ARTICLE

Validity Reassessment of Developmental Eye Movement (DEM) Test in the Italian Population

Alessio Facchin, MA[1]; Silvio Maffioletti, FAILAC[2]; Tony Carnevali, OD, FAAO[3]

[1]Department of Psychology - Università degli Studi di Milano Bicocca, Italy; [2]Degree Course in Optics and Optometry - Università degli Studi di Milano Bicocca, Italy; [3]Southern California College of Optometry, Fullerton, California, USA

ABSTRACT

Background. This study confirms and expands the original DEM validity data initially published, but for the Italian population. We not only use the same measures as in the original study, but also new methods of psychometric and statistical analyses as well.

Methods: Four experimental groups of children, a total of 152 subjects, were evaluated with the DEM test and various other psycho-educational tests to assess DEM validity.

Results: The results confirm the suitability of the four types of validity given in the DEM manual in the Italian population. Moreover, by using a construct validity scheme, the results show: a significant developmental trend for all variables ($p<0.001$), a significant internal correlation ($p<0.01$), a correlation with reading tests ($p<0.05$), a significant difference in vertical time ($p<0.005$) and adjusted horizontal time ($p<0.05$) in a group of LD children. The results also show a convergence validity with a test of visual exploration ($p< 0.01$), but a divergence validity with subjective evaluation of ocular movements, and the factorial analysis accurately demonstrates saturation to three main factors.

Conclusions: These more extensive analyses confirm the validity of the DEM test to assess ocular motility in the developmental age in the Italian population and permits differentiation of ocular motility and naming problems. The confirmation of validity has created a more appropriate baseline for a future standardization of the DEM test in the Italian population.

Keywords: children's vision, DEM test, eye movements, ocular motility, saccades, validity

Introduction

The Developmental Eye Movement (DEM) test is an inexpensive, practical and easy method of assessing and quantifying ocular motor skills in children. Its purpose is to make a quantitative measurement of ocular-movement abilities in a simulated reading condition by naming numbers in a simple and easy task without using a complex instrument. Clinical utilization of this test is justified if it has good validity. That is, does the DEM measure ocular motor skills and how well does it do so? There are several methods that can be used to evaluate validity.

In the original DEM article[1] the qualities of the test (validity, reliability and standard results) are reported, but there is very little discussion about test validity. In addition to this article, Richman and Garzia did present some data of validity for the DEM test in the DEM manual.[2,3] The authors utilized four kinds of validity: raw scores and chronological age, internal consistency, relationship to the achievement test, and results of learning-disabled children.

The DEM manual reports that the regularity of the trend exhibited by the four variables measured by the test as a function of chronological age is an indication of validity.[1] The Vertical Time (VT), Adjusted Horizontal Time (AHT), and accuracy measured by Errors show improvement over time. The Ratio Score,

Correspondence regarding this article should be emailed to silvio.maffioletti@virgilio.it or sent to Silvio Maffioletti at the Corso di Laurea in Ottica e Optometria, Università degli Studi di Milano Bicocca, Via Cozzi 53 20125 Milano (Italy).

Facchin A. Validity reassessment of developmental eye movement (DEM) test in the italian population. Optom Vis Dev. 2011;42(3):155-167

the factor that explains ocular movements, does not present a clear developmental trend (developmental improvement) because the data reported shows a higher value in the seven-year age group than in the six-year group. This finding is also confirmed by Fernandez-Velasquez and Fernandez-Fidalgo.[4] Nonetheless, the tendency of ratio demonstrated in Jimenez et al[5] exhibits a linear developmental trend. And in fact, in the original American data, if we draw a regression line through the data points, we can also demonstrate a clear trend of improvement by age.

The second parameter of validity shown by Richman and Garzia[3] is internal consistency. As reported in the manual "the correlation between the components of DEM provides some indication of its internal consistency and validity". However, the use of internal consistency is not proper when evaluating the relationship of the DEM subtests for two reasons. First, internal consistency is used for a test that measures one single skill or factor;[6] the DEM measures four different variables (VT, AHT, Ratio and Error). Second, to assess the internal consistency, the Cronbach alpha correlation coefficient is most commonly used[7] as in the evaluation of a questionnaire test. Thus it is more accurate to refer to the internal correlation of the test, instead of the internal consistency, by correctly using the Pearson r correlation coefficient.

The third aspect of the validity of the DEM considered in the manual is the relationship with an achievement test. The relationship between the DEM test and the achievement test is complex. The test used for evaluation, the Wide Range Achievement Test, measures word reading, sentence comprehension, together with spelling and solution of math problems. As shown by Richman and Garzia,[3] there is a good correlation between reading words and vertical time ($r=-0.79$ $p<0.001$) and horizontal time ($r=-0.78$ $p<0.001$), but a modest correlation with ratio ($r=-0.55$ $p<0.001$), the variable that can measure ocular movements most effectively (the negative correlation is relative to the measure adopted to classify reading). However, the WRAT measures only reading accuracy, not reading speed, and thus the comparison with eye movements is only partial.

The fourth aspect of validity is based on the results of learning-disabled children on the DEM test.

The definition of dyslexia is complex[8] and has been frequently revisited over the years.[9] It is worth noting that in Italy there is a precise definition given by Consensus Conference updated in 2007. Dyslexia is a specific deficit which concerns only the domain of reading in individuals with normal intelligence. The clinical evaluation must be done with standard tests that measure reading words, non-words and sentences. Dyslexia may be diagnosed if reading speed is below 2 SD from the mean expected for each grade and/or the accuracy is below the 5th percentile.[8] This new definition is much different from an older model.[1,3] However, the basic problem remains the same. Dyslexia is a developmental disorder that compromises the achievement of reading in children with normal intelligence.[9]

Richman and Garzia[1,3] showed a significant difference between normal and dyslexic readers in all the variables measured by the DEM test; thus the classification of the DEM results can be attributed correctly as naming speed problems (Behavior III) or to a combination of naming speed problems and ocular motility problems (Behavior IV). Griffin et al[10] correctly affirm that children with low results in the DEM (Behavior III and IV) present with a problem; but it is not correct to conclude that those with low results necessarily have a reading problem. Number naming is not the same as word naming or sentence reading. The naming of letters, words, numbers, each with different visuo-spatial conditions, represent different kinds of naming that require different cognitive demands and therefore cannot be directly compared.[11,12] It is clear that reading and eye movements are related. A good reader, one with excellent fluency and speed, and minimal or no errors, shows good ocular motility and manifests high achievement results and rapid naming. Patients with different kinds of learning disabilities[13] could perform differently in a number-naming, word-naming and ocular-motility tests however.[14,15] For this reason, to evaluate the reading process, it is necessary to perform specific reading tests.

