

# Assessing Decoding Ability: The Role of Speed and Accuracy and a New Composite Indicator to Measure Decoding Skill in Elementary Grades

Journal of Learning Disabilities  
2015, Vol. 48(2) 176–195  
© Hammill Institute on Disabilities 2013  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0022219413495298  
journaloflearningdisabilities.sagepub.com  
SAGE

Isabella Morlini, PhD<sup>1</sup>, Giacomo Stella, PhD<sup>1</sup>,  
and Maristella Scorza, PhD<sup>1</sup>

## Abstract

Tools for assessing decoding skill in students attending elementary grades are of fundamental importance for guaranteeing an early identification of reading disabled students and reducing both the primary negative effects (on learning) and the secondary negative effects (on the development of the personality) of this disability. This article presents results obtained by administering existing standardized tests of reading and a new screening procedure to about 1,500 students in the elementary grades in Italy. It is found that variables measuring speed and accuracy in all administered reading tests are not Gaussian, and therefore the threshold values used for classifying a student as a normal decoder or as an impaired decoder must be estimated on the basis of the empirical distribution of these variables rather than by using the percentiles of the normal distribution. It is also found that the decoding speed and the decoding accuracy can be measured in either a 1-minute procedure or in much longer standardized tests. The screening procedure and the tests administered are found to be equivalent insofar as they carry the same information. Finally, it is found that speed and accuracy act as complementary effects in the measurement of decoding ability. On the basis of this last finding, the study introduces a new composite indicator aimed at determining the student's performance, which combines speed and accuracy in the measurement of decoding ability.

## Keywords

latent factors, reading tests, variable distribution

An essential and important component of the teaching–learning process is the early detection of students with decoding disabilities (e.g., Blachman, Ball, Black, & Tangel, 1994; O'Connor & Jenkins, 1999; Torgersen, Wagner, & Rashotte, 1997; Vadasy & Sanders, 2011; Vadasy, Sanders, & Abbott, 2008). Teaching these students in the school should respect their pace and learning methods and should include a system of assessment that takes into account their different performances. Numerous studies have shown that there is a strong relationship between decoding and reading comprehension (e.g., Catts, Hogan, & Adolf, 2005; Francis, Fletcher, Catts, & Tomblin, 2005; Hoover & Tunmer, 1993; Keenan & Betjemann, 2006; Keenan, Betjemann, & Olson, 2008; Nation & Snowling, 1997). Fraser, Goswami, and Conti-Ramsden (2010) recently showed that there is substantial overlap between reading disability and specific language impairment at the level of phonological skills and auditory processing. Designing early reading screening instruments is complicated, as the age at which they are administered may

influence significantly their predictive value, even within a short time frame or within the same school year (O'Connor & Jenkins, 1999). The appropriate age to begin screening young children has been debated. In some cases, the literature argues that for young children screening of early literacy skills should occur before formal literacy instruction begins (Badian, 1982, 2000; Invernizzi, Justice, Landrum, & Booker, 2004; Justice, Invernizzi, & Meier, 2002). In regular orthographic systems, like those in Italy, Spain, and Finland, at the end of Grade 1, children are typically close to the ceiling in terms of reading accuracy (Goswami, Gombert, & Barrera, 1998; Seymour, Aro, & Erskine, 2003). Italian studies on reading acquisition show that at the beginning of primary school children are highly

<sup>1</sup>University of Modena and Reggio Emilia, Modena, Italy

## Corresponding Author:

Isabella Morlini, Department of Economics, University of Modena and Reggio Emilia, Via Berengario 51, 41.121 Modena, Italy.  
Email: isabella.morlini@unimore.it

heterogeneous because of their different level of maturity and their different cognitive and linguistic ability. In kindergarten, Italian children are not trained systematically in reading acquisition and phonology, and this leads to a heterogeneous situation among children at the beginning of the first grade. Nevertheless, because of the consistency of Italian orthography, at the end of the first grade the majority of children successfully perform in reading and decoding. Decoding ability in elementary grades in Italy is currently assessed with the aid of tests requiring the students to read aloud a selected list of words, or a selected list of nonwords or a text. These lists are taken from standardized tests. The most widely used standardized tests have been introduced and illustrated by Sartori, Job, and Tressoldi (1995, 2007). Recently, a new screening procedure for identifying impaired decoders in elementary grades in Italy was proposed by Stella, Scorza, and Morlini (2011). This screening requires the students to read a selected text for exactly 1 minute. What is important in the use of tests and screening procedures is the way the results are interpreted. One of the defining characteristic of a skilled decoder is that he or she not only is able to spell written words (or nonwords) accurately, but also does so rapidly and automatically (Bowers & Wolf, 1993). An individual who spells accurately but very slowly cannot be considered a skilled decoder. Slow rate of word reading is then characteristic of impaired decoding as well as low accuracy (Compton & Carlisle, 1994), especially in transparent languages (e.g., Jiménez & Hernández-Valle, 2000; Jiménez & Ramirez, 2002; Jiménez, Rodríguez, & Ramirez, 2009; Serrano & Defior, 2008; Sprenger-Charolles, Colé, Lacert, & Serniclaes, 2000; Sprenger-Charolles, Siegel, Jiménez, & Ziegler, 2011; Ziegler, Perry, Ma-Wyatt, Ladner, & Schulte-Korne, 2003). In Italy, some examiners assessing decoding ability do not take into account both factors, and an individual can be classified as impaired because he or she is able to read words (or nonwords) very rapidly, even though he or she misspells a fairly large number of words (or nonwords). Some other examiners do not consider speed, and individuals with weak decoding skills who are able to read a large number of words, provided they are given ample time, are erroneously classified as adequate decoders by these examiners. Many authors have outlined the necessity of considering both speed and accuracy for a valid assessment of decoding skills and have recommended to include both these factors in standardized test procedures in learning disabilities research (e.g., Dennis & Evans, 1996; Torgersen et al., 1997; Vernon, 1990). Therefore, studying the role of speed and accuracy in assessing decoding ability and developing a composite indicator of decoding efficiency that incorporates measures of speed as well as of accuracy are two main goals in reading research. In the literature, only some few indicators have been proposed (Dennis & Evans, 1996; Joshi & Aaron, 2002; Olson, Gillis, Rack, DeFries, & Fulker, 1991;

Sinatra & Royer, 1995). Another goal is to define suitable threshold values in variable measuring speed and accuracy for classifying a student as a normal decoder or as a “too slow” or “too inaccurate” decoder. At the present in Italy, these thresholds have been defined assuming the Gaussian distribution of the variables. Our study shows that all variables measuring speed and accuracy in the administered tests are not Gaussian and therefore the threshold values should be estimated on the basis of the empirical distribution of these variables rather than on the basis of the percentiles of the Gaussian distribution. In synthesis, the present study has the following aims:

1. The first aim is to examine the empirical distribution of the variables measuring speed and accuracy in two widely used Italian standardized tests and in a new screening procedure proposed by Stella et al. (2011) and to see whether the normative threshold values for the variables measuring speed and accuracy in the standardized tests lead to a percentage of students classified as impaired reader in agreement with the expected one. Although there have been no studies to indicate an accurate percentage, the Italian Dyslexia Association (<http://www.aiditalia.org>) believes that learning disability can affect between 3% and 4% of the Italian scholastic population.
2. The second aim is to study the bivariate relationships among variables to investigate whether speed and accuracy are correlated or act as independent elements in the assessment of the decoding ability.
3. The third aim is to investigate whether the speed and the accuracy measured in the new screening procedure (which is exactly 1 minute long) are correlated with the speed and the accuracy measured in the much longer standardized tests.
4. The fourth aim is to propose a new composite indicator of decoding efficiency that incorporates a measure of speed as well as a measure of accuracy. This indicator can be used in any test or screening procedure, both in countries with transparent languages, as well as Italian, and in countries with deep languages. Indeed, even though a shallow orthography facilitates decoding, Paulesu et al. (2001) found that dyslexic readers in shallow orthography read better than dyslexic readers in deep orthography, but they perform significantly worse than their nondyslexic counterpart readers in shallow orthography. The indicator requires only the noncorrelation of the two measures. It may be used for classifying children as impaired or not impaired on the basis of a score that takes into account both the speed and the accuracy. Singularly using a measure of speed and a measure of accuracy, a person may be classified as impaired with one measure but as not impaired with the other measure.

**Table 1.** Information About the Participants.

	Grade				
	I	II	III	IV	V
Number of students	333	384	200	276	276
Mean age (in years)	6.8	7.5	8.3	9.3	10.1

The article is organized as follows. In the second section we illustrate the study conducted. Drawing from the results obtained by administering the standardized tests and the screening procedure to about 1,500 students attending elementary grades in Italy, in the third section we report univariate and bivariate analyses of the variables measuring speed and accuracy, examine the empirical distribution of these variables, and discuss the choice of the threshold values for classifying the students' performance. In the fourth section we propose the new composite indicator. Finally, in the fifth section we give some concluding remarks.

## The Study

A new screening procedure and two standardized tests were administered to 1,469 students in elementary grades in the Lombardia and Emilia Romagna regions of northern Italy. The tests and the procedure were administered to students attending Grades II, III, IV, and V in February and to students attending Grade I in May. Since Italian is a language with transparent or shallow orthography, where the letters of the alphabet, alone or in combination, are in most instances uniquely mapped to each of the speech sounds occurring in the language, in general at the end of the first school year students are able to read (Goswami et al., 1998; Seymour et al., 2003). The tests and the screening procedure were administered to all students attending the schools selected for the study, but results relative to foreign students and to students with a certified learning disability were dropped from the analysis. Information about the participants is shown in Table 1. The administered screening procedure is called SPILLO (Stella et al., 2011). In this procedure, the student is asked to read aloud a text for exactly 1 minute. The text presented tells a story and is composed of 181 words. When reading a text, semantic and morphosyntactic information can enhance the reader's understanding, whereas reading a list of words or nonwords is primarily a measure of intact or impaired lexical or sub-lexical process. Reading a text allows the evaluation of the efficiency of the so-called instrumental reading, that is, the ability to quickly and correctly recognize and spell words present in a text. The chosen text tells a story. This story has a score of 71 on the Gulpease Index, which is similar to the Flesch Index (Flesch, 1948). It compares the words in the text with those in the base vocabulary of the Italian

language of De Mauro (1997; [www.eulogos.net/it/censor](http://www.eulogos.net/it/censor)). The Gulpease Index uses a predefined scale from 0 (*minimum readability*) to 100 (*maximum readability*). This scale is based on the evaluation of real understanding of a body of text by readers of different school ages. A score of 71 indicates that the text is "difficult to read" for students in primary school and that children in elementary grades are not capable of reading and understanding the text by themselves. The ability to decode is measured by the following variables, computed by the examiner at the end of the procedure (after 1 minute):

- $Y_1$ : number of words read in a minute
- $Y_2$ : number of syllables per second read in a minute
- $Y_3$ : number of incorrect pronunciations in a minute

