ARTICLE

The Developmental Eye Movement (DEM) Test: Normative Data for Italian Population

Alessio Facchin, MA;[1,2] Silvio Maffioletti;[3] Tony Carnevali, OD[4]

[1]Department of Psychology - Università degli Studi di Pavia, Italy; [2]Optometry Private Practice, Varese, Italy;
[3]Degree Course in Optics and Optometry - Università degli Studi di Padova, Italy; [4]Southern California College of Optometry, Fullerton, California, USA

ABSTRACT

Background: The purpose of this study is to provide normative data for the Italian population on the Developmental Eye Movement Test (DEM) and to compare these norms to those published for the American, Spanish, Cantonese, Japanese and Portuguese populations.

Methods: 1122 children from 6 to 14 years of age participated in the study. All subjects were obtained through school vision screenings. The DEM test was administered as outlined in manual. The results were compared with other norms published.

Results: Normative data for the Italian children is provided for ages 6 to 14 years. Significant differences were found among all norms published, but a larger difference was found for the Cantonese and Japanese in which vertical and horizontal time was faster than all others. And, when comparing the Italian to the American and second Spanish norms fewer differences were found.

Conclusion: We have presented norms for the DEM test for the Italian population. The specific normative data provide an optimal and reliable clinical application of the DEM test. The comparison with other normative data suggests the application of a specific norm for each population because the DEM test appears to be a language and a cultural dependent test for all variables investigated.

Keywords: automaticity, DEM test, Developmental Eye Movement test, oculomotor norms, oculomotor testing, saccadic eye movements

Correspondence regarding this article should be emailed to alessiopietro.facchin@gmail.com or sent to Alessio Facchin, MA, Dipartimento di Psicologia, Università degli Studi di Pavia, Piazza Botta 1, 27100 Pavia, Italy.

Facchin A, Maffioletti S, Carnevali T. The developmental eye movement (DEM) test: normative data for Italian population. Optom Vis Dev 2012;43(4):162-179

Introduction

Since its introduction over 20 years ago, the Developmental Eye Movement (DEM) test has achieved wide acceptance and is being used extensively in many countries. The test is a practical, easy and inexpensive method of assessing and quantifying ocular motor skills in children. Without using a complex instrument, the purpose of the DEM test is to make a quantitative measurement of ocular movement skills by naming numbers in a simple and easy simulated reading task[1].

The design of the DEM is straightforward. The test comprises a pretest card and three test cards: two vertical tests (A and B) and one horizontal (C). 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 dividing the adjusted horizontal time by the vertical time[2]. The total error reflects the total number of Errors in the C card (i.e. omission, addition, substitution and transposition).

The interpretation of the test is also relatively uncomplicated. Ratio is the main measure used to differentiate ocular motility dysfunction[3]. The DEM test can also be used to determine a problem with automaticity[2] in number reading that may indirectly

reflect a dysfunction in sustained attention, visuo-spatial attention, number recognition, speaking time, phonological process or other cognitive problem. 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 but VT normal. This pattern is characteristic of ocular movement dysfunction.
- Type III: Abnormally increased values for VT and AHT, but Ratio normal. 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 saccadic eye movement.

Despite the fact that the DEM test has achieved such wide use, in the minds of some, there is still some question of its validity / reliability. The validity of the DEM test is generally well established. The DEM test may be considered a psychometric test[4,5] and its relative characteristics, validity, reliability and normative data represent the primary factors used to select a test and determine the quality of the results obtained.[6] Validity of the test is reported in the DEM manual and recently was retested and expanded upon in the Italian population[5] reconfirming the validity of DEM test to assess ocular motility in developmental age.

When comparing the DEM test to objective recording of eye movements, a great deal of variability may be found. Lack[7] found a modest relationship between the Visagraph Number Test (VGN) and DEM test. The author concluded that the modest relationship between VGN variables and DEM data was probably due to the psychophysical differences between the stimuli used even if the two tests appear to be similar. Ayton et al.[8], using a more sophisticated eye-tracker instrument, found no significant correlation between DEM subtests and quantitative eye movement parameters such as gain, latency, asymptotic peak velocity, and number of corrective saccades. Webber et al.[9], retested the same hypothesis using the DEM, Visagraph and a reading comprehension test. They concluded that the DEM test was a poor predictor of reading rate; however their data does show a significant relationship ($p<0.01$) between Adjusted Horizontal Time (AHT) when compared to the number of fixations and $p<0.05$ when number of regressions and span of recognition are taken into account. These measures are more closely related to eye movements than reading. Thus, it appears that the relationship between objective and psychometric evaluation of ocular movements remains an open question and different studies show conflicting results depending on stimuli used or parameters tested.

The reliability of the test was also reported in the manual and in several other studies. In two of these studies, the results are partially divergently and do not clearly answer the question of repeatability of the DEM test[10,11]. However, a recent research study has shed more light on the reliability of the DEM test with a better study design and methodology[12]. The results show that the DEM test has a within-session and between-session high reliability for VT and AHT but poor reliability for Ratio (the factor used to differentiate ocular movement dysfunction) and Errors. The authors conclude that the DEM test is a reliable test; but suggest that caution is necessary when interpreting data from a single test administration and when monitoring treatment for saccadic dysfunction. The same conclusion was also suggested by the authors of the DEM test in the last edition of manual[2]. Moreover, the authors underline the positive role of AHT as a reliable measure of visual processing speed to predict reading problems. The role of AHT as a global measure of functioning (naming, ocular movement and other cognitive factors) was also identified as being important in a study of validity of the DEM test[5].

Besides these concerns, another major issue that is still in question about the DEM test is the role of language on test outcomes. Given the simplicity of design, ease of use and the validation of the test, the DEM test has been used and studied in a number of countries. The effect of language on DEM test results has been of a particular interest and some controversy. Historically, in a study of normative data of the DEM test in a Spanish speaking population, it was reported that there are no differences in relative normative values for children from 7 years of age and above for a Spanish and English speaking population thus suggesting that the DEM results are independent of language[13].

Subsequent studies, however, have rejected this hypothesis. In fact, research with other languages

Table 1: Demographic data of subjects by age and grade.