The aforementioned validity characteristics, the reliability and the norms of the DEM test can be viewed as psychometric properties.[6] Although ocular movements can be measured in a precise and objective manner, in the DEM test there are influences of several subjective variables such as sustained attention, number recognition and retrieval, visual verbal integration time, speaking time,[1] visuo-spatial attention[16] and other cognitive variables. Therefore, a full psychometrical evaluation can be applied to the DEM. The psychometrical characteristics are the primary factors used to select the test and determine the quality of the results.[17] Validity in psychometric testing can be subdivided in three main categories: content, criterion and construct.

The content validity involves the systematic examination of the test content to ascertain whether it covers a representative sample of the behavior domain to be measured.[17] A test has content validity built into it by careful selection of which items or tasks to include. When applied to the DEM test, three measured variables can be observed: naming speed, horizontal saccadic speed and accuracy. The criterion validity reflects the success of the measurements used for prediction or estimation of outcomes in other disciplines not measured directly, for example academic performance, in a previous undergraduate school, predicted by reading speed.

Construct validity is defined as a sum of empirical evidence that represents a check of the construct.[17] That is, construct validity seeks to show that a test actually measures the construct that it claims to be measuring by an accumulation of empirical evidence.[6] Ocular movements are not a conceptual construct but they are real movements of the eye in space. This is true when the ocular movements are measured using objective methods; however, in the DEM test there are influences of several psychological and cognitive variables. Therefore, it can be considered to be a full construct with which to examine the validity. Despite this older division of test validity in three categories, the more recent studies on validity place all aspects into a construct validity, thus bypassing the older division of criterion and content.[18]

In this context, the validity of the DEM test was reanalyzed in subjects from the Italian population by using the methods of analysis reported in the DEM manual[3] as well as this psychometric concept of construct validity to confirm and expand the validity of the DEM.

This new and original approach is being applied 20 years after the original American data were obtained. Since different languages, educational programs and the concept of educational agreement can modify these values; all these factors will be considered in reassessing validity of the DEM in the Italian population.

Methods

Subjects. All subjects were selected randomly from several interdisciplinary school screening programs and were required to have parental authorization to participate in the study. The screenings were performed by senior optometric interns, optometrists and psychologists.

Table 1: Group descriptions and tests administered to each group.

Group	Description	Test Administered
1	46 children from 6 to 10 year: 6y (4); 7y (10); 8y (12); 9y (11); 10y (9)	DEM SDO[5] KITAP[17]
2	40 children from 7 to 11 year: 7y (1); 8y (9); 9y (10); 10y (10); 11y (10)	DEM NSUCO[19] Subtest 4 and 5 of DDE-2 Battery[19]
3	42 children from 7 to 13 year: 7y (5); 8y (10); 9y (3); 10y (6); 11y(4); 12y (7); 13y(7)	DEM Subtest 4 and 5 of DDE-2 Battery
LD	24 Dyslexic children from 7 to 13 year: 7y (1); 8y (5); 9y (7); 10y (4); 11y (2); 12y (3); 13y (2)	DEM Subtest 4 and 5 of DDE-2 Battery

All subjects had to meet certain inclusion criteria: to use their habitual lenses for testing (as indicated by parent; if a subject required correction but did not have one, the child was excluded); to have a binocular visual acuity at near of better than 0.8 decimal acuity (20/25); and to have no manifest or obvious binocular anomalies such as strabismus or high phorias as assessed by cover test.

A total of 164 children were screened; two were eliminated because they did not have their habitual eyeglasses and 10 for not completing the entire battery of tests required in this study. A total of 152 subjects met the selection criteria as well as complete all the necessary tests. These subjects came from 4 different screenings. Three groups were recruited from three small schools in northern Italy: Group 1, from a private school in a city; Group 2, from a public school in a small town; and Group 3, from a public school in a city. The LD (Learning Disabled) Group was recruited from a specific center: the Istituto Medea, Bosisio Parini, LC. Group 1 consisted of 46 children ranging in age from 6 to 10 years, 22 female and 24 male; Group 2 consisted of 40 children from ages 7 to 11, 15 female and 25 male; and Group 3 consisted of 42 children from 7 to 13 years, 18 female and 24 male. The LD Group consisted of 24 dyslexic children from 7 to 13 years, 6 female and 18 male. Each group is further detailed in Table 1.

The LD Group consisted only of dyslexic children as diagnosed by a multidisciplinary team and in accord with a more recent definition of dyslexia.[8] The diagnostic work-up was conducted at the Istituto Medea by a neurologist, a psychologist and other professionals, as necessary, to exclude other conditions such as neurological problems that present reading difficulties but could not be classified as dyslexia.

Tests and Procedures

All testing for this study was performed by the author A.F and senior optometric students from the Optics and Optometry program at the Università degli Studi di Milano Bicocca, Italy. They were assisted and supervised by the authors A.F and S.M.. Each of the examiners was well trained in all aspects of the study methodology, test procedures, and subject instructions to ensure standardization and consistency in administering the tests and making observations. To further ensure consistency, each test was administered by two examiners-one optometry student and one of the authors.

The testing was conducted at each school in a room specifically set up for the evaluations. The room was chosen to minimize noise and distractions and provide uniform illumination with greater than 400 lux.

Different tests were administered to each group: Group 1-DEM, Saccade Direct Observation (SDO), and KITAP; Group 2-DEM, NSUCO, and Subtest 4 and 5 of the DDE-2 Battery; and Group 3-DEM and Subtest 4 and 5 of the DDE-2 Battery. The tests given to each group are indicated in Table 1.

The tests were administered as reported in the test manual or as described in specific articles about the tests. A brief description of the tests and administration procedures is outlined below:

DEM Test: The DEM comprises a pretest card and three test cards: two vertical tests (A and B) and one horizontal (C). The test is administered precisely in accordance with the protocols outlined in the DEM Manual. The Vertical Time (VT) is a sum of the time for completion of the two cards A and B. The VT reflects the time it takes to read aloud 80 numbers arranged vertically. The Adjusted Horizontal Time (AHT) is the time of card C adjusted for omission or addition errors. The AHT reflects the total time to read aloud the same 80 numbers arranged in a horizontal pattern and the time to perform saccadic eye movements from number to number. A ratio score is calculated by dividing the adjusted horizontal time by the vertical time. The total error reflects the total number of errors in the C card (i.e. omission, addition, substitution and transposition).

