

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

5,800

Open access books available

142,000

International authors and editors

180M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Screening Young Children at Risk for Reading Failure

Sotiria Tzivinikou

Abstract

Reading and reading difficulties are some of the most researched topics in the literature in regard to psychology and education. Additionally, some specific subjects such as prediction and prevention attract research interest as well. These issues are discussed in the present chapter that focused on the screening measures and their characteristics towards significance and effectiveness. More specifically, discrimination accuracy, sensitivity, and specificity as well as validity and reliability were taken into consideration. Some well-known studies were examined revealing a range of methodological issues, which affected the effectiveness of using measures in the extant research. Although the findings were consistent with literature, they continued to be scant and not widely accepted, affected by several limitations regarding the sampling and the experimental design.

Keywords: reading difficulties, screening, discrimination accuracy, sensitivity, specificity

1. Introduction

The reading struggling and prevention of reading failure are among the most important and well-studied subjects in the relevant literature. Two decades earlier, Joseph Torgesen, in his influential article “Catch Them Before They Fall: Identification and Assessment to Prevent Reading Failure in Young Children” argued that *“The best solution to the problem of reading failure is to allocate resources for early identification and prevention. The goal is to describe procedures ... to identify children who need extra help in reading before they experience serious failure...”* [1].

Actually, in the following years, great emphasis has been placed on the issue of screening for at-risk children and important research findings have emerged, such as Ref. [2] findings showing that most children at risk for early reading difficulties could be effectively identified at the beginning of kindergarten. As the literature review shows, a lot of effective and precise screening tools and procedures have been developed in order to locate the at-risk children as soon and as precisely as possible.

2. Considerations on effectiveness of screening

It is widely accepted that diagnostic assessment is not practical for assessing all children for academic risk, while screening procedures could provide reliable and

valid information regarding children's current academic skills and meet financial and time constraints [3]. However, screening is a preliminary process of identification that could identify those children who may be at risk of future difficulty in school and in need of further individual diagnostic testing. More specifically, it is a brief assessment that provides predictive information about a child's development in a specific academic area, in order to identify at-risk children that need extra support through early intervention. The screening measure is administered to all children and is used to identify an initial risk pool of children suspected of being at risk of developing reading disabilities. Screening information leads to the decision of risk for each child screened. Risk decisions are made by selecting a critical cut-point along a continuum of scores on a single or group of screening measures [4].

Screening may include parent interviews or written questionnaires and checklists, observation of the child, or use of specific screening tests. Because the earlier a learning disability is detected, the better chance a child will have of succeeding in school and in life, it is used mainly at the kindergarten or at the beginning of the first grade. Often, early identification is delayed, and as a result, the at-risk children might experience significant problems in learning to read. The consequences of these delays for the child include prolonged frustration, missed opportunities for special instructional interventions, and cumulative academic deficiencies, as well as lifelong secondary psychological problems.

From early years until now, there has been a common understanding of characteristics of effective developmental screening tests. These characteristics are an adequate standardization sample, low cost, ease of administration, appropriate content, and adequate validity and reliability (e.g., see [5, 6]). However, predictive validity or instrument reliability has also been cited as a major problem in screening for children at risk [7–10]. Ref. [11] stated "... a test with a low predictive value is unlikely to be either efficient or useful..." (p. 1583). An effective framework is usually appreciated based on the measures of relevance and utility. Relevance of the measures relates to the relationship between the measure and the purpose of the assessment on the one hand, and the utility of the measures on the other hand, which is usually evaluated by cost-effectiveness [12].

Screening studies discussed the outcome results as poor or good, with poor indicating a subject who exhibits the target disorder and good a subject who does not. The measurement is realized in two points of time. Based on the measurement results, four placements may occur; the subject may be placed in cell A: failed screen and poor outcome = true positive; cell B: failed screen and good outcome = false positive; cell C: passed screen and poor outcome = false negative; and cell D: passed screen and good outcome = true negative. The matrix is deceptively simple and easy to misinterpret, because cell information varies in relation to rows, columns, or the entire matrix [7, 13].

On the other hand, a vast majority of the studies recommended the assessment of accuracy in terms of sensitivity and specificity as appropriate indices to identify the capacity of an examined screening instrument (**Table 1**). These indices can be calculated using the formula: Sensitivity = $TP / (TP + FN)$ and Specificity = $TN / (TN + FP)$. Sensitivity and specificity are two sides of a coin. Sensitivity is related to the probability that a result of a test will be positive, when the criterion—in this case, disability—is present. Expressed as a percentage, sensitivity measurement results in a true positive rate. On the contrary, specificity produces a true negative rate expressed as a percentage, referring to the probability that a test result will be negative when the criterion—in this case, disability—is not present. The overall classification accuracy can be estimated using the Eq. $(TP + TN) / (TP + FP + FN + TN)$ [5]. Positive likelihood ratio is the ratio between the probability of a positive test result given the presence of the disease and the probability of a positive test result given the absence of risk (e.g., [4, 8, 12, 14–28]).

| Predictor (screen) | Poor outcome (criterion) | Good outcome (criterion) |
|-----------------------|--------------------------|--------------------------|
| Poor (Fail to screen) | (TP) True positive | (FP) False positive |
| Good (Pass to screen) | (FN) False negative | (TN) True negative |

Sensitivity = $TP / (TP + FN)$
 Specificity = $TN / (TN + FP)$
 Classification accuracy = $(TP + TN) / (TP + FP + FN + TN)$

Table 1.
Screening results table.