The standardized tests are the *Batteries for the Diagnosis of Reading and Spelling Disabilities* of Sartori et al. (1995, 2007). These tests are the most widely used diagnostic tests in Italy. While the student reads, the examiner times the reading and makes a note of the mistakes. In the first test the student is asked to read a list of words and in the second test a list of nonwords. The list of words contains the same proportion of words as a function of familiarity (one third of the words are "very familiar," one third are "familiar," and one third are "not familiar") and as a function of the number of syllables (one third of the words have two syllables, one third have three syllables, and one third have four syllables). The list of nonwords contains the same proportion of words as a function of the number of syllables. In the first test the variables measuring performances of decoding are the following:

- $X_1$ : time (in seconds) in reading the list of words
- $X_2$ : number of syllables per second read in the list of words
- $X_3$ : number of words mispronounced in reading the list of words

In the second test the variables are these:

- $X_4$ : time (in seconds) in reading the list of nonwords
- $X_5$ : number of syllables per second read in the list of nonwords
- $X_6$ : number of incorrect pronunciations in reading the list of nonwords

Although the number of syllables per second seems to be more common in text or sentence reading tests whereas the time in seconds seems to be preferred for word and nonword reading tests, the two scores are provided interchangeably for the Italian standardized tests. Intuitively, both measures have exactly the same meaning. However, when one derives the syllables per second or the time in

**Table 2.** Univariate Statistics for Variables Measuring the Decoding Speed in the Word Reading Standardized Test.

	$X_1$					$X_2$				
	I	II	III	IV	V	I	II	III	IV	V
N	333	384	200	276	276	333	384	200	276	276
Min	92	69	52	50	43	0.24	0.47	0.97	0.80	1.36
Max	1148	595	289	348	206	3.04	4.06	5.38	5.60	6.51
$x_{0.05}$	145	96	70	59	54	0.39	0.73	1.34	1.83	2.31
$x_{0.25}$	232	135	90	73	64	0.66	1.20	1.92	2.59	3.11
$x_{0.50}$	314	177	109	88	74	0.89	1.58	2.56	3.16	3.78
$x_{0.75}$	424	234	146	108	90	1.21	2.08	3.12	3.84	4.38
$x_{0.95}$	723	381	209	153	121	1.39	2.92	4.00	4.75	5.19
M	360	198	120	96	79	0.98	1.67	2.59	3.20	3.77
SD	190	90	44	34	23	0.48	0.66	0.81	0.89	0.92
CV	0.5	0.4	0.4	0.4	0.3	0.49	0.40	0.31	0.28	0.24
Skewness	1.7	1.5	1.3	2.5	1.8	1.48	0.74	0.34	0.16	0.03
Kurtosis	3.3	2.8	1.8	12.3	5.5	3.32	0.41	-0.16	-0.22	-0.18

**Table 3.** Univariate Statistics for Variables Measuring the Decoding Speed in the Nonword Reading Standardized Test.

	$X_4$					$X_5$				
	I	II	III	IV	V	I	II	III	IV	V
N	333	384	200	276	276	333	384	200	276	276
Min	58	49	33	34	31	0.23	0.41	0.60	0.54	1.07
Max	542	311	210	234	118	2.17	2.57	3.82	3.71	4.06
$x_{0.05}$	89	68	52	49	38	0.37	0.61	0.80	1.16	1.30
$x_{0.25}$	120	89	69	59	50	0.62	0.92	1.29	1.47	1.73
$x_{0.50}$	151	110	83	71	59	0.83	1.15	1.51	1.76	2.14
$x_{0.75}$	204	137	98	86	73	1.05	1.42	1.83	2.14	2.52
$x_{0.95}$	342	206	157	109	97	1.42	1.85	2.43	2.82	3.32
M	173	120	88	74	63	0.86	1.18	1.59	1.85	2.17
SD	82	45	30	24	18	0.33	0.37	0.49	0.53	0.62
CV	0.5	0.4	0.3	0.3	0.3	0.38	0.32	0.31	0.29	0.29
Skewness	1.8	1.6	1.4	2.2	0.6	0.54	0.45	0.83	0.72	0.69
Kurtosis	3.8	3.5	2.7	11.5	0.0	0.38	0.24	1.87	0.65	0.17

seconds from a same test performance, the pronounced nonlinearity of the function relating the two measures can produce important discrepancies between results obtained with each score, both in clinical practice and in experimental research. This fact is illustrated in the next sections. The time requested for administering the tests depends on the student’s ability and usually varies between 1 and 20 minutes. The screening procedure and the tests were administered in a quiet environment outside the classroom. The text and the list of words and nonwords were printed in Times New Roman font. Each student was asked to read aloud as quickly as possible but to be as accurate as possible at the same time. The procedure and the tests required the examiner not to intervene when the student made a mistake.

### Univariate, Bivariate, and Multivariate Analyses

#### Speed: Univariate Analyses

Speed is measured by the variables  $X_1, X_2, X_4, X_5, Y_1$ , and  $Y_2$ . In Tables 2, 3, and 4 we list the values of some univariate statistics computed for these variables. In Figure 1 we show the empirical distributions through histograms. Performances in the decoding speed improve from Grade I to Grade V: Both the median (indicated with  $x_{0.50}$  in the tables) and the mean values  $X_1$  and  $X_4$  decrease, whereas the mean and the median values of  $X_2, X_5, Y_1$ , and  $Y_2$  increase. Variables measuring the number of words and the number of syllables read in a second have a similar pattern: The average values of  $Y_1$  and  $Y_2$  across the five grades behave

**Table 4.** Univariate Statistics for Variables Measuring the Decoding Speed in the Screening Procedure.

	$Y_1$					$Y_2$				
	I	II	III	IV	V	I	II	III	IV	V
N	333	384	200	276	276	333	384	200	276	276
Min	8	21	48	9	57	0.30	0.78	1.77	0.32	2.07
Max	127	157	164	180	180	4.23	5.25	5.53	5.98	5.98
$x_{0.05}$	16	34	64	78	96	0.58	1.27	2.30	2.80	3.32
$x_{0.25}$	30	54	77	108	127	1.10	1.97	2.77	3.63	4.23
$x_{0.50}$	41	66	97	125	146	1.48	2.38	3.33	4.18	4.87
$x_{0.75}$	53	81	115	140	164	1.95	2.90	3.86	4.65	5.53
$x_{0.95}$	16	34	64	78	96	2.65	3.82	4.55	5.70	5.98
$M$	43	69	98	123	144	1.56	2.46	3.36	4.16	4.82
SD	20	24	24	27	25	0.68	0.77	0.73	0.81	0.79
CV	0.5	0.3	0.2	0.2	0.2	0.44	0.31	0.22	0.19	0.16
Skewness	1.2	0.6	0.2	-0.4	-0.8	1.06	0.45	-0.09	-0.34	-0.69
Kurtosis	2.9	0.4	-0.7	0.9	0.4	2.10	0.20	0.68	1.04	0.05

similarly to the average values of  $X_2$  and  $X_5$ . Dispersion, measured by the coefficient of variation (CV), always decreases with the grade level. The larger variability in Grades I, II, and III may be explained by considering that many covariates (e.g., cultural level, experiences in day nursery, etc.) have a great influence on decoding performances. Beginning with Grade III, these covariates become less relevant and the scholastic population becomes more homogeneous.

**Skewness.** Variables measuring the decoding speed on the tests ( $X_1, X_2, X_4, X_5$ ) have a positive skew and present outlying values higher than  $x_{0.75} - 1.5(x_{0.75} - x_{0.25})$ , in all grades. These characteristics are desirable for  $X_1$  and  $X_4$ , which have a positive direction of pathology (impaired readers are children with high values for these variables), but not for  $X_2$  and  $X_5$ , which have a negative direction of pathology (impaired readers are children with low values for these variables).  $Y_1$  has a positive skew in Grades I, II, and III but a negative skew in Grades IV and V.  $Y_2$  has a positive skew in Grades I and II and a negative skew in Grades III, IV, and V. In Grades IV and V these variables also have outlying values smaller than  $x_{0.25} + 1.5(x_{0.75} - x_{0.25})$ . These features show that  $Y_1$  and  $Y_2$  have a more discriminative power in the last school years of primary school, when the students read more fluently.

**Normality.** Determining whether variables measuring speed are normally distributed is an important goal since the currently normative threshold values on the standardized tests, used for discriminating impaired and normal decoders, are based on this assumption (Sartori et al., 1995, 2007). Table 5 reports the  $p$  values of four nonparametric tests: the Shapiro–Wilk (Shapiro & Wilk, 1965), Anderson–Darling

(Anderson & Darling, 1952), Lilliefors (Lilliefors, 1967), and Jarque–Bera (Jarque & Bera, 1980, 1981, 1987) tests for the normality;  $p$  values greater than .01 are presented in bold. With the exception of  $X_2$  in Grades III, IV, and V and  $Y_2$  in Grade III, all variables seem far from the normal distribution. The null hypothesis of Gaussian distribution is rejected with all tests considered ( $\alpha = .01$ ) for all the following variables:  $X_1$  in all grades,  $X_2$  in Grades I and II,  $X_4$  in all grades,  $X_5$  in Grades III, IV, and V,  $Y_1$  in Grades I, II, III, and V, and  $Y_2$  in Grades I, II, IV, and V. Regarding  $X_5$  in Grades I and II, the null hypothesis of Gaussian distribution is rejected, for  $\alpha = .01$ , with the Shapiro–Wilk, Anderson–Darling, and Jarque–Bera tests whereas it is accepted with the Lilliefors test. For  $Y_1$  in Grade IV, the null hypothesis is rejected with the Shapiro–Wilk and Jarque–Bera tests whereas it is accepted with the Anderson–Darling and Lilliefors tests.

