

Strengths and Weaknesses of Screening Reading Difficulties

Sotiria Tzivinikou

Department of Early Childhood Education, Aristotle University of Thessaloniki

Greece

tzivisot@nured.auth.gr

Abstract

Early identification of problem readers and appropriate interventions would reduce subsequent failure and would enhance reading skills. Screening instruments measure the risk/probability that reading problems will occur. In this article we discuss certain aspects with regard to the most important indices of screening measures: predictive accuracy, sensitivity and specificity. Despite the existence of excellent research in the area, development of a widely accepted, effective screening procedure continues to be a scientific challenge.

Keywords: early identification, reading difficulties, effective screening, sensitivity, specificity.

Introduction

Making predictions about the future reading achievement of successful students is rarely a concern. Of greater interest is the screening of young school children who have fallen behind in learning to read (Scarborough, 1998). What we wish from any screening program is to identify students in need of further individual diagnostic testing. A test or series of tests are given to an individual or a group of students who have something in common, age, grade level, or signs of a special problem, such as deficient fine motor co-ordination or poor reading performance. Results from screening tests provide a first look at the students in order to determine temporary grouping. There is consensus based on research evidence (Sears & Keogh, 1993) that early identification of reading problems, followed by appropriate, effective interventions could enhance reading ability. The important questions are what to identify and how to identify.

From early years until recently, researchers noted that the earlier the diagnosis, the better the chances of remediation. Early screening can increase the likelihood of the intervention's success if it is implemented at an earlier stage in the child's development, when behaviour is more malleable (Strag, 1972, Sears & Keogh 1993, Hurford, et al. 1994, 2002). Thomson (1980) noted that even with support at the age of 8 or so, reading difficulties would never be fully remediated. On the other hand, enhancing the family's coping mechanisms as soon as possible can aid their adjustment to and acceptance of the disability (Mercer et al. 1988).

However, potential labelling and stigmatizing of children is also an important aspect under consideration as a possible negative consequence.

Screening operates under the assumption that symptoms of a handicapping condition can be detected and measured. A screening instrument measures the risk, defined as the probability that an event-handicapped condition will occur (Last, 1983). Snow et al. (1998) argued that three types of risk factors for learning disabilities can be mentioned: (a) child-based risk factors, including severe cognitive deficiencies, language and hearing impairments, and attention-deficit / hyperactivity disorder; (b) family-based risk factors, including family history of reading difficulties, and home literacy environments; and (c) neighbourhood-, community-, and school-based risk factors, including ineffective schools and teaching methods (for a more comprehensive review, see Tzivinikou, 2002).

Prediction studies were divided into three different age categories: kindergarten children, primary school children and adults. But the vast majority of these studies with the mildly handicapped were initiated at kindergarten. At the pre-school stage, many children are already showing early signs of their disorder that can be detected by those with experience in this area. The key is usually an uneven developmental profile, particularly in cases where there is a family history of speech or literacy difficulties, or where there is evidence of significant birth difficulties. Characteristic difficulties include one or more of the following: (a) delays in the development of speech and language, (b) difficulties in learning simple patterns of sequential activity, such as remembering the order of simple instructions or reproducing a pattern of coloured beads or bricks, (c) difficulties in fine or gross motor co-ordination, (d) high distractibility and poor concentration (Singleton, et al. 1995).

Prediction techniques may include a battery of tests, a single instrument, and/or teacher perception or ratings. A battery of tests may consist of any combination of tests, subtests, and single-variable measures. A global score or a pattern of scores is used for prediction. There is evidence that batteries containing multiple tests generally provide better prediction than single instruments, but the increased effectiveness of multi-test batteries is generally not large enough to warrant the extra time and resources required to administer them (Scarborough, 1998, Torgesen, 1998). Predictive studies are utilized to assess risk; a screening instrument or predictive variable is administered at one point in time and the outcome or criterion is measured at a second point in time. Results from such studies may then be used to (a)

validate and determine the predictive validity and reliability of the instrument, (b) determine the risk status of the subject, and (c) formulate analytic intervention studies (Carran, Scott, 1992).

Indices of Screening Accuracy

A screening procedure should be a quick, efficient method that permits evaluation of each child. It does not provide a diagnosis but rather functions as a system which designates children who are at greater risk for subsequent difficulties. Thus, it should not be confused with diagnosis.

