

# Application, limits, scoring and improvements of Groffman Visual Tracing test

Alessio Facchin<sup>1,2</sup>, Lavinia Giordano<sup>1</sup>, Giovanni Brebbia<sup>2</sup> and Silvio Maffioletti<sup>2,3</sup>

<sup>1</sup>Materials Science Department & COMiB Research Centre in Optics and Optometry, University of Milano-Bicocca, Milano, Italy.

<sup>2</sup>Institute of Research and Studies in Optics and Optometry, Vinci, Italy.

<sup>3</sup>Degree course in Optics and Optometry, University of Torino, Torino, Italy.

Received April 14, 2020, accepted June 13, 2020.

Correspondence: [alessiopietro.facchin@gmail.com](mailto:alessiopietro.facchin@gmail.com)

## Abstract

The Groffman Visual Tracing (GVT) test is a psychometric oculomotor test comprising two cards with five contorted and intersected lines for the clinical evaluation of ocular movement. The participant starts from the one of letters at the top of the page, follows the line from the letter, and reports the corresponding number at the bottom of the page. The aim of this study is to evaluate two claims made by the original author of the test: “it is a developmental test”, and “the feasibility of its application from primary school children up to adults”. This was achieved by using the GVT test and a simplified version of it.

In two consecutive experiments, two groups of children and adults were tested. In the first experiment, 75 children (1st, 3rd, and 5th grade) and 25 adults underwent the GVT test. In the second experiment, 115 children from 1st to 5th grade underwent a simplified version of the GVT test. Total scoring, accuracy and execution time were evaluated.

In the first experiment, a developmental trend was found, but 24% of children in the 1st and 3rd grades did not follow any lines correctly due to the difficulty of the test. In the second experiment, all participants were able to perform the test, and the accuracy improved significantly with age ( $p < 0.0001$ ), a sign of an evolutionary trend. The time required to follow the lines was found to decrease with age ( $p < 0.0001$ ), and the accuracy of simplified version was better than the standard version ( $p < 0.0001$ ).

A developmental trend was found, but the standard version of the GVT test has proven to be too difficult for younger children. The modified version provides best results. Children at or below the 5th grade should be tested using the modified version. Older children and adults can be tested with the standard version. Specific norms based on execution times and accuracy should be established.

*Keywords:* eye movements, Groffman test, visual tracing, saccade

## Riassunto

Il Groffman Visual Tracing (GVT) è un test psicomotorio costituito da due schede, sulle quali sono riprodotte cinque linee contorte che si intersecano e si sovrappongono più volte. I partecipanti iniziano da una lettera nella parte alta del foglio, seguono la linea e denominano il corrispondente numero, nella parte bassa del foglio. L'obiettivo di questo studio è di verificare due affermazioni dell'autore del test: “È un test per l'età dello sviluppo” e “Può essere adottato dall'inizio della scuola primaria fino all'età adulta”. Tale valutazione è stata effettuata attraverso l'uso della versione standard e di una versione semplificata del test.

In due esperimenti sono stati testati due gruppi di bambini

e di adulti. Nel primo esperimento, a 75 bambini (della 1°, 3° e 5° anno della scuola primaria) e 25 adulti è stato somministrato il GVT test. Nel secondo esperimento, a 115 bambini (dal 1° al 5° anno della scuola primaria) è stata somministrata una versione semplificata del test. Sono stati valutati il punteggio, l'accuratezza e il tempo di esecuzione.

Nel primo esperimento è stato trovato che la performance migliora con l'età, ma il 24% dei bambini del 1° e 3° anno non è riuscito a seguire correttamente alcuna linea, data la difficoltà del test. Nel secondo esperimento, tutti i partecipanti sono stati in grado di eseguire il test e l'accuratezza cresce significativamente con l'età ( $p < 0.0001$ ), a conferma di un trend evolutivo. Il tempo di esecuzione ha evidenziato un decremento con l'età ( $p < 0.0001$ ) e l'accuratezza della versione semplificata del test è risultata migliore rispetto alla versione standard ( $p < 0.0001$ ).

Un miglioramento dei risultati in relazione all'età è stato trovato, ma la versione standard del GVT test si è rivelata troppo difficile per i bambini più piccoli. La versione modificata del GVT test si è invece rivelata più adeguata. Pertanto, con i bambini dal 1° al 5° anno della scuola primaria, è opportuno utilizzare la versione semplificata del test. Nei bambini più grandi e per gli adulti, la versione standard è risultata adeguata. Sarà necessario in seguito definire i valori normativi di riferimento riguardanti il tempo di esecuzione e l'accuratezza.

*Parole chiave:* movimenti oculari, Groffman test, visual tracing, saccadi

## Introduction

In a policy statement defined by the American Academy of Optometry and the American Optometric Association (1997), it was stated that, in the visual testing of children with learning-related visual problems, a complete visual examination should be made (Scheiman & Rouse, 2006; Scheiman & Wick, 2019). In particular, visual pathway integrity, visual efficiency, and visual information processing need to be investigated. Specifically, the second area of visual efficiency included accommodation, binocular vision, and eye movements.

From a clinical point of view, for the evaluation of eye movements there are few standardized oculomotor tests available. Indeed, a precise and objective evaluation of ocular movement should be made objectively using an eye tracker (Scheiman & Wick, 2019, chapter 1). However, its cost and the length of time required for implementation, examination, and analysis restrict its primary use to research and it is not appropriate for use in clinical practice. Moreover, clear interpretation of the data is not easy because of the requirement for updated language-specific text and norms.

Consequently, other instruments, such as psychometric tests, were developed for the evaluation of eye movements (Richman & Garzia, 1987).