Age	Female	Male	Total
6	30	35	65
7	94	102	196
8	91	109	200
9	81	107	188
10	81	103	184
11	39	38	77
12	44	35	79
13	43	42	85
14	21	27	48
Total	522	600	1122

Grade	Female	Male	Total
1st	35	38	73
2nd	119	138	257
3rd	79	103	182
4th	108	128	236
5th	56	62	118
6th	33	39	72
7th	48	36	84
8th	33	48	81
9th	11	8	19
Total	522	600	1122

has found different DEM normative values for each of the populations studied[14-16] concluding therefore that the DEM test results are substantially dependent on language-not only as speech but also as it relates to the culture of each population. Different reading achievement, length of word numbers, development curve and many other factors have been implicated resulting in different test norms specific for each language[14]. Thus, based upon these studies and somewhat debatable language influence, from a clinical perspective, it would be most appropriate to use language-specific norms to ensure valid and reliable DEM test results.

The purpose of this paper, therefore, is two-fold: first, to develop DEM normative values for the Italian population; and second, to compare these values to other normative groups (one USA-English, two Spanish, one Chinese, one Japanese, and one Portuguese) to investigate whether language is an important consideration in evaluating DEM test results and whether language-specific norms are necessary to properly interpret DEM outcomes.

Methods

Subjects

All subjects were selected from several school screening programs and were required to have parental authorization to participate in the study. The screenings were performed in 6 different public and 1 private schools in the Lombardy region of northern Italy. Three of these schools were located in a city and three in the surrounding countryside. The grade levels tested were from 1st to 9th grade based upon the Italian educational organization. The schools are typical of those throughout Italy since they all use a national curriculum for each grade.

The inclusion criteria for the study were as follows: subjects were required to wear their habitual optical correction if available, have a binocular visual acuity at near equal to or better than 0.8 decimal acuity or 20/25 Snellen acuity, no obvious binocular anomalies (such as strabismus or high phorias as assessed by cover test), a positive response on the TNO stereotest, and pass the DEM pretest. No other selection or inclusion criteria were applied.

A total of 1153 children were screened with 1122 subjects meeting the inclusion criteria. The demographic characteristics of subjects are outlined in Table 1. The rejected subjects did not meet the inclusion criteria or were not able to complete the pretest. This research was conducted in accordance with the guidelines outlined in the Declaration of Helsinki and approved by the local ethics committee.

Procedures

All testing for this study was performed by the authors A.F. and S.M., other qualified optometrists and senior optometric interns from the Optics and Optometry Program at the Università degli Studi di Milano Bicocca, Italy. They were all assisted and supervised by the authors A.F. and S.M.

To ensure that testing was performed in a standardized and consistent manner, each of the examiners were thoroughly trained in all aspects of the study, methodology, test procedures and subject instructions. Moreover, the test was conducted at each school in a room chosen to minimize noise and distractions and to provide uniform illumination at a level greater than 400 lux.

The DEM test was administered exactly as instructed in the test manual[1,2]. The subjects were seated in front of a desk suitable for the child's age

Table 2: Mean results of DEM subtests for each age group. Standard deviation in parenthesis.

Age	VT	AHT	Ratio	Error
6	72.29 (20.99)	108.12 (30,49)	1.53 (0.29)	14.9 (8.3)
7	52.74 (10.17)	75.01 (19.33)	1.43 (0.25)	7.9 (7.6)
8	45.77 (9.68)	59.91 (14.87)	1.31 (0.20)	4.0 (4.6)
9	41.98 (7.89)	52.04 (12.78)	1.24 (0.18)	2.6 (3.8)
10	38.13 (6.35)	44.72 (8.08)	1.18 (0.12)	2.0 (2.6)
11	35.06 (6.41)	39.49 (8.44)	1.13 (0.12)	1.7 (2.0)
12	31.55 (5.74)	35.34 (6.47)	1.12 (0.09)	1.1 (1.8)
13	29.71 (4.58)	33.16 (6.57)	1.12 (0.12)	1.2 (1.9)
14	29.01 (4.91)	32.33 (5.29)	1.12 (0.07)	0.6 (0.9)

and the test cards were placed on a 20º tilted lectern to maintain the proper distance. Also to make certain recording time was accurate and consistent a sports chronograph was used. Time started only after subject instructions were given and the subject began to actually read the numbers; time ended when the last digit was read.

All data acquired (time and errors) were recorded on a specific score sheet, similar to the original, but in Italian. To improve precision, all times data were recorded to two decimal digits. This same level of precision was used throughout testing and data collection and in the statistical analysis.

As referenced in the manual[1,2], the vertical time (VT) was the sum of A and B cards without correction of errors. The adjusted horizontal time (AHT) was corrected for errors of omission and addition, in exactly the same manner as specified in the manual.

Statistical analysis

For each age group, the mean and standard deviations for all variables (VT, AHT, Ratio, and Error) were calculated. Also to confirm the developmental trend over time and to assess differences between genders a factorial ANOVA was performed.

The DEM data collected were then compared to the original American norms (EN[1,2]), Spanish-1 (SP1[13]), Spanish-2 (SP2[17]), Chinese (CH[14]), Japanese (JP[15]) and Portuguese (PT[16]).

To assess the difference between norms, a one-way ANOVA was utilized. But, since variances between EN, SP1, SP2 and CH were found to be not homogeneous in the study for the Portuguese norms[16] and since we added two other groups (JP and IT) to this same dataset, it was important to apply the Welch approach to ANOVA.

For the 6 to 11 years of age, the analysis was performed for 7 different population norms (IT, EN, SP1, SP2, CH, JP, PT). For the 12-year-old group, the analysis was done with 5 norms (IT, EN, SP2, JP, PT). For the 13-year-old group, 4 norms were compared (IT, EN, JP, PT). But for the 14- year-old group, there are only norms for the Italian and Japanese populations, thus a direct comparison between the two was performed using the t-test with Welch correction[18].

Because the different groups of data present different variances, in order to make a comparison between DEM results from the Italian to the other groups, a Games-Howell post-hoc test[19] was performed.

And finally, for direct clinical use, we also developed a mean and standard deviation graph and the percentile distribution tables for the Italian population from 6 to 14 years of age.

Results

Mean and standard deviations for VT, AHT, Ratio and Errors for all age groups in the Italian population are reported in Table 2. This same data is also presented in graph form and in percentile distribution tables for VT, AHT, Ratio and Error in the Appendix for faster clinical application.

Even though the developmental trend has been well established[2,5] we determined the difference between age and gender in the same analysis. To accomplish this, a factorial ANOVA for each DEM component (VT, AHT, Ratio, Error) was performed with the factor Age with nine levels (from age 6 to 14) and the factor Gender with two levels (Male and Female).