Using a risk index can serve as a good alternative to single cut scores. This index includes calculations as probability of being classified as at risk or not at risk. A weighted regression formula of predictors to a specific outcome determines the classification and the construction of the risk index. Moreover, the ability of a test to discriminate diseased cases from normal cases is evaluated using a receiver operating characteristic (ROC) curve analysis. ROC curves can also be used to compare the diagnostic performance of two or more screening tests [5, 29].

An ROC curve is provided by a screen that cannot discriminate between cases and non-cases. This is a straight line passing through the origin with unit slope, and effective screens will provide a convex curve above this line. Area under the curve (AUC), that is, the ROC curve, provides a measure of the screening test performance. This measure goes beyond sensitivity and specificity at a single threshold, integrating the full range of scores that need to be taken into account for making a decision about a threshold in order to separate illness from health. This practically means that a value of 0.5 (that is under the straight line of unit slope) indicates a lack of effectiveness, whereas a value very close to 1.0 is indicative of a very good screen.

Ref. [3] noted that the AUC is an indicator of a screening tool's overall ability to differentiate between children with lower-than-average emergent literacy skills and children with average or better emergent literacy skills, and it is calculated at all possible cut scores. Using optimal cut score statistics allows examination of the utility of the screening tool under the circumstances in which it would typically be used. Ref. [4] suggested that AUC values above 0.90 represent excellent diagnostic accuracy, between 0.80 and 0.90 represent good, 0.70–0.80 fair, and values below 0.70 are considered poor.

3. Single or multiple predictors and criterion measures

Large amounts of predictors have been proposed by researchers. Several pre-reading measures, when administered in kindergarten, are predictors of later reading abilities. These measures include letter name and letter sound knowledge, phonological awareness, verbal short-term memory, and rapid automatized naming [6].

Two related studies [23, 24] found that measures of letter naming, phonological awareness, rapid object naming, and non-word repetition at the beginning of kindergarten were very good predictors of reading outcomes at the end of the first grade. Ref. [2] further has shown that measuring at-risk children's response to supplemental intervention during kindergarten can improve accuracy of identification beyond that of early screening. Even when predicting performance on the state assessment in the third grade, Ref. [5] found that a comprehension measure was the best predictor. In addition, the review [34] revealed that risk factors associated with speech and language delay were male gender, family history, and low parental education.

Moreover, phonological awareness was recognized by Refs. [16, 17] as an important risk factor. However, Ref. [8], proposed as risk factors the letter-name knowledge, and the rapid serial naming, reference [20], proposed the Initial Sound Fluency task of the DIBELS, reference [21], proposed the rapid naming objects, reference [22], proposed the Word Identification and Passage Comprehension subtests and the Word Attack subtest of the WJ-R., and final reference [19], proposed as risk factors the Letter-Name Fluency (LNF), and the Nonsense Word Fluency (NWF).

Additionally, most of the screening studies used multiple predictors, and all of them used phonological processing measures [8, 16–22]. Some of them used the total or part of a specific screening test in order to test their validity and reliability [20–22]. Some others used measures such as pre-reading behaviors, reading habits [18], or working memory [30]. Others used parents or self-reported questionnaires and checklists [31, 32], and finally some used teacher ratings [28, 33].

Similar risk indicators have been used in the context of the newest screening studies. For example, a multivariate screening battery was administered by Ref. [4] to 252 beginning first-grade children. The children had low initial reading abilities, and their reading outcomes were measured at the end of the second grade. Logistic regression analyses showed a high degree of accuracy concerning the prediction of reading outcomes. This screening model, which proved to be highly accurate, included measures of phonological awareness, rapid digit naming, and oral vocabulary.

Ref. [28] examined 240 fourth-grade children and they were classified as not-at-risk or at-risk readers based on a three-factor model reflecting reading comprehension, word recognition/decoding, and word fluency. More specifically, participants were assessed using measures of reading comprehension, oral language, word recognition, word decoding, phonological processing, auditory memory, and spelling.

As criterion measures, all of them used reading ability tested by a number of standardized and normalized reading tests. The most popular of them were the Woodcock Diagnostic Reading Battery; Woodcock-Johnson Psycho-Educational Battery-Revised; CTOPP; Reading-Gray Oral Reading Test; WRAT Spelling; and Peabody Individual Achievement Test.

4. Research design considerations and findings

Regarding the experimental design of the screening studies, it could be noted that a lot of these had longitudinal or follow-up designs and the other half had a cross-sectional one. Commonly, the follow-up studies had two phases with one-year interval. Others had different designs, for example, Ref. [21] included three phases and 16-month interval and Ref. [17] presented two phases and 4–6-week interval. These studies administered the set of predictors (tests or part of tests or single measures) and at the second phase, the criterion measures were administered, that is, the reading ability measures. The studies with cross-sectional designs administered the predictors and the reading measures at the same time.