Several tests are available. The NSUCO is an observational structured scale test in which the examiner evaluates different aspects during the execution of a standard test of saccades and pursuit (Maples & Ficklin, 1990). The King-Devick is a visuo-verbal screening test to evaluate eye movements. It is based on the measurement of the speed of rapid number naming ((A. T. King, 1976)) and recently was applied as a concussion screening tool (Galetta et al., 2016; D. King et al., 2013). Similarly, the developmental eye movement test (DEM) is a visuo-verbal test to examine ocular movement in a reading like condition. It is widely used in developmental age groups and has norms for different languages (Baptista et al., 2011; Facchin et al., 2012;

Jimenez et al., 2003; Okumura & Wakamiya, 2010; Pang et al., 2010; Richman & Garzia, 1987; Serdjukova et al., 2016; Xie et al., 2016). In these “paper and pencil tests”, the functioning of ocular movement is derived indirectly from the overall performance. Other than eye movements, several cognitive functions are involved in the overall performance of these tests: sustained attention, number recognition and retrieval, visual verbal integration time, speaking time, visuo-spatial attention, and other cognitive skills. For these reasons, these tests are not pure oculomotor tests (Ayton et al., 2009), but their ease of application and usefulness have been well demonstrated (Facchin et al., 2011; Maples & Ficklin, 1990; Moiroud et al., 2018; Richman, 2009; Richman & Garzia, 1987). Over the years, the DEM and King-Devick tests have been applied in a large number of investigations of their application and also to assess their psychometrical properties (Facchin et al., 2011; Facchin & Maffioletti, 2018; Moiroud et al., 2018; Rizzo et al., 2016; Tjarks et al., 2013).

The DEM test was largely used, but one of its limitations (also valid for the King-Devick) is the presence of verbal naming that takes from between 64% and 90% of the total horizontal time (Facchin et al., 2011). For this reason, there is a requirement for an oculomotor test without the naming component and the Groffman Visual Tracing (GVT) test acts in this manner (Groffman, 1966). The GVT test is an oculomotor “tracing” test in which a participant is required to follow a line in a group of five crowded lines from a letter at the top of the page to a number at the bottom. The number of times the line is followed correctly and the time spent to achieve this produce the final score (Groffman, 1966). Since the author reported that there are no naming skills required, this represents a simple performance test related to ocular movement.

In the first article by the author, visual tracing was correctly defined as “the oculomotor skill used to follow a continuous stimulus from one point to another” (Groffman, 1966) and it is related to both saccades and pursuits. The GVT test is not a real test of pursuits because of the lack of moving objects.

For the purposes of the author (Groffman, 1966), the GVT test complies with the following requirements and criteria: it provides a quantitative measure of oculomotor ability; it is a purely visual test (without other senses); there are no language factors; it is independent of cognitive factors; it is a developmental test; and it is applicable from children in kindergarten to adults.

Although mentioned in several textbooks (Chinn, 2014; Levi & Carney, 2009; Press, 2008; Press & Moore, 1993; Scheiman & Wick, 2019; Solan, 1982), the GVT test has received little attention in the literature. Only a few studies (Cui et al., 2017; Groffman, 1993; Langaas et al., 2002; Smaakjær et al., 2018) and some dissertations have cited this test and its application.

Specifically, GVT was used for the clinical assessment of eye movements in children with reading disabilities and with developmental coordination disorder (Langaas et al., 2002). The authors found that children with deficits perform poorly on the GVT test compared with controls, and a high number of children failed the GVT. In stroke patients, GVT was used for the assessment of oculomotor dysfunction before and after vision therapy. The results show an improvement of GVT score after vision therapy (Smaakjær et al., 2018).

As reported by Scheiman and Wick, 2019, in chapter 1, no studies indicating the psychometric properties and application of the GVT test have been published. Our main aims were to examine some of these properties. Specifically, we wanted to test the two last assertions of the author: “it is a developmental test” and “it is applicable from primary school to adults”, together with performing a clinical evaluation of the test. We tested it in two consecutive experiments, as described below.

## Experiment 1: standard version of the GVT test

The aim of the first experiment was to test the application of the GVT test in three groups of children and in one group of adults in order to evaluate the developmental trend and the feasibility of its application to participants ranging from primary school children to adults.

## Methods

### Participants

Children were recruited during a school screening program, and adult participants enrolled informally as volunteers. Only children with written informed consent from their relatives permitting them to take part in the study were enrolled. A total of one hundred and four participants were initially enrolled, but four (3 children and 1 adult) did not meet the inclusion criteria and were excluded. Finally, four groups of 25 participants were created, demographic characteristics are shown in Table 1. Adult participants were recruited randomly among patients attending an optometric office. Inclusion criteria were the presence of normal binocular vision assessed by cover test, the absence of ocular diseases reported by the participant or relatives, no history of refractive surgery, strabismus or amblyopia reported by children or relatives, and a visual acuity equal to or greater than +0.1 logMAR in each eye at near using a LEA symbols logMAR chart (Goodlite 250800, Elgin, IL, USA). All participants had no current or previous neurological or psychiatric disorders. Participants wore their own glasses or contact lenses (if needed) during testing. The study was carried out in accordance with the guidelines given in the Declaration of Helsinki and it was approved by the Board of Optics and Optometry of the University of Milano-Bicocca (January 14, 2019).

Table 1: Demographic characteristics for the four groups of participants in Experiment 1.

Group	n	Mean age	SD	Range
1st class	25	6.5	0.5	6 – 7
3rd class	25	8.2	0.4	8 – 9
5th class	25	10.5	0.5	10 – 11
Adults	25	28.9	5.9	21 – 39
Total	100			

### Groffman Visual Tracing Test

According to the original paper (Groffman, 1966), the Groffman Visual Tracing test was an oculomotor test based on two cards of 216 × 279 mm (i.e. US letter size). Each card comprised five separate continuous lines that intersected one another in a tangled pattern. The task was to “follow” each line as rapidly as possible without losing the line pattern. Each participant was asked to follow each line from a letter at the top of the page (A, B, C, D, and E) to a number (1 to 5) at the bottom. Execution times and final recognized numbers were recorded. Firstly, a demonstration card was placed on the lectern, and the instructions about the start, intersections, and ends were given. If the participant did not understand, the instructions for a demonstration card were repeated up to three times. After three repetitions, if the participant could not follow a single line on the demonstration card correctly, testing was terminated because of the failure to attain the minimum level of skill required for the execution of the test.

According to the original paper, the instruction was: “This is a test to see how quickly and accurately you can follow a line using only your eyes. Look at the line that starts at the letter A. Follow it with your eyes. When it reaches another line (point to the first intersection),

follow it through the gap (point to the broken line). This line goes under the whole line and continues through." (Groffman, 1966). After the demonstration card, card A and B were administered in this order. The instruction for each card and line was: "Now we are going to trace five more lines. Your score will depend on accuracy and speed so work quickly, but try not to make a mistake." (Groffman, 1966).

