

# The Wold Sentence Copy Test



## Academic Performance

■ W.C. Maples, O.D., M.S.

### Abstract

*The Wold Sentence Copy Test (Wold) is a standardized visual motor test, that was designed to be used as a screening device for teachers, according to the author. It is easily administered in a group setting. In this study we sought to verify and possibly update the previously published school grade specific Wold norms.*

*Some 2561 administrations of the Wold were given to 540 children, grades one through five, in the course of six examination periods over three consecutive school years. Standardized academic performance data (Iowa Test of Basic Skills-ITBS) were furnished by the school system. The various demographic variables, of age, gender, grade and race were analyzed to develop Wold norms. Correlations were calculated between the Wold score and academic performance as measured by the ITBS. Several arbitrary failure criteria were analyzed. Relative rates of risk for the different ITBS were calculated and failure criteria were developed based upon these relative risks. Normative data from this study showed that previously published norms for the reported grades were low. Girls scored higher than boys. The Wold scores were significantly and directly correlated to most of the ITBS scores when all grades and gender scores were considered. Relative risk of failure for each academic subject was calculated, and the best predictor of academic performance by the Wold score was ascertained. Relative risk generally decreased as grade/age increased. A failure criterion for each grade*

*using the new Wold norms represented a score greater than  $\frac{1}{2}$  standard deviation below the mean, for a particular grade level. This criterion was a reliable predictor of performance on the ITBS of greater than five months below the established mean grade score for specified ITBS subtests. The new Wold norms are presented.*

### Key Words

*failure criteria, grade norms, Iowa Test of Basic Skills, visual motor integration, Wold sentence Copy Test*

### INTRODUCTION

The Wold Sentence Copy Test (Wold) was first described in 1970 as one of a battery of tests recommended to the classroom teacher to identify visual motor difficulties.<sup>1,a</sup> The test consists of a sentence at the top of a page which is to be copied by the subject at the bottom half of the page. (See Appendix A.) The test has subsequently been referenced in several texts and is one of the tests employed by some optometrists to evaluate visual motor skills.<sup>2-4</sup> Wold recommended that the examiner observe the patient's posture, number of fixations, spacing, vocalization, concentration, attention, fatigue, symbol formation and frustration level while the patient performs the test.<sup>1</sup> Wold

did not elaborate on how this information was noted or what one would do with this data. Norms for grades two through eight have been published.<sup>1,2</sup>

A multi-year prospective epidemiological study investigated the impact of vision and visual skills on elementary school children in Oklahoma.<sup>5,6</sup> A total of 540 children were enrolled as subjects. As part of this project, the Wold was administered on six different occasions, to the same children who were available at the time of testing, over the course of three years. A total of 2,561 evaluations were performed. The Wold test was found to be a predictor of some standardized academic scores, specifically, the Iowa Test of Basic Skills (ITBS).<sup>5</sup> This standardized test (ITBS) is often used by public education in Oklahoma to evaluate the student's academic performance. The norms are updated periodically.<sup>6</sup> In an evaluation of the Wold and ITBS data, the correlations of the Wold with the ITBS were found to be statistically significant in nine of 21 academic sub-categories.<sup>6</sup> These categories included four aspects of math, the core battery, verbal receptive (listening), verbal expressive (language expression), reading, and overall academic performance. The utility of the Wold is further evident because it can be administered in a group setting and requires only one minute to administer. The score is the number of symbols completed in 60 seconds and allows for the use of parametric statistical methods to analyze the data.

The easily administered and scored Wold test has the potential for the class-

room teacher to identify at risk children in the early grades. Visual maturity impacts academic performance.<sup>7-10</sup> Such a test would be very helpful for identifying children early in the year before the first standardized tests are administered. It would also be helpful to schools who do not routinely give standardized tests. It would be a very cost effective method of identifying these children so that remedial programs to address the child's potential deficits can be then designed and implemented. This would be particularly true when standardized testing is not available and would have the potential of decreasing the negative experience of failure. Children who encounter negative reinforcement from being a non-achiever may ultimately be helped to improve his/her quality of life. Early identification leading to early remediation has the benefit of increasing his/her potential earning power as an adult.<sup>6</sup> Alternately, children who fail in school are more likely to have difficulty competing in the workplace as adults.<sup>5,6</sup> They are more likely to need public assistance and to be socially delinquent.