**Discussion of the normative threshold values on the standardized tests.** The currently used thresholds for  $X_1, X_2, X_4, X_5$  are based on the assumption of normality. The thresholds are used for classifying students as normal readers or impaired readers. They have been specified on the basis of the mean and the variance, assuming a normal distribution (Sartori et al., 1995, 2007). The thresholds have been obtained as  $\mu + 2\sigma$  (for  $X_1$  and  $X_4$ ) and as  $\mu - 2\sigma$  (for  $X_2$  and  $X_5$ ), where  $\mu$  indicates the mean and  $\sigma$  the standard deviation, considering that in a Gaussian distribution these values exclude about 2% of the population. The estimated values of  $\mu$  and  $\sigma$ , reported in Sartori et al. (2007) and currently used as normative values in Italy (also for obtaining the  $z$  scores), have been estimated only for Grades II, III, IV, and V. Using the  $t$  test for the means and the nonparametric test of Levene for the variances, the means and the variances

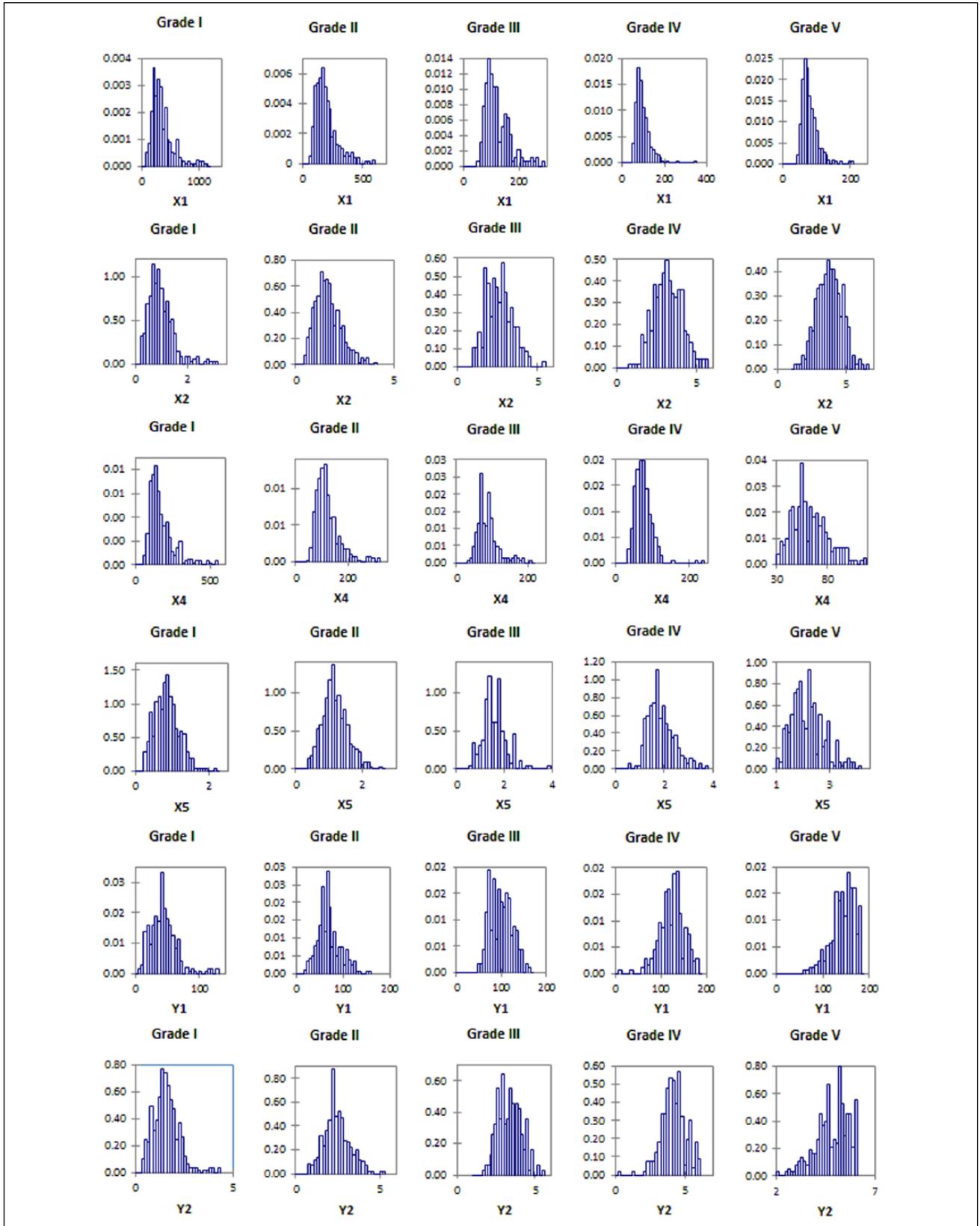


Figure I. Histograms of variables measuring decoding speed:  $X_1, X_2, X_4, X_5, Y_1, Y_2$ .

**Table 5.** *p* Values of Some Nonparametric Tests for Normality.

	$X_1$					$X_2$				
	I	II	III	IV	V	I	II	III	IV	V
Shapiro–Wilk	.00	.00	.00	.00	.00	.00	.00	<b>.05</b>	<b>.62</b>	<b>.94</b>
Anderson–Darling	.00	.00	.00	.00	.00	.00	.00	<b>.14</b>	<b>.69</b>	<b>.89</b>
Lilliefors	.00	.00	.00	.00	.00	.00	.00	<b>.22</b>	<b>.62</b>	<b>.75</b>
Jarque–Bera	.00	.00	.00	.00	.00	.00	.00	<b>.13</b>	<b>.41</b>	<b>.78</b>
	$X_4$					$X_5$				
	I	II	III	IV	V	I	II	III	IV	V
Shapiro–Wilk	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Anderson–Darling	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
Lilliefors	.00	.00	.00	.00	.00	<b>.12</b>	<b>.03</b>	.00	.00	.00
Jarque–Bera	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
	$Y_1$					$Y_2$				
	I	II	III	IV	V	I	II	III	IV	V
Shapiro–Wilk	.00	.00	.00	.00	.00	.00	.00	<b>.13</b>	.00	.00
Anderson–Darling	.00	.00	.00	<b>.11</b>	.00	.00	.00	<b>.12</b>	.00	.00
Lilliefors	.00	.00	.00	<b>.13</b>	.00	.00	.00	<b>.03</b>	.00	.00
Jarque–Bera	.00	.00	.00	.00	.00	.00	.00	<b>.18</b>	.00	.00

Note. *p* values greater than .01 are presented in bold.

**Table 6.** Percentages of Students Classified as Impaired Readers With the Normative Thresholds on the Standardized Tests.

Grade	$X_1$	$X_2$	$X_4$	$X_5$	$Y_1$	$Y_2$
Grade II	7.8	0.5	5.7	0.0	5.0	6.2
Grade III	3.0	0.0	4.5	0.5	5.5	5.0
Grade IV	2.5	1.1	1.4	0.7	4.3	4.3
Grade V	2.2	0.7	1.1	0.0	4.7	6.2

reported in Sartori et al. (2007) are significantly different ( $\alpha = .05$ ) from the means and the variances obtained in our study (see Table 4). The normative thresholds in our sample lead to percentages of students classified as impaired readers that vary greatly not only across the grades but also across the variables (see Table 6). In Grade II, for example, 30 students are classified as impaired decoders with  $X_1$  but only 2 with  $X_2$ . Variable  $X_5$  with the normative threshold classifies as impaired decoders in each grade a percentage of students far from 2% and also from the expected value (between 3% and 4%). Students with a mixed profile in the word reading speed (students who are classified as impaired with  $X_1$  but are in the normal range with  $X_2$  or vice versa) are numbered 28 in Grade II, 6 in Grade III, and 4 in Grades

IV and V. Students classified as impaired with both measures of word reading speed are only 2 in Grade II, zero in Grade III, 3 in Grade IV, and 2 in Grade V. Students with a mixed profile in the nonword reading speed (students who are classified as impaired with  $X_4$  but are in the normal range with  $X_5$  or vice versa) are numbered 22 in Grade II, 9 in Grade III, 2 in Grade IV, and 3 in Grade V. Students classified as impaired with both measures of nonword reading speed are numbered zero in Grade II, 1 in Grade III, 2 in Grade IV, and zero in Grade V. Because of the nonnormality of the variables, the presence of outliers, and the level of asymmetry, which is different from one grade to another, more accurate thresholds should be defined in terms of the percentiles. The thresholds for  $Y_1$  and  $Y_2$  have been set equal to  $x_{0.05}$  to discriminate a percentage of people higher than the expected one. The procedure is not intended as a diagnostic test for learning disorders, but it is a screening procedure for detecting students with heavy difficulties in decoding. The causes of these difficulties are to be defined by subsequent more detailed analyses. In our sample, the percentages of students classified as impaired readers with  $Y_1$  and  $Y_2$  are similar in all grades and are not far from 5%. Students with a mixed profile in the text reading speed (students who are classified as impaired with  $Y_1$  but not with  $Y_2$  or vice versa) are only 5 in Grade II, zero in Grade III, 3 in

**Table 7.** Univariate Statistics for Variables Measuring the Decoding Accuracy on the Standardized Tests.

	$X_3$					$X_6$				
	I	II	III	IV	V	I	II	III	IV	V
N	333	384	200	276	276	333	384	200	276	276
Min	0	0	0	0	0	0	0	0	0	0
Max	84	55	18	19	22	38	31	20	24	27
$x_{0.25}$	5	4	2	1	1	4	5	3	3	2
$x_{0.50}$	10	6	4	3	2	8	7	6	5	4
$x_{0.75}$	17	10	6	5	4	13	10	8	8	6
$x_{0.95}$	38	19	11	10	9	24	17	14	14	12
M	13.6	8.0	4.5	3.6	2.8	9.8	8.2	6.3	5.7	4.8
SD	12.3	6.8	3.5	3.3	3.1	7.2	5.1	4.0	4.3	4.1
CV	0.9	0.9	0.8	0.9	1.1	0.7	0.6	0.6	0.8	0.8
Skewness	2.1	2.6	1.3	1.6	2.5	1.3	1.2	1.1	1.3	1.9
Kurtosis	6.0	11.9	1.9	3.6	8.9	1.7	2.0	1.4	2.0	6.0

**Table 8.** Univariate Statistics for Variable Measuring the Decoding Accuracy in the Screening Procedure.

	$Y_3$				
	I	II	III	IV	V
N	333	384	200	276	276
Min	0	0	0	0	0
Max	9	10	6	11	11
$x_{0.25}$	0.0	0.5	0.5	0.9	0.5
$x_{0.50}$	1.0	1.5	1.5	1.0	1.0
$x_{0.75}$	2.0	2.5	2.0	2.5	2.0
$x_{0.95}$	4.0	5.0	4.0	5.0	4.0
M	1.3	1.8	1.4	1.7	1.5
SD	1.4	1.5	1.2	1.5	1.5
CV	1.1	0.9	0.8	0.9	1.0
Skewness	1.8	1.3	1.0	1.8	2.3
Kurtosis	4.8	3.0	1.1	5.5	8.6

Grade IV, and 4 in Grade V. Students classified as impaired with both variables regarding text reading speed are numbered 19 in Grade II, 10 in Grade III, 12 in Grade IV, and 13 in Grade V.