There are two approaches to the study of reading disabilities. Firstly, the most common approach to reading assessment is to separate children into groups based on their reading scores. Consequently, it is important to determine if variables thought to be related to the development of reading skills are predictive of group membership, that is, they predict if the child belongs to the at-risk group or not. Secondly, the alternative approach is to consider reading as a continuum of abilities. Based on that, it is significant to determine if the variables thought to influence the development of reading abilities can predict the full range of the child's reading scores obtained. Concerning the significant discriminant function models regardless of which language measure was used, classification accuracy was about as good

or better for the typical reading group as it was for the poor reading groups [34]. Screening studies mainly used t-tests, ANOVAs, MANOVAs; correlations; logistic regression; and discriminant analysis. Often, the cutoff scores used by the studies were arbitrary, usually recommended by the literature (e.g., [16]) or revealed by the statistic multiple analyses to give the best results [20, 31, 32].

Screening procedures that result in sensitivity levels at or above 90% and specificity levels of at least 80% are generally deemed acceptable ([29]). An alternative index of accuracy is the area under the receiver operating characteristic (ROC) curve. According to Ref. [29], an ROC curve is a plot of the true positive rate (sensitivity) against the false positive rate (specificity) for each of the cut points of a decision-making instrument. Therefore, the area under the curve (AUC) may be used as an overall estimate of the accuracy of an assessment. Values above 0.80 are considered good, while values above 0.90 are excellent [29]. Ref. [25] found that AUC was 0.84 when reading outcome was based on individual component measures of reading and 0.86 when reading outcome was based on a composite score for reading.

Ref. [3] had administered at two time points two screening tools to 176 preschoolers. Specifically, the study used the Revised Get Ready to Read! (GRTR-R) tool, the Individual Growth and Development Indicators (IGDIs), and a diagnostic measure. Comparing the two screening tools based on a receiver operating characteristic curve analysis, it emerged that, at optimal cut scores, IGDIs provided less accurate classification of children's overall emergent literacy skills than GRTR-R. However, neither measure was particularly good at classifying specific emergent literacy skills.

On the other hand, Ref. [23] examined if kindergarten measures of language ability predicted reading comprehension difficulties independently of direct word reading measures. In addition, they investigated if response to language intervention in kindergarten added to the prediction of third-grade reading comprehension. The participants were 263 kindergarten children at risk and 103 children for control group matched in age.

Ref. [26] examined and evaluated if and to what extent R-CBM and CBM maze were technically adequate to inform their use in the context of a universal screening program of reading in fourth and fifth grades. The results of the study suggest evidence of short- and long-term alternate forms of reliability, criterion validity, and predictive validity for both R-CBM and CBM maze. It is also supported that possibly the two measures are comparable for use in universal screening at those grade levels. Therefore, the study suggests that R-CBM and CBM maze could be used interchangeably for screening of reading outcomes.

Ref. [34] was a review aimed to update the evidence on screening and treating children for speech and language delay in children through 5 years of age. In 23 studies evaluating the accuracy of screening tools, sensitivity ranged between 50 and 94%, and specificity ranged between 45 and 96%. As noted above, 12 treatment studies improved various outcomes in language, articulation, and stuttering. There has been restricted evidence concerning interventions that provided other improved outcomes or adverse effects of treatment. Male gender, family history, and low parental education were the main risk factors that were related to speech and language delay. The use of various screening tools can lead to accurate identification of children who need/undergo diagnostic evaluations and interventions. Evidence, on the other hand, is not adequate concerning their applicability in primary care settings. In addition, some treatments for young children, who have been identified with speech and language delays and disorders, may be effective.

The recent study of Ref. [35] aimed at dyslexia's early detection via machine by observing how people interact in the context of a linguistic computer-based game. In order to train a statistical model that predicts readers with and without dyslexia using measures derived from the game, they examined 267 children and adults.

Specifically, the model was trained and evaluated in a 10-fold cross experiment. Using the most informative features, it reached an 84.62% of accuracy.

Another recent study of Ref. [12] focused on a year-end state reading assessment in two states. The study examined the predictive validity and classification accuracy of individual- and group-administered screening measures related to student performance. A total of 321 students participated in the study, and in the fall of fourth grade, they were assessed regarding word-level, text fluency, and reading comprehension. Logistic regression results, applying a multivariate approach, revealed minimal to no increase in classification accuracy over the single comprehension measure. Receiver operating characteristic (ROC) curve analyses determined local cut scores to maintain sensitivity constantly at 0.90; this resulted in a large number of false positives.

Referring to predictive accuracy, Ref. [16] in accordance with findings of the past decade found that both phonological awareness and letter identification yielded the highest overall results. Moreover, all the constructs were promising as far as the accuracy rates are concerned. The false positive rate ranged from 13 to 27%, depending on the construct. The false negative rate ranged from 0.06 to 0.21%. Researchers continue to struggle with high hit and miss rates in predictive accuracy. Most importantly, researchers must address the high rate of false negatives. As funds and resources to provide reading interventions are limited, this is of particular practical importance to ensure that the most appropriate students are served.