Requirements of effective developmental screening tests are: an adequate standardization sample, low cost, ease of administration, appropriate content and adequate validity and reliability (Gredler, 1997). However, predictive validity or instrument reliability has also been cited as a major problem in screening for children at risk (Carran, Scott 1992). Cadman et al. (1984) state, “a test with a low predictive value is unlikely to be either efficient or useful . . .” (p. 1583). One aspect of an effective framework is the relevance and utility of measures. Relevance refers to the degree of relationship between the measure and the purpose of the assessment; utility of the measures is usually evaluated by cost-effectiveness (Messick, 1989).

The screening results will be discussed as being poor or good, with poor indicating a subject who exhibits the target disorder and good a subject who does not. Based on measurement at two points of time, a subject may be identified as: (A) true positive: failed screen and had poor outcome, (B) false positive: failed screen and had good outcome, (C) false negative: passed screen and had poor outcome, (D) true negative: passed screen and had good outcome.

Meisels (1991, 1993) and Wenner (1995) recommended sensitivity and specificity indices as appropriate for identifying an instrument’s capacity to predict children to be recommended for further testing and remediation. These indices can be calculated using the formula: $Sensitivity = A / (A + C)$ and $Specificity = D / (B + D)$ (A, B, C, D as described above). Sensitivity is the proportion of children at risk who are correctly identified as at risk (true positives); in other words, it provides information regarding a test’s ability to identify people at risk, so that they might receive a full diagnosis. Specificity refers to the proportion of children not at risk who are correctly excluded from intervention (true negatives).

Screening tests with high sensitivity give valuable information on the importance of a negative screening result; if a student is said not to be at risk, there is a high likelihood that he or she is truly not at risk. However, knowing that a test is highly sensitive gives little information on the utility of a positive screening result, because many students could still be falsely positive. To better define the usefulness of a positive screening result, one must also examine the specificity or likelihood ratio. A test with a high specificity makes the probability of the child being truly at risk very high. Likewise, a test with a high likelihood ratio indicates that the test is very good at increasing the certainty of a positive identification of at-risk children. Looking at all these test characteristics concurrently allows for a clear definition of the merits and weaknesses of the screening test.

The accuracy of screening measures is important given the concern of either mislabeling a child or failing to detect a delay. This accuracy is determined by comparing children's performance on screening to a battery of diagnostic tests and to standards for screening tests. Glascoe and Byrne (1993) noted that the standards include sensitivity of 80% and specificity of 90% as preferable, and that positive predictive value (of children who fail the screening test and are found to have true developmental problems on diagnostic testing) of 70%, or about 3 out of every 4 referrals, is also preferable. In consistency, Carran, Scott, (1992), and Meisels (1991) supported that values above .80 are considered acceptable for these indicators.

Nevertheless, Badian et al. (1990) noted that even when predictive validity is relatively high (e.g. correlation coefficient of 0.6 to 0.7 between a predictor and reading), identification of children who will fail in reading is usually low. A high correlation coefficient may be due primarily to accurate prediction of good reading, rather than poor. A possible reason for the difficulty in predicting which children will fail in reading may be the assumption that such children will be the low scorers on a predictive test battery. In fact, low scorers are likely to be a heterogeneous group, and may include children with an attention deficit disorder or youngsters who are uncooperative, from deprived backgrounds or of below average intelligence. Such children may later have difficulties in reading or mathematics, but experience has shown that many of them will not have difficulty learning to read, so will be predictive false positives.

On the other hand Mattison, et al. (1982) noted that knowing the true positive and the true negative rates (which they term respectively *sensitivity* and *specificity*) for a diagnostic

test is insufficient for know what value it has in a large, unselected population. Thus it is important to ascertain the predictive value of a test, i.e., the likelihood that a person with a positive test actually has what is being diagnosed.

Other important indices are validity and reliability. The majority of predictive studies considered the validity concept either directly or indirectly. Examining the items in a measure, the correlation between the measure and other validated measures, and the underlying theoretical construct generally is undertaken in order to determine if a measure is valid. Evidence of a measure's validity is generally provided through examination of content, criterion-related validity, and construct validity.

Test validity is prerequisite to test reliability. 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.

Educational Implications

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. But a concern of false positives is that they place an increased demand on scarce resources (Fletcher et al. 2002).

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, making their problems more difficult to remediate at a later time (Fletcher et al. 2002). 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 (Mantzicopoulos & Morrison 1994).