Ratio is the main measure used to differentiate ocular motility dysfunction. To determine the percentile or standard score for VT, AHT, Ratio and Error, it is necessary to consult normative tables. Considering the variables measured, four possible behavior types can be found with the DEM test:

Type I: Average normal values for all variables

Type II: Abnormally increased values for AHT and Ratio. This pattern is characteristic of ocular movement dysfunction

Type III: Abnormally increased values for VT and AHT, but normal Ratio. There is a difficulty in automaticity of number naming

Type IV: Abnormally increased values for VT, AHT and Ratio. This behavior is a combination of type II and III and represents difficulties in number naming and eye movement.

SDO: The saccade direct observation (SDO) test is a subjective evaluation of saccades on a scale from 1 (poor) to 5 (best) administered as described by Jiménez et al:[5] two fixation targets were positioned 40 cm away and 10 cm apart. The subject was instructed to look from one target to the other 10 times. Scoring depended on the precision of the fixation movement of the eye: 4- smooth and precise; 3- slight undershooting; 2- pronounced undershooting or increased latency; 1- inability to perform the task.

NSUCO: The NSUCO/Maples Oculomotor Test is a standardized method of scoring eye movements.[19] It allows observation of the patient's eye movement ability and accuracy along with head movement and body movement without requiring a subjective response of the patient. Administration of the test was in accordance with the manual. For purposes of this study only the accuracy parameter of the saccade test was considered. The accuracy of saccades is scored based on a scale from 1 (large over- or under shooting is noted one or more time) to 5 (no over or under shooting). The mean accuracy score is recorded for 5 round trips between two targets (10 saccades).

KITAP battery of tests for attention performance for children: For this evaluation only the visual exploration subtest of the KITAP battery[20] is used. In this computerized task, one will see a group of 25 witches flying on their brooms in one direction. Occasionally a witch makes a mistake and flies in the wrong direction. This error has to be detected pressing key one or two. The total number of matrixes presented is 100. The reaction time to identify a target is measured in seconds and the median is calculated for the test.

DDE-2 Battery of tests for evolutionary dyslexia and dysorthographia: For this evaluation only the subtest 4 and 5 of the DDE-2 battery was used. The Subtest 4 of DDE-2 Battery is composed of four vertical lists each of 28 words of different lengths, while Subtest 5 is composed of three vertical lists each of 16 legal non-words (words orthographically valid, but are non-existent and have no meaning). Both of these subtests use lists of words/non-words of different lengths in the Italian language. It is administered as described by Sartori, Job, Tressoldi.[21] The subject reads all the words in a column as fast as possible and the examiner records time and accuracy. The variables recorded are the total time spent to read the lists and the errors performed during reading for each subtest, for a total of four variables: Subtest 4 time, Subtest 4 errors, Subtest 5 time, Subtest 5 errors. Each subject was required to read Subtest 4 and Subtest 5, in order, as reported in the test manual.[21]

Analytical methods

Since different tests were performed for each group, some of the analysis was considered separately for each group. But whenever possible the analysis was conducted for all subjects combined. For each group, the analysis pertinent to the test and collected data was performed. The Learning Disabled (LD) group was taken into account only for the validity evaluation of pathological group.

The following extended analytical methods were used to verify the several aspects of construct validity.[18]

Evolutionary (developmental) trend: The evolutionary trend reflects the developmental improvement of skills as a function of age. The mean and standard deviation for each component factor of the DEM was determined and plotted against age. Following each of the component variables of the DEM test and determining whether a true trend over time is present can be used to verify the construct validity. The quantitative analysis was performed using a one factor ANOVA.

Internal correlation: The relationship of the four variables (VT, AHT, RATIO and ERRORS) measured by the DEM to each other was determined by using the Pearson r correlation.

To better demonstrate the relationship to developmental age, it is necessary to remove the component due to age from the correlation. For this purpose, the same analysis was performed by using the partial correlation,[22] in which the correlation due to age was removed:

$$r_{xy|z} = \frac{r_{xy} - r_{xz} r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$

Where x and y are the two variables of interest and z the variable that must be removed from the first two (for this case, age).

Relationship with reading test: The relationship between the DEM test and the achievement test is complex. Subjects with reading difficulties in the DEM test present high VT and AHT[23] values but the highest VT and AHT values are typical of dyslexia.[3]

In normal readers it is possible to have a better understanding of the relationship between DEM and reading by comparing the DEM test to a reading test which uses words and legal non words.

From the standpoint of internal correlation, a high correlation between all subtests is not expected; but there should be at least some correlation between a few variables of the DEM and reading test.

The expectation being that a higher correlation should be found between AHT or Ratio and a word reading subtest (Subtest 4 of DDE-2 Battery) than a non word reading subtest (Subtest 5 of DDE-2 Battery). The analysis to test this hypothesis was performed with the same partial correlation used for internal correlation, but removing the correlation due to age.

Evaluation of a pathological group: Another evaluation of construct validity can be obtained from a pathological group, in this case dyslexic children, whose performance is anticipated to be worse than a normal reference group. Richman and Garzia[1,3] showed a significant difference between normal and dyslexic readers in all the variables measured by the test, thus attributing dyslexia to the classifications of DEM behavior III (naming problem) or behavior IV (naming and ocular motility problems). For this evaluation, the performances of Group 3 and the LD Group were compared. The statistical test used was the t-test.

Convergence and divergence validity: To test for this validity each of the various methods of testing are compared to one another. Convergence validity is said to occur when different tests or instruments measure the same variables and demonstrate a significant correlation among the variables. Conversely, when there is no correlation, divergence validity is said to occur.[17] There cannot be a relationship when tests or instruments measure other constructs that are not theoretically related (divergence). A low correlation between two tests may indicate divergence, the two tests may be measuring different variables or that the sensitivity of the two tests may be very different or operate at different levels. A modest correlation may result from measuring the same variable with different methods. And a higher correlation occurs generally between tests that measure the same variable by the same method. The tests are compared using the Pearson r correlation.