The study of Ref. [17] examined the convergent and concurrent validity of two recently developed measures of phonological processing, the TOPA and the CTOPP. Both of these instruments used in combination appear to be useful in the early identification of children at risk for difficulty in learning to read. Based on the results, however, the use of either, or both, of these instruments as sole predictors of reading outcome cannot be supported.

The study of Ref. [20] compared DIBELS test with CTOPP. Specifically, the concurrent validity and diagnostic accuracy of the published test DIBELS was examined and was compared to the well-documented published test of CTOPP. Results suggest that the DIBELS strongly correlates with subtest and composite scores of the CTOPP that are designed to measure phonological awareness and memory, and less strongly with rapid naming tasks.

The findings of Ref. [18] indicated that the accuracy of the discrimination was high, 89.7%, with a 6.2% false negatives rate. However, using the calibration data from the reference group to identify at-risk status in a different sample, the accuracy fell to 80.2% with a 10.2% false negative rate.

Ref. [31] found that the Adult Reading History Questionnaire (ARHQ) was valid. This was demonstrated by the high correlation between the ARHQ and diagnostic measures for adults ($r_s = 0.57-0.70$). However, not every familial case is perfectly detected by ARHQ. Therefore, it would be more preferable and appropriate if clinicians and researchers used this questionnaire less as a diagnostic tool and more as a screening instrument.

The findings of Ref. [8] supported that letter name knowledge and rapid serial naming were most important in predicting later RD. The study had a sensitivity of 0.49 and specificity of 0.76. The findings of Ref. [21] were not consistent with the initial findings of the designers that the DEST was significantly and strongly correlated with later reading ability. Specifically, the rapid naming of objects variable emerged as a consistent predictor of later attainment, which predicted significant amounts of variability in reading and spelling, and the correlation coefficient were 0.344 ($p \leq 0.05$).

Ref. [22] examined the relations among standardized reading achievement tests, phonological awareness measures (CTOPP), and fluency rates (CBM, subtest of

Woodcock-Johnson Tests of Achievement-Revised) and how these measures relate to teacher ratings. The authors supported that measures of phonological awareness and reading fluency that provide further information may be included as part of reading assessment in addition to traditional norm-referenced measures of reading achievement.

Ref. [19] examined whether the measures could accurately identify poor readers in first grade. The sensitivity of phonological awareness was 42.9 and 66.7% for ORF and the WJ-R Word Attack, respectively, missing one-half and one-third of the students who later demonstrated reading problems. In addition, measures of letter name knowledge and letter sound knowledge were not sensitive in identifying students who were performing poorly on either first-grade reading criteria, with sensitivity of 57.1%.

Ref. [32] constructed a parent report checklist including information about the development history of the child and some indicators for reading problems. The author supported that this checklist was valid and reliable and it could be screened between RD and NRD with 97.2% discriminative accuracy.

In the study of Ref. [30], phonological awareness, distinctness of phonological representations, and phonological working memory were captured in the context of a series of tasks. Furthermore, a questionnaire was designed including two scales of self-reports: (a) one concerned with typical dyslexic symptoms and (b) one concerned with reading interest. The findings noted that the most powerful discriminator was the self-report data.

Ref. [36] examined the accuracy of teacher ratings. Therefore, kindergarten children identified by their teachers as making substandard progress toward one or more academic objectives performed significantly less well than a matched group of no identified children on tests of word reading, spelling, mathematics, and knowledge of letter names and letter sounds. Furthermore, by the end of the third school year, greater proportions of identified children than no identified children were receiving special learning assistance.

Another study examining teachers' rating was Ref. [33]. Kindergarten teachers appear to be better predictors of students who will not develop academic difficulty, as negative predictive values were consistently high regardless of the predictive variable. Variables associated with learning rather than behavioral or social variables may be better indicators of future academic achievement. The authors proposed that effective academic screening measures be used in conjunction with teacher ratings in order to maximize specificity in identifying children who are at risk for later learning disability early in their academic years.

More recently, Ref. [28] compared teacher ratings and reading factors as predictors for future reading competence. Specifically, they administered multiple measures of reading to 230 fourth-grade children. Teachers rated children's reading skills, academic competence, and attention. A three-factor model including reading comprehension, word recognition/decoding, and word fluency was used, in order to classify children as not-at-risk or at-risk readers. Predictors of reading status included group-administered tests of reading comprehension, silent word reading fluency, and teacher ratings of reading problems. The receiver operating characteristic curve (ROC) analysis yielded an area under the curve index of 0.90.

5. Screening in RTI context

The goal of universal screening is to promote the early identification of reading difficulties or potential reading difficulties. In order to prevent further difficulties,

screening measures that detect a large proportion of at-risk students would be desirable so that appropriate remedial support can be provided to students.

Screening and identification of students with/at-risk for reading difficulties represent an important first step in RTI models, for k-2 grades, and, in addition, for students in upper elementary grades where there is a particularly large percentage of struggling readers [12].