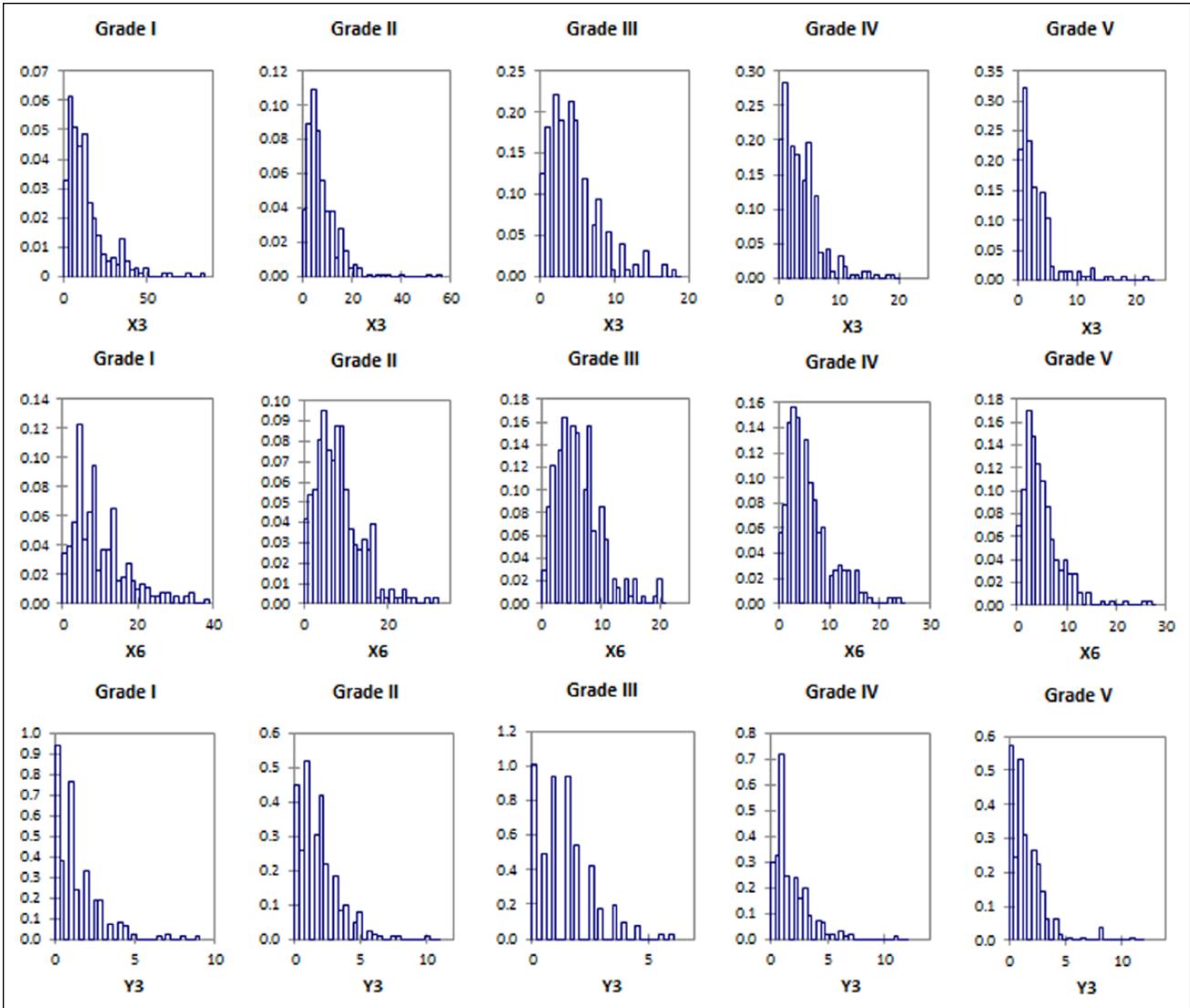
**Accuracy: Univariate Analyses**

In Tables 7 and 8 we list the values of some univariate statistics computed for the variables measuring the decoding accuracy ( $X_3, X_6, Y_3$ ). Figure 2 shows the empirical distribution of these variables through histograms. As well as the variables measuring speed, the variables  $X_3, X_6$ , and  $Y_3$  have an empirical distribution that is asymmetric and far from the Gaussian distribution. Although the mean and median values of  $X_3$  and  $X_6$  have a decreasing pattern from Grade I to Grade V, the mean and the median values of  $Y_3$  are roughly

constant across grades. This different pattern is a result of the fact that the time in the screening procedure is always equal to 1 minute, whereas it depends on the ability of the student on the standardized tests. In the screening, if one student increases his or her performance from one grade to the subsequent grade, he or she increases the speed of reading without being penalized on reading accuracy. Outliers are all in the “direction of pathology,” and this is a desirable property. The normative threshold values for  $X_3$  and  $X_6$  are the 95th percentiles obtained in the study of Sartori et al. (2007). These values are similar to  $x_{0.95}$  obtained in our sample (and reported in Table 3). The threshold for  $Y_3$  is the 95th percentile as well. The percentages of students classified as impaired readers with the currently used thresholds are reported in Table 9. These percentages (with the exception of the percentage of students classified as impaired with  $X_6$  in Grade IV) are in agreement with the expected value.

**Bivariate Analyses**

Tables 10, 11, 12, 13, and 14 report the matrices of the Pearson correlation coefficients between pairs of variables, in each grade. In all grades, the pairwise correlations between variables measuring speed are all significantly different from zero ( $\alpha = .01$ ). Even though the transformation from  $X_1$  to  $X_2$  and from  $X_3$  to  $X_4$  is not linear, these pairs of variables are highly correlated. The transformation from  $Y_1$  to  $Y_2$  is not a perfect linear transformation (since the words have a different number of syllables), but the correlation is nevertheless always equal to 1. The fact that  $Y_1$  and  $Y_2$  are highly correlated with  $X_1, X_2, X_4$ , and  $X_5$  provides evidence that all these variables are a measure of the same aspect of the decoding skill. Although  $X_3$  and  $X_6$  are highly correlated,  $Y_3$  is not correlated with  $X_3$  nor with  $X_6$ . This is a



**Figure 2.** Histograms of variables measuring decoding accuracy:  $X_3$ ,  $X_6$ ,  $Y_3$ .

**Table 9.** Percentages of Students Classified as Impaired Readers With Variables Measuring Decoding Accuracy.

Grade	$X_3$	$X_6$	$Y_3$
Grade II	2.9	4.4	5.2
Grade III	1.5	4.0	4.5
Grade IV	4.3	9.4	4.0
Grade V	3.6	4.0	5.8

result of the fact that the time is fixed on the screening procedure. The number of mistakes has a different pattern from the number of mistakes on a reading test where the time depends on the ability of the respondent. To further investigate if  $Y_3$  is a measure of the same aspect measured by  $X_3$  and  $X_6$ , we analyzed the association between the

classifications of the student’s performances. Even though two variables are not correlated, they may both lead to the conclusion that a student is a normal decoder (or to the conclusion that he or she is an impaired decoder). The values of the  $\chi^2$  statistic in the contingency tables obtained by considering, for each grade and for each couple of variables, Category 1 (the student has a value equal to or bigger than  $x_{0.95}$  and is classified as an impaired decoder) and Category 0 (the student has a value smaller than  $x_{0.95}$  and is classified as a normal decoder) all have  $p$  values smaller than .05. Therefore, for an alpha of 5%, we may reject the null hypothesis of independence between pairs of categories in each grade. To summarize results, Table 15 reports the contingency table obtained by considering students in all grades. These bivariate statistical analyses show that decoding speed and decoding accuracy can be measured with

**Table 10.** Correlations for Grade I.

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$Y_1$	$Y_2$	$Y_3$
$X_1$	1	-.79	.53	.91	-.81	.48	-.72	-.73	.05
$X_2$	-.79	1	-.46	-.72	.90	-.42	.91	.91	-.11
$X_3$	.53	-.46	1	.45	-.43	.80	-.50	-.50	.37
$X_4$	.91	-.72	.45	1	-.86	.42	-.66	-.66	-.02
$X_5$	-.81	.90	-.43	-.86	1	-.41	.82	.82	-.06
$X_6$	.48	-.42	.80	.42	-.41	1	-.46	-.46	.38
$Y_1$	-.72	.91	-.50	-.66	.82	-.46	1	1.00	-.10
$Y_2$	-.73	.91	-.50	-.66	.82	-.46	1.00	1	-.10
$Y_3$	.05	-.11	.37	-.02	-.06	.38	-.10	-.10	1

**Table 11.** Correlations for Grade II.

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$Y_1$	$Y_2$	$Y_3$
$X_1$	1	-.87	.57	.79	-.75	.49	-.77	-.79	.09
$X_2$	-.87	1	-.53	-.71	.83	-.46	.89	.88	-.11
$X_3$	.57	-.53	1	.39	-.41	.69	-.54	-.56	.43
$X_4$	.79	-.71	.39	1	-.89	.35	-.65	-.66	.00
$X_5$	-.75	.83	-.41	-.89	1	-.36	.75	.75	-.02
$X_6$	.49	-.46	.69	.35	-.36	1	-.48	-.48	.39
$Y_1$	-.77	.89	-.54	-.65	.75	-.48	1	1.00	-.10
$Y_2$	-.79	.88	-.56	-.66	.75	-.48	1.00	1	-.10
$Y_3$	.09	-.11	.43	.00	-.02	.39	-.10	-.10	1

**Table 12.** Correlations for Grade III.

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$Y_1$	$Y_2$	$Y_3$
$X_1$	1	-.92	.45	.75	-.70	.36	-.69	-.69	.17
$X_2$	-.92	1	-.49	-.73	.79	-.41	.75	.75	-.19
$X_3$	.45	-.49	1	.31	-.34	.68	-.45	-.45	.31
$X_4$	.75	-.73	.31	1	-.89	.32	-.58	-.58	.14
$X_5$	-.70	.79	-.34	-.89	1	-.36	.65	.65	-.13
$X_6$	.36	-.41	.68	.32	-.36	1	-.41	-.40	.33
$Y_1$	-.69	.75	-.45	-.58	.65	-.41	1	1.00	-.12
$Y_2$	-.69	.75	-.45	-.58	.65	-.40	1.00	1	-.12
$Y_3$	.17	-.19	.31	.14	-.13	.33	-.12	-.12	1

**Table 13.** Correlations for Grade IV.

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$Y_1$	$Y_2$	$Y_3$
$X_1$	1	-.89	.62	.84	-.71	.45	-.76	-.75	.24
$X_2$	-.89	1	-.60	-.77	.83	-.44	.77	.76	-.27
$X_3$	.62	-.60	1	.51	-.49	.61	-.55	-.54	.34
$X_4$	.84	-.77	.51	1	-.88	.41	-.61	-.61	.19
$X_5$	-.71	.83	-.49	-.88	1	-.44	.61	.60	-.23
$X_6$	.45	-.44	.61	.41	-.44	1	-.35	-.34	.28
$Y_1$	-.76	.77	-.55	-.61	.61	-.35	1	1.00	-.07
$Y_2$	-.75	.76	-.54	-.61	.60	-.34	1.00	1	-.06
$Y_3$	.24	-.27	.34	.19	-.23	.28	-.07	-.06	1

**Table 14.** Correlations for Grade V.

	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$Y_1$	$Y_2$	$Y_3$
$X_1$	1	-.93	.64	.78	-.70	.48	-.82	-.81	.28
$X_2$	-.93	1	-.55	-.81	.80	-.44	.78	.78	-.24
$X_3$	.64	-.55	1	.49	-.43	.61	-.62	-.61	.35
$X_4$	.78	-.81	.49	1	-.94	.48	-.67	-.67	.27
$X_5$	-.70	.80	-.43	-.94	1	-.43	.60	.60	-.26
$X_6$	.48	-.44	.61	.48	-.43	1	-.42	-.42	.30
$Y_1$	-.82	.78	-.62	-.67	.60	-.42	1	1.00	-.23
$Y_2$	-.81	.78	-.61	-.67	.60	-.42	1.00	1	-.23
$Y_3$	.28	-.24	.35	.27	-.26	.30	-.23	-.23	1