## PURPOSE

This study has been undertaken to verify and possibly improve the Wold norms for the grades and ages represented in the present study. A second goal was to ascertain if a failure criterion on the Wold could be predicative of overall classroom performance. This simple test would provide the teacher with an early opportunity to confirm which children might be at academic risk.

## SUBJECTS

Some 2,561 evaluations were performed on 540 students during the three years of the study. Ideally, the total number of evaluations should have been 3,240 but children who were not available the day of the evaluation, had moved away or for various other reasons did not participate in each evaluation, caused the number to be less than the potential number. The ages varied according to the school grade: thus, those in first grade were about age 6, those in the second grade were about age 7, etc.

## METHODS

The classroom teachers were trained to administer the Wold test. The Wold was administered twice per year for each class: once at the beginning of the school year, in

the fall (September/October), and once toward the end, in the spring (March/April) for three consecutive years, from grades one through three, two through four and three through five. The children were allowed to perform the copy task for one minute, as recommended by Wold and Treganza.<sup>2</sup> The only exception to the one minute time was in testing the first grade, where the child was allowed two minutes. This extra time was given to ensure that the child, just beginning the writing task in the classroom, would have adequate time to perform the test. An average symbols/minute value was then calculated for the first grade scores. Although size constancy of the symbols and spacing were subjectively judged in the analysis of the child's performance, only the symbols/minute are reported here.

We averaged the Wold data by grade, gender and ethnic origin. These means were compared by a student *t*-test. We arbitrarily set various failure criteria for the Wold and the nine sub-categories of the ITBS.<sup>6</sup> Relative risk rates were calculated for each of these criteria. Relative risk is an epidemiological concept wherein individuals who share a particular condition are used to calculate a pass/fail ratio on a second condition.<sup>11,12</sup> The individuals were broken down into two groups: one group who scored above and a second group who scored below a designated Wold failure criteria. In these two groups (Table 1) the relative risk was computed by taking the percent of individuals who scored above the failure criteria for the ITBS category and dividing it by the percent of individuals who scored below the failure criteria for the ITBS. A ratio was then formulated by taking the relative risk for those who fell above the Wold failure criteria and dividing it by the relative risk for those that fell below the Wold failure criteria. This final ratio was defined as the relative risk of failure. This ratio is illustrated in Table 1.

## RESULTS

A total of 2,561 Wold evaluations were performed on 540 students during the three years of the study. There was no statistical difference in the ethnic scores between groups on the Wold test. However girls scored significantly better than boys and as the subjects increased in age the Wold scores improved. There was not a significant difference between the sub-

Table 1. Relative Risk Ratio Calculation		
	<-0.5GL ITBS Score	>-0.5GL ITBS Score
<-0.5s Wold Score	A	B
>-0.5s Wold Score	C	D

*A, B, C and D represent these row percentages  
Relative Risk of Failure = (B/A)/(D/C)*

ject's Wold scores by age and their Wold scores by grade therefore only Wold grade scores will be presented here.

Table 2 contains the total number of children who took each of the designated ITBS tests.<sup>6</sup> The same child is represented several times in this table, during the same test period, yielding a higher number than the 2,561 total evaluations. Two sub-categories, listening and word analysis, were not administered in the third or fourth grades (not included in the ITBS battery for these grades). The fifth grade was excluded from the analysis because of the low number of students in the different categories. Since fifth grade is not represented, the subject numbers for the individual grades do not sum to the combined subject total.