Factorial analysis: Factorial analysis is a statistical method used to explain variability among observed variables in terms of fewer unobserved variables called factors.[6] Factor analysis estimates how much of the variability of data is due to common unobserved factors or errors. Thus the purpose of factor analysis is to determine the number of common factors influencing a set of measures and the strength of the relationship between each factor and each observed measure. It is important to relate the factors obtained with the theoretical construction of the test to better explain the patterns of relationship among the test variables. Factor analysis is performed by examining the pattern of correlations between the observed measures. Measures that are highly correlated either positively or negatively are likely influenced by the same factors, while those that are relatively uncorrelated are likely influenced by different factors.

For example, variations in four observed variables may reflect variations in a single unobserved variable or factor, or in a reduced number of unobserved variables (two or three factors), or in four unobserved variables or factors, directly linked to the single observed variables. Applied to the DEM test, it is important to observe relationships with respect to three factors: one factor possibly related to a common part of naming on VT and AHT; the second factor related to Ratio, independent from naming; and the third factor perhaps related to Errors, independent from the others.

Factorial saturation is the correlation between factors found in the analysis and the real variables. For this analysis, the Varimax rotation of the factorial axis is used to maximize the factor loads.[24]

Table 2: Mean results (seconds) of all groups (SD in parenthesis).

Age	n.	VT	AHT	Ratio	Error
6	4	59.02 (11.80)	86.17 (17.90)	1.47 (0.21)	11.0 (9.52)
7	16	52.80 (9.31)	79.39 (22.07)	1.50 (0.31)	8.25 (5.89)
8	31	48.09 (7.48)	62.59 (12.70)	1.31 (0.23)	6.16 (7.66)
9	24	42.11 (4.91)	51.70 (8.59)	1.23 (0.13)	2.54 (3.90)
10	25	36.58 (4.90)	43.43 (6.52)	1.19 (0.12)	2.40 (3.24)
11	14	33.17 (3.69)	38.67 (7.88)	1.17 (0.21)	1.50 (1.60)
12	7	32.70 (5.37)	36.55 (3.37)	1.13 (0.09)	1.57 (2.93)
13	7	30.10 (7.69)	32.09 (6.12)	1.08 (0.10)	0.71 (0.95)

Results

Evolutionary trend: The evolutionary trend of the various components of the DEM test reflects the improvement of skills as a function of age. Mean and standard deviations are listed in Table 2 and plotted in Figures 1 to 4.

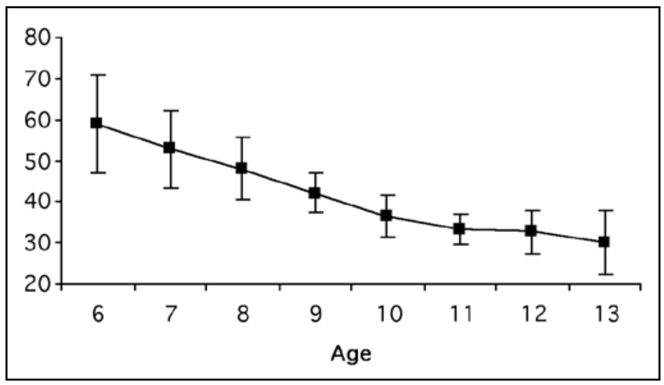

Figure 1: *Evolutionary trend of VT shown by time decreasing as a function of age. The dots indicate the mean value and the bars +/-1SD for all groups combined.*

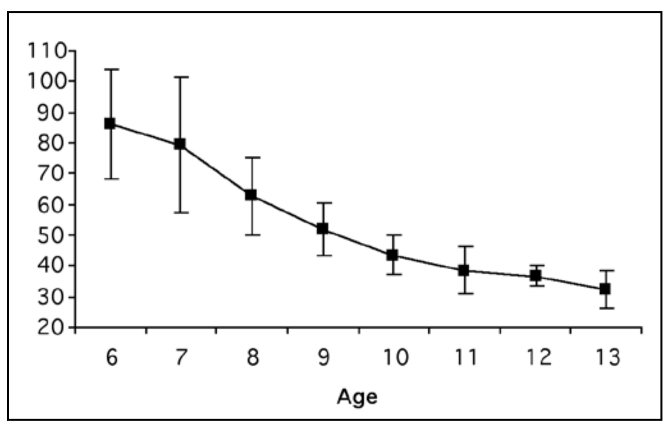

Figure 2: *Evolutionary trend of AHT shown by time decreasing as a function of age. The dots indicate the mean value and the bars +/-1SD for all groups combined.*

To confirm the difference between years a single factor ANOVA was performed with eight levels (age from 6 to 13) for each DEM component. The

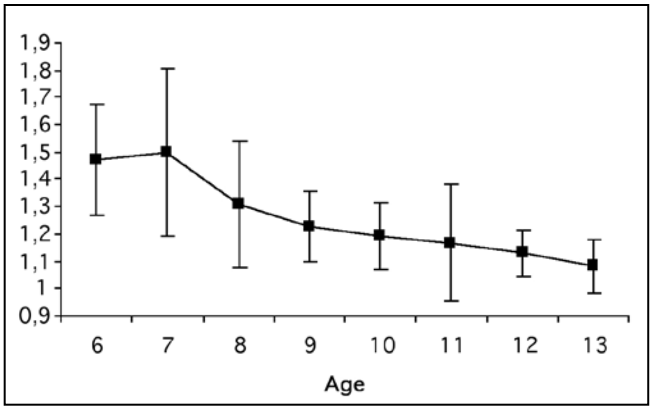

Figure 3: *Evolutionary trend of Ratio as a function of age. Except for age 7, this variable shows a clear evolutionary trend. The dots indicate the mean value and the bars +/-1SD for all groups combined.*

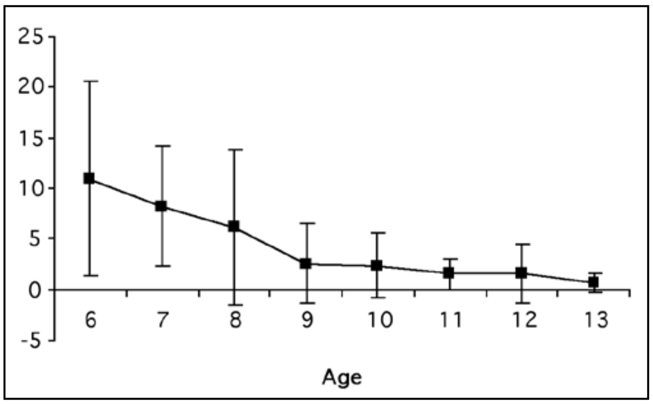

Figure 4: *Evolutionary trend of errors shown by reducing numbers as a function of age. The dots indicate the mean value and the bars +/-1SD for all groups combined.*

results show significant differences for VT ($F(7, 120)=24.866$, $p<0.0001$); AHT ($F(7, 120)=28.804$, $p<0.0001$); Ratio ($F(7, 120)=6.6957$, $p<0.0001$) and errors ($F(7, 120)=5.0032$, $p<0.0001$).