**Table 15.** Contingency Tables.

	$Y_3 < x_{0.95}$	$Y_3 \geq x_{0.95}$	Total		$Y_3 < x_{0.95}$	$Y_3 \geq x_{0.95}$	Total
$X_3 < x_{0.95}$	1,327	55	1,382	$X_6 < x_{0.95}$	1,326	57	1,383
$X_3 \geq x_{0.95}$	65	22	87	$X_6 \geq x_{0.95}$	66	20	86
Total	1,392	77	1,469	Total	1,392	77	1,469

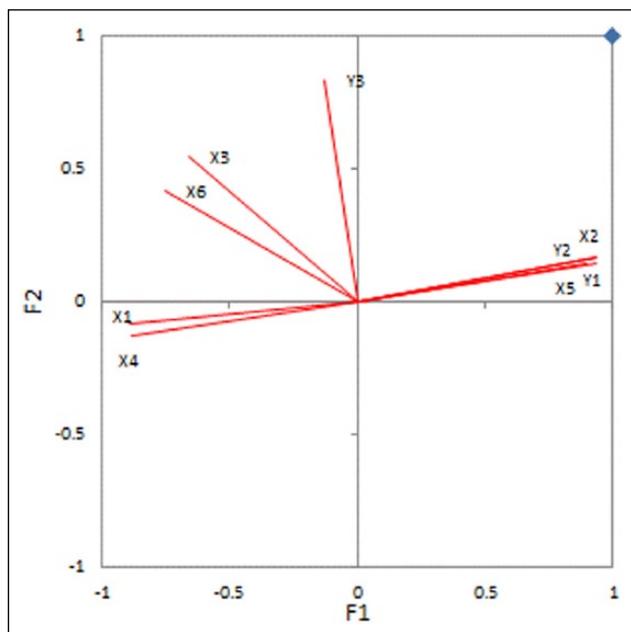
either the 1-minute procedure or the much longer standardized tests. The screening procedure and the tests seem to be equivalent insofar as they carry the same information.

The diagnosis of developmental dyslexia and learning disability requires measures of decoding ability encompassing accuracy and speed. However, the use of a measure of speed and a measure of accuracy even when these are derived from a same single test or procedure may lead to diagnostic decisions that can be very different if one relies on one measure or the other. Considering the time in seconds and the number of errors on the list of words ( $X_1$  and  $X_3$ ), 55 children have a score in the pathological range for one measure but are in the normal range for the other one (25 in Grade II, 9 in Grade III, 11 in Grade IV, 10 in Grade V) and only 15 children are classified as impaired with both measures (8 in Grade II, 0 in Grade III, 4 in Grade IV, 3 in Grade V). Considering the syllables per second and the number of errors in the list of words ( $X_2$  and  $X_3$ ), the number of mixed profiles is 39 (13 in Grade II, 3 in Grade III, 13 in Grade IV, 10 in Grade V), whereas the number of children classified as impaired with both measures is 2 (0 in Grades II and III and 1 in Grades IV and V). Using the list of non-words and considering the time in seconds and the number of errors ( $X_4$  and  $X_6$ ), the children classified as impaired with one measure but not with the other one are numbered 95 (33 in Grade II, 18 in Grade III, 30 in Grade IV, 14 in Grade V), whereas the number of students classified as impaired with both measures is only 3 (3 in Grade II and 0 in all the other grades). Considering the syllables per second and the number of errors ( $X_5$  and  $X_6$ ), 70 students have a score in the pathological range for one measure but are in

the normal range for the other one (17 in Grade II, 9 in Grade III, 30 in Grade IV, 14 in Grade V) and none of the students are impaired with both measures. Using the screening procedure and considering the number of words and the number of errors ( $Y_1$  and  $Y_3$ ), the number of mixed profiles is 76 (24 in Grade II, 13 in Grade III, 19 in Grade IV, 20 in Grade V) and the number of children classified as impaired with both measures is 5 (2 in Grade II, 1 in Grades III, IV, and V). Using the screening procedure and considering the number of syllables and the number of errors ( $Y_2$  and  $Y_3$ ), the number of mixed profiles is 86 (29 in Grade II, 13 in Grade III, 20 in Grade IV, 24 in Grade V) and the number of children classified as impaired with both measures is 5 (2 in Grade II, 1 in Grades III, IV, and V). These serious discrepancies emerge since speed and accuracy are two different aspects of decoding skill, as is illustrated in the next section. Rather than choosing one measure over the other, the solution we suggest is to use a composite indicator combining a measure of speed and a measure of accuracy.

### Multivariate Analyses

An exploratory factor analysis was performed to investigate the multivariate relationships among variables used on the tests and in the screening procedure. The analysis, performed on the correlation matrix (reached with the values of all variables in all grades), shows the presence of two main latent orthogonal factors, both with the principal components (PC) and with the common factors method (CF). The first factor is highly correlated with variables measuring speed and, to a less degree, with  $X_3$  and  $X_6$ . With



**Figure 3.** Biplot resulting from factor analysis applied to the correlation matrix. The factors are extracted with the principal component method and are unrotated.

the PC method, the eigenvalue of this factor is equal to 6 and the percentage of explained variance is 66.7%. With the CF method, the eigenvalue is 5.83 and the percentage of variance is 64.8%. The second factor is highly correlated with  $Y_3$  and, to a less degree, with the other variables measuring accuracy ( $X_3$  and  $X_6$ ). With the PC method, the eigenvalue of this factor is 1.28 and the percentage of explained variance is 14.3%. With the CF method, the eigenvalue is 0.86 and the percentage of explained variance is 9.6%. Figure 3 and Table 16 summarize the results obtained with the PC method. Rotating the factors does not improve the percentage of variance explained by the first two factors. These results indicate that speed and accuracy are two different components of decoding skill. We estimate the degree to which the set of variables  $X_1, X_2, X_4, X_5, Y_1, Y_2$  measures a single unidimensional latent construct (the decoding speed) and the set of variables  $X_3, X_6, Y_3$  measures another unidimensional latent construct (the decoding accuracy). We estimate the internal consistency of each set of variables by means of the coefficient  $\omega$  (McDonald, 1999; Zinbarg, Revelle, Yovel, & Li, 2005; Zinbarg, Yovel, Revelle, & McDonald, 2006), considering the correlation matrix. For the variables regarding speed,  $\omega = .86$ . For the variables regarding accuracy,  $\omega = .64$ . Since these variables all have positive pairwise correlations, we may also calculate the  $\alpha$  coefficient of Cronbach (1951) and the  $\rho^*$  coefficient (Brown, 1910; Spearman, 1910). We obtain  $\alpha = .85$  and  $\rho^* = .70$ . Regarding the decoding speed, if we select variables having positive pairwise correlations

**Table 16.** Correlations Between Factors and Variables.

Factor	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$Y_1$	$Y_2$	$Y_3$
$F_1$	-.89	.93	-.65	-.88	.90	-.75	.93	.94	-.13
$F_2$	-.09	.15	.54	-.13	.15	.42	.17	.17	.83

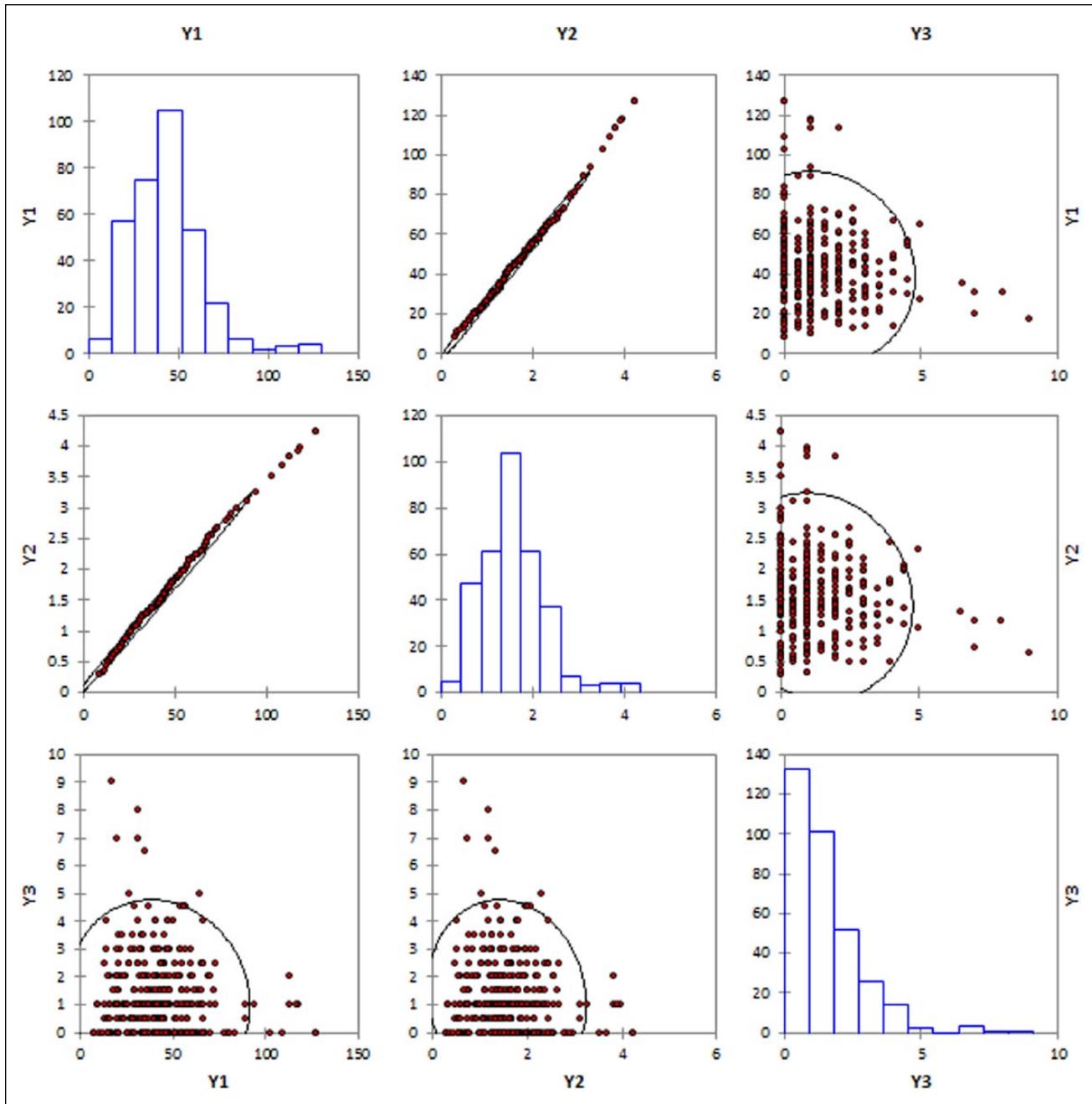
Note. The factors were extracted with principal components methods and are unrotated.

(viz.,  $X_2, X_5, Y_1, Y_2$ ),  $\omega = .94$  and both the  $\alpha$  coefficient and the  $\rho^*$  coefficient are .98. All indexes show a high intercorrelation among variables belonging to each set. These findings show that the number of words and the number of syllables per second read in 1-minute procedure are measures of the same feature measured by the variables regarding decoding speed on the standardized (much longer) tests. In addition, the number of errors on the screening procedure is a measure of the same feature measured by the number of errors on the list of words and nonwords.