The answer key was identical for cards A and B and was reported on a scoresheet. Scoring was performed using the original procedure, as reported in the test manual. Since the original score cannot differentiate between lower but accurate tracing (e.g. participant #4, 4 lines correctly followed slowly, 20 points) and fast but inaccurate tracing (e.g. participant #12, 2 lines correctly followed faster, 20 points), in order to have a better explanation of the results, we decided to take into account the accuracy (number of lines followed correctly that ranges from 0 to 5) and execution times separately, as many recent performance tests do. Consequently, in addition to the original scoring, the execution times of the lines correctly followed and the overall accuracy for each card were used for the analysis.

### Procedure

For the children the GVT test was administered during a school screening program. The overall evaluation was performed in a quiet and well-illuminated room (approximately 400 lux). Firstly, children performed the basic screening program that included visual acuity at far and near, objective refraction, stereopsis, near point of convergence, objective observation, and cover test. After these tests, the child was seated at the desk wearing the proper refractive correction (if necessary), and the different cards were positioned on a lectern at 40 cm. A stopwatch was used to record the execution time. The card was positioned on a lectern and lines were covered by a blank sheet in order to avoid the child following lines before the start of the test. In this phase, only letters at the top of the page were visible. The examiner then spoke the letter, removed the blank sheet, and started the time. The examiner stopped the stopwatch when the participant gave the corresponding number. The number and the execution time were recorded on a scoresheet. If the number reported was incorrect, scoring was zero. If the number was correct, the execution time was recorded. If the participant lost the line, scoring was zero. The original score of the GVT test was computed using the table reported in the test manual. Adult participants were tested in an office under the same conditions as described above.

### Statistical analysis and scoring

For comparative purposes, the data for the adults were analysed separately before they were included in the analysis of the children's data. Original scoring, accuracy, and execution times were analysed with a general linear model by using different structures of ANOVA to assess the evolutionary trend and to perform specific comparisons. Post-hoc analyses were performed with Tukey correction for multiple comparisons. Effect size was evaluated using partial eta squared. Since the execution times were available only for the lines followed correctly, we used these times for the analysis of the participants that have these data. Where appropriate, 95% confidence levels (CI) were reported. Statistical analyses and figures were performed with R statistical environment (R Core Team, 2019).

### Results

The clinical application of the GVT test in 1st-grade children showed that execution of the task was very difficult for children at this age (6-7 years old), and the larger part refused to perform the second card B. For this reason, we decided to apply and consider only card A to the overall group of children. The second card "B" was administered only to adult participants. Conse-

quently, the comparisons between groups were performed only for card "A" and cards "A" and "B" were compared only in adults.

### Original Score

The original score was analysed using a one-way ANOVA with the between-participants factor Group with 4 levels (1st grade, 3rd grade, 5th grade, Adults). The results show that the factor Group was significant [ $F(3,96) = 15.05, p < 0.0001, \eta_p^2 = 0.32$ ]. Post-hoc analyses showed significant differences between 1st and 5th grade ( $p < 0.001$ ), between 1st grade and Adults ( $p < 0.0001$ ), between 3rd and 5th grade ( $p < 0.05$ ) and between 3rd grade and Adults ( $p < 0.0001$ ). There was an improvement in performance with grade. The data are plotted in Figure 1.

On examining the raw data with respect to accuracy (see Table 2), we found that 24% of children in the 1st and 3rd grades could not correctly follow any lines and only 4% followed five lines correctly. In the 1st grade, 52% at best could follow only one line out of 5 on the first card "A" correctly.

Table 2: Accuracy of the GVT test according to group.

Group	Lines	Frequency	Percent	Cumulative percent
1st grade	0	6	24.0	24.0
	1	7	28.0	52.0
	2	6	24.0	76.0
	3	3	12.0	88.0
	4	2	8.0	96.0
	5	1	4.0	100.0
	Total	25	100.0	
3rd grade	0	6	24.0	24.0
	1	3	12.0	36.0
	2	9	36.0	72.0
	3	5	20.0	92.0
	4	0	0.0	92.0
	5	2	8.0	100.0
	Total	25	100.0	
5th grade	0	0	0.0	0.0
	1	3	12.0	12.0
	2	3	12.0	24.0
	3	6	24.0	48.0
	4	8	32.0	80.0
	5	5	20.0	100.0
	Total	25	100.0	
Adults	0	0	0.0	0.0
	1	1	4.0	4.0
	2	5	20.0	24.0
	3	2	8.0	32.0
	4	6	24.0	56.0
	5	11	44.0	100.0
	Total	25	100.0	

Note. For each group, a frequency, percent, and cumulative percent was reported.

The accuracy between groups was analysed using a one-way ANOVA with the between-participants factor Group with 4 levels (1st grade, 3rd grade, 5th grade, Adults). The results show a significant effect of Group [ $F(3,96) = 16.13, p < 0.0001, \eta_p^2 = 0.34$ ]. Post-hoc comparisons showed significant differences between 1st and 5th grade ( $p = 0.0001$ ), between 1st grade and Adults ( $p < 0.0001$ ), between 3rd and 5th grade ( $p < 0.001$ ),

and between 3rd grade and Adults ( $p < 0.0001$ ). The mean accuracy improved with grade. The data are listed in Table 2 and plotted in Figure 2.

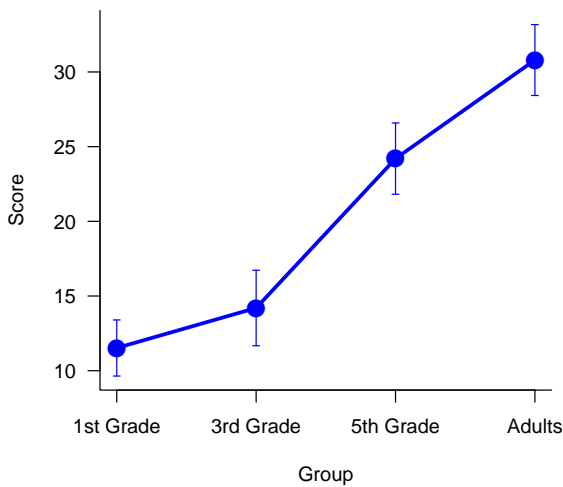


Figure 1: The score of GVT as a function of the group. Bars represent  $\pm 1$  standard error of the mean (SEM).