Is it possible for a screening measure to have a 0 false negative rate? Hurford, et al. (2002) answered 'no'. Their explanations have to do with the different levels of readiness in

children upon entering school. In any case, scientific efforts will be continued in order to decrease the false rates of screening.

Methodological Considerations

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. Satz and Fletcher (1988) argued that any validation of an early screening instrument should incorporate (a) longitudinal design, (b) independent assessments of kindergarten performance and learning ability separated by a temporal interval of at least 3 years, (c) random sampling of children in a validation/cross-validation design and (d) systematic assessment of predictive utility and validity. Taking this into consideration, an examination of early identification studies reveals that three major designs are used: one common format for evaluation involves administering a screening (i.e., prediction) instrument in kindergarten and a criterion (i.e., performance) instrument at a later date. (Grogan, 1995, Glascoe & Byrne, 1993, Sears & Keogh, 1993, Catts, 1991, Nicolson & Fawcett, 1997, Badian, et al 1990, Wenner, 1995, Näslund, 1990). The correlation scores between these measures are then evaluated as evidence of screening test or instrument utility.

A second type of instrument validation study involves concurrent administration of a prediction and a validation instrument. Scores of each test are correlated to 'validate' the prediction instrument. (Fawcett & Nicolson, 1996, Horn & O'Donnell, 1984, Hurford et al 1994, Rafoth, 1988). However, the utility of this practice is limited by the validity and reliability of the validation instrument. A third type of prediction study establishes prediction information (i.e., scores on a prediction instrument) and, after an intervention period, obtains criterion information (i.e., performance data).

As Horn and Packard (1986) noted, although many prediction studies were well-designed and presented prediction rather than associations between predictor(s) and later reading ability, they did not mention the percentage of correct classification rate between the groups examined. Some examples of such studies were Näslund, (1990), Sears and Keogh, (1993), and Rafoth (1988).

According to Coleman and Dover (1993), the major methodological problem in devising early identification measures is how best to assess predictive validity. Typically, researchers collect a variety of measures during kindergarten that are subsequently used to predict 1st or 2nd grade reading achievement based on the results of standardised tests. Results are summarised through either correlational or classificational approaches.

The correlational approach yields multiple correlations between predictor and criterion variables, which indicate the amount of variance in reading scores that can be explained by the screening measures. This approach provides evidence of the relationship of the screening tools to subsequent reading levels across all levels of reading ability, but it does little to suggest which children in particular are at risk for school failure (Lefly & Pennington, 2000; Gottesman et al., 1991).

Regarding sensitivity and specificity, these indices offer an enhanced interpretation of the results of a screening test, extending beyond the relatively non-specific information provided by correlations. For instance, a highly significant relationship may be revealed through simple correlations, but this finding tells the researcher little about the ability of the test to correctly classify individuals as at risk or not at risk (Limbos & Geva, 2001).

Assuming that screening tests have the explicit purpose of assigning individuals a status (at risk or not at risk), classification approaches to predictive validity establish a cut-off score on the criterion measure, below which the child is said to be at risk. They then attempt to use the screening results to identify subjects that ultimately fall into the risk group. Predictions are usually generated through discriminant function screening variables to maximise the differences between risk and non-risk groups on a linear vector of the original items. Predictive validity is then judged in terms of the proportion of subjects whose group membership (at risk or non at risk) is correctly identified, as well as the pattern of false positive and false negative identifications.

Although classification matrices provide useful information about the predictive validity of screening measures, they must be analysed carefully. Many studies present high accuracy rates that are misleading with regard to the value of the instrument. Because the number of children who are at risk for educational difficulties represents a small proportion of the

entire school population, it is possible that a screening measure that never identified any child as being at risk could still have a respectable overall accuracy of prediction.

A second concern in using discriminant function approaches to prediction is the stability of weights given to variables in determining the prediction equation. To ensure the validity of the prediction equation, multiple samples must be employed. One sample serves to calibrate the equation, and the resulting weights are then used to make predictions on a second sample. Only when the accuracy of the prediction equation is comparable across the two independent samples can it be said to have population validity (Lefly & Pennington, 2000, Coleman & Dover, 1993).