The results show significant differences for all DEM variables for the main factor Age: for VT ($F(8, 1104)=177.63$, $p<0.0001$), AHT ($F(8, 1104)=240.63$, $p<0.0001$), Ratio ($F(8, 1104)=63.27$, $p<0.0001$) and errors ($F(8, 1104)=78.54$, $p<0.0001$)- demonstrating a clear and definite improvement with age. However, for the main factor Gender, and the interaction between Age and Gender, there was no significant difference for each of the DEM subtests (VT, AHT, Ratio, Error). For the Ratio, we found no evidence in the Italian population of the slight increase at age 7 years that was seen in other studies[1,5,13]. In Table 2, and in the figure reported in the Appendix, we note that with increasing age not only an improvement of performance but also a general reduction in the variability of data.

Table 3 presents the cumulative normative data for the seven populations (IT, EN, SP1, SP2, CH, JP,

Table 3: Mean and standard deviation for the norms of DEM test for Italian (IT), American (EN), Spanish (SP1, SP2), Chinese (CH), Japanese (JP) and Portuguese (PT). In the last column are reported the results of ANOVA (with correction for multiple comparison). The 14 years old groups were analyzed by t-test with Welch correction.

Age Group	Sub test	Mean (SD in parenthesis)							differences
		IT	EN	SP1	SP2	CN	JP	PT	p
6	VT	72.29 (20.99)	63.11 (16.59)	86.3 (23.02)	65.2 (19.07)	50.98 (13.16)	69.7 (24.0)	72.44 (17.49)	<0.001
	AHT	108.12 (30.49)	98.26 (32.61)	146.9 (41.6)	97.79 (24.59)	71.27 (18.45)	88.1 (29.3)	107.99 (24.49)	<0.001
	Ratio	1.53 (0.29)	1.58 (0.45)	1.67 (0.37)	1.56 (0.33)	1.41 (0.25)	1.27 (0.11)	1.53 (0.34)	<0.001
	Errors	14.9 (8.3)	15.22 (11.49)	7.7 (4.43)	13.54 (9.01)	9.66 (7.51)	6.0 (6.6)	8.41 (6.97)	<0.001
7	VT	52.74 (10.17)	54.83 (9.20)	58.27 (13.04)	53.89 (11.00)	43.34 (8.82)	42.3 (8.5)	60.39 (14.48)	<0.001
	AHT	75.01 (19.33)	87.94 (28.18)	81.38 (26.91)	73.48 (21.44)	57.14 (14.21)	62.5 (14.4)	81.47 (21.84)	<0.001
	Ratio	1.43 (0.25)	1.60 (0.41)	1.70 (0.29)	1.36 (0.26)	1.32 (0.16)	1.50 (0.31)	1.36 (0.24)	<0.001
	Errors	7.9 (7.6)	12.50 (12.91)	6.97 (6.41)	8.50 (7.88)	5.59 (6.72)	5.72 (6.2)	5.00 (5.84)	<0.001
8	VT	45.77 (9.68)	46.76 (7.89)	49.60 (9.65)	48.77 (13.08)	38.50 (7.99)	38.2 (7.1)	50.97 (10.61)	<0.001
	AHT	59.91 (14.87)	57.73 (12.32)	68.30 (19.57)	60.92 (17.81)	47.55 (10.60)	54.2 (10.9)	65.67 (17.62)	<0.001
	Ratio	1.31 (0.2)	1.24 (0.18)	1.40 (0.31)	1.25 (0.18)	1.24 (0.15)	1.43 (0.23)	1.30 (0.27)	<0.001
	Errors	4.0 (4.6)	4.61 (6.91)	5.55 (5.97)	6.37 (6.95)	3.96 (3.74)	2.21 (3.7)	2.33 (3.31)	<0.001
9	VT	41.98 (7.89)	42.33 (8.20)	42.37 (7.46)	43.00 (7.16)	36.53 (6.61)	36.4 (6.4)	48.10 (10.45)	<0.001
	AHT	52.04 (12.78)	51.13 (13.3)	54.54 (11.60)	50.84 (10.31)	43.03 (7.83)	47.3 (9.5)	57.72 (12.11)	<0.001
	Ratio	1.24 (0.18)	1.21 (0.19)	1.31 (0.22)	1.18 (0.14)	1.18 (0.12)	1.32 (0.21)	1.22 (0.21)	<0.001
	Errors	2.6 (3.8)	2.17 (4.10)	1.69 (2.66)	4.40 (5.31)	2.98 (3.28)	2.29 (2.7)	1.72 (3.67)	<0.001
10	VT	38.13 (6.35)	40.28 (7.43)	40.19 (7.65)	38.84 (7.00)	29.38 (6.00)	32.8 (6.0)	43.70 (9.85)	<0.001
	AHT	44.72 (8.08)	47.64 (10.11)	47.24 (10.37)	44.87 (10.58)	33.69 (8.72)	40.3 (7.7)	49.30 (9.91)	<0.001
	Ratio	1.18 (0.12)	1.19 (0.17)	1.25 (0.16)	1.15 (0.13)	1.14 (0.15)	1.26 (0.18)	1.14 (0.13)	<0.001
	Errors	2.0 (2.6)	1.91 (2.68)	0.81 (1.76)	2.65 (3.36)	1.56 (2.16)	0.97 (2.1)	0.76 (2.47)	<0.001
11	VT	35.06 (6.41)	37.14 (5.42)	36.16 (6.32)	36.32 (6.30)	29.83 (5.36)	29.7 (6.1)	38.33 (5.51)	<0.001
	AHT	39.49 (8.44)	42.62 (7.61)	43.49 (10.04)	40.63 (7.87)	32.87 (7.03)	36.1 (6.8)	42.98 (6.35)	<0.001
	Ratio	1.13 (0.12)	1.15 (0.13)	1.18 (0.15)	1.12 (0.12)	1.10 (0.11)	1.23 (0.14)	1.13 (0.12)	<0.001
	Errors	1.7 (2.0)	1.68 (2.34)	0.51 (1.68)	1.84 (2.96)	1.7 (2.72)	1.26 (3.8)	0.45 (1.63)	<0.001
12	VT	31.55 (5.74)	35.14 (5.87)		35.81 (7.34)		28.4 (6.5)	35.47 (5.57)	<0.001
	AHT	35.34 (6.47)	39.35 (8.11)		40.36 (8.97)		32.9 (7.6)	39.29 (5.20)	<0.005
	Ratio	1.12 (0.09)	1.12 (0.10)		1.13 (0.09)		1.16 (0.11)	1.12 (0.13)	n.s.
	Errors	1.1 (1.8)	1.11 (1.17)		2.61 (3.01)		1.15 (2.7)	0.43 (1.36)	<0.001
13	VT	29.71 (4.58)	33.75 (6.53)				26.5 (4.3)	34.13 (4.81)	<0.001
	AHT	33.16 (6.57)	37.56 (7.23)				31.4 (5.6)	37.46 (6.10)	<0.001
	Ratio	1.12 (0.12)	1.12 (0.12)				1.19 (0.14)	1.10 (0.10)	<0.05
	Errors	1.2 (1.9)	1.61 (2.15)				0.95 (2.1)	0.48 (1.70)	n.s.
14	VT	29.01 (4.91)					25.6 (4.7)		<0.01
	AHT	32.33 (5.29)					28.9 (5.7)		<0.05
	Ratio	1.12 (0.07)					1.13 (0.11)		n.s.
	Errors	0.6 (0.9)					0.51 (1.4)		n.s.