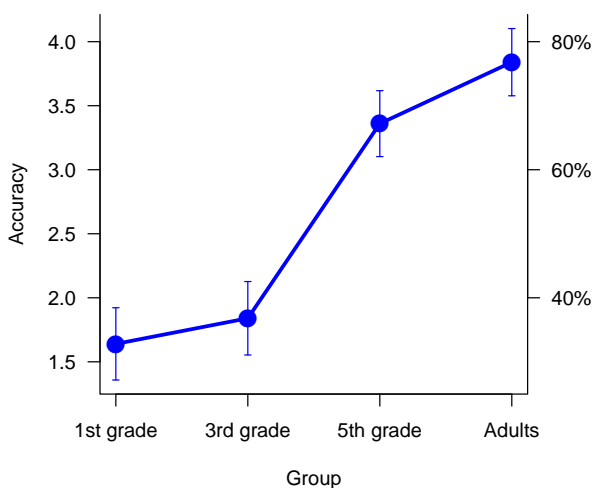


Figure 2: Evolution of accuracy during Grade for the GVT test. Bars represent  $\pm 1$  SEM.

The average score reported in the original paper (Groffman, 1966) was separated according to age. In order to compare our results with original norms for each child, we grouped all children participating according to their specific age. The comparisons for each age group with average score (t-test) showed a non-significant difference for all children and adults [in this last case the highest age available was used (12 and adults)]. However, large variability in our data explains these results. In fact, the comparison of variance (F-test) shows a significant difference ( $p < 0.05$ ) for all ages except for 11-year old children.

**Execution times**

Execution times were analysed with a factorial ANOVA, with the factor Group with four levels (1st grade, 3rd grade, 5th grade, Adults), and the factor Line with 5 levels (A, B, C, D, and E). The results show a significant result for Group [ $F(3,247) = 6.01, p < 0.001, \eta_p^2 = 0.07$ ], a significant effect for Line [ $F(4,247) = 7.59, p < 0.0001, \eta_p^2 = 0.11$ ] and no significant interaction between Group and Line ( $p = 0.97$ ). Post-hoc comparisons for the factor Group showed significant differences between 1st grade and Adults ( $p < 0.001$ ) and between 5th grade

and Adults ( $p < 0.05$ ). The data separated by Group are plotted in Figure 3. Execution times improved with group, but this improvement was small, and there were significant differences in the execution time for each line.

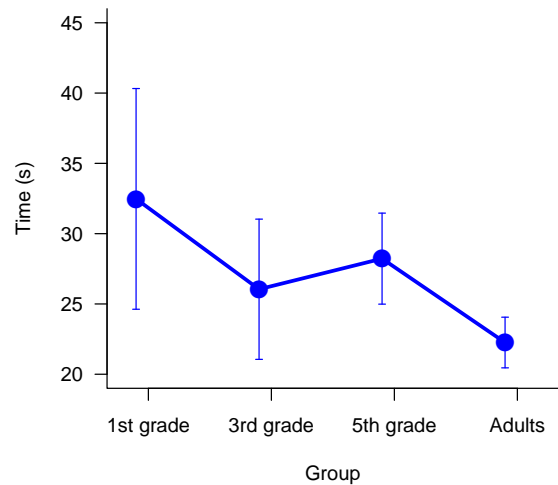


Figure 3: Execution times for the GVT test for card A as a function of Grade. Bars represent  $\pm 1$  SEM.

**Correlations between original score, accuracy and execution times**

Since all parameters that were evaluated (original score, accuracy, and times) improved significantly with grade, we tested their relationship. The results show a high positive correlation between original score and accuracy [ $r = 0.938 (0.909 - 0.958), p < 0.0001$ ] and a medium negative correlation between original score and time [ $r = -0.418 (0.229 - 0.577), p < 0.0001$ ]. The original GVT score was highly related to the accuracy, with a similar evolutionary trend as shown in Figures 1 and 3.

**The GVT test in adults**

Adult participants were able to perform both cards; their accuracy is reported in Table 3.

Table 3: Accuracy of the GVT test in Adults.

Card	Lines	Frequency	Percent	Cumulative percentage
A	1	1	4.0	4.0
	2	5	20.0	24.0
	3	2	8.0	32.0
	4	6	24.0	56.0
	5	11	44.0	100.0
	Total	25	100.0	
B	1	1	4.0	4.0
	2	4	16.0	20.0
	3	4	16.0	36.0
	4	4	16.0	52.0
	5	12	48.0	100.0
	Total	25	100.0	

Note. For each card, a frequency, percent, and cumulative percent was reported.

The accuracy was analysed using a paired sample t-test. The result reveals no significant difference in accuracy between cards ( $p = 0.87$ ).

For the execution times, data were analysed using a factorial ANOVA with the factor Card with two levels (A, B) and the factor Line with 5 levels (A – E). The results show a significant re-

sult for the factor Card ( $F(1, 183) = 7.63, p < 0.01, \eta_p^2 = 0.04$ ), for the factor Line [ $F(4, 183) = 4.50, p < 0.01, \eta_p^2 = 0.09$ ], and the interaction Card  $\times$  Line [ $F(4, 183) = 4.05, p < 0.01, \eta_p^2 = 0.08$ ]. Post-hoc comparisons for card A showed significant differences between line A and line C ( $p < 0.001$ ), between line B and line C ( $p < 0.05$ ), between line C and line D ( $p < 0.05$ ) and between line C and line E ( $p < 0.001$ ). For card B no significant differences were found. Each line on different cards had a different execution time. The results are plotted in Figure 4.

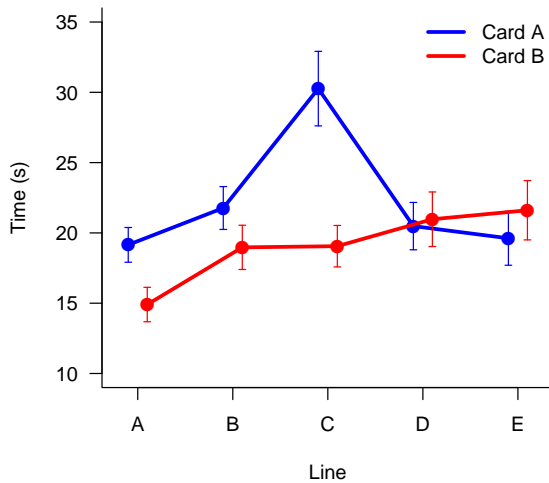


Figure 4: Execution times for Card A and B in adult participants. Bars represent  $\pm 1$  SEM.

