross observation of eye movements is considered to have a different accuracy (see validity paragraph below). The NSUCO (Northeastern State University College of Optometry)\textsuperscript{26} ocular motor test is a standardized method and scoring of observational eye movement testing (saccades and pursuit). A comparison between the NSUCO accuracy scale and the DEM test in a group of children did not find a significant relationship between these tests because they measure oculomotor movements in a different way and with different accuracy.\textsuperscript{28} The DEM test is one that is designed for oculomotor movement in a specific reading-like condition and a simple exterior validity confirms this point. The DEM test was constructed with the aim of testing oculomotor movement in a reading-like condition in children, and the VT score takes into account the naming skill and allows the correct detection of problems. Based on the results of the four scores (VT, AHT, Ratio, and Errors), four possible classifications can be made using the cut-off score of the 30th or 15th percentiles.\textsuperscript{13,29–31} Normal VT, AHT, and Ratio (in percentiles) express the behavior of a normal subject. Low VT, low AHT, and normal Ratio express the behavior of a naming problem. Normal VT, low AHT, and low Ratio express an oculomotor disorder and, a low VT, a low AHT, and a low Ratio express comorbidity of naming and oculomotor problems. Despite the four types of problems that could be found being clear, a continuum evaluation of each subtest using percentile scores may be the best way to express this result.\textsuperscript{30}

Conversely, the DEM test cannot detect reading problems or other conditions for different reasons.\textsuperscript{32} Even if naming numbers does express one part of language processing,\textsuperscript{33} it does not represent the final process of reading, which is more complex.\textsuperscript{34} The reading process could only be tested by direct reading of the sequence of words/non-words, sentences, and paragraphs. The different reading levels need to be adapted to the level of development and skill acquisition of the child. Specific reading tests were available for each specific step of developing reading, for example, three for each year from 2nd to 5th grade.\textsuperscript{35} For example, the vertical numbers of the DEM test are similar to the subtest 1B of the DDE battery.\textsuperscript{36} The DDE battery is an Italian test for the assessment of reading and it is aimed at assessing different aspects of reading skills. Since reading batteries are very different from each other and also language specific, other tests in particular languages could have included vertical number reading as a component of the examination. In summary, vertical number naming represents only one of the aspects of the multi-facetted process of reading.

A study that investigates the relationship between the diagnosis of dyslexia and the results of DEM found no significant relationship.\textsuperscript{32} Not all dyslexic children (and even more learning disabled) are equal. Even though the diagnosis was the same (eg based on DSV IV-TR; DSM-V or ICD-10-CM), there are different cognitive skills profiles for every single case.\textsuperscript{11} As a result, there are dyslexic children with normal oculomotor movement and those with poor oculomotor movement.\textsuperscript{21,27,37–39} This point explains the lack of a strong relationship between these two aspects and, at the same time, underlines the need to test the oculomotor movements because they could be normal or poor in each dyslexic child. Another source of variability between children is the fact that not only pathological dyslexic children need testing and care but also poor readers. This is a problem that is also defined with reading difficulties: child below normal range but not severe enough to be diagnosed as having learning disabilities. Finally, specific LD is not represented only by dyslexia but also by dysgraphia, dysorthographia, and dyscalculia,\textsuperscript{40,41} making the heterogeneity between cases more widely present and easy to detect. DEM test was also used in testing the oculomotor behaviour of DCD children.\textsuperscript{24}

**Psychometrical Properties of DEM Test**

The clinical usefulness of specific tests needs to be determined not subjectively by clinicians but it should be based on results of scientific studies on validity, repeatability, normative data, and other measurement properties.\textsuperscript{42,43}
Validity

Validity expresses the ability of the test to measure what it is supposed to measure. There is no unique form of validity, and even recent studies have attempted to put the different aspects of validity into the construct of validity, going over the boundary between content and criterion. This point could be obtained by using a discrete amount of evidence that represents a check of the construct.

The validity of the DEM test has been tested several times and in several ways over the years. Initially, the first DEM study reported four methods that were originally used to check validity: raw scores and chronological age, internal consistency, relationship with an achievement test, and the relationship with education level. The first analysis showed that DEM scores improve with age, and this result was subsequently confirmed and evident in all normative studies.

The second aspect is internal consistency. This aspect is peculiar and requires a deeper analysis. Garzia et al reported that most internal correlations were significant and the lack of significance between VT and Ratio is a sign of the independence of the two factors. Subsequently, a replication study confirmed the previous results using partial correlation (age removed). In addition, an exploratory factor analysis was also performed. Exploratory factor analysis is a statistical method used to identify the underlying relationships between measured variables. The results support the structure of the test and the previous internal correlations, showing saturation to three main factors. The first factor is highly related to VT and AHT and represents the naming component. The second factor is related to AHT and Ratio and could represent the overall process of eye movement. The third factor is related only to Errors and represents this score.