With the possible exception of Ratio, each of the variables demonstrates a clear and definite improvement with age. With Ratio there is a slight increase at age 7 years even though each of the components (VT and AHT) show improvement with age, but then shows the typical evolutionary changes with age shown by the other variables.

Internal Correlation: The DEM test was given to each of the three groups; but the data and results have been combined for all groups. Table 3 tabulates the Pearson r correlation for each of the component variables of the test. The highest correlation is between VT and AHT ($r=0.85$, $p<0.01$) and between Ratio and AHT ($r=0.76$, $p<0.01$). The least correlation is between Ratio and VT ($r=0.33$, $p<0.01$). All were statistically significant at $p<0.01$.

To better demonstrate the relationship among the component variables, the correlation was performed

Table 3: Internal correlation of all groups combined (all $p<0.01$)

	VT	AHT	RATIO
AHT	0.85		
RATIO	0.33	0.76	
ERROR	0.50	0.58	0.45

Table 4: Partial internal correlation (age removed) of all groups (* $p<0.05$; ** $p<0.01$)

	VT	AHT	RATIO
AHT	0.66**		
RATIO	–0.06	0.69**	
ERROR	0.30*	0.43**	0.31**

Table 5: Partial correlation (age removed) between reading skill and DEM variables for Groups 2 and 3 combined (* $p<0.05$; ** $p<0.01$).

	VT	AHT	Ratio	Err
Subtest 4 (words) time	0.40**	0.65**	0.39**	0.16
Subtest 4 error	0.18	0.37**	0.26*	0.21
Subtest 5 (non words)time	0.47**	0.43**	0.10	0.10
Subtest 5 error	0.30*	0.27*	0.03	0.28*

using the partial correlation (age removed) and the results are listed in Table 4.

Without the co-variation due to age, the relationship between the various components of the DEM appears clearer. The highest correlations are still between VT and AHT ($r=0.66$) and between AHT and Ratio ($r=0.69$). These are statistically significant at $p<0.01$. Interestingly, the lowest correlation is between VT and Ratio ($r=-0.06$), which is not statistically significant, followed by VT and Error and Ratio and Error which are significant at the $p<0.05$ and $p<0.01$ respectively.

Relationship with reading test: For this analysis, Groups 2 and 3 were combined and the partial correlation has been used to remove the influence of age. The results are summarized in Table 5.

The highest correlation occurs between AHT and Subtest 4 time for reading words ($r=0.65$, $p<0.01$), and is larger than AHT and Subtest 5 time for reading non-words ($r=0.43$, $p<0.01$).

The higher correlation between AHT and Subtest 4 time ($r=0.65$, $p<0.01$) is also larger than VT and Subtest 4 time ($r=0.40$, $p<0.01$).

Table 6: Mean results for DEM and DDE-2 Battery in Group 3 (Control) and LD Group (Dyslexic)

	Dyslexic	Control	p-value
Age	9.8	10.02	n.s.
DDE-2 battery:			
Time 4	184.96	104.18	<0.005
Error 4	11.33	2.02	<0.0001
Time 5	141.71	71.59	<0.0005
Error 5	13.04	3.02	<0.0001
DEM:			
VT	50.41	39.90	<0.005
AHT	61.96	50.63	<0.05
Ratio	1.25	1.25	n.s.
Error	6.04	4.24	n.s.

Table 7: Results of saccade direct observation in Group 1

Age	N	Mean	Std. Dev.
6	4	2.75	0.50
7	9	3.50	1.00
8	14	3.58	1.16
9	10	3.54	1.04
10	9	2.88	0.93

The correlation between VT and Subtest 4 time and Subtest 5 time are close to the same at 0.40 and 0.47 respectively. Both are statistically significant at p<0.01.

With respect to Ratio, the highest correlation occurs with Subtest 4 time reading a list of words (r=0.39, p<0.01) compared to Subtest 5 time reading non words (r=0.10, p=n.s.).

Evaluation of a pathological group: Comparing the LD Group to Group 3 as control, age was statistically the same ($t_{(64)}$=0.59 p=n.s.), but there were significant differences between the two groups when comparing the different subtests of the DDE-2 Battery and the DEM test (Refer to Table 6).

If reading is considered, there are differences observed in speed and accuracy measured by DDE-2 Battery test in two groups (dyslexic subjects vs. control). Reading speed for words (Subtest 4 of DDE-2 Battery): dyslexic subjects are slower than control ($t_{(64)}$=4.2, p<0.0001) and perform more errors ($t_{(64)}$=6.9, p<0.0001). Reading speed for non words (Subtest 5 of DDE-2 Battery): dyslexic subjects are slower than control ($t_{(64)}$=3.8, p<0.0005) and perform more errors ($t_{(64)}$=8.4, p<0.0001).

Table 8: Correlation between SDO and subtest of DEM. All results are not significant.

	SDO
VT	–0.03
AHT	–0.10
Ratio	–0.16
Error	–0.05

Table 9: Results of NSUCO saccade test accuracy in Group 2.

Age	N	Mean	Std. Dev.
8	10	3.50	0.53
9	10	3.50	0.71
10	10	3.20	0.79
11	10	3.50	0.53

Table 10: Correlation between SDO and subtest of DEM. All results are not significant.

	NSUCO Saccade Test Accuracy
VT	0.00
AHT	–0.07
Ratio	–0.05
Error	0.09

Examining the results of the DEM in the two groups, it can be observed that there is a significant difference for VT ($t_{(64)}$=3.40, p<0.005) and AHT ($t_{(64)}$=2.10, p<0.05). But there are no significant differences for Ratio and Error. The difference in the DEM test in dyslexic children is only in VT and AHT, not in accuracy or Ratio.

Convergence and divergence validity: The convergence method to test validity can be applied to the DEM test by using different tests that examine ocular movements. Tables 7 and 8 show the results of the saccade direct observation (SDO) test in Group 1.