## Discussion

The aim of this first experiment was to test the developmental trend and the feasibility of the GVT test in ages ranging from primary school children to adults.

Compared to the original norms, the children tested generally performed in the mean values, but with large variability. Using the score reported in the original paper, it is not possible to differentiate between fast and inaccurate and slower and accurate participants. For these reasons, we decided to assess the accuracy and execution times separately. Without this division, the results below could not have been evaluated.

Observing accuracy and execution times, the GVT test shows an evolutionary trend, and we can confirm the statement made by the author that the GVT test is a developmental one.

Conversely, considering the second question, the task is too difficult for children in the 1st and 3rd grade. In order to allow this test to be applied at an optimum level in young children, we have performed a simplified version, as described below for the second experiment.

## Experiment 2: simplified version of the GVT test

Based on the results of Experiment 1, with the aim of producing a better test for children, cards A and B were simplified by deleting lines B and D from both of them. Higher accuracy is expected compared to the original five lines and this raises the possibility of administering both cards even in younger children. The aim of this second experiment was to assess the performance of this modified version.

## Material and methods

### Participants

A different group of 115 children from 1st to 5th grade participated in the second experiment. They were equally subdivided into 23 participants for each grade. Inclusion and exclusion criteria were the same described earlier for Experiment 1.

## Tests

### Simplified GVT test

In order to simplify the GVT test, the original cards A and B were modified by deleting lines B and D from both of them. Consequently, the accuracy score was calculated over three lines for each card. For comparison between the original and simplified version, the accuracy was calculated in percent. Original and simplified version GVT cards are shown in Figure 5.

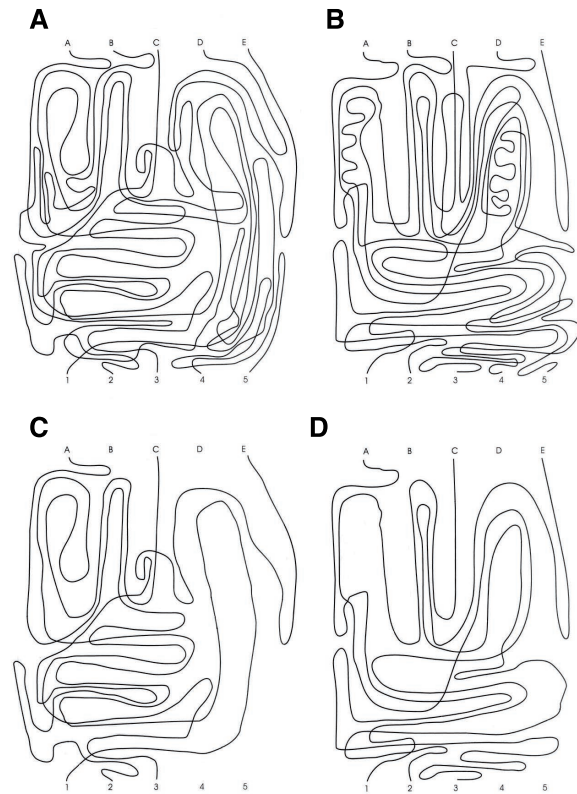


Figure 5: Original and simplified versions of the GVT test (demonstration card was not shown). Panels A and B show the original visual tracing test cards "A" and "B" respectively. Panels C and D show the simplified version of the GVT test cards "A" and "B".

## Procedure

The procedure was the same as that described earlier for Experiment 1.

## Results

### Accuracy

The accuracy for both cards is reported in Table 4.

In order to compare the accuracy of the simplified GVT test between grades, a repeated measure ANOVA was performed using the within factor Card with two levels (A and B) and the between factor Grade with five levels (1st – 5th). The results show a significant result only for the main factor Grade [ $F(4, 110) = 9.6, p < 0.0001, \eta_p^2 = 0.26$ ]. Post-hoc pairwise comparisons for the factor Grade showed significant differences between 1st and 4th grade ( $p < 0.0001$ ), between 1st and 5th grade ( $p < 0.0001$ ), between 2nd and 5th grade ( $p < 0.01$ ) and between 3rd and 5th grade ( $p < 0.05$ ). Both cards present the same accuracy and developmental trend. The data are plotted in Figure 6.

Table 4: The frequency relative to the accuracy of the simplified version of the GVT test with three lines for cards A and B.

Grade	Line	Card A			Card B		
		Frequency	Percent- age	Cumulative percentage	Frequency	Percent- age	Cumulative percentage
1	0	5	21.7	21.7	4	17.4	17.4
	1	7	30.4	52.2	9	39.1	56.5
	2	5	21.7	73.9	4	17.4	73.9
	3	6	26.1	100.0	6	26.1	100
	Total	23	100.0		23	100.0	
2	0	0	0.0	0.0	3	13.0	13.0
	1	6	26.1	26.1	7	30.4	43.5
	2	7	30.4	56.5	4	17.4	60.9
	3	10	43.5	100.0	9	39.1	100
	Total	23	100.0		23	100.0	
3	0	1	4.3	4.3	2	8.7	8.7
	1	6	26.1	30.4	5	21.7	30.4
	2	5	21.7	52.2	8	34.8	65.2
	3	11	47.8	100.0	8	34.8	100
	Total	23	100.0		23	100.0	
4	0	0	0.0	0.0	2	8.7	8.7
	1	3	13.0	13.0	1	4.3	13.0
	2	2	8.7	21.7	4	17.4	30.4
	3	18	78.3	100.0	16	69.6	100
	Total	23	100.0		23	100.0	
5	0	1	4.3	4.3	0	0.0	0.0
	1	1	4.3	8.7	2	8.7	8.7
	2	1	4.3	13.0	2	8.7	17.4
	3	20	87.0	100.0	19	82.6	100
	Total	23	100.0		23	100.0	

no interaction between Experiment and Grade. The accuracy of the simplified version of the GVT test is significantly higher than for the original GVT test for children in 1st, 3rd and 5th grade. The results are plotted in Figure 7.