Some Recent Prediction Studies

A well-designed study measuring validity and reliability was performed by Coleman & Dover (1993). In order to determine that all five RISK factors were significantly related to school placement decisions, a series of discriminant function analyses were conducted to assess the predictive validity of the inventory. These analyses allowed for the construction of a prediction-performance matrix from which the accuracy of RISK predictions to final student outcomes could be judged (placement in resource classes or regular classes). Establishing a stable discriminant function requires consensual validation; that is, the function must be calibrated on one sample and then fitted to a second sample to determine if it has generality. For this purpose, the four Kindergarten cohorts were divided into two groups. One group was used as the calibration sample, and the other became the target sample. Finally, the calibration discriminant function loading was used, with the entire sample collapsed into a single group.

Bishop (2003) found that both phonological awareness and letter identification yielded the highest overall results, supporting findings reported over the past decade. Moreover, the accuracy rates of all the constructs were promising. The false positive rate ranged from 13% to 27%, depending on the construct, while the false negative rate ranged from .06% to .21%.

Havey, et al. (2002) examined the convergent and concurrent validity of two recently developed measures of phonological processing, the TOPA and the CTOPP. Both instruments would, therefore, appear to be useful in the early identification of children at-risk for diffi-

culty in learning to read. Results, however, do not support the use of either, or both, of these instruments as sole predictors of reading outcome.

Hintze, et al. (2003) examined the concurrent validity and diagnostic accuracy of the published test, DIBELS, compared to another well documented published test, the 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.

Using the same test (CTOPP), Sofie and Riccio (2002) examined relationships among standardized reading achievement tests, phonological awareness measures and fluency rates (CBM, subtest of Woodcock-Johnson Tests of Achievement-Revised (WJ-R) and the relatedness of these measures to teacher ratings). The authors supported that in addition to traditional norm-referenced measures of reading achievement, measures of phonological awareness and reading fluency that provide further information may be included as part of reading assessment. The classification rate was out of the study interests, so there was not any evidence about the accuracy of identification.

The findings of Hurford, et al. (2002) indicated that the accuracy of discrimination was high, 89.7%, with a 6.2% false negatives rate; when the calibration data from the reference group were used to identify at-risk status in a different sample, the accuracy fell to 80.2% with a 10.2% false negative rate.

Lefly and Pennington (2000) found that the Adult Reading History Questionnaire (ARHQ) was valid, being demonstrated by the high correlation between the ARHQ and diagnostic measures for adults ($r_s = .57-.70$). But the ARHQ does not perfectly detect every familial case. Thus, clinicians and researchers should use this questionnaire as a screening instrument and not as a diagnostic tool. On the other hand, findings from Pennington and Lefly (2001) (sensitivity .49 and specificity .76) supported that letter-name knowledge and rapid serial naming were most important in predicting later RD.

The Taylor et al. (2000) study examined the accuracy of teacher ratings. 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 nonidentified

children on tests of word reading, spelling, mathematics, and knowledge of letter names and letter sounds. Furthermore, greater proportions of identified children than nonidentified children were receiving special learning assistance by the end of the third school year.

Another study examining teachers' ratings was Teisl, et al. (2001). 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 may be better indicators of future academic achievement than behavioral or social variables. 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.

Tzivinikou (2002) constructed a parent report checklist, including a large amount of information about the child's development history, with problems often referred to in the literature as being indicators for reading problems. The author supported that this checklist was valid and reliable and could screen between reading disabled children (RD) and non-disabled (NRD) with 97.2% discriminative accuracy.

Conclusions

Summarizing aspects discussed in this article, we could note that screening tests could be accepted as effective if they have some specific characteristics such as norm-referenced sample, appropriate content, validated validity and reliability, ease of administration and interpretation, and also if they are quick and cost-effective. Additional criteria are related to its discrimination accuracy with emphasis on false negative and false positive rates (Cadman et al., 1984; Carran & Scott, 1992; Mercer et al., 1988; Messick, 1991).

Screening instruments strive to detect children at risk for handicapping conditions early enough for definitive diagnosis of the underlying condition to be made and for prompt intervention to be initiated. Some studies have questioned the accuracy of screening instruments for classifying students for special education programs. Misidentification is avoided when a screening procedure correctly identifies and refers at-risk children to an early intervention program, and when they correctly exclude from intervention children who are not at risk (Wenner, 1995).

Developmental screening tests are widely used for early identification but not all are studied for their accuracy. The percentage of children correctly detected as with and without problems is not always known. The absence of such data makes it difficult for professionals to choose measures wisely and to avoid those that under-detect or over-refer.