The correlation of the SDO test with each component of the DEM is low and not significant. The best agreement is between SDO and Ratio, but it is also not significant (r=-0.16, p= n.s).

In Group 2, the NSUCO Oculomotor Test was used to assess eye movements. Tables 9 and 10 summarize the results of the NSUCO test. The correlation between the accuracy of the subtest of saccades of NSUCO and AHT of the DEM was low (r=-0.07) and not significant.

Another convergence comes from a comparison of a test that examines visual exploration skill. In Group 1, the visual exploration subtest of the KITAP battery of attention has been used. Tables 11 and 12 summarize the results of the exploration subtest.

For the KITAP test of exploration, the median response time to detect the stimulus was compared to each component of the DEM test. The results shows only significant correlation between AHT and the exploration time (r=0.29, p<0.01).

Table 11: Results of Subtest of visual exploration of KITAP battery of attention in Group 2. RT=response time in seconds.

Age	Median RT (sec.)	Std. Dev.
6	6.77	2.19
7	7.15	1.89
8	5.08	2.18
9	4.88	1.24
10	3.94	0.84

Table 12: Correlation between DEM subtest and KITAP test of exploration. (** p<0.01).

	Median RT KITAP
VT	0.22
AHT	0.29**
Ratio	0.08
Error	0.18

Table 13: Factorial analysis that shows saturation to three factors; bolded values show the high correlation (>0.6) between factors and variables. The last row reports the proportion of variance explained by the single factor to the total of variance of data.

	Factor 1	Factor 2	Factor 3
VT	**0.963**	0.118	0.236
AHT	**0.749**	**0.604**	0.265
Ratio	0.171	**0.963**	0.203
Errors	0.262	0.222	**0.939**
Prp. Totl.	0.397	0.339	0.262

Factorial analysis: The results of the factorial analysis are listed in Table 13.

Results show saturation to three main factors. The first factor is related to VT and AHT, the second factor is related to AHT and Ratio and the third to Errors. Globally the results of the factorial analysis demonstrate and differentiate correctly VT and AHT from Ratio and Errors.

This model presents high efficacy because it does explain a very high proportion of variance. In fact the three factors account for 99.8% (the sum of the variance explained by the three factors) of the total variance of the raw data.

Discussion

Evolutionary trend: Given the limitation that there were only four subjects in the 6 year old group, several aspects of the evolutionary trend measured by the DEM in the groups examined can be confirmed. Each of the variables of VT, AHT, and Error demonstrates a clear improvement with age. But for Ratio, this study provides additional evidence of increasing ratio in the seven-year age group compared with that in the six. These findings also confirm a trend in the original Garzia et al[1] norms and in the study by Fernandez Velasquez and Fernandez Fidalgo.[4] The evolutionary trend does not provide direct confirmation of validity, but can only confirm measurement of an evolutionary variable. Overall the data does confirm this trend with some limitations to the Ratio score.

With respect to Ratio for the 7-year old group, a plausible justification might be found in the mathematical structure of ratio AHT/VT. Although each of the component variables may improve, the ratio may still increase if AHT does not improve at the same rate as VT. With that in mind, there is a very dramatic improvement in language skills from age 6 years to age 7 years. In the American educational system, and similarly in the Italian system, generally children from first to second grade transform into true readers. They learn to read aloud with much greater fluency, accuracy and understanding. Thus, if naming skill improves faster than saccades, and because naming is equal in both vertical and horizontal reading, the Ratio is going to increase. After 7 years of age, there is a less dramatic development in language skills and thus both naming and saccades may change at the same rate, and Ratio demonstrates the typical improvement with age. Further study is obviously needed to confirm this explanation.

Internal Correlation: In analyzing the internal correlation of the DEM test, the results show a significant relationship between some variables of the DEM test. The correlation between horizontal (AHT) and vertical time (VT) is important to observe and this reflects the common naming component in the two examined variables. It can be seen that the correlation is 0.85 (p<0.01), which is greater than reported in the DEM manual (r=0.75, p< 0.001). Similarly the correlation between Ratio and AHT is large and significant. If the results are compared with those of Garzia et al[1] a higher correlation is found in our study, but the same trend is exhibited.

Using partial correlation and removing the co-variation component due to age, the internal correlation appears more clearly established. The high correlation between VT and AHT (r= 0.66, p<0.01) can be explained by the larger component of time associated with the naming process, equal in vertical and horizontal time. To illustrate this point, two reference values can be taken from the DEM test manual for six- and thirteen-year old children. For the former, the mean VT was 63.11 seconds and the mean

AHT was 98.26 seconds; for the latter, VT was 33.75 seconds and the AHT 37.56 seconds. Consequently, it is possible to calculate the percentage of horizontal time theoretically due to naming. For the six-year olds, the contribution of naming to the total time was 64% (and 36% on saccades), and for the thirteen-year olds it was 90% (and 10% on saccades). These large values for the contribution of naming, which are common for vertical and horizontal time, can explain the high correlation between VT and AHT.

The correlation between horizontal time and Ratio is 0.69 ($p<0.01$), which can be explained by the mathematical construct of Ratio. As indicated by Garzia et al[1] and confirmed by this study, the low correlation between VT and Ratio suggests the small contribution of naming to the ocular movement results. The errors correlate with other variables; with increasing age, errors are reduced, but also time to execute the AHT and VT is reduced.

Taken together, these results confirm the validity of several variables (VT, AHT, Ratio and Errors) measured by the DEM test. It is not sufficient to consider only the results with the highest correlation values, but it is necessary to evaluate every single correlation with each component of the DEM test.

Relationship with reading test: As observed for the internal correlation of the test, the partial correlation has been used to remove the influence of age. The correlation between AHT and time of Subtest 4 for reading words ($r=0.65$, $p<0.01$) is larger than AHT and time for Subtest 5 for reading non-words ($r=0.43$, $p<0.01$); this can be explained by the two-way reading models.[25] This model explains that reading of non-words is executed by single conversion from grapheme to phoneme and this kind of reading is slower. The reading of words takes advantage of a lexical retrieval and, for this reason, is faster.

The correlation between AHT – Subtest 4 time ($r=0.65$, $p<0.01$) is larger than VT – Subtest 4 time ($r=0.40$, $p<0.01$) suggesting that saccades play a significant role in performing the reading task. With respect to the relationship to Ratio, the value for ocular movement presents the highest relationship with a list of words ($r=0.39$, $p<0.01$) compared to non words ($r=0.10$, p= n.s.). These results are partially in contrast with the correlation reported in the DEM test manual. The latter shows the same correlation between AHT or VT, and the WRAT reading subtest. This study on the other hand, confirms that the naming of words correlates better to AHT of the DEM test[23] that is there is a greater correlation between horizontal time and reading words perhaps as a result of the contribution made by saccades.