As Ref. [37] noted, during the last decade, responsiveness to intervention (RTI) has become popular among many practitioners. Specifically, it has been used as a means of transforming schooling into a prevention system with multiple levels. In order to be implemented successfully, RTI requires ambitious intent, a comprehensive structure, and coordinated service delivery. The level of its effectiveness also relies on building-based personnel that has specialized expertise at all levels of the prevention system.

In that context, a direct route approach to screening is typically employed by schools. Based on this approach, students identified as at risk by a screening process are directly placed in intervention. Direct route approaches require screening decisions to be highly accurate. However, few studies that have examined the predictive validity of reading measures report achieving recommendations concerning classification accuracy.

Ref. [5] compared two approaches that aimed at improving the classification accuracy of predictors of third-grade reading performance. Findings indicated that relying on single screening measures does not result in high levels of classification accuracy. Classification accuracy improved by 2% when a combination of measures was employed and by 6% when a predicted probability risk index was used.

On the other hand, from an RTI perspective, Ref. [24] investigated whether measures of language ability and/or response to language intervention in kindergarten uniquely predicted reading comprehension difficulties in third grade. A total of 366 participants were administered a battery of screening measures at the beginning of kindergarten and progress monitoring probes across the school year. A subset of participants also received a 26-week Tier 2 language intervention. Participants' achievement in word reading was assessed at the end of second grade, and their performance in reading comprehension was measured at the end of third grade. Results showed that measures of language ability in kindergarten significantly added to the prediction of reading comprehension difficulties over and above kindergarten word reading predictors and direct measures of word reading in second grade.

6. Discriminative accuracy-sensitivity-specificity-ROC analysis

A screening test could be perceived as effective in case it is norm-referenced, and it has appropriate content, validity and reliability, and ease of administration and interpretation. It also needs to be quick and cost-effective. An additional criterion is related to its discrimination accuracy with emphasis on false negative and false positive rates [7, 11]. The accuracy of screening measures is important given the concern of either mislabeling a child or failing to detect a delay.

Continuous efforts for improvement of accuracy of screening instruments have been reported in the relevant literature. These include using a combination of assessments and assessing risk on a continuum rather than as "fixed" cut scores. In addition, the use of probabilities based on multiple assessments has the potential to enhance the accuracy of the screening process by making screening decisions based on multiple indicators as well as on what is known about the prevalence of the condition under question.

However, according to Ref. [38], the concept of validity has expanded beyond the traditional correlation coefficient between a criterion and the new measure. It was defined as not only the degree with which the measure assesses the construct but also “the adequacy and appropriateness of the inferences and actions taken on the basis of the scores” (p. 13). Validity thus includes social consequences and relevance/utility in addition to more traditional concepts. Furthermore, the same reference, [38], included reliability, content, and criterion validity as part of construct validity. So, even though only a few of the reviewed studies were interested in reliability of testing measures, in accordance to Ref. [38], a larger number of these studies were interested in the other aspects (e.g., [19]).

If a test is not valid, then, reliability is moot. In other words, if a test is not valid, there is no point in discussing reliability, because test validity is required before reliability can be considered in any meaningful way. The studies that had emphasized reliability after validity’s validation were Refs. [31, 32].

The validity of any predictive instrument depends in part on two key factors: sensitivity and specificity. To compute sensitivity and specificity using the formula mentioned above, the performance of each child on the assessments was first classified as above or below the cutoff score. A cutoff score is a value below which poor school performance may be suspected [14].

Ideally, the determination of an appropriate cutoff score should be based upon locally developed norms. Ref. [39] supported the use of local cutoff points as well: “in order to differentiate those ‘at-risk’ children a cutoff may use local norms for the best predictability for future achievement in that school system” (p. 15). Nevertheless, Ref. [40] argued “the cut-off point(s) between normal reading and disabled reading is always arbitrary” (p. 30). In addition, Ref. [7] agreed that often the cutoff point is an arbitrary value that has been adjusted to achieve the best results in predictive accuracy. Once outcome data have been collected, the cutoff score may be altered to achieve the best results.

Emphasis is placed on interpretation of sensitivity and predictive value, both of which reflect a screen’s ability to accurately identify or predict subjects who will have a poor outcome. Reported values above 0.80 are considered acceptable for these indicators [7, 14].

From RTI’s perspective, researchers have argued that high levels of sensitivity are necessary for universal screening measures [12, 37]. Although consensus has not been reached regarding optimal levels of sensitivity, acceptable sensitivity values noted in the literature range from 0.70 to 0.90 [12]. Relatedly, specificity levels of at least 0.70 are generally considered adequate for screening measures.

Related to the labeling issue is the false positive rate, the number of children identified in kindergarten who were not poor readers in first grade. This means that children who do not need intervention may be identified as in need for it. Administrators may be more concerned with false negative rates as in [9], but another negative consequence related to false positive cases is the additional cost of the intervention.

However, Ref. [1] supported a different point of view and noted that schools should provide this intervention to as many children as possible, if they desire to maximize their chances for early intervention with the most impaired children. This may seem as a waste of resources at first glance. On the other hand, many of the falsely identified children receiving intervention are likely to be below-average readers even if they may not be among the most seriously disabled readers.