PT). To be noted, all populations were studied from ages 6 to 11 years; for age 12, norms are reported for the IT, EN, SP2, JP, and PT populations; for age 13, for the IT, EN, JP, and PT populations, and for age 14, norms are reported only for the IT and JP populations.

The last column in Table 2 indicates the statistical significance assessed with Welch ANOVA of the difference between norms for each variable for each population. As calculated, the p-value indicates almost one statistically significant difference in norms, age and DEM subtest, except for ages 12 and 14 in which ratio is not statistically significant for all populations in which it was reported and for age 13 and 14 in which there were no differences in errors reported. These results suggest that, for each group comparison, at least one normative value is statistically different than all other datasets.

Figure 1: *Comparison of DEM test Vertical Time with results of other studies.*

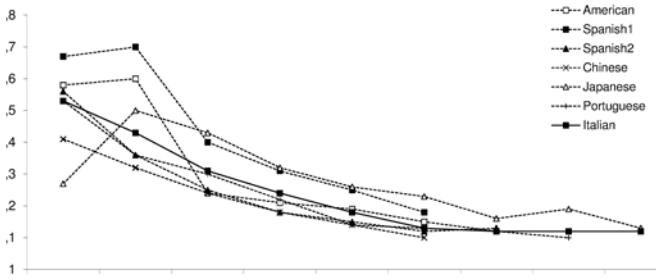

Figure 3: *Comparison of DEM test Ratio with results of other studies.*

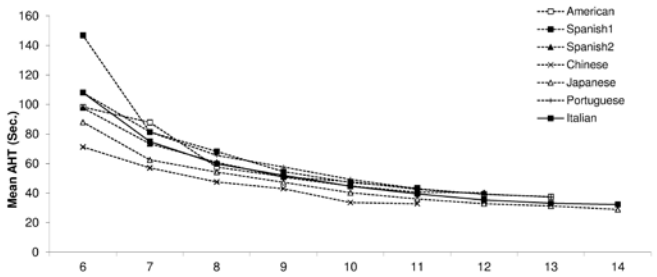

Figure 2: *Comparison of DEM test Adjusted Horizontal Time with results of other studies.*

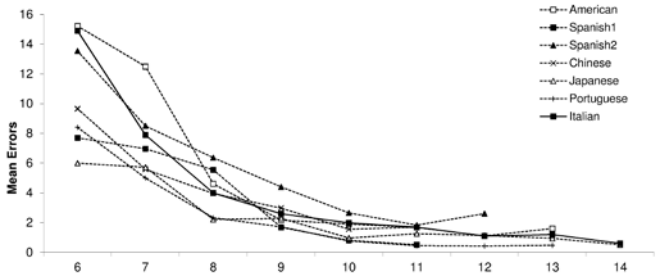

Figure 4: *Comparison of DEM test Errors with results of other studies.*

Table 4: Games-Howell Post-Hoc comparison between Italian and other DEM norms, p-value.

	Sub test	Age 6	7	8	9	10	11	12	13	14
IT-EN	VT	<0.01	n.s.	n.s.	n.s.	<0.05	<0.05	<0.01	<0.01	
	AHT	n.s.	<0.001	n.s.	n.s.	<0.05	<0.05	<0.01	<0.01	
	Ratio	n.s.	<0.01	<0.01	n.s.	n.s.	n.s.	n.s.	n.s.	
	Error	n.s.	<0.01	n.s.	n.s	n.s.	n.s.	n.s.	n.s.	
IT-SP1	VT	<0.001	<0.001	<0.001	n.s.	<0.01	n.s.			
	AHT	<0.001	<0.01	<0.001	<0.05	<0.01	<0.001			
	Ratio	<0.01	<0.001	<0.001	<0.001	<0.001	<0.01			
	Error	<0.001	n.s.	<0.01	<0.01	<0.001	<0.001			
IT-SP2	VT	<0.05	n.s.	<0.05	n.s.	n.s.	n.s.	<0.001		
	AHT	<0.05	n.s.	n.s.	n.s.	n.s.	n.s.	<0.001		
	Ratio	n.s.	<0.01	<0.01	<0.001	<0.05	n.s.	n.s.		
	Error	n.s.	n.s.	<0.001	<0.001	n.s.	n.s.	<0.01		
IT-CH	VT	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001			
	AHT	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001			
	Ratio	<0.05	<0.001	<0.01	<0.01	n.s.	n.s.			
	Error	<0.001	<0.05	n.s.	n.s.	n.s.	n.s.			
IT-JP	VT	n.s.	<0.001	<0.001	<0.001	<0.001	<0.001	<0.01	<0.001	<0.001
	AHT	<0.001	<0.001	<0.01	<0.01	<0.001	<0.05	n.s.	n.s.	<0.01
	Ratio	<0.001	n.s.	<0.001	<0.05	<0.01	<0.001	<0.05	<0.01	n.s.
	Error	<0.001	<0.05	<0.01	n.s.	<0.01	n.s.	n.s.	n.s.	n.s.
IT-PT	VT	n.s.	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	
	AHT	n.s.	<0.05	<0.01	<0.001	<0.001	<0.01	<0.001	<0.001	
	Ratio	n.s.	<0.05	n.s.	n.s.	<0.05	n.s.	n.s.	n.s.	
	Error	<0.001	<0.001	<0.001	n.s.	<0.001	<0.001	<0.01	<0.01	

Each of the figures 1 to 4 graphically represents the comparative results for each of the variables of the DEM test for each of the populations. Clearly for VT (Fig 1) and AHT (Fig 2) the Japanese and/or the Chinese have a faster response time at all age groups than any of the other populations. VT and AHT for the Italian population are generally faster than SP1 but slower when compared to CH and JP. However, it is necessary to emphasize that the Chinese norms were based upon an adjusted vertical time.