Evaluation of a pathological group: On examining the results of the DEM in the LD Group, it can be observed that there is a significant difference for VT ($t_{(64)}=3.40$ $p<0.005$) and AHT ($t_{(64)}=2.10$ $p<0.05$). There are no significant differences for Ratio and Error. The difference in the DEM test in dyslexic children is only in VT and AHT, not in accuracy or Ratio. These results confirm that the type III behavior identified by the DEM test is a primary problem in naming numbers.

Although in the first DEM manual norms are given for every year or grade, there are no explicit indications of which value can be considered normal or pathological. Observing the example reported in the manual, it can be deduced that the cutoff value used to differentiate normal vs. pathological was at the 30th percentile. Successive studies confirmed this hypothesis.[26] However, in the second edition of the DEM manual,[2] this point has been expanded with a more precise criteria. There are two kinds of selection criteria: one for screening and one for clinical use.

The purpose of the DEM as a screening test is to identify a potential problem and refer without significant over referrals. The screening criteria uses the cutoff limit of 16th percentile with no reported symptoms, with a further suggestion to consider the results between 16th and 35th percentile to be at risk or a borderline result. On the other hand, for clinical use the selection criteria is the same with the cutoff limit set to the 16th percentile, but now the presence of symptoms linked to ocular movement dysfunction is necessary. These criteria are well aligned with other psycho-educational tests such as reading tests.

Convergence and divergence validity: The correlation between the two observation tests of saccades, SDO and the NSUCO, and the DEM is low and not significant. The best agreement is between SDO and Ratio, but it is also not significant ($r=-0.16$, p= n.s). The correlation between the accuracy of the subtest of saccades of the NSUCO and AHT of the DEM was also low ($r=-0.07$) and not significant. The results of the SDO and NSUCO tests in this study support the conclusion of Jiménez et al: there is no improvement in saccades by age when saccades are tested by observation.[5] Although Maples et al reports that the NSUCO subtest of saccade accuracy shows improvement with age,[27] our data did not support this finding. Perhaps this difference may be attributed to the limited number of subjects in our study for each age group. But since only the accuracy of saccades was used for the evaluation, the fact that both tests have in common the subjective evaluation of eye movements in free space might very well explain the similarities in the results.

Moreover this outcome may also be explained by the characteristics of the tests themselves: one is a measure of ocular movements in a reading-like condition (DEM), while the others are a subjective observation of ocular movements between two non-accommodative targets in free space. Clearly, the DEM requires a high level of accuracy and visuo-spatial demand while the SDO and NSUCO tests are created to observe large anomalies of ocular movements. Moreover, the DEM test results improve as a function of age, while the tests with direct observation of saccades in our study reveals no significant differences between ages and their relationship obviously diverge. Indeed, the DEM test and the SDO or NSUCO saccade tests operate at different levels and cannot be easily compared.

The results of the KITAP test have been compared to those of the DEM test. The results show only a slight but significant correlation between AHT and the exploration time ($r=0.29$ $p<0.01$) suggesting that better eye movement skills can enhance the speed of exploration of the matrix and thus parallel the performance of the horizontal subtest of the DEM. However, the two tasks and their relative responses are different thus explaining the low (but significant) relationship between the DEM and the KITAP.

Factorial analysis

The results of factorial analysis accurately show saturation to three main factors. The first factor is related to the common naming component between VT and AHT. The second factor is related to AHT and Ratio and the third to Errors. These three factors account for 99.8% of the total variance of the DEM test data.

Thus globally the results of the factorial analysis demonstrate and differentiate correctly naming from ocular movements and errors.

The higher correlation found between VT and AHT reflected on internal correlation and factorial analysis may be related to the fact that naming is common to both of these two variables.

A large part of the analysis performed in this study is dedicated to construct validity, because this aspect can be seen as a unified concept of every kind of validity,[7] and can be a robust test of what it was designed to measure.[6,17,18]

However, there are several issues that may be considered limitations of the study. Two aspects of construct validity that were not analyzed are the relationship of the DEM test to an objective or physiological measurement of ocular movements and comparing the DEM test results before and after a specific training program. Nevertheless, Ayton et al,[28] has compared the DEM with an objective eye movement measure. They conclude that the DEM test performance is not correlated with saccadic eye movement skills, but it is related to reading performance. Then again another recent study by Manzoli et al,[29] found a medium to high correlation between the DEM and an objective measurement of eye movements. These different results can be attributed to the difficulty in obtaining a consistent outcome from eye movement recordings and to an absence of a well defined construct of "good ocular movement", easier for the DEM (small Ratio), but not as clear for objective recording.

Further confirmation of construct validity can be provided by evidence of variation in the results after completion of particular training tasks. Applied to the DEM, test and retest will be necessary before and after oculo-motor training. Obviously this will require further study considering the reliability of the test,[26,30] or a more complex research paradigm comparing a treated and an untreated group.

Finally, our subjects come from small groups of different ages and numbers and the evaluation or testing was done by different examiners for each group. Although this may be seen as a limitation, since construct validity is a sum of empirical evidence, each of the different groups, even if small in numbers, can be considered as a single experiment to check the construct validity. There is also a possibility that using different examiners may introduce some variability in the data collection especially in a test that requires experience and fine perceptual discrimination like SDO and NSUCO. However, because these tests were always administered by two well trained examiners, this possibility is minimized.

The confirmation of validity has created a more accurate foundation for a future standardization of the DEM test in the Italian population. Currently the DEM test does not have specific norms for the Italian population. Previous research made in a Spanish speaking population has shown that the DEM is independent of language and can be used in any population regardless of language.[4] This is due to the mathematical construct of Ratio. Even if VT and AHT change, the ratio remains the same independently of language used. Moreover, Fernandez Velasquez, Fernandez Fidalgo[4] found significant differences only in 6 year old children for VT, AHT and Error between Spanish and original American normative values. Jimenez et al[5] in another study also concluded that his results were more similar to those obtained in the original American group.[1] These results suggest that norms are independent of language or population. Conversely, in a Cantonese speaking population significantly faster results were found for all age groups of children in VT (for this study, the authors used the Adjusted Vertical Time) and AHT, but not in Ratio and Error.[31] Thus the question about normative data remains open without correct data for each population. In this study, no comparison was made between the American original data and the Italian data because the number of subjects for each age group was not adequate for statistical analyses.