To evaluate the reliability of the word reading test, the nonword reading test, and the new screening procedure, we may consider the percentages of variance explained by the first two factors of the correlation matrices, since all these reading tasks measure two latent constructs of the decoding skill. In the word reading test, the first two factors (extracted with both the PC and the CF method) of the correlation matrix of variables  $X_1, X_2$ , and  $X_3$ , explain 92.7% of the variance. On the nonword reading test, the first two factors (extracted with both the PC and CF methods) of the correlation matrix of variables  $X_3, X_4$ , and  $X_5$  explain 93.2% of the variance. In the screening procedure, the first two factors (extracted with both the PC and the CF method) of the correlation matrix of variables  $Y_1, Y_2$ , and  $Y_3$  explain 99.97%. These percentages indicate a very high reliability of all the three reading tasks.

### The New Composite Indicator

In each grade, the scatterplots matrix of  $Y_1, Y_2$ , and  $Y_3$  clearly shows noncorrelation and also independence between the variable measuring accuracy and any of the variables measuring speed. Figures 4, 5, 6, 7, and 8 show that for the couples of variables  $Y_1, Y_3$  and  $Y_2, Y_3$ , the sample points are distributed around the center of mass in a spherical manner. The fact that speed and accuracy are two different factors of decoding ability was also highlighted in the previous section, by means of the factor analysis performed over all the variables. An explorative factor analysis on the three variables involved in the screening procedure confirms this idea: The first factor is highly correlated with  $Y_1$  and  $Y_2$  and explains 67% of the total variance; the second factor is uniquely correlated with  $Y_3$  and explains 33% of the total variance (both with the PC and the CF method). Drawing from these results, we may chose  $Y_1$  for speed (but



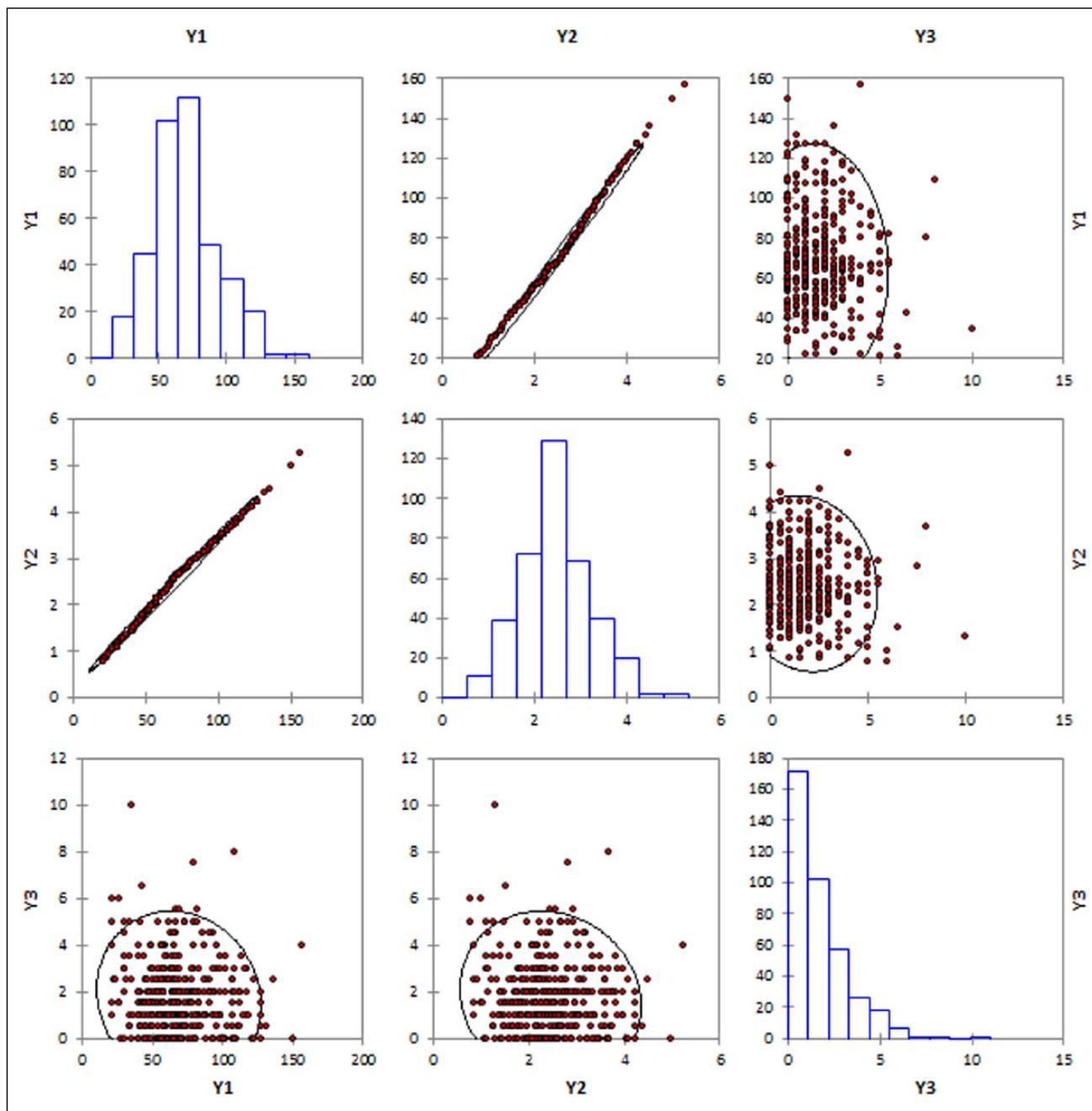
**Figure 4.** Scatterplot matrix of variables  $Y_1$ ,  $Y_2$ , and  $Y_3$  in Grade I.

we may choose  $Y_2$  as well), and we propose the following composite indicator:

$$MD = Z_1^2 + Z_3^2,$$

Here  $Z_1$  is the  $z$  score of  $Y_1$  and  $Z_3$  is the  $z$  score of  $Y_3$ . For each observation,  $MD$  is the Mahalanobis distance from the center of mass of the sample points in the two-dimensional Euclidean space spanned by  $Y_1$  and  $Y_3$ . Since  $Y_1$  and  $Y_3$  are

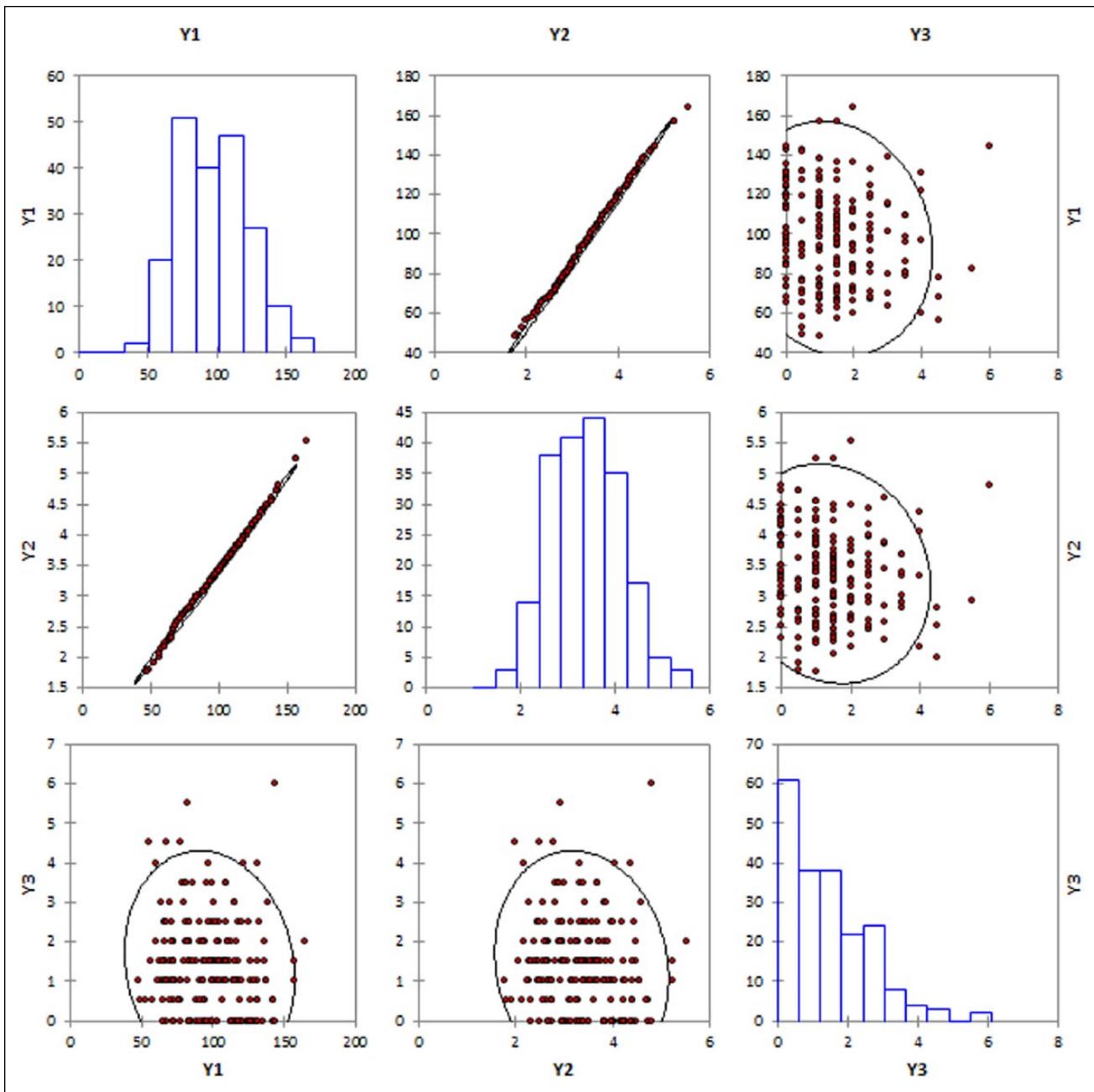
noncorrelated,  $MD$  coincides with the Euclidean distance calculated on the  $z$  scores. Observations with high values of  $MD$  can be considered atypical, without making any assumptions on the bivariate distribution of  $Y_1$  and  $Y_3$ . However, if  $Y_1$  and  $Y_3$  were Gaussian, observations with the same value of  $MD$  would have the same density in the bivariate normal distribution and, asymptotically,  $MD \sim \chi^2_{g=2}$ . Therefore, the value of 6 would discriminate 5% of the population, since  $P(\chi^2_{g=2} \geq 6) = 0.05$ . Our aim is to determinate a cutoff value