Interestingly, however, CH ratios are lower than IT for ages 6 to 9 and the JP Ratios are significantly higher for age 6, but lower for ages 8 to 13; SP1 Ratios are lower than IT for all age groups. Only ratio appears to be similar in the older children when the mean value reached is approximately 1.12. One other interesting observation is the disappearance of the "hump" in ratio at age 7 (an increase from age 6 to age 7) first noted in the American study and later found also in the Japanese and in the Spanish 1 studies; in the IT, SP2, CH and PT studies there is a gradual decrease in ratio from age 6 through age 7 and for older ages as well. (Fig 3).

As to errors (Fig 4), the Italians start with the largest number of errors at age 6, but diminish rapidly by age 7 to age 14. In comparison, the Portuguese have the least number of errors for all ages.

In comparing the Italian population with each of the other countries, the single comparison with Games-Howell post-hoc analysis most clearly demonstrates that there are no norms totally equal to the Italians norms. The level of significance of the differences in norms, as reported in Table 4, for each variable for each age group when comparing each country in figures 1 to 4 indicates that the greater the differences in norms from one country to the other, the greater the statistical significance; conversely, the smaller the differences, the less the statistical significance. Therefore, in this table, n.s. signifies that the norms are similar and not statistically significant.

Upon closer scrutiny of Table 4, for VT, the most similar data are between IT-EN and between IT-SP2- with some variability by age. When comparing AHT for the IT and other populations, the most similar results are between IT and SP2 from 7 to 11 years- again with some variability by age. The SP1, CH, JP norms up to age 11 and PT from age age 7 or greater appear significantly different. IT values for Ratio are similar for PT, except for ages 7 and 10, and for EN, except for ages 7 and 8. IT Errors are similar to EN, except for age 7group, but more dissimilar to the PT population. Other normative data present some differences and similarities depending on age.

Within the framework of statistical significance, it is clear that the EN and SP2 norms appear more closely aligned to the Italian norms and the JP, CH and SP1 the most different.

Discussion

One of the purposes of this study was to develop norms for the DEM test for the Italian population. Herein attached as an Appendix are the Mean and Standard Deviations for age-appropriate VT, AHT, Ratio, and Error, with their corresponding percentile rank. The tables presented are for ages 6 to 14 years and are designed for ease of use in direct clinical applications. The norms acquired in this study include children up to age 14 years. This represents an extension of the original norms and provides an opportunity to create a link between child and adult norms. In fact, preliminary adult norms for the DEM test in the American population[20] shows that adults generally perform faster than a 13 year old child, and thus require a specific norm for them as well.

With respect to ratio, in this study we did not find a higher value for the 7-year old group than the ratio for age 6 as was reported in the American, Spanish 1 and Japanese norms. The outcome of our present study is in line with other norms (SP2, CH, and PT) which clearly demonstrate a developmental trend in ratio. It is possible that different educational programs applied subsequent to the original American study in 1987 and the Spanish 1 study in 1995 could have modified this outcome, but it does not explain the Japanese results. The JP Ratio for the 6-year-old group may be an anomaly because it is similar to the ratio for the age 10 group. Consequently if we exclude the ratio for the age 6 group, the developmental trend is evident.

In the first edition of the DEM manual, the cut-off value that defines a pathological result was selected to be the 30th percentile. But the second edition of the DEM manual[2] shifted the criterion to the 16th percentile. This is in line with other psycho-educational tests[21]. This more restricted criterion clearly indicates the need for language and population specific norms in order to obtain more realistic and valid results.

The second purpose of this study was to investigate the role of language in the DEM test. The DEM may be considered a visuo-spatial test whose purpose is to evaluate ocular movements in a reading like condition; but if it were only a purely visuo-spatial test, the DEM would theoretically be independent of language. Clearly that is not the case. Since there is a large component of naming in the test, quantifiable between 64% and 90%[5], the DEM is therefore more than a visual-motor test and thus would suggest that language should play a significant role in test outcome.

Historically, the first study relative to an application of the DEM test in another language, other than American English, found that the test is independent of language[13]. However, succeeding studies[14,16], including this one, do not support this assertion.

The single comparison analysis using the Games-Howell post-hoc statistic to compare the Italian norms with those from other countries demonstrated that the Italian norms were most comparable to the original American and the second Spanish norms, but very dissimilar to the norms of all other countries. Clearly the Chinese and the Japanese groups present faster results than the other groups in the rapid naming of numbers as demonstrated by faster VT and AHT results. This outcome is also reported by Pang et al.[14]. In explaining the results, Pang hypothesized that this difference could not be due only to the difference in word length and its relative pronunciation. The very difference in the reading speed between CH and other populations may be related to the differences in the educational systems of the countries in question. Chinese children start the formal reading instruction early, typically by the age of three to four years[14]. In fact Chinese children learn to read one to two years earlier than the other children in the studies being compared in our research and this earlier development and training could very well explain the faster response times in the DEM.

Theoretically, VT and AHT are directly related to language; but Ratio is mathematically relatively independent and therefore should not change among different populations. This affirmation would hold true if language were taken strictly as speech. But since language is widely accepted as being not only speech but also related to other factors including age of learning to read, educational systems and programs, developmental curves, and more generally to culture; it is not inconceivable that ratio may differ among various populations. Thus it is not farfetched to find different normative values among different populations even with the same language as for example in UK (English) or in Latin America (Spanish or Portuguese).

Finding a difference between populations may also depend on the statistical method applied. As a matter of fact, when the Spanish (SP1 and SP2) to American comparison was performed[13,17] the simple t-test was used without considering the absence of homogeneity of variance and unequal sample numbers[18,19]. When the disparity in variances and in sample size is taken into account by using the appropriate statistical methods such as the Welch ANOVA and Games-Howell Post-Hoc comparison, the differences between populations becomes more noticeable. Consequently, it is possible that if different statistics were used in the first study[13], a more significant degree of difference would have been found between the American and Spanish populations.

Other factors that may possibly affect DEM results and indirectly the DEM norms could be related to the reliability of the test. As reported in three studies[10-12], using different methods, and in the DEM manual[2], the DEM test has good reliability for VT and AHT and poor reliability for the Ratio and Error. A *learning effect* has been observed in that the second presentation of the test generally yields faster results than the first[10,11] and some subjects change classification of behavioral types as defined by the DEM manual when using a pass-fail criteria[12]. However, this last study demonstrated a change in classification by using different criteria than the 16th percentile rank recommended in the manual. This different criterion could easily cause subjects to move from one classification to another and affect test outcomes.