The relationship with reading and/or achievement tests, as shown by Garzia et al, is complex and it does not have a simple and direct explanation. The purpose of this relationship in terms of an indication of validity concerns the fact that LD children diagnosed with reading tests perform worst in the DEM test. A priori, this seems to be legitimate and different results show a trend in this direction. However, over time there is evidence that not all LD children present deficits in ocular movements, since they have a different cognitive profile. Therefore, the general comparison between LD and non-LD children in DEM results and their correlation does not perfectly represent an indication of validity. However, the original study reported a significant difference in all aspects of DEM between lower-achieving and TD children. A comparison between DD and TD children using different reading tests showed differences in the VT and AHT scores, but not in Ratio and Errors. Taking into account the correlation between tests, the AHT showed a positive relationship (partial correlation, age removed) to a test of reading words. A large relationship was found between the time required to read a series of words and AHT (r = 0.65 p < 0.01). A subsequent study has confirmed this result. VT and AHT are good predictors of academic performance and reading rate. Consequently, VT and, in particular, AHT itself can still be useful in clinical settings when a quick evaluation is required.

After taking into account the convergent and divergent validity with other tests, only a moderate relationship was found with AHT (r = 0.29) for comparison with a test of visual exploration. Conversely, the relationship with the NSUCO oculomotor test (accuracy score) was low and not significant. Both test eye movement but at two different levels.

The final step of validity seems to be reached in relating DEM to objective eye movement. At least four articles have reported this analysis. In the first study, the authors found no significant correlation between DEM test performance and specific eye movement parameters such as peak velocity, gain, latency, and corrective saccades. Conversely, they did find a significant correlation between reading and visual processing speed. They concluded that DEM may be useful and should retain its diagnostic role in the clinical practice. However, they did not take into account the total number of saccades performed during the test, which is the most simple indication of the quality of eye movement.

The second study investigated the relationship between DEM results, eye movement parameters, and standardized reading achievement scores in normally-sighted children. The results showed that there was a positive and significant correlation between AHT and the number of fixations (0.41), and between AHT and span recognition (−0.42). In addition, it was found that there was a significant correlation between reading rate and VT (−0.41), and between AHT and reading rate (−0.55). It is necessary to emphasise that eye movement recording was performed using standard text and not during the DEM test.
The third study specifically recorded eye movement during the execution of DEM in DD, DD age-equivalent, and TD children, other than in reading tests. The results showed a significant correlation between DEM C and reading test but, interestingly, a positive and significant relationship between mean duration of fixation and time of execution of DEM C in each group tested. The authors concluded that the DEM test is a useful tool to test eye movement abilities in dyslexic children. They also explained that the lack of a relationship with saccades is to be expected because the participants had not read any text.\(^{38}\)

The authors of the fourth study have performed deep eye tracking measurement and analysis during the execution of a digital version of DEM. The authors also added a test of visual processing speed. After analysis of each specific eye movement parameter, they concluded that the number, the amplitude, and direction of the saccades were related to DEM performance on the horizontal, but not to the vertical subtest. Conversely, the time spent on the execution of saccades was small compared to the time spent on fixations. The duration of fixation was positively correlated with performance on vertical and horizontal DEM subtest, and with visual processing speed. In summary, the DEM test is not a saccade test, but expresses a whole score of a set of eye movement parameters, together with visual verbal and visual processing speed skills.\(^{55}\)

Broadly speaking, DEM seems not only to be measuring saccades, but a general oculomotor behavior. This is because the underlying process seems not only to be linked to the saccades themselves, but also to fixations in which saccade preparation is made.\(^{38,48,54,55}\)

There is a minor point in question regarding the (face) validity of the DEM test which is related to the lack of correction for errors in vertical time. This problem arises from a mistake in the first edition of the manual and relies on the difference between the reported scoresheet and the description given in the manual. This was clarified in the second version of the manual. In the first edition, there is space in the scoresheet to calculate and report the adjusted vertical time, even though elsewhere in the manual and in the associated article that was published it was highlighted that the adjustment of vertical time should be avoided because errors in this part of the test are very low. This has also been reflected in different interpretations between reported norms because some use unadjusted time\(^{5,16,30}\) whilst others use adjusted VT.\(^{15,56}\) However, there are few errors performed in VT and this only became relevant for very inaccurate participants with a severe naming problem.

For adults, a specific variation of DEM, the A-DEM was developed. The A-DEM was constructed with the aim of increasing the cognitive visuo-verbal retrieval demand.\(^{20}\) However this is contrary to the basic principle of reducing the cognitive demand on the vertical test. Consequently, the primary goal of the original test was not achieved. Nevertheless, the results of A-DEM showed that some parameters including the Ratio were consistent between age groups. Finally, a panel conference concluded that the DEM was preferable to the A–DEM.\(^{23}\)

**Norms**

Since the DEM test is a language-specific test and there is a different learning curve for each country which depends on socio-demographic, political, and cultural differences, the normative data in the first years of school become important. Normative values are actually available for nine languages and countries: English,\(^{3}\) two for Spanish,\(^{14,27}\) Cantonese,\(^{15}\) Japanese,\(^{56}\) Portuguese,\(^{58}\) Italian,\(^{30}\) Mandarin,\(^{59}\) Latvian,\(^{16}\) and French\(^{60}\) languages.