**Figure 5.** Scatterplot matrix of variables  $Y_1$ ,  $Y_2$ , and  $Y_3$  in Grade II.

discriminating nearly 5% of the population. Since learning disability can affect between 3% and 4% of the Italian scholastic population, a screening procedure should identify a higher percentage of impaired students: Subsequent analysis will determine the causes of this impairment. For identifying observations that are atypical, we must consider that the variables are not Gaussian and that a high value of  $MD$  may identify not only impaired readers but also readers with very good performances. To detect atypical values in the

direction of pathology, we have to observe the sign of  $Z_1$  and  $Z_3$ . If  $Z_1$  is positive and  $Z_3$  is negative, the reader is atypical because he or she has a very good performance. In all the other cases, the reader is atypical because he or she is impaired. In our sample, a threshold value equal to 6 classifies as impaired readers a percentage of students equal to 2.7% in Grade I, equal to 4.4% in Grade II, equal to 4.5% in Grade III, equal to 6.1% in Grade IV, and equal to 6.5% in Grade V. We are currently conducting a new study to

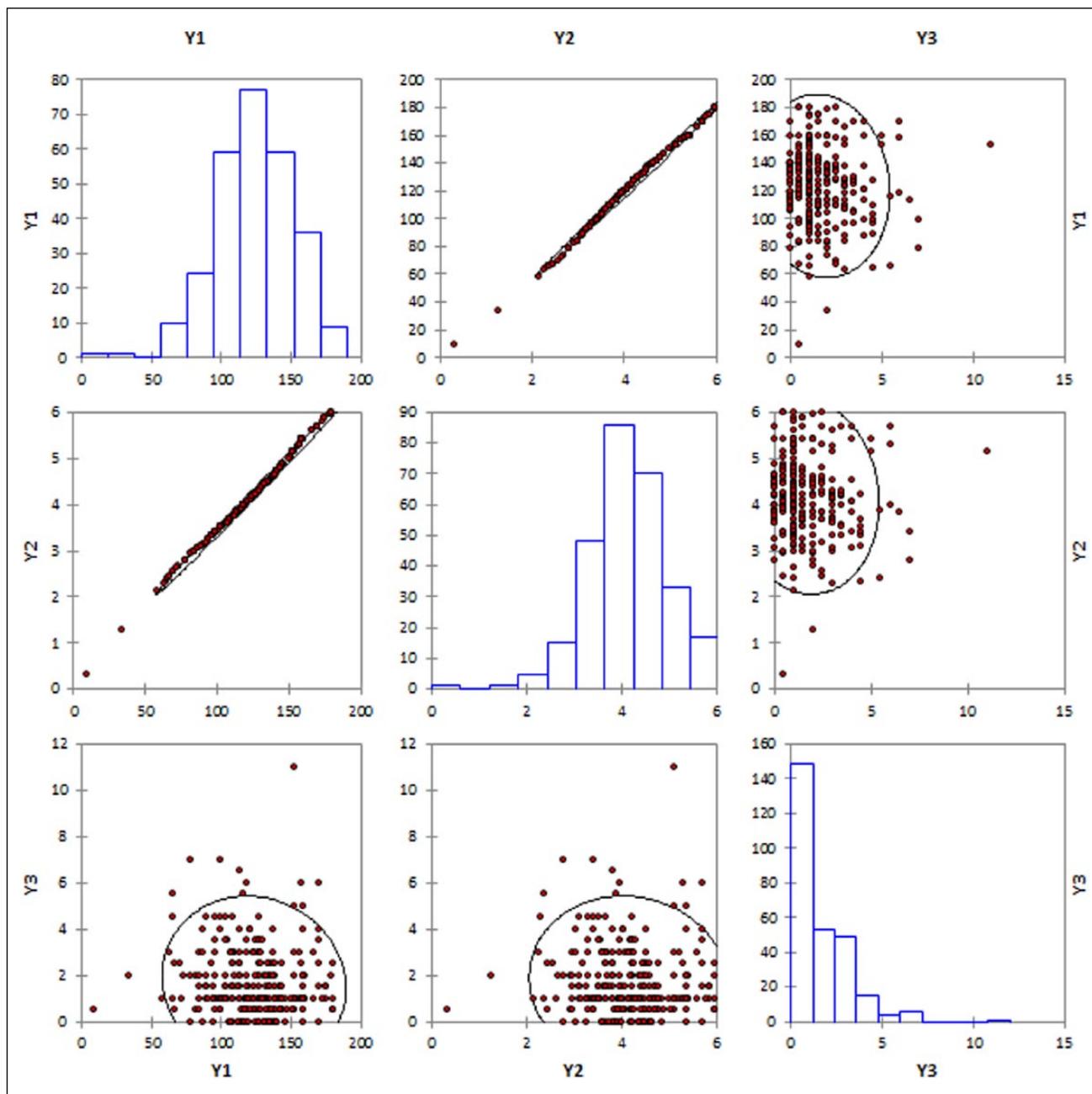


**Figure 6.** Scatterplot matrix of variables  $Y_1$ ,  $Y_2$ , and  $Y_3$  in Grade III.

investigate the validity of this threshold for the screening procedure. For this study we are selecting a large number of children to administer the screening procedure to, computing the index  $MD$ , estimating the empirical distribution of  $MD$ , and estimating the 95th percentile of this distribution.

The composite indicator  $MD$  may be used in any reading test where the variables measuring speed have high pairwise correlations and the variables measuring accuracy have high pairwise correlations but there is only a

slight correlation between a variable measuring speed and a variable measuring accuracy. Thus, the indicator can be used also on tests requiring students to read a selected list of words or nonwords. In our sample the indicator works well. Using  $X_2$  and  $X_3$ , with a threshold equal to 6, it classifies as impaired readers a percentage of children equal to 3.9% in Grade I, 3.1% in Grade II, 4.5% in Grade III, 4.7% in Grade IV, and 5.1% in Grade V. Using  $X_5$  and  $X_6$  the percentages of students classified as impaired readers are

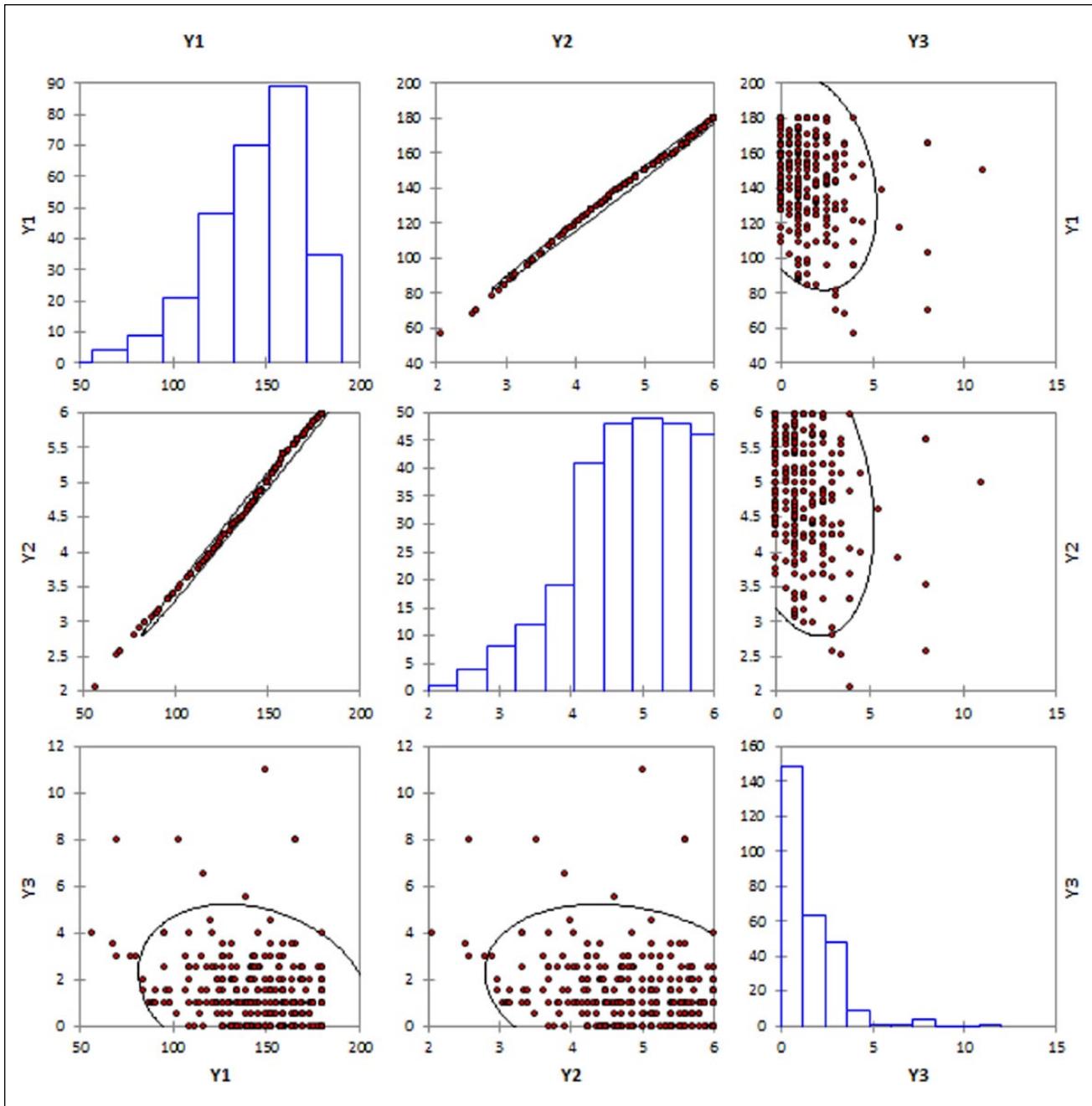


**Figure 7.** Scatterplot matrix of variables  $Y_1$ ,  $Y_2$ , and  $Y_3$  in Grade IV.

as follows: 4.2% in Grade I, 4.4% in Grade II, 4.0% in Grade III, 4.0% in Grade IV, and 5.5% in Grade V. All these percentages are similar across grades and are not far from the expected value (5%). Moreover, the percentages remain similar while changing the test. The use of this indicator reduces the discrepancies between results obtained with the list of words and the list of nonwords, and this has important desirable consequences in clinical practice.

### Conclusions

In this study we have proposed some statistical analyses of the variables measuring speed and accuracy in two standardized tests of reading and in a new screening procedure used in Italy to detect impaired decoders in elementary grades. We believe that our findings have important implications for both research and clinical assessment. We also think that most of the findings may be extended to reading



**Figure 8.** Scatterplot matrix of variables  $Y_1$ ,  $Y_2$ , and  $Y_3$  in Grade V.

tests used in different countries. For clinical assessment, the most important findings can be summarized as follows:

- Variables measuring decoding speed (both in terms of time or number of words and in terms of syllables read in a second) are not Gaussian, are asymmetric, and present many outliers. The skew is always positive in variables measuring the time and syllables per second

on tests where the time depends on the ability of the reader. Variables measuring the syllables per second are bounded in the pathology direction and assume a small range of values in this direction. On the contrary, variables measuring time in seconds have no limits in the deficit direction and assume a wide range of values in the pathology domain. For these reasons, the time in seconds has a more discriminative power

than the number of syllables and should be preferred in clinical practice. For variables measuring the number of words or syllables read in a second in a procedure where the time is fixed, the skew is positive in the first school years and becomes negative in later school years. The negative skewness and the presence of outliers indicate that these variables have more discriminative power in the later years of primary school, when students read more fluently.