Table 3 contains the previously published means of the Wold.<sup>1,2</sup> It also includes the pooled (when both genders are combined) and gender averages, and their respective standard deviations, as found in the present study. The pooled score (both genders) in symbols per minute for first grade was lowest with 21 symbols/minute, improving to almost 82 symbols/minute by the fifth grade. A symbol here includes not only the letters but all punctuation marks. Analysis of race was not found to be significantly different in this study but gender, grade and age were found to be significantly different.

Means were further calculated to reflect scores at the beginning of the year and once again at the end of the year. Table 4 contains these pooled and gender scores for each season of the school year. Scores increased with age revealing the developmental nature of the Wold test. The one exception to this trend was between the spring of the third year and the fall of the fourth academic year, where the Wold score decreased slightly. The increasing score trend was true for the pooled as well as for the individual gender scores. Females scored statistically sig-

**Table 2. Total Number of Children Administered the ITBS in Previous Research<sup>6</sup>**  
**(Total Number of Children in each Grade and Total Number Pooled)**  
**( Pooled = the test results of all subjects in the study)**

Grade	1	2	3	4	Pooled (1-5)
Math Total	172	178	173	42	567
Math Computation	131	177	173	42	525
Math Problems	131	179	173	59	544
Math Concepts	172	179	173	92	618
Listening	190	267	=	=	457
Language Expression	190	267	322	216	1091
Reading Total	189	266	322	214	1085
Core Battery Total	161	173	173	38	545
Word Analysis	172	178	=	=	350
= Indicates Not Tested at This Grade					

**Table 3.**  
**Previously Published Means and Present Study's Means and SD for the Wold Sentence Copy Test (Symbols/Minute)**  
**( Pooled = both genders combined)**

Grade	Previously Published Symbols/Minute <sup>1,2</sup>		Maples (present study)					
	Wold/Freeman	Wold/Ayres	Pooled	SD	Male	SD	Female	SD
1			21.0	13.2	19.3	11.7	22.8	14.4
2	30	39.7	44.4	20.6	42.3	20.2	46.7	20.9
3	40	42.0	61.6	22.8	57.4	22.6	66.0	22.3
4	50	45.8	67.9	21.5	65.1	21.3	71.4	21.3
5	60	50.5	81.9	19.6	78.5	20.3	86.0	18.0
6	67	54.5						
7	74	58.9						
8	80	62.8						

**Table 4.**  
**Means and SD for the Wold Sentence Copy Test (Symbols/Minute)**  
**Based upon Two Evaluations (Fall=f & Spring=s)**  
**( Pooled = both genders combined)**

Grade N	Mean Pooled	SD	N	Female	SD	N	Male	SD	P
1 F	15.7	9.7	117	16.2	10.0	127	14.7	9.3	0.091
1 S	27.0	14.1	103	29.7	15.6	111	24.5	12.0	0.006
2 F	35.0	13.8	161	37.1	13.6	177	33.2	13.7	0.010
2 S	54.4	22.0	151	57.1	22.3	168	51.9	21.4	0.036
3 F	58.2	22.2	183	62.0	21.9	197	54.5	22.0	0.001
3 S	65.2	22.9	177	70.2	21.9	186	60.4	22.9	< 0.001
4 F	62.1	21.0	114	65.8	21.2	139	59.1	19.6	0.011
4 S	74.1	20.3	103	77.7	18.3	134	71.4	21.3	0.017
5 F	76.6	17.7	48	81.3	16.1	61	73.0	18.1	0.015
5 S	87.9	20.2	46	91.1	18.9	49	84.8	21.1	0.036

F = Fall Session; S = Spring Session

nificantly higher ( $p < .05$ ) than males in every testing except for the first grade fall. Age and grade Wold scores were almost identical, and, since previous norms have been reported by grade, only grade norms are reported here.