Future Research Suggestions

Published research demonstrates the efficacy of early intervention, supporting the need for carefully designed and accurate screening measures. Despite the recent interest and research on screening reading disabilities, the body of research on the effectiveness of these measures still remains methodologically problematic and the findings are scant. Thus, development of a cost-effective, equitable screening, a diagnostic and supportive method, that is acceptable to government, educational authorities, school, children and parents, still remains a scientific challenge.

References

- Badian, N.A., McAnulty, G.B., Duffy, F. & Als, H. (1990). Prediction of dyslexia in kindergarten boys. *Annals of dyslexia*, vol. 40, 152-169.
- Bishop, A. G. (2003). Prediction of first-grade reading achievement: a comparison of fall and winter kindergarten screenings. *Learning Disability Quarterly*, vol. 26(summer), 189-200.
- Cadman, D., Chamber, L., Feldman, W., & Sackett, D. (1984). Assessing the effectiveness of community screening programs. *Journal of the American Medical Association*, vol. 251, 1580–1585.
- Carran, D. T., Scott, K. G. (1992). Risk assessment in preschool children: Research implications for the early detection of educational handicaps. *Topics in Early Childhood Special Education*, vol. 12(2), 196-222.
- Catts, H.W. (1991). Early identification of dyslexia: Evidence from a follow up study of speech-language impaired children. *Annals of dyslexia*, vol. 41, 163-177.
- Coleman, M.J. & Dover, M.G. (1993). The RISK Screening Test: Using Kindergarten Teachers' Rating to Predict Future Placement in Resource Classrooms. *Exceptional Children*, vol. 59(5), 468-477.

- Fawcett, A. J., & Nicolson, R. I. (1996). *Dyslexia Early Screening Test*. London: The Psychological Corporation.
- Fletcher J. M., Foorman B. R., Boudousquie A., Barnes M A., Schatschneider C. & Francis, D. J. (2002) Assessment of Reading and Learning Disabilities. A Research-Based Intervention-Oriented Approach. *Journal of School Psychology*, vol. 40(1), 27–63.
- Glascoe, F.P. & Byrne, K.E. (1993). The Accuracy of Three Developmental Screening Tests. *Journal of Early Intervention*, vol. 17(4), 368-379.
- Gottesman, R.L.; Cerullo, F.M.; Bennett, R.E. & Rock, D.A. (1991). Predictive Validity of a Screening Test for Mild School Learning Difficulties. *Journal of School Psychology*, vol. 29, 191-205.
- Gredler, G. R. (1997). Issues in early childhood screening and assessment. *Psychology in the Schools*, Vol. 34(2), 99-106.
- Grogan, S.C. (1995). Which cognitive abilities at age four are the best predictors of reading ability at age seven? *Journal of Research in Reading*, vol. 18(1), 24-31.
- Havey, J.M., Story, N., & Buker, K. (2002). Convergent and concurrent validity of two measures of phonological processing. *Psychology in the Schools*, vol. 39(5), 507-514.
- Hintze, J. M., Ryan, A. L. & Stoner, G. (2003). Concurrent Validity and Diagnostic Accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review*, vol. 32(4), 541-557.
- Horn, W.F. & O'Donnell, J.P. (1984). Early Indication of Learning Disabilities: A comparison of two methods. *Journal of Educational Psychology*, vol. 76(6), 1106-1118.
- Horn, W.F. & Packard, T. (1986). Early Identification of Learning Problems: A Meta-Analysis. *Journal of Educational Psychology*, vol. 77(5), 597-607.
- Hurford, D. P., Potter, T. S. & Hart, G. S. (2002). 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*, 23, 159–180.
- Hurford, D.P., Schauf, J.D., Bunce, L., Blauch, T. & Moore, K. (1994). Early identification of children at risk for reading disabilities. *Journal of learning disabilities*, vol. 27, 371-382.
- Last, J.M. (1983). *A dictionary of epidemiology*. New York: Oxford University Press.
- Lefly, L. & Pennington, B. F. (2000). Reliability and Validity of the Adult Reading History Questionnaire. *Journal of Learning Disabilities*, vol. 33(3), 286-297.