In any case, a possible solution to the over-identification rate was proposed by Ref. [40] by using a two-stage screening process or to provide small-group diagnostic interventions in the first grade. Consistent with them, Ref. [1] reported a significant reduction in the percentage of false negative errors within the same

sample of children by doubling the number of children they identified as at risk. About 10% of the children, who scored lowest on their predictive tests, resulted in a 42% false negative rate, while by using 20% of the children who scored lowest on their measures, the false negative rate was reduced to 8%.

Almost all of the studies used as predictors a battery of tests or multiple screening measures as Refs. [1, 9] proposed. However, some of the studies (e.g., see Ref. [18]) had used so many variables that the requisite general characteristics of the effective screening could be affected [7, 11]. So, there must be a balance between the demand of quickness, ease, cost-effectiveness, and other characteristics and the accuracy rate in order for a screening procedure to be possibly developed and accepted by the reading scientific community and educators, parents, and children.

A major contributor to the aspect of the discriminate accuracy is that often only a correlation coefficient between a group's scores on a preschool screening instrument and a later achievement measure is provided in the literature as evidence of the test's effectiveness. Such data, although important, provide information only on the similarity of the group's performance on both tests. A correlation coefficient provides no information as to the specific identification of the at-risk and not-at-risk children and the relationship between such status and the projected outcome of a group or poor reader [13].

Lack of discriminative accuracy data [17, 21, 22, 30] contributes to the difficulty of interpreting their findings in terms of screening effectiveness. Some studies had focused on these aspects and reported a range of accuracy and false positives, false negatives, and sensitivity and specificity. Better results (predictive accuracy over the 80%) regarding these aspects were reported by Refs. [18, 32]. Furthermore, Refs. [19, 33] reported a large number of cases; so, it was unclear which the best one was.

In terms of intervention programs designed to remediate deficiencies in at-risk students, false positives, although undesirable, are not critical. These children will receive a training program that they do not actually require. In some cases, the instruction could actually benefit the child's performance. Nevertheless, a concern of negative positives is that they place an increased demand on scarce resources [25].

On the other hand, a false negative error is more serious because these children do not receive the additional assistance they require at the earliest possible time, which makes their problems more difficult to remediate later [25]. A false negative classification will most likely deprive children of the benefits of early intervention because their test results incorrectly suggest that they are not at risk for learning difficulties. In such cases, the cost to the children may be devastating because they are likely to experience repeated failures and frustrations with academic tasks before they are actually identified and placed appropriately.

Is it possible for a screening measure to have a 0 false negative rate? Ref. [18] answered "no." Their explanations regard the different levels of readiness of children on their entry in school. In any case, scientific efforts will be continued in order to decrease the false rates of screening.

7. Conclusions

This chapter referred to the early identification and prediction of future low reading achievement and discussed the important aspects regarding effective predictors, the discrimination rate, and the sensitivity and specificity of the screening measures. However, because screening studies have usually used inconsistent measurement of risk factors, including heterogeneous patient populations, and inconsistently adjusted for confounders in multivariate models [34], their findings were not comparable.

For the best single or multiple predictors, there is evidence that batteries containing multiple tests generally provide better prediction than single instruments, but the increase in efficiency of multi-test batteries is generally not large enough to warrant the extra time and resources required to administer them [1, 5, 9]. Additionally, vocabulary measures proved to be one of the best unique predictors [23]. Moreover, Ref. [23] found that a measure of expressive vocabulary was a good predictor of reading comprehension status.

The most often measures that could be used as effective predictors were the letter name and letter sound knowledge, phonological awareness, verbal short-term memory, and rapid automatized naming [2, 4, 6, 23]. Very often, screeners were based on reading comprehension, word recognition/decoding, and word fluency [24, 28]. Additionally, some studies found as significant predictors the familial risk, and the child's specific characteristics, as well as his/her developmental and school history [32].

On the other hand, although Refs. [33, 36] found that teacher rating was a significant predictor that is consistent with a number of other studies, these ratings cannot substitute for early identification tests. Therefore, they proposed that combining test and teacher data would improve identification of kindergarten children at risk for reading failure. Recently, Ref. [28]'s findings were consistent with the above-mentioned studies.

A method used for validation of an early screening instrument should incorporate: (a) longitudinal design [6, 27], (b) independent assessments of kindergarten performance and learning ability separated by a temporal interval of specific time, [2, 21, 23, 24], (c) random sampling of children in a validation/cross-validation design, and (d) systematic assessment of predictive utility and validity [12]. There is clear evidence that early screening is a viable process, but this effort will only reach fruition, if research is conducted with appropriate rigor. However, there is a low incidence of educational handicaps, especially in the early grades. This means that a large sample size should be included for screening, and the formative evaluations should be age- and/or grade-specific and valid across grade levels for outcome comparisons.

More than a lot of the screening studies had longitudinal designs, and, the vast majority of the included studies did not adopt their proposed random sampling of participants. Therefore, a number of limitations emerged regarding the generalizability of the findings to other populations. The sampling of the studies was mainly constructed by self-selection of the participants or was a volunteer sample [8]. As Ref. [17] noted, the number of participants was modest and the sample was not selected randomly. Although the samples seemed representative of the school district from which they were selected, results may not be generalized to the larger population of young children or to specific subgroups. Quite a lot of the research was conducted with those methodological problems.