Correlation coefficients were calculated for all 21 sub-categories of the ITBS as compared to the Wold symbols/minute score for all subjects (pooled) and between genders. The correlation scores and their significance values are listed in Table

5. Nineteen of the 21 sub scores were found to be positively and significantly correlated to the pooled Wold score. Additional correlation coefficients were calculated for gender. Fourteen of the ITBS sub scores were found to be positively correlated to the ITBS with males and 17 were positively correlated in females. The higher the score on the Wold, the higher the score tended to be on the academic sub-category.

Comparisons were then made between different variations of the pooled (all subjects in the study) means on the Wold test. Relative rates of failure of the Wold were calculated for each grade level. Among the failure criteria investigated were: those students who fell below 1 standard deviation (s) from the mean ( $> -1s$ ), those who scored  $> -0.5s$  and, those who scored anywhere below the mean. These Wold criteria were each compared to those students who scored more than 1 grade level below their respective grade ( $> -1GL$ ),  $> -0.5GL$  below their respective grade and lastly, those who were at all below grade level for their respective grade. Each of the nine academic areas reported earlier were investigated in this way.<sup>6</sup> The failure criterion which best predicted academic performance was greater than  $\frac{1}{2}$  standard deviation below the mean grade value ( $> -0.5s$ ) on the Wold for all subjects. This criterion best predicted children who also scored more than five months (in a 10 month school year) below their grade levels ( $> -0.5GL$ ) in the academic categories of the ITBS. The incidence of failure for the nine academic subject areas, using  $> -0.5GL$  is represented in Table 6. Both the total sample and grade failure percentages are listed. The pooled (all subjects in the study) failure percentages varied from a high of 21.7% (*word analysis*) to a low of 6.5% (*math computation*). The grade scores varied by grade and by subject. The highest percentage of failure among the grades was found in the second grade with 22.9% (*math problems*). The lowest percentage of failure was in the first grade with only 0.8% failure (*math computation*).

The relative rates of failure for the nine academic sub-categories can be found in Table 7. This table contains the pooled (all subjects in the study) and grade, relative risk ratios. In the pooled data, the relative risk of academic failure ( $> -0.5GL$ ) was highest for *word analysis* (3.5), followed

closely by *math total* (3.4). The lowest relative rates were for *math problems*, *math concepts* and *reading*, all with 1.8. Grade one relative risk ranged from a high of 10.5 for *reading* to a low of 1.8 for *math problems*. The second grade relative risk varied from 5.1 for *math computation* to 0.7 for *reading* and *core battery*. Third grade relative risk ranged from 2.4 with *language expression* to a low of 0.8 for *math problems* and *math concepts*. The fourth grade relative risk varied from a high of 3.0, *core battery*, to 1.1 for *math concepts*. In Table 7, the total math score is not calculated by an average of the math computation, math concepts and math problems but by another method not discussed in the ITBS manual.<sup>6,13,14</sup> Table 8 contains pooled and gender means for the first through the fifth grades. Thus, using the pooled score for third grade, any score less than (61.6 minus 11.4) 50.2 constitutes a failure. The 0.5s for each category are also found next to the means.

## DISCUSSION

The data in Table 2 shows that the number of subjects in each of the academic cells was substantial. A larger number will likely show that small differences will be significant. Large samples, therefore, reveal associations that could be hidden by the analysis of a smaller sample. This accounts for the relatively low correlation values (Table 5, *r* values) but very high significance (*p* values). However, the fact that the scores on the Wold are directly correlated with the academic scores (the higher the Wold, the higher the academic score), should not lead one to conclude that there is a causal relationship between the two. The percent of variance (square of the correlation) has previously been reported.<sup>6</sup> Another research project with both an experimental and control group, which offers vision therapy to address visual motor delays/dysfunctions might help to answer the question of causation. If, after vision therapy, the Wold scores and academic scores on an experimental sample improved, while at the same time, the Wold scores and academic scores on a matched control group that did not receive vision therapy, but rather a sham therapy, did not improve, then evidence for a cause and effect would be stronger. The present study, however, does not indicate that poor visual motor skill, as tested by the Wold, is a cause of