- Limbos, M.M. & Geva, E. (2001). Accuracy of teacher assessments of second-language students at risk for reading disability. *Journal of Learning Disabilities*, vol. 34(2), 136-151.
- Mantzicopoulos, P., & Morrison, D. (1994). Early prediction of reading achievement: Exploring the relationship of cognitive and noncognitive measures to inaccurate classifications of at risk status. *Remedial and Special Education*, 15(4), 244-251
- Mattison, R.E., Cantwell, D.P., & Baker L. (1982). A practical Method for Screening Psychiatric Disorders in Children with Speech and Language Disorders. *Journal of Abnormal Child Psychology*, vol. 10(1), 25-32.
- Meisels, S. (1991). Dimension of early identification. *Journal of Early Intervention*, vol. 15, 26-34.
- Meisels, S. (1993). Remaking classroom assessment with the work sampling systems. *Young Children*, vol. 28, 34-40.
- Mercer, C.D., Algozzine, B., & Trifiletti, J. (1988). Early identification: An analysis of the research. *Learning Disability Quarterly*, 11, 176-188.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., p. 13-103). New York: Macmillan.
- Näslund, J.C. (1990). The interrelationships among Preschool Predictors of Reading Acquisition for German children. *Reading and Writing*, vol. 2, 327-360.
- Nicolson, R. & Fawcett, A.J. (1997). Development of objective procedures for screening and assessment of dyslexic students in higher education. *Journal of Research in Reading*, vol. 20 (1), 77-83.
- Pennington, B.F. & Lefly, D.L. (2001). Early reading development in children at family risk for dyslexia. *Child Development*, vol. 72(3), 816-833.
- Rafoth, M.A. (1988). Early identification of learning disabilities using the Meeting Street School Screening. *Journal of Learning Disabilities*, vol. 21, 186-188.
- Satz, P. & Fletcher, J.M. (1988). Early Identification of Learning Disabled Children: An Old Problem Revisited. *Journal of Consulting Clinical Psychology*, vol. 56(6), 824-829.
- Scarborough, H.S. (1998). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming and IQ. *Annals of Dyslexia*, vol. XLVII, 115-136.
- Sears, S. & Keogh, B. (1993) Predicting reading performance using the Slingerland procedures. *Annals of dyslexia*, vol. 43, 78-89.

- Singleton, C.H; Thomas, K.V. & Leedale, R.C. (1995). *CoPS 1: Cognitive Profiling System for the diagnosis of dyslexia and assessment of special educational needs*. An extract from the Final Report on the Humberside Early Screening Research Project, Department of Psychology. Hull: University of Hull.
- Snow, C., Burns, M.S., & Griffin, P. (1998). Preventing reading difficulties in young children. Washington, DC: National Academy Press.
- Sofie, C. A. & Riccio, C. A. (2002). A Comparison of Multiple Methods for the Identification of Children with Reading Disabilities. *Journal of Learning Disabilities*, vol. 35(3), 234-244.
- Strag, G.A. (1972). Comparative Behavioral Ratings of Parents with Severe Mentally Retarded, Special Learning Disability, and Normal Children. *Journal of Learning Disabilities*, vol. 5(10), 631-635.
- Taylor, H.G. Anselmo, M., Foreman, A.L., Schatschneider, C. & Angelopoulos, J. (2000). Utility of kindergarten teacher judgments in identifying early learning problems. *Journal of Learning Disabilities*, vol. 33(2), 200-210.
- Teisl, J.T., Mazzocco, M.M.M. & Myers G.F. (2001). The utility of kindergarten teacher ratings for predicting low academic achievement in first grade. *Journal of Learning Disabilities*, vol. 34(3), 286-293.
- Thomson, M.E. (1980). *The Development of Written Language Skills in Dyslexic Children: A Reading and Spelling Error Analysis*. Paper presented to British Psychological Society, Development Section Annual Conference, Edinburgh. Sept. 5th - 7th 1980.
- Torgesen, J.K. (1998). Catch them before they fall: Identification and Assessment to Prevent Reading Failure in Young Children, *American Educator*, vol. Spring/Summer, 32-39.
- Tzivinikou, S. (2002). *Potential Discriminative factors for Dyslexia: A predictive statistical model based on the Pavlidis questionnaire distinguishing 8-9 year-old dyslexic and non-dyslexic-control Greek children. Considerations of validity and potential predictive efficiency*. Unpublished doctoral thesis.
- Wenner, G. (1995). Kindergarten screens as tools for the early identification of children at risk for remediation or grade retention. *Psychology in the schools*, vol. 32, 249-254.

[This page left blank intentionally]