In summary, effective screening tools demonstrate high levels of sensitivity in correctly identifying those students who will actually encounter difficulties, as well as high levels of specificity in the accurate identification of those who are not likely to demonstrate reading difficulties. Ultimately, the goal is to maximize classification accuracy, a summative measure of the overall proportion of students who were correctly identified as at-risk or not at-risk on a screening measure.

8. Future research suggestions

The importance of early intervention has been proven by a large amount of research findings. In this context, the need for carefully designed and accurate screening measures emerges as crucial. Despite the recent interest and research on

screening reading disabilities, the body of research on the effectiveness of these measures remains problematic in terms of methodology and the findings seem to be scant. Therefore, the development of a cost-effective and equitable screening, diagnostic, and supportive method that is acceptable by government, educational authorities, school, children, and parents still remains a scientific challenge.

Therefore, it would be useful to design a large longitudinal study with 3 years' interval. Existing research has often used small and non-representative group sizes; thus, there remains a need for further research emphasizing on appropriate sampling in order to make it easy to extrapolate findings to other sampling and generally other situations.

The development of screening tools that are valid, reliable, easy to manage and interpreted by educators with the highest accuracy, sensitivity and specificity, remains an extremely important necessity.

IntechOpen

Author details

Sotiria Tzivinikou
University of Thessaly Volos, Greece

*Address all correspondence to: sotitzivi@uth.gr

IntechOpen

© 2018 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Torgesen J. Catch them before they fall: Identification and assessment to prevent reading failure in young children. *American Educator*. 1998;22:32-39
- [2] Vellutino F, Scanlon D, Zhang H, Schatschneider C. Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading and Writing*. 2008;21(4):437-480. DOI: 10.1007/s11145-007-9098-2
- [3] Wilson S, Lonigan C. Identifying preschool children at risk of later reading difficulties: Evaluation of two emergent literacy screening tools. *Journal of Learning Disabilities*. 2010;43(1):62-76. DOI: 10.1177/0022219409345007
- [4] Compton D, Fuchs D, Fuchs L, Bouton B, Gilbert J, Barquero L, et al. Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*. 2010;102(2):327-340. DOI: 10.1037/a0018448
- [5] Johnson E, Jenkins J, Petscher Y. Improving the accuracy of a direct route screening process. *Assessment for Effective Intervention*. 2010;35(3):131-140. DOI: 10.1177/1534508409348375
- [6] Ozernov-Palchik O, Norton E, Sideridis G, Beach S, Wolf M, Gabrieli J, et al. Longitudinal stability of pre-reading skill profiles of kindergarten children: Implications for early screening and theories of reading. *Developmental Science*. 2017;20(5):e12471. DOI: 10.1111/desc.12471
- [7] Carran D, Scott K. Risk assessment in preschool children: Research implications for the early detection of educational handicaps. *Topics in Early Childhood Special Education*. 1992;12(2):196-211. DOI: 10.1177/027112149201200205
- [8] Pennington B, Lefly D. Early reading development in children at family risk for dyslexia. *Child Development*. 2001;72(3):816-833. DOI: 10.1111/1467-8624.00317
- [9] Scarborough H. Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia*. 1998;48(1):115-136. DOI: 10.1007/s11881-998-0006-5
- [10] Sears S, Keogh B. Predicting reading performance using the slingerland procedures. *Annals of Dyslexia*. 1993;43(1):78-89. DOI: 10.1007/bf02928175
- [11] Cadman D, Chamber L, Feldman W, Sackett D. Assessing the effectiveness of community screening programs. *JAMA: The Journal of the American Medical Association*. 1984;251(12):1580-1585. DOI: 10.1001/jama.251.12.1580
- [12] Kent S, Wanzek J, Yun J. Screening in the upper elementary grades: Identifying fourth-grade students at-risk for failing the state reading assessment. *Assessment for Effective Intervention*. 2018;1:1-13. DOI: 10.1177/1534508418758371
- [13] Gredler G. Issues in early childhood screening and assessment. *Psychology in the Schools*. 1997;34(2):99-106. DOI: 10.1002/(sici)1520-6807(199704)34:2<99::aid-pits3>3.0.co;2-n
- [14] Meisels S. Dimensions of Early Identification. *Journal of Early Intervention*. 1991;15(1):26-35. DOI: 10.1177/105381519101500105