Are the norms for each language similar or different? Some comparisons that have been made with seven of them showed that some are different, particularly for the younger group and in the lower percentile rank.\(^{15,30,58,59}\) Coherently, each specific norm has to be used for each language. Conversely, other direct comparisons did not find differences.\(^{14,57,60}\)

In summary, since some specific norms were different and others appeared similar to the original American ones, those for DEM seem to be related to differences in languages, cultures, and education systems. Therefore, specific norms need to be used in order to avoid errors in the classification of results.

For adults, the situation is relatively complex. Adults are interesting because they have a higher risk of acquired brain injury with the need for an evaluation of eye movement,\(^{19}\) other than for adult dyslexics who have special needs. Some studies have taken into account this age group but without recognizing normative data.\(^{22}\) This is reflected in the requirement for adult norms for the DEM test.\(^{17,19,23}\) However, the results of A-DEM showed that some parameters, including the Ratio, were constant between age groups. Based on this information, higher tier values could be used in the absence of specific norms.
Repeatability

The two terms reliability and repeatability have been used in the same way in the case of the DEM test, even though they themselves express different concepts.61 In test-retest studies, ICC or others correlations were used for repeatability. In combination with this approach, the agreement analyses and Bland Altman plot and calculation were used and the mean bias between sessions compared.

Garzia et al12 showed that the DEM test presents the following correlation coefficients: (Pearson r): r = 0.89, p <0.001; r = 0.86, p <0.01; r = 0.57 p <0.05 and r = 0.07 n.s. respectively for VT, AHT, Ratio, and errors. These data showed that the DEM test has good test-retest reliability for vertical and horizontal time, but medium for Ratio, and low for Errors. A subsequent study has substantially replicated these results.62 Lower results were found by Rouse et al63 in testing 3rd-grade children. It was found that VT and AHT presented fair to good repeatability, whereas the Ratio score was poor. Limits of agreement were added. Tassinari and DeLand29 found that the correlation coefficients were higher than those previously found. They added the comparison of the pass-fail classification and good agreement was found between test and retest.

Orlansky et al31 performed a comprehensive evaluation of reliability. Subjects were tested three times in two sessions using parallel versions of the DEM test. Correlation coefficients showed fair to good correlation for the VT and AHT scores and poor results for Ratio and Error scores. Interestingly, they found general improvements in performance in all test scores in the retest session. Coherently, they suggested being careful in the use of DEM results for diagnosis or the assessment of treatment results.

Taking into account all previous observations, Facchin and Maffioletti28 performed another test-retest comparison, using the clinical version of the test, and not the parallel form of DEM used in the previously cited study.31 They found high values for VT and AHT, medium to high for Ratio, and medium for Errors. Conversely, they confirmed the presence of a significant improvement in performance (mean bias) in the second session for VT, AHT, and Ratio and provided the minimum change in percentile for a significant result. However, the so-called learning effect (an improvement in the second session) was lower than was previously found.

In summary, the DEM test shows high correlation values but conversely poor agreement between sessions due to an improvement of performance over session. Even though this learning effect appears clear in different studies, the second edition of the manual gives a solution. Richman suggests a retest of the vertical and/or horizontal part in the case of a low score.

This problem of agreement affects not only repeatability itself but also normative data. The suggestion of Richman is clinically correct, but it is necessary to take into account the fact that normative data were built for the first administration. The last clinical suggestion is to take into account that this limitation and interpretation of results (and improvement after session or therapy) has to be performed with prior knowledge of this problem of agreement.

This learning-effect problem also affects the King Devick test because it is very similar.64 The repeatability and agreement of the KD test show the same general behavior. Correlations (ICC) were high, but there is a significant improvement in performance in the second administration (reduced time of execution). The authors concluded that learning effects occur for the familiarity with the test. They recommended limiting the comparison of the results that should be made with the individual baseline performance, rather than with normative data.

Final Remarks

The selection of an appropriate instrument must take into account its strengths and limitations and also with awareness these points.42 Different instruments have been developed for a specific purpose, but this could evolve with time and experience. The DEM test has demonstrated its use in developmental-age applications, compared to King Devick which has been used in stroke, sport, and concussion.65–67 Nevertheless, since research and application cannot be limited, these differences in clinical use have become less significant,22,68 and DEM and KD could coexist in the same field of application.

Future Direction

Based on these considerations, the future use of the DEM test could involve stroke and other neurological patients.69,70 For these reasons, there is a requirement for specific adult norms and the investigation of the application and usefulness in these specific deficits.71 In longer term, taking into account the Scheiman and Rouse three-component model of vision,72,73 the DEM test could not only establish its position but also become part of a wide psychometric battery of eye movement tests together with other tests, such as NSUCO, Groffman visual tracing74 and others.
Disclosure

The author reports no conflicts of interest in this work.

References