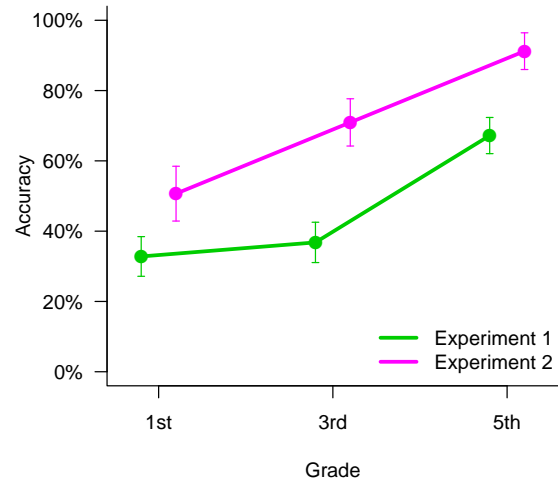


Figure 7: Comparison of accuracy between the standard (Experiment 1) and the simplified version (Experiment 2) of the GVT test. Bars represent ±1 SEM.

**Execution times**

In order to evaluate the difference in execution times, a mixed ANOVA was performed with the between factor Grade with five levels (1st – 5th), the within factor Card with two levels (A and B) and the factor Line with three levels (A, C, E). A significant result was found for the main factor Grade [ $F(4, 465) = 14.61, p < 0.0001, \eta_p^2 = 0.11$ ], for the main factor Line [ $F(2, 465) = 11.15, p < 0.0001, \eta_p^2 = 0.05$ ] and the interaction Card × Line [ $F(2, 465) = 7.05, p < 0.001, \eta_p^2 = 0.03$ ]. Post-hoc analyses for the factor Grade showed significant differences between 1st and 3rd grade ( $p < 0.001$ ), between 1st and 4th grade ( $p < 0.0001$ ), between 1st and 5th grade ( $p < 0.0001$ ), between 2nd and 4th grade ( $p < 0.0005$ ) and between 2nd and 5th grade ( $p < 0.0001$ ). For Card A, post-hoc analyses showed significant differences between line A and line B ( $p < 0.001$ ) and between line C and line E ( $p < 0.0005$ ). For card B post-hoc analyses showed significant differences between line A and line C ( $p < 0.05$ ) and between line A and line E ( $p < 0.01$ ). Execution times decrease with Grade and are different between cards and lines. The results are plotted in Figure 8.

**Discussion**

The aim of the second experiment was to assess the performance of the simplified version of the GVT test compared to the original one.

With the simplification of the test, the overall accuracy increased significantly between grades. Qualitatively, using the original test, the majority of young children refused to perform the second card of test (B) due to its difficulty. With the simplified version, all the children were able to perform both cards. Even modified cards were performed in the same order (as is necessary for a clinical application), and both exhibited a similar accuracy. These results imply that the accuracy does not improve between cards, and consequently, there is not a learning effect, as occurs in other tests (Facchin & Maffioletti, 2018).

Conversely, each line on each card has a different and a specific execution time. Based on this result, a single scoring system cannot be applied to the execution time of all lines. Each line on each card requires specific scoring parameters.

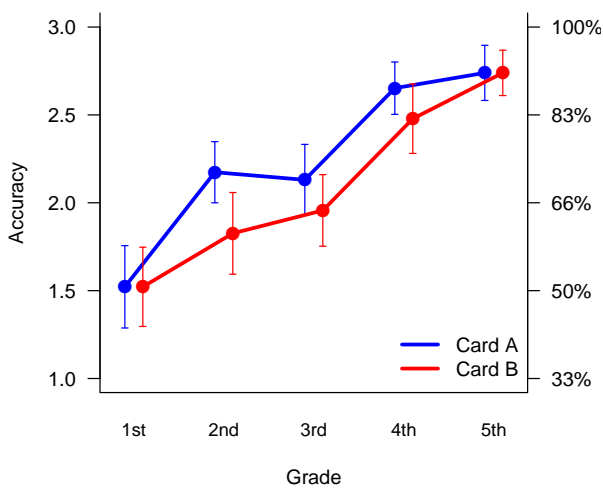


Figure 6: The accuracy between grades for the simplified version of the GVT test. Bars represent ±1 SEM.

More importantly, the relative accuracy for the 1st, 3rd and 5th grade from the two versions of the GVT test (original and simplified) scored in percent were compared using a factorial ANOVA with the factor Experiment with two levels (Experiment 1 and Experiment 2) and Grade with three levels (1st, 3rd, 5th). The results show a significant difference for the main factor Experiment [ $F(1, 138) = 26.1, p < 0.0001, \eta_p^2 = 0.16$ ], for the main factor Grade [ $F(2, 138) = 19.8, p < 0.0001, \eta_p^2 = 0.22$ ], but