Independent of reliability itself, what is clear is that the DEM test results need to be considered as a part of the totality of all clinical findings and patient history, and not simply based on a single pass or fail threshold on a single test[2,12] . Even the DEM manual recommends that in case of a pathological result the test should be administered twice. This procedure is also recommended in case of suspected learning difficulties[21] in which a full evaluation is necessary and not rely only on the DEM test for a definitive diagnosis.

Moreover, because the VT and the AHT appear to be a more valid[5] and reliable[12] measure than ratio, they can be taken as a global and reliable measure of visual processing speed to predict reading problems. For this reason, specific norms are required for these variables, in terms of mean value, standard deviation and percentile rank.

In addition to language or culture differences, test reliability only offers a partial explanation for the variability observed in the DEM norms for the different populations (all data acquired are for the first administration of the test). However, taking into account the limitations observed, further study is obviously needed to confirm this hypothesis.

Given these factors, and in spite of debatable reliability, the DEM test appears to be a language and culture dependent test for all variables investigated. The results of this study, with its expanded age range and statistical analyses, confirm the findings of Baptista et al.[16] while at the same time rejecting the older affirmation of Fernandez-Velasquez and Fernandez-Fidalgo[13] that the DEM was independent of language. With the addition of this present study of the Italian population, the DEM test now has normative data for seven different languages thus expanding its application as a clinical and screening test to many more countries all over the world.

Conclusion

We have presented norms for the DEM test for an extended age range of 6 to 14 years for the Italian population. The specific normative data provide an optimal and reliable clinical application of the DEM test. Comparison with other normative data suggests the application of a specific norm for each population because the DEM test appears to be a language and cultural dependent test for all variable investigated.

Acknowledgements

We would like to thank all examiners, students and screeners who have assisted in the collection of data.

References

1. Richman JE, Garzia RP. Developmental Eye Movement Test, Examiners booklet, version 1. South Bend, IN: Bernell Corp.; 1987.
2. Richman JE. Developmental Eye Movement Test, Examiner's manual, version 2.0. South Bend, IN: Bernell Corp.; 2009.
3. Garzia RP, Richman JE, Nicholson SB, Gaines CS. A new visual-verbal saccade test: the development eye movement test (DEM). J Am Optom Assoc 1990;61:124-35.
4. Maples WC. Oculomotor dysfunctions: Classification of saccadic and pursuit dysfunctions. In: Press LJ, editor. Applied concepts in vision therapy. St. Louis: Mosby, 1997: 120-36.
5. Facchin A, Maffioletti S, Carnevali T. Validity reassessment of Developmental Eye Movement (DEM) Test in the Italian population. Optom Vis Dev 2011;42:155-67.
6. Anastasi A, Urbina S. Psychological Testing, Seven Ed. Upper Saddle River NJ: Prentice Hall; 1997.
7. Lack D. Comparison of the developmental eye movement test, the visagraph numbers test with a test of the English language arts. J Behav Optom 2005;16:1.
8. Ayton LN, Abel LA, Fricke TR, McBrien NA. Developmental eye movement test: what is it really measuring? Optom Vis Sci 2009;86:722-30.
9. Webber A, Wood J, Gole G, Brown B. DEM test, visagraph eye movement recordings, and reading ability in children. Optom Vis Sci 2011;88:295-302.
10. Rouse MW, Nestor EM, Parot CJ, Deland PN. A reevaluation of the Developmental Eye Movement (DEM) test's repeatability. Optom Vis Sci 2004;81:934-8.
11. Tassinari JT, DeLand P. Developmental Eye Movement Test: reliability and symptomatology. Optometry 2005;76:387-99.
12. Orlansky G, Hopkins KB, Mitchell GL, Huang K, Frazier M, Heyman C, Scheiman M. Reliability of the developmental eye movement test. Optom Vis Sci 2011;88:1507-19.
13. 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;72:902-6.
14. 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;93:213-23.
15. Okumura T, Wakamiya E. Visual Skills in Children with Learning Difficulties: Meijitosho Shuppan Corporation 2010.
16. Baptista AM, de Sousa RA, de Morais Guerra Casal CC, Marques RJ, da Silva CM. Norms for the developmental eye movement test for portuguese children. Optom Vis Sci 2011;88:864-71.
17. Jimenez R, Gonzalez MD, Perez MA, Garcia JA. Evolution of accommodative function and development of ocular movements in children. Ophthalmic Physiol Opt 2003;23:97-107.
18. Welch BL. The significance of the difference between two means when the population variances are unequal. Biometrika 1938:350-62.
19. Games PA, Howell JF. Pairwise multiple comparison procedures with unequal N's and/or variances: a Montecarlo study. J Educ Stats 1976;1:113-25.
20. Powell JM, Birk K, Cummings EH, Ciol MA. The need for adult norms on the developmental eye movement test (DEM) J Behav Optom 2005;16:38-41.
21. AID. Consensus conference, Disturbi evolutivi specifici di apprendimento. Raccomandazioni per la pratica clinica definite con il metodo della Consensus Conference. In. Bologna: Associazione Italiana Dislessia; 2007.

(Figures and Tables pages 171- 179)

Figures and Tables

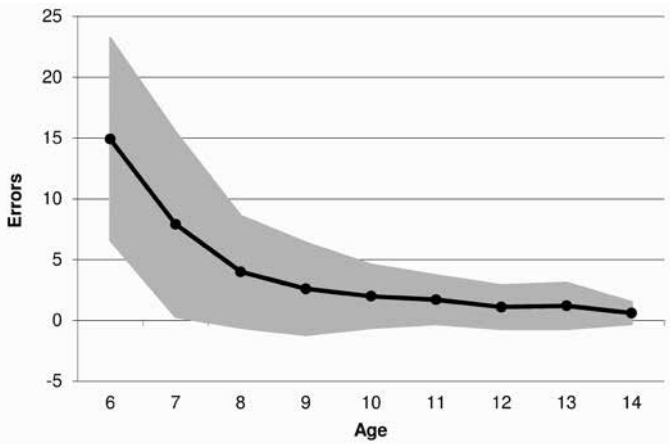

Figure A1: *Italian normative data: the dots indicate the mean value and shadows +/- 1 standard deviation for Vertical Time.*