**Table 5.**  
**Correlation Coefficients and Statistical Significance (p = .05) of the**  
**ITBS Sub-Category Scores to the Wold Scores**  
**(\* = Not Significant) ( Pooled = both genders combined)**

ITBS	Pooled		Males		Females	
	p	r	p	r	p	r
Vocab	<.001	.433	<.001	.441	<.001	.456
Comp	<.001	.444	<.001	.474	<.001	.432
Total Read	<.001	.484	<.001	.509	<.001	.479
Listening	<.001	.423	<.001	.383	<.001	.475
Spelling	.004	.163	*.111	.127	.008	.212
Language	<.001	.459	<.001	.433	<.001	.499
Capital	<.001	.137	.034	.168	*.192	.105
Punctuation	.016	.136	*.068	.145	*.127	.123
Total Grade	<.001	.526	<.001	.516	<.001	.545
Concepts	<.001	.458	<.001	.468	<.001	.496
Problems	<.001	.368	<.001	.408	<.001	.368
Total Math	<.001	.514	<.001	.539	<.001	.525
Core Total Grade	<.001	.493	<.001	.495	<.001	.506
Social Studies	.023	.023	*.987	.001	.033	.152
Science	.002	.158	*.068	.131	.002	.218
Word Analysis	<.001	.368	<.001	.414	<.001	.324
Maps	*.178	.091	*.468	.071	*.127	.144
Reference	*.117	.106	*.188	.128	*.229	.114
Total Info	.001	.166	.044	.144	.004	.205
Composite	.050	.100	*.417	.059	.006	.198
Math Comp	<.001	.478	<.001	.476	<.001	.497

**Table 6.**  
**Percent Incidence of Failure (Greater than 5mos below grade level**  
**(>0.5GL) for each category) ( Pooled = all grades combined)**

Subject	Pooled	1st Grade	2nd Grade	3rd Grade	4th Grade
Math Total	10.8	14.0	7.9	8.1	19.0
Math Computation	6.5	0.8	4.0	9.2	19.0
Math Problems	18.4	10.7	22.9	21.4	11.9
Math Concepts	17.2	19.8	15.6	12.1	22.8
Listening	12.9	16.3	10.5	—	—
Language Expression	15.3	15.3	12.4	13.4	19.4
Reading	17.7	14.8	15.4	19.9	18.7
Core Battery	12.1	12.4	11.0	12.1	15.8
Word Analysis	21.7	22.1	21.3	—	—

academic failure although it appears to be associated.

This study has expanded the normative data for the Wold test in two ways. First, in previously published norms (Table 3), first grade was not included, but is included in the present study. Secondly, this research has revealed that in each grade, the pooled (all subjects in the study) means in the present study were higher than previously published means.

Using the norms found in previous studies, one would expect that a child might score in the adequate range with the older norms but would be found inadequate with the present norm level. Consider the Wold/Freeman norms (Table 3). At grade eight, the Wold/Freeman norms are 80 symbols per minute; yet, the pooled average for this new data indicates that this level of performance should be reached at the fifth grade level.

**Table 7.**  
**Relative Risk of Failure using the Wold Sentence Copy Test as the**  
**Screening Instrument**  
**(Students who pass Wold {>.5s} compared to students**

Subject	Pooled	1st Grade	2nd Grade	3rd Grade	4th Grade
Math Total	3.4	6.8	1.8	2.2	1.7
Math Computation	3.0	*	5.1	1.2	1.7
Math Problems	1.8	1.8	4.0	0.8	2.0
Math Concepts	1.8	5.0	1.9	0.8	1.1
Listening	2.6	3.8	1.6	—	—
Language Expression	1.9	2.8	1.6	2.4	1.2
Reading	1.8	10.5	0.7	1.2	1.6
Core Battery	2.0	5.2	0.7	1.4	3.0
Word Analysis	3.5	7.0	1.5	—	—