- [15] Wenner G. Kindergarten screens as tools for the early identification of children at risk for remediation or grade retention. *Psychology in the Schools*. 1995;**32**(4):249-254. DOI: 10.1002/1520-6807(199510)32:4<249::aid-pits2310320402>3.0.co;2-v
- [16] Bishop A. Prediction of first-grade reading achievement: a comparison of fall and winter kindergarten screenings. *Learning Disability Quarterly*. 2003;**26**(3):189-200
- [17] Havey J, Story N, Buker K. Convergent and concurrent validity of two measures of phonological processing. *Psychology in the Schools*. 2002;**39**(5):507-514. DOI: 10.1002/pits.10056
- [18] Hurford D, Potter T, Hart G. Examination of three techniques for identifying first-grade children at risk for difficulty in word identification with an emphasis on reducing the false negative error rate. *Reading Psychology*. 2002;**23**(3):159-180. DOI: 10.1080/02702710290061300
- [19] Speece D, Mills C, Ritchey K, Hillman E. Initial evidence that letter fluency tasks are valid indicators of early reading skill. *The Journal of Special Education*. 2003;**36**(4):223-233. DOI: 10.1177/002246690303600403
- [20] Hintze J, Ryan A, Stoner G. Concurrent validity and diagnostic accuracy of the dynamic indicators of basic early literacy skills and the comprehensive test of phonological processing. *School Psychology Review*. 2003;**32**(4):541-557
- [21] Simpson J, Everatt J, Cheney K. Phonological skills and naming speed as predictors of future literacy deficits. In: Presentation presented at 5th International Conference of BDA. 2001
- [22] Sofie C, Riccio C. A comparison of multiple methods for the identification of children with reading disabilities. *Journal of Learning Disabilities*. 2002;**35**(3):234-244. DOI: 10.1177/002221940203500305
- [23] Catts H, Nielsen D, Bridges M, Liu Y. Early identification of reading comprehension difficulties. *Journal of Learning Disabilities*. 2016;**49**(5):451-465. DOI: 10.1177/0022219414556121
- [24] Catts H, Nielsen D, Bridges M, Liu Y, Bontempo D. Early Identification of Reading disabilities within an RTI Framework. *Journal of Learning Disabilities*. 2015;**48**(3):281-297. DOI: 10.1177/0022219413498115
- [25] Fletcher J, Foorman B, Boudousquie A, Barnes M, Schatschneider C, Francis D. Assessment of reading and learning disabilities a research-based intervention-oriented approach. *Journal of School Psychology*. 2002;**40**(1):27-63. DOI: 10.1016/s0022-4405(01)00093-0
- [26] Graney S, Martínez R, Missall K, Aricak O. Universal screening of reading in late elementary school. *Remedial and Special Education*. 2009;**31**(5):368-377. DOI: 10.1177/0741932509338371
- [27] McNamara J, Scissons M, Gutknecht N. A longitudinal study of kindergarten children at risk for reading disabilities: The poor really are getting poorer. *Journal of Learning Disabilities*. 2011;**44**(5):421-430. DOI: 10.1177/0022219411410040
- [28] Speece Deborah L et al. Identifying children in middle childhood who are at risk for reading problems. *School Psychology Review*. 2010;**39**(2):258
- [29] Schoonjans F. MedCalc statistical software [Internet]. MedCalc. 2016. Available from: <https://www.medcalc.org/> [cited: August 27, 2018]
- [30] Wolff U, Lundberg I. A technique for group screening of dyslexia among adults. *Annals of Dyslexia*.

2003;**53**(1):324-339. DOI: 10.1007/s11881-003-0015-3

[31] Lefly D, Pennington B. Reliability and validity of the adult reading history questionnaire. *Journal of Learning Disabilities*. 2000;**33**(3):286-296. DOI: 10.1177/002221940003300306

[32] Tzivinikou S. Potential Discriminative factors for Dyslexia [doctoral thesis]. London: Brunel University; 2002

[33] Teisl J, Mazzocco M, Myers G. The utility of kindergarten teacher ratings for predicting low academic achievement in first grade. *Journal of Learning Disabilities*. 2001;**34**(3):286-293. DOI: 10.1177/002221940103400308

[34] Wallace I, Berkman N, Watson L, Coyne-Beasley T, Wood C, Cullen K, et al. Screening for speech and language delay in children 5 years old and younger: A systematic review. *Pediatrics*. 2015;**136**(2):e448-e462. DOI: 10.1542/peds.2014-3889

[35] Rello L, Romero E, Rauschenberger M, Ali A, Williams K, Bigham J, et al. Screening dyslexia for english using HCI measures and machine learning. In: *International Conference on Digital Health*. Lyon: ACM; 2018. pp. 80-84

[36] Taylor H, Anselmo M, Foreman A, Schatschneider C, Angelopoulos J. Utility of kindergarten teacher judgments in identifying early learning problems. *Journal of Learning Disabilities*. 2000;**33**(2):200-210. DOI: 10.1177/002221940003300208

[37] Fuchs D, Fuchs L, Compton D. Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children*. 2012;**78**(3):263-279. DOI: 10.1177/001440291207800301

[38] Messick S. Validity and washback in language testing. *Language Testing*. 1996;**13**(3):241-256. DOI: 10.1177/026553229601300302

[39] Harrington R, Jennings V. A comparison of three short forms of the McCarthy scales of children's abilities. *Contemporary Educational Psychology*. 1986;**11**(2):109-116. DOI: 10.1016/0361-476x(86)90001-9

[40] Flynn J, Rahbar M. Improving teacher prediction of children at risk for reading failure. *Psychology in the Schools*. 1998;**35**(2):163-172. DOI: 10.1002/(sici)1520-6807(199804)35:2<163::aid-pits8>3.0.co;2-q