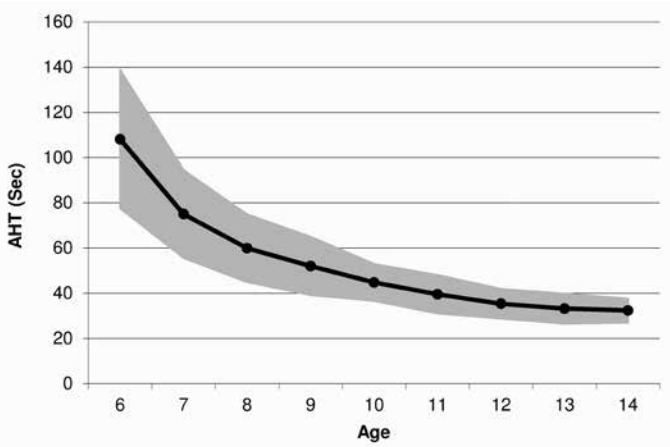

Figure A2: *Italian normative data: the dots indicate the mean value and shadows +/- 1 standard deviation for Adjusted Horizontal Time.*

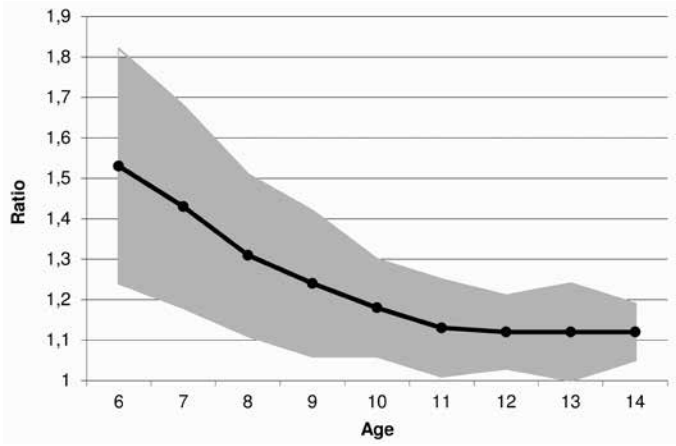

Figure A3: *Italian normative data: the dots indicate the mean value and shadows +/- 1 standard deviation for Ratio.*

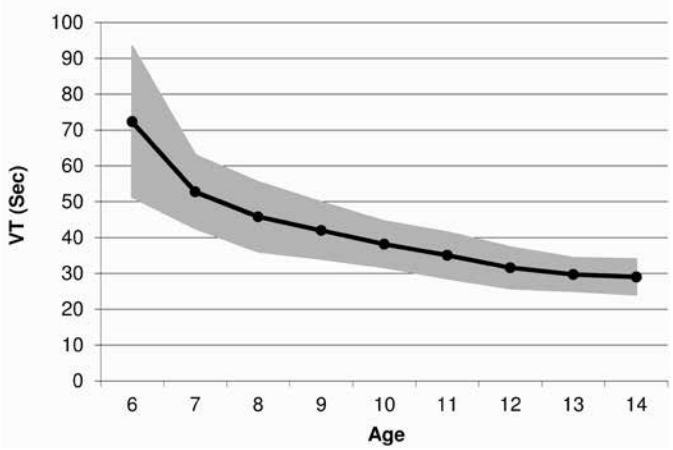

Figure A4: *Italian normative data: the dots indicate the mean value and shadows +/- 1 standard deviation for Errors.*

Table A1: Mean and standard deviation for the Italian norms of DEM test, separated for each age group.

Age	VT	AHT	Ratio	Error
6	72.29 (20.99)	108.12 (30,49)	1.53 (0.29)	14.9 (8.3)
7	52.74 (10.17)	75.01 (19.33)	1.43 (0.25)	7.9 (7.6)
8	45.77 (9.68)	59.91 (14.87)	1.31 (0.20)	4.0 (4.6)
9	41.98 (7.89)	52.04 (12.78)	1.24 (0.18)	2.6 (3.8)
10	38.13 (6.35)	44.72 (8.08)	1.18 (0.12)	2.0 (2.6)
11	35.06 (6.41)	39.49 (8.44)	1.13 (0.12)	1.7 (2.0)
12	31.55 (5.74)	35.34 (6.47)	1.12 (0.09)	1.1 (1.8)
13	29.71 (4.58)	33.16 (6.57)	1.12 (0.12)	1.2 (1.9)
14	29.01 (4.91)	32.33 (5.29)	1.12 (0.07)	0.6 (0.9)

Table A2: DEM norms for Italian-speaking from 6 years to 6 years and 11 months.

(n = 65)	Mean	SD
Vertical time (sec.)	72.29	20.99
Adjusted Horizontal time (sec.)	108.12	30.49
Ratio	1.53	0.29
Errors	14.9	8.3

Vertical time (sec.)	Percentile	Adjusted Horizontal Time (sec.)	Ratio	Percentile	Errors
41	99	52	1.13	99	0
44	95	64	1.16	95	0
50	90	74	1.22	90	3
53	85	77	1.24	85	5
55	80	80	1.26	80	7
57	75	89	1.32	75	9
59	70	92	1.36	70	9
61	65	94	1.38	65	11
64	60	98	1.42	60	13
66	55	100	1.44	55	14
69	50	102	1.47	50	15
71	45	105	1.51	45	16
72	40	108	1.53	40	19
73	35	113	1.58	35	20
78	30	121	1.63	30	20
83	25	125	1.69	25	21
89	20	132	1.72	20	23
96	15	138	1.79	15	24
102	10	153	1.92	10	26
119	5	179	2.12	5	27
135	1	192	2.59	1	33

Table A3: DEM norms for Italian-speaking from 7 years to 7 years and 11 months.

(n = 196)	Mean	SD
Vertical time (sec.)	52.74	10.17
Adjusted Horizontal time (sec.)	75.01	19.33
Ratio	1.43	0.25
Errors	7.9	7.6

Vertical time (sec.)	Percentile	Adjusted Horizontal Time (sec.)	Ratio	Percentile	Errors
34	99	41	0.99	99	0
39	95	52	1.07	95	0
42	90	53	1.16	90	0
44	85	56	1.19	85	0
45	80	59	1.22	80	1
46	75	61	1.24	75	1
47	70	64	1.28	70	2
48	65	66	1.30	65	4
48	60	67	1.32	60	5
49	55	70	1.36	55	5
51	50	71	1.38	50	6
52	45	74	1.42	45	7
53	40	75	1.44	40	10
54	35	79	1.50	35	10
57	30	81	1.53	30	11
59	25	83	1.56	25	11
60	20	90	1.61	20	13
62	15	94	1.67	15	15
66	10	101	1.76	10	17
74	5	113	1.92	5	23
84	1	150	2.15	1	34

Table A4: DEM norms for Italian-speaking from 8 years to 8 years and 11 months.