*No child who passed the Wold was below ½ grade level, academically.*

**Table 8.**  
**Minimum and 0.5s Scores for the Wold Sentence Copy Test**  
**(Symbols/Minute) ( Pooled = both genders combined)**

Grade	Pooled	0.5s	Male	0.5s	Female	0.5s
1	21.0	6.6	19.3	5.9	22.8	7.2
2	44.4	10.3	42.3	10.1	46.7	10.5
3	61.6	11.4	57.4	11.3	66.0	11.2
4	67.9	10.8	65.1	10.7	71.4	10.8
5	81.9	9.8	78.5	10.2	86.0	9.0

Since the norms presented here have both a beginning of school year and an end of school year criteria (Table 4), one could feel more confident in saying the child is below, at, or above his grade level in visual motor performance at specific times during the school year. These more specific criteria would mean that the educational analysis could be more specific and accurate. Likewise, if one wished to be even more specific, the gender norms clearly demonstrate that females perform, on the whole, better than males. This is not surprising since it is generally accepted that generally, girls develop faster than boys in many of the motor and educational skills during the age/grade represented in this study.<sup>7-10</sup>

It was found that the Wold's best overall predictive value for failure on the different academic areas was when the failure criteria for the Wold was set at >-0.5s and the academic failure was set at >-0.5 GL. Incidence and prevalence rates for learning problems are estimated at between 2 and 10%.<sup>15</sup> Five percent are diagnosed with learning disabilities; however,

an equal or greater number have milder learning problems. Solan and Ficcara,<sup>16</sup> and Solan and Suchoff<sup>17</sup> have recommended that scores below ½ standard deviation from the mean (>-.5s) be used as the failure indicator for performance testing. This ½ standard deviation criterion would likely identify the moderately disabled in the classroom. This seems to be a reasonable benchmark since it would encompass the lower 31% as opposed to the lower 16% (-1s) of performance. Table 6 shows that for >-0.5GL on the ITBS failure in the academic areas is less than the 31% expected. This may indicate a better local educational experience for these children, since these scores are obtained from standardized educational tests given nation wide.

We believe that the Wold failure criterion of >-0.5s is appropriate because it best identified our subjects who were potentially in academic trouble as evidenced by their performance on the ITBS. Thus, the quickly and easily administered Wold has the potential, particularly in the early grades, of raising a "red flag" in identify-

ing academically at risk children so that further confirmatory and remedial measures can be taken.

Table 8 contains the pooled and gender Wold norms as well as the ½ standard deviation (0.5s). We strongly recommend that for the various grades any Wold score that is lower than the gender -0.5s symbols/minute numerical value be considered as a failure indicating a grapho-visual motor delay or dysfunction.

## CONCLUSIONS

1. The norms in this study are more specific and accurate than the norms previously reported. The present norms for each grade were developed by testing at the beginning and end of the school year. The data included a significantly larger pool of subjects. Additionally, the present study included the first grade.
2. The Wold is a predictor of some areas of academic performance. These areas include overall classroom performance (core battery), math performance, reading, expressive and receptive language skills.
3. The Wold failure criterion >-0.5s is appropriate for all grades included in this study.
4. The Wold could be used by the classroom teacher as a quick and easy screening tool to identify students who may be at risk for grapho-visual motor delay.
5. Early identification will enable the school system to pursue solutions for the disability at an earlier time and will lessen the potential for damage to the child's educational experience.

## Acknowledgements

The author appreciates the significant contributions of Mr. Richard Hoenes for his statistical assistance. This project was funded in part by the Australasian College of Behavioral Optometry; College of Optometrists in Vision Development; Optometric Extension Program Foundation Inc; American Foundation for Vision Awareness; and Mountain States Congress of Optometry.

The author has no financial or other interests in the Wold test.

## Sources

- a. Optometric Extension Program Foundation, Inc. 1821 E. Carnegie Avenue, Suite 3-L Santa Ana, CA 92705-5510, (949) 250-8070.