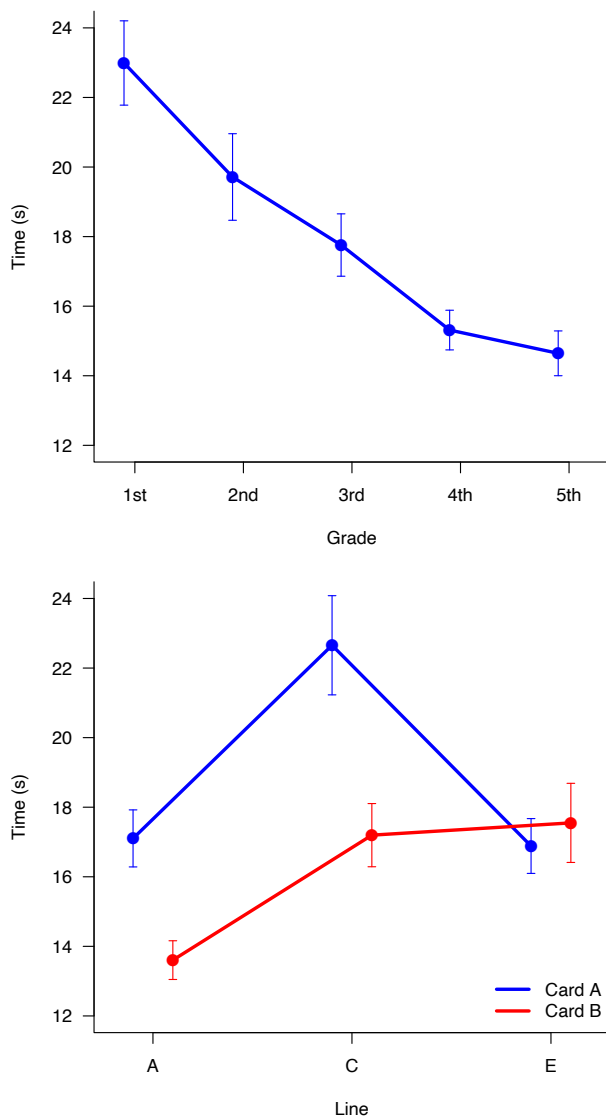


Figure 8: Execution times of the simplified test GVT. In the top graph, it was represented the mean execution times between Grades. In the bottom graph, the mean time of execution for each card and each line was reported. Bars represent  $\pm 1$  SEM.

We found that accuracy improves, and execution times decrease almost linearly according to grade in children. Consequently, the simplified GVT test is undoubtedly a developmental test.

### General discussion

In this study we have aimed to apply the GVT test in a group of children and adults in order to test the two last assertions of the author of the test: “it is a developmental test”, and “it is applicable from primary school to adults”, together with a clinical evaluation of the test. We have found several interesting results.

With respect to the application of a standard test in young children, examining in detail the data with respect to accuracy in card A, we found a very poor result. A total of 24% were unable to follow a single line, 28% could correctly follow one line, and only 4% perform all lines correctly. Children in the 3rd grade performed in a similar manner, and only in the 5th grade were there discrete results (no one failed to follow all lines, and 20% followed all lines correctly). Young patients with reading disabilities (Langaas et al., 2002) and adult patients with stroke seem to have the same problems (Smaakjær et al., 2018).

Using the original score, no differentiation between styles of execution was possible. In some cases, accurate but slower children received a score equal or lower than faster but inaccurate

children. The standard score was reported in the paper published in 1966, but there was no description of how this score was made. It was stated that 120 participants were tested, and the respective mean and standard deviation were reported. The original score was highly related to accuracy and moderately to time of execution.

Moreover, the time taken to execute the test was found to be different for each card and line. Consequently, a unique point – the scoring system seems not to be valid. For these reasons, we opted to analyse accuracy and execution times separately, as for the most part psychometric – performance tests do [e.g. in the domain of oculomotor test NSUCO (Maples & Ficklin, 1990) and DEM test (Richman & Garzia, 1987)].

Based on the overall results of the two experiments, we can review the criteria provided by the author, specifically, “GVT is a developmental test”. We confirm that the ability improves during grades, as shown clinically by an increase of accuracy and in decreasing the time of execution during grades, in particular using the simplified version of the test.

“GVT was applicable from primary school to adults”. This claim was partially correct. The GVT test was applicable in all ages but only in different forms or versions. In fact, the original version was too hard for the youngest children, and our modified and simplified version was found to be easier by children in primary school.

This study represents a baseline for future work that should take into account these observations. Future research on the GVT test could take into account its psychometric properties (validity, test-retest, and inter-examiner repeatability) and the development of specific norms, taking into account the overall accuracy (over 6 or 10 lines depending on age) and execution times separated for each card and line, preferably with modern scoring (i.e. percentile rank). The use of two separate scorings for execution times and accuracy permits clinicians to discriminate between different strategies that the participant may use.

When specific norms of GVT test are available, its first application will undoubtedly be in the field of oculomotor dysfunction, specifically in children with learning disabilities or other deficits. This test, together with other psychometric tests such as DEM and NSUCO could represent a valid and specific battery for oculomotor testing (Langaas et al., 2002).

In this study, we have reported that the GVT test demonstrates a clear evolutionary trend as an indication of validity, but the application of this test in adults could also be useful. It represents an oculomotor – performance test, and with the specific norms, it could be a valuable test in the evaluation of patients with special needs (Taub et al., 2012), in particular those with specific oculomotor problems such as patients with brain-injury (Gallaway et al., 2017; Scheiman et al., 2017; Smaakjær et al., 2018) and adults with learning disability.

### Acknowledgements

We would like to thank Silvia Tavazzi for helpful comments on an early version of the manuscript.

© Copyright Facchin, A., et al. This article is distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use and redistribution provided that the original author and source are credited.

### References

- Ayton, L. N., Abel, L. A., Fricke, T. R., & McBrien, N. A. (2009). Developmental eye movement test: what is it really measuring? *Optometry and Vision Science*, *86*, 722–730. <https://doi.org/10.1097/OPX.0b013e3182195dae>
- Baptista, A. M., de Sousa, R. A., de Moraes Guerra Casal, C. C., Marques, R. J., & da Silva, C. M. (2011). Norms for the developmental eye movement test for Portuguese children. *Optometry and Vision Science*, *88*, 864–871. <https://doi.org/10.1097/OPX.0b013e3182195dae>