(n =200)	Mean	SD
Vertical time (sec.)	45.77	9.68
Adjusted Horizontal time (sec.)	59.91	14.87
Ratio	1.31	0.20
Errors	4.0	4.6

Vertical time (sec.)	Percentile	Adjusted Horizontal Time (sec.)	Ratio	Percentile	Errors
31	99	35	0.95	99	0
34	95	42	1.03	95	0
36	90	45	1.07	90	0
38	85	47	1.11	85	0
39	80	48	1.14	80	0
40	75	50	1.17	75	0
41	70	51	1.20	70	1
42	65	53	1.22	65	1
43	60	55	1.25	60	1
43	55	56	1.27	55	2
44	50	58	1.29	50	2
45	45	59	1.31	45	3
47	40	60	1.32	40	4
48	35	63	1.35	35	5
49	30	65	1.39	30	5
51	25	67	1.43	25	6
52	20	70	1.46	20	8
54	15	74	1.51	15	10
55	10	78	1.59	10	11
58	5	85	1.68	5	13
78	1	114	1.95	1	21

Table A5: DEM norms for Italian-speaking from 9 years to 9 years and 11 months.

(n = 188)	Mean	SD
Vertical time (sec.)	41.98	7.89
Adjusted Horizontal time (sec.)	52.04	12.78
Ratio	1.24	0.18
Errors	2.6	3.8

Vertical time (sec.)	Percentile	Adjusted Horizontal Time (sec.)	Ratio	Percentile	Errors
27	99	31	0.94	99	0
29	95	34	1.01	95	0
33	90	38	1.07	90	0
34	85	41	1.10	85	0
36	80	43	1.12	80	0
37	75	44	1.13	75	0
38	70	44	1.15	70	0
39	65	46	1.17	65	0
40	60	47	1.18	60	0
41	55	49	1.19	55	1
41	50	51	1.20	50	1
42	45	52	1.23	45	1
43	40	53	1.25	40	2
44	35	54	1.27	35	2
45	30	55	1.30	30	3
46	25	57	1.31	25	4
48	20	62	1.34	20	5
50	15	65	1.38	15	7
53	10	69	1.43	10	7
57	5	73	1.52	5	10
68	1	99	1.87	1	20

Table A6: DEM norms for Italian-speaking from 10 years to 10 years and 11 months.

(n = 184)	Mean	SD
Vertical time (sec.)	38.13	6.35
Adjusted Horizontal time (sec.)	44.72	8.08
Ratio	1.18	0.12
Errors	2.0	2.6

Vertical time (sec.)	Percentile	Adjusted Horizontal Time (sec.)	Ratio	Percentile	Errors
28	99	29	0.97	99	0
30	95	33	1.01	95	0
31	90	35	1.04	90	0
32	85	37	1.05	85	0
33	80	38	1.07	80	0
34	75	39	1.08	75	0
34	70	40	1.10	70	0
35	65	41	1.11	65	0
36	60	42	1.13	60	0
36	55	43	1.14	55	1
37	50	44	1.16	50	1
38	45	44	1.17	45	1
38	40	45	1.19	40	2
39	35	46	1.20	35	2
41	30	48	1.22	30	3
41	25	50	1.26	25	3
43	20	52	1.27	20	4
45	15	54	1.31	15	5
47	10	57	1.35	10	6
49	5	59	1.41	5	8
61	1	67	1.65	1	11

Table A7: DEM norms for Italian-speaking from 11 years to 11 years and 11 months.

(n = 77)	Mean	SD
Vertical time (sec.)	35.06	6.41
Adjusted Horizontal time (sec.)	39.49	8.44
Ratio	1.13	0.12
Errors	1.7	2.0

Vertical time (sec.)	Percentile	Adjusted Horizontal Time (sec.)	Ratio	Percentile	Errors
23	99	23	0.94	99	0
25	95	27	0.98	95	0
28	90	29	1.00	90	0
29	85	30	1.01	85	0
30	80	32	1.03	80	0
30	75	35	1.06	75	0
32	70	36	1.07	70	0
33	65	37	1.08	65	0
33	60	37	1.09	60	1
34	55	38	1.10	55	1
34	50	38	1.11	50	1
35	45	39	1.12	45	1
35	40	40	1.13	40	2
36	35	42	1.14	35	2
38	30	42	1.16	30	2
38	25	44	1.18	25	3
40	20	45	1.22	20	3
41	15	47	1.23	15	4
43	10	49	1.27	10	5
47	5	56	1.31	5	5
52	1	64	1.49	1	8

Table A8: DEM norms for Italian-speaking from 12 years to 12 years and 11 months.

(n = 79)	Mean	SD
Vertical time (sec.)	31.55	5.74
Adjusted Horizontal time (sec.)	35.34	6.47
Ratio	1.12	0.09
Errors	1.1	1.8

Vertical time (sec.)	Percentile	Adjusted Horizontal Time (sec.)	Ratio	Percentile	Errors
22	99	24	0.93	99	0
24	95	26	0.96	95	0
25	90	27	1.01	90	0
26	85	29	1.03	85	0
27	80	30	1.04	80	0
28	75	31	1.06	75	0
28	70	32	1.08	70	0
29	65	33	1.09	65	0
29	60	33	1.10	60	0
30	55	34	1.10	55	0
30	50	35	1.12	50	0
32	45	35	1.13	45	1
32	40	36	1.14	40	1
33	35	36	1.15	35	1
34	30	37	1.17	30	1
34	25	38	1.18	25	2
36	20	40	1.19	20	2
38	15	41	1.21	15	2
39	10	45	1.24	10	3
43	5	48	1.31	5	5
48	1	53	1.38	1	8

Table A9: DEM norms for Italian-speaking from 13 years to 13 years and 11 months.

(n = 48)	Mean	SD
Vertical time (sec.)	29.01	4.91
Adjusted Horizontal time (sec.)	32.33	5.29
Ratio	1.12	0.07
Errors	0.6	0.9

Vertical time (sec.)	Percentile	Adjusted Horizontal Time (sec.)	Ratio	Percentile	Errors
21	99	22	0.95	99	0
23	95	25	0.97	95	0
24	90	27	1.00	90	0
25	85	28	1.03	85	0
26	80	28	1.04	80	0
26	75	30	1.06	75	0
26	70	30	1.08	70	0
27	65	31	1.10	65	0
27	60	31	1.12	60	0
27	55	31	1.13	55	0
28	50	32	1.13	50	0
29	45	32	1.13	45	0
30	40	32	1.14	40	1
30	35	33	1.17	35	1
31	30	34	1.17	30	1
31	25	35	1.17	25	1
32	20	35	1.18	20	1
33	15	36	1.19	15	2
34	10	38	1.20	10	2
37	5	44	1.23	5	2
44	1	49	1.23	1	3