Conclusion

The results of this study not only confirm the original validity of the DEM test but further enhance the validity by presenting additional data and analyses not originally evaluated. By doing so, this study also confirms the validity of the DEM test in normal and learning-disabled children in the Italian population. The DEM test does indeed provide a valid assessment of ocular movements as well as allows the proper differentiation of ocular movement deficits and rapid naming number deficits.

Acknowledgement

We would like to thank the other examiners who have collected the various groups of data, Alessandro Peraldo, Sabrina Prudenzano and Annalisa Usuelli. We would also like to thank Silvia Tavazzi for her comments on an earlier version of this paper.

References:

1. Garzia RP, Richman JE, Nicholson SB, Gaines CS. A new visual-verbal saccade test: the development eye movement test (DEM). J Amer OptomAssoc. 1990 Feb;61(2):124-35.
2. Richman JE. Developmental Eye Movement Test, Examiner's manual, version 2.0. South Bend, IN: Bernell Corp. 2009.
3. Richman JE, Garzia RP. Developmental Eye Movement Test, Examiners booklet, version 1. South Bend, IN: Bernell Corp. 1987.
4. Fernandez-Velazquez FJ, Fernandez-Fidalgo MJ. Do DEM test scores change with respect to the language? Norms for Spanish-speaking population. Optom Vis Sci. 1995 Dec;72(12):902-6.
5. Jimenez R, Gonzalez MD, Perez MA, Garcia JA. Evolution of accommodative function and development of ocular movements in children. Ophthalmic Physiol Opt. 2003 Mar;23(2):97-107.
6. Kline P. The handbook of psychological testing. Second ed. London: Routledge 1999.
7. Cronbach LJ. Five perspectives on validity arguments. In: Wainer H, Braun HI, eds. Test Validity. Hillsdale, NY: Erlbaum 1988.
8. AID. Consensus conference, Disturbi evolutivi specifici di apprendimento. Raccomandazioni per la pratica clinica definite con il metodo della Consensus Conference. Bologna: Associazione Italiana Dislessia 2007.
9. APA. Diagnostic and Statistical Manual of Mental Disorders, DSM IV Tr. Washington D.C.: American Psychiatric Association 2000.
10. Griffin JR, Lung J, Nguyen M. The developmental eye movement test and detection of dyslexia. Optom Vis Sci. 1998;75(12):17.
11. Damian MF. Asymmetries in the processing of Arabic digits and number words. Memory Cog. 2004 Jan;32(1):164-71.
12. Ischebeck A. Differences between digit naming and number word reading in a flanker task. Memory Cog. 2003 June 2003;31(4):529-37.
13. Spinelli D, Brizzolara D, De Luca M, Gasperini F, Martelli M, Zoccolotti P. Subtypes of developmental dyslexia in transparent orthographies: A comment on Lachmann and Van Leeuwen (2008). Cognitive Neuropsych. 2009;26(8):752-8.
14. Menghini D, Finzi A, Benassi M, Bolzani R, Facoetti A, Giovagnoli S, et al. Different underlying neurocognitive deficits in developmental dyslexia: a comparative study. Neuropsychologia. 2010 Mar;48(4):863-72.
15. Semrud-Clikeman M, Guy K, Griffin JD, Hynd GW. Rapid naming deficits in children and adolescents with reading disabilities and attention deficit hyperactivity disorder. Brain Lang. 2000 Aug;74(1):70-83.
16. Facoetti A, Ruffino M, Peru A, Paganoni P, Chelazzi L. Sluggish engagement and disengagement of non-spatial attention in dyslexic children. Cortex. 2008 Oct;44(9):1221-33.
17. Anastasi A, Urbina S. Psychological Testing. Seven ed. Upper Saddle River NJ: Prentice Hall 1997.
18. Urbina S. Essentials in psychological testing. Hoboken NJ: John Wiley and Sons 2004.
19. Maples WC. NSUCO Oculomotor test. Santa Ana, CA: Optometric Extension Program Foundation 1995.
20. Fimm V, Zimmermann P. KITAP: Tests of Attentional Performance for Children. Herzogenrath: Psytest 2003.
21. Sartori G, Job R, Tressoldi PE. DDE-2: Batteria per la valutazione della dislessia e della disortografia evolutiva. Firenze: Organizzazioni speciali - Giunti 2007.
22. Spiegel MR, Stephens LJ. Schaum's Outline of Theory and Problems of Statistics. Third ed. New York: McGraw Hill 1999.
23. Palomo-Alvarez C, Puell MC. Relationship between oculomotor scanning determined by the DEM test and a contextual reading test in schoolchildren with reading difficulties. Graefe's archive for clinical and experimental ophthalmology = Albrecht von Graefes Archiv fur klinische und experimentelle Ophthalmologie. 2009 Sep;247(9):1243-9.
24. Kaiser H. The varimax criterion for analytic rotation in factor analysis. Psychometrika. 1958;23(3):187-200.
25. Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J. DRC: a dual route cascaded model of visual word recognition and reading aloud. Psychol Rev. 2001 Jan;108(1):204-56.
26. Tassinari JT, DeLand P. Developmental Eye Movement Test: reliability and symptomatology. Optometry. 2005 Jul;76(7):387-99.

27. Maples WC, Atchley J, Ficklin T. Northeastern State University College of Optometry's oculomotor norms. J Behav Optom. 1992;3(6):143-50.

28. Ayton LN, Abel LA, Fricke TR, McBrien NA. Developmental eye movement test: what is it really measuring? Optom Vis Sci. 2009 Jun;86(6):722-30.

29. Manzoli MV, Facchin A, Ravasi A, Pregliasco R, Maffioletti S. Simultaneous objective and subjective evaluation of ocular movement using DEM and Readalyzer European Academy of Optom Optics. Cophenhagen 2010.

30. Rouse MW, Nestor EM, Parot CJ, Deland PN. A reevaluation of the Developmental Eye Movement (DEM) test's repeatability. Optom Vis Sci. 2004 Dec;81(12):934-8.

31. Pang PC, Lam CS, Woo GC. The Developmental Eye Movement (DEM) test and Cantonese-speaking children in Hong Kong SAR, China. Clin Exp Optom. 2010 Jul 1;93(4):213-23.