- Variables measuring the number of errors always have a positive skew. This is desirable since the direction of pathology in these variables is positive (impaired readers are students with high values in these variables). Although the average number of errors on a test where the time depends on the ability of the student decreases from Grade I to Grade V, the number of errors on a procedure where the time is fixed has a roughly constant pattern. This means that for a screening with fixed time if one student increases performance from one grade to the subsequent grade, he or she increases the speed of decoding without diminishing decoding accuracy.
- Because of the nonnormality of the variables measuring speed and accuracy, the threshold values must be estimated on the basis of the percentiles of the empirical distribution of these variables, rather than by using the percentiles of the normal distribution. With the currently used thresholds in Italy (based on the assumption of normality), we have contrasting results. For example, in Grade II, 30 students are classified as impaired on time (in seconds) in reading the list of words, whereas only 2 are classified as impaired for the number of syllables per second in reading the list of words. Using the same test (the list of words or the list of nonwords), many students are classified differently when using time in seconds or the number of syllables per second. This is because the variables are not Gaussian and are asymmetric. The  $z$  scores are therefore not Gaussian. The 95th percentile of the  $z$  scores of these variables does not coincide with the 95th percentile of the standard normal distribution (the same for the 5th percentile). This finding presents many doubts about the validity of the normative thresholds currently used in Italy. This is an important issue for clinicians given that a diagnosis of decoding disability can provide access to treatment and to other facilities that are reserved for individuals with dyslexia at school.

The following findings are relevant for researchers:

- Speed and accuracy are two orthogonal latent factors of decoding skill. The consequence is that scores that

are obtained from a same reading performance will necessarily be incongruent.

- Rather than using a measure of speed and a measure of accuracy, we may consider a composite indicator. A person may be impaired for speed but in the normal range for accuracy, or vice versa. This can have undesirable consequences: Studies using different scores may support different theories of decoding ability, and decoding ability classifications may change according to the measure used. In this work we have proposed a new composite indicator,  $MD$ , that considers a variable measuring speed and a variable measuring accuracy. For each observation, the value of  $MD$  is the Mahalanobis distance from the center of mass of the sample points in the two-dimensional Euclidean space spanned by the two variables. Observations with high values on this indicator can be considered atypical, without making any assumptions about the bivariate distribution of the variables. If these variables were found to be Gaussian, observations with the same value of  $MD$  would have the same density in the bivariate normal distribution and, asymptotically,  $MD \sim \chi^2_g = 2$ . The use of this indicator requires only the noncorrelation of the two measures. Therefore,  $MD$  can be used in any test and in any screening procedure where speed and accuracy are found to be orthogonal factors.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### References

- Anderson, T. W., & Darling, D. A. (1952). Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. *Annals of Mathematical Statistics*, *23*, 193–212.
- Badian, N. A. (1982). The prediction of good and poor reading before kindergarten entry: A 4-year follow-up. *Journal of Special Education*, *16*, 309–318.
- Badian, N. A. (2000). Do preschool orthographic skills contribute to prediction of reading? In N. Badian (Ed.), *Prediction and prevention of reading failure* (pp. 31–56). Timonium, MD: York Press.
- Blachman, B. A., Ball, E., Black, R., & Tangel, D. (1994). Kindergarten teachers develop phoneme awareness in low-income, inner-city classrooms: Does it make a difference? *Reading and Writing: An Interdisciplinary Journal*, *6*, 1–17.
- Bowers, P. G., & Wolf, M. (1993). Theoretical links among naming-speed, precise timing mechanisms and orthographic skill in dyslexia. *Reading and Writing*, *5*, 69–85.

- Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology*, 3, 296–322.
- Catts, H. W., Hogan, T. P., & Adolf, S. M. (2005). Developmental changes in reading and reading disabilities. In H. W. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 25–40). Mahwah, NJ: Lawrence Erlbaum.
- Compton, D. L., & Carlisle, J. F. (1994). Speed of word recognition as a distinguish characteristic of reading disabilities. *Educational Psychology Review*, 6, 115–140.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
- De Mauro, T. (1997). *Guida all'uso delle parole* [Handbook for using words]. Rome, Italy: Editori Riuniti.
- Dennis, I., & Evans, J. B. T. (1996). The speed-error trade off problem in psychometric testing. *British Journal of Psychology*, 87, 105–129.
- Flesch, R. F. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221–223.
- Francis, D. J., Fletcher, J. M., Catts, H. W., & Tomblin, J. B. (2005). Dimensions affecting the assessment of reading comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 369–394). Mahwah, NJ: Lawrence Erlbaum.
- Fraser, J., Goswami, U., & Conti-Ramsden, G. (2010). Dyslexia and specific language impairment: The role of phonology and auditory processing. *Scientific Studies of Reading*, 14, 8–29.
- Goswami, U., Gombert, J. E., & Barrera, L. F. (1998). Children's orthographic representations and linguistic transparency: Nonsense word reading in English, French and Spanish. *Applied Psycholinguistics*, 19, 19–52.
- Hoover, W. A., & Tunmer, W. E. (1993). The components of reading. In G. B. Thompson, W. E. Tunmer & T. Nicholson (Eds.), *Reading acquisition processes* (pp. 1–19). Adelaide, Australia: Multilingual Matters.
- Invernizzi, M., Justice, L., Landrum, T. J., & Booker, K. (2004). Early literacy screening in kindergarten: Widespread implementation in Virginia. *Journal of Literacy Research*, 36, 479–500.
- Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6(3), 255–259.
- Jarque, C. M., & Bera, A. K. (1981). Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence. *Economics Letters*, 7(4), 313–318.
- Jarque, C. M., & Bera, A. K. (1987). A test for normality of observations and regression residuals. *International Statistical Review*, 55(2), 163–172.
- Jiménez, J. E., & Hernández-Valle, I. (2000). Word identification and reading disorders in the Spanish language. *Journal of Learning Disabilities*, 32, 267–275.
- Jiménez, J. E., & Ramírez, G. (2002). Identifying subtypes of reading disability in the Spanish language. *Spanish Journal of Psychology*, 5, 3–19.
- Jiménez, J. E., Rodríguez, C., & Ramírez, G. (2009). Spanish developmental dyslexia: Prevalence, cognitive profile and home literacy experiences. *Journal of Experimental Child Psychology*, 103, 167–185.
- Joshi, R. M., & Aaron, P. G. (2002). Naming speed and word familiarity as confounding factors in decoding. *Journal of Research in Reading*, 25(2), 160–171.
- Justice, L. M., Invernizzi, M. A., & Meier, J. D. (2002). Designing and implementing an early literacy screening protocol: Suggestions for the speech-language pathologist. *Language, Speech, and Hearing Services in Schools*, 33, 84–101.
- Keenan, J. M., & Betjemann, R. S. (2006). Comprehending the Gray Oral Reading Test without reading it: Why comprehension tests should not include passage-independent items. *Scientific Studies of Reading*, 10, 363–380.
- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, 12, 281–300.
- Lilliefors, H. (1967). On the Kolmogorov-Smirnov Test for Normality with mean and variance unknown. *Journal of the American Statistical Association*, 62, 399–402.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum.
- Nation, K., & Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, 67, 359–370.
- O'Connor, R. E., & Jenkins, J. R. (1999). The prediction of reading disabilities in kindergarten and first grade. *Scientific Studies of Reading*, 3, 159–197.
- Olson, R. K., Gillis, J. J., Rack, J. P., DeFries, J. P., & Fulker, D. W. (1991). Confirmatory factor analysis of word recognition and process measures in the Colorado Reading Project. *Reading and Writing: An Interdisciplinary Journal*, 3, 225–248.
- Paulesu, E., Demonet, J. F., Fazio, F., McCrory, E., Chanoine, V., Brunswick, N., & Frith, U. (2001). Dyslexia: cultural diversity and biological unity. *Sciences*, 291, 2165–2167.
- Sartori, G., Job, R., & Tressoldi, P. E. (1995). *DDE Batteria per la valutazione della dislessia e della disortografia evolutiva* [DDE battery for the evaluation of dyslexia and disorthography]. Florence, Italy: Giunti O.S.
- Sartori, G., Job, R., & Tressoldi, P. E. (2007). *DDE-2 Batteria per la valutazione della dislessia e della disortografia evolutiva-2* [DDE-2 battery for the evaluation of dyslexia and disorthography]. Florence, Italy: Giunti O.S.
- Serrano, F. D., & Defior, S. (2008). Dyslexia speed problems in a transparent orthography. *Annals of Dyslexia*, 58, 81–95.
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, 94, 143–174.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52, 591–611.
- Sinatra, G. M., & Royer, I. M. (1995, April). *Investigating a new method of combining response time and accuracy scores on measures of reading ability*. Paper presented at the annual meeting of the Society for Scientific Studies of Reading, San Francisco, CA.
- Spearman, C. C. (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271–295.
- Sprengr-Charolles, L., Colé, P., Lacert, P., & Serniclaes, W. (2000). On subtypes of developmental dyslexia: Evidence from processing time and accuracy scores. *Canadian Journal of Experimental Psychology*, 54, 88–104.

- Sprenger-Charolles, L., Siegel, L. S., Jiménez, J. E., & Ziegler, J. C. (2011). Prevalence and reliability of phonological, surface, and mixed profiles in dyslexia: A review of studies conducted in languages varying in orthographic depth. *Scientific Studies of Reading, 15*, 498–521.
- Stella, G., Scorza, M., & Morlini, I. (2011). *SPILLO Strumento per l'identificazione della lentezza nella lettura orale. Software per la verifica delle abilità di lettura dei bambini della scuola primaria (dalla prima alla quinta)* [SPILLO. Tool for the identification of slowness in reading aloud. Software for evaluating the reading ability of children in primary school (from first to fifth grade)]. Florence, Italy: Giunti Scuola.
- Torgersen, J. K., Wagner, R. K., & Rashotte, C. A. (1997). Prevention and remediation of severe reading disabilities: Keeping the end in mind. *Scientific Studies of Reading, 1*, 217–234.
- Vadasy, P. F., & Sanders, E. A. (2011). Efficacy of supplemental phonics-based instruction for low-skilled first graders: How language minority status and pretest characteristics moderate treatment response. *Scientific Studies of Reading, 15*, 471–497.
- Vadasy, P. F., Sanders, E. A., & Abbott, R. D. (2008). Effects of supplemental early reading intervention at 2-year follow up: Reading skill growth patterns and predictors. *Scientific Studies of Reading, 12*, 51–89.
- Vernon, P. A. (1990). An overview of chronometric measures of intelligence. *School Psychology Review, 19*, 399–410.
- Ziegler, J. C., Perry, C., Ma-Wyatt, A., Ladner, D., & Schulte-Korne, G. (2003). Developmental dyslexia in different languages: Language-specific or universal? *Journal of Experimental Child Psychology, 86*, 169–193.
- Zinbarg, R., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's, Revelle's, and McDonald's: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika, 70*, 123–133.
- Zinbarg, R., Yovel, I., Revelle, W., & McDonald, R. (2006). Estimating generalizability to a universe of indicators that all have an attribute in common: A comparison of estimators for  $\omega$ . *Applied Psychological Measurement, 30*, 121–144.