- Chinn, S. (2014). *The Routledge International Handbook of Dyscalculia and Mathematical Learning Difficulties*. Taylor & Francis.
- Cui, J., Zhang, Y., Cheng, D., Li, D., & Zhou, X. (2017). Visual Form Perception Can Be a Cognitive Correlate of Lower Level Math Categories for Teenagers. <https://doi.org/10.3389/fpsyg.2017.01336>
- Facchin, A., Maffioletti, S., & Carnevali, T. (2011). Validity reassessment of developmental eye movement (DEM) test in the Italian population. *Optometry and Vision Development*, 42, 155–167. <https://doi.org/10.3389/fpsyg.2018.01279>
- Facchin, A., & Maffioletti, S. (2018). The reliability of the DEM test in the clinical environment. *Frontiers in Psychology*, 9.
- Facchin, A., Maffioletti, S., & Carnevali, T. (2012). The developmental eye movement (DEM) test: normative data for Italian population. *Optometry and Vision Development*, 43.
- Galetta, K. M., Liu, M., Leong, D. F., Ventura, R. E., Galetta, S. L., & Balcer, L. J. (2016). The King-Devick test of rapid number naming for concussion detection: meta-analysis and systematic review of the literature. *Concussion*, 1(2). <https://doi.org/10.2217/cnc.15.8>
- Galloway, M., Scheiman, M., & Mitchell, G. L. (2017). Vision therapy for post-concussion vision disorders. *Optometry and Vision Science*, 94(1), 68–73. <https://doi.org/10.1097/OPX.0000000000000935>
- Groffman, S. (1966). Visual tracing. *Journal of American Optometric Association*, 37, 139–141.
- Groffman, S. (1993). Correlation between cognitive processing and ocular motility. *Optometry and Vision Science*, 70(5), 380–383.
- Jimenez, R., Gonzalez, M. D., Perez, M. A., & Garcia, J. A. (2003). Evolution of accommodative function and development of ocular movements in children. *Ophthalmic and Physiological Optics*, 23, 97–107. <https://doi.org/10.1046/j.1475-1313.2003.00093.x>
- King, A. T. (1976). *The proposed King-Devick Test and its relation to the Pierce Saccade and reading level*. Available from the Carl Shepherd Memorial Library.
- King, D., Brughelli, M., Hume, P., & Gissane, C. (2013). Concussions in amateur rugby union identified with the use of a rapid visual screening tool. *Journal of the Neurological Sciences*, 326(1-2), 59–63. <https://doi.org/10.1016/j.jns.2013.01.012>
- Langaas, T., Aadne, R., Dahle, E. C., Hansen, A. M. H., Horgen, G., Myhren, D. J. T., Myklebust, A., Ronnestad, T., Storengen, T., & Wiik, M. S. (2002). Clinical assessment of eye movements in children with reading disabilities and with Developmental Coordination Disorder (DCD). *Investigative Ophthalmology & Visual Science*, 43(13), 4661.
- Levi, D. M., & Carney, T. (2009). Crowding in Peripheral Vision: Why Bigger Is Better. *Current Biology*, 19(23), 1988–1993. <https://doi.org/10.1016/J.CUB.2009.09.056>
- Maples, W. C., & Ficklin, T. (1990). Comparison of eye movement skills between above average and below average readers. *Journal of Behavioural Optometry*, 1(4), 87–91.
- Moiroud, L., Gerard, C. L., Peyre, H., & Bucci, M. P. (2018). Developmental Eye Movement test and dyslexic children: A pilot study with eye movement recordings. *PLoS One*, 13(9), e0200907. <https://doi.org/10.1371/journal.pone.0200907>
- Okumura, T., & Wakamiya, E. (2010). *Visual Skills in Children with Learning Difficulties*. Meijitoshu Shuppan Corporation.
- Pang, P. C., Lam, C. S., & Woo, G. C. (2010). The Developmental Eye Movement (DEM) test and Cantonese-speaking children in Hong Kong SAR, China. *Clinical and Experimental Optometry*, 93, 213–223. <https://doi.org/10.1111/j.1444-0938.2010.00470.x>
- Press, L. J. (2008). *Applied Concepts in Vision Therapy*. Optometric Extension Program Foundation.
- Press, L. J., & Moore, B. D. (1993). *Clinical Pediatric Optometry*. Butterworth-Heinemann. <https://books.google.it/books?id=YNpsAAAAMAAJ>
- Richman, J. E. (2009). *Developmental Eye Movement Test, Examiner's Manual, version 2.0*. Bernell Corp.
- Richman, J. E., & Garzia, R. P. (1987). *Developmental Eye Movement Test, Examiners Booklet, version 1*. Bernell Corp.
- Rizzo, J.-R., Hudson, T. E., Dai, W., Birkemeier, J., Pasculli, R. M., Selesnick, I., Balcer, L. J., Galetta, S. L., & Rucker, J. C. (2016). Rapid number naming in chronic concussion: eye movements in the King-Devick test. *Annals of Clinical and Translational Neurology*, 3(10), 801–811. <https://doi.org/10.1002/acn3.345>
- Scheiman, M., & Rouse, M. W. (2006). *Optometric management of learning-related vision problems*. Elsevier Health Sciences.
- Scheiman, M., Talasan, H., Mitchell, G. L., & Alvarez, T. L. (2017). Objective assessment of vergence after treatment of concussion-related CI: a pilot study. *Optometry and Vision Science*, 94(1), 74. <https://doi.org/10.1097/OPX.0000000000000936>
- Scheiman, M., & Wick, B. (2019). *Clinical Management of Binocular Vision: Heterophoric, Accommodative, and Eye Movement Disorders* (5th). Lippincott Williams & Wilkins.
- Serdjukova, J., Ekimane, L., Valeinis, J., Skilters, J., & Krumina, G. (2016). How strong and weak readers perform on the Developmental Eye Movement test (DEM): norms for Latvian school-aged children. *Reading and Writing*, 1–20. <https://doi.org/10.1007/s11145-016-9671-7>
- Smaakjær, P., Tødtén, S. T., & Rasmussen, R. S. (2018). Therapist-assisted vision therapy improves outcome for stroke patients with homonymous hemianopia alone or combined with oculomotor dysfunction. *Neurological Research*, 40(9), 752–757. <https://doi.org/10.1080/01616412.2018.1475321>
- Solan, H. A. (1982). *The treatment and management of children with learning disabilities*. Charles C Thomas Pub Ltd.
- Taub, M. B., Bartuccio, M., & Maino, D. (2012). *Visual Diagnosis and Care of the Patient with Special Needs*. Wolters Kluwer Health.
- Tjarks, B. J., Dorman, J. C., Valentine, V. D., Munce, T. A., Thompson, P. A., Kindt, S. L., & Bergeron, M. F. (2013). Comparison and utility of King-Devick and IMPACT® composite scores in adolescent concussion patients. *Journal of the Neurological Sciences*, 334(1-2), 148–153. <https://doi.org/10.1016/j.jns.2013.08.015>
- Xie, Y., Shi, C., Tong, M., Zhang, M., Li, T., Xu, Y., Guo, X., Hong, Q., & Chi, X. (2016). Developmental Eye Movement (DEM) test norms for Mandarin Chinese-speaking Chinese children. *PLoS One*, 11, e0148481. <https://doi.org/10.1371/journal.pone.0148481>