National Center for Learning Disabilities
The power to hope, to learn, and to succeed

The NCLD *Get Ready to Read!* Screening Tool Technical Report

Grover J. Whitehurst, Ph.D.

Applied Research Partners, Inc.

and

The State University of New York at Stony Brook

Background

A recent National Research Council report on preventing reading problems in young children (Snow, Burns, & Griffin, 1998) concluded that the majority of reading problems could be prevented, primarily by reducing the number of children who enter school inadequately prepared. Inadequate preparation refers to weak literacy skills, particularly in the domains of linguistic awareness and knowledge about print. At the pre-reader level, linguistic awareness refers to a child's developing sensitivity to the sound structure of his or her oral language, e.g., that words are made from syllables; that spoken utterances are made from individual words; that compound words are made from single words. Knowledge about print refers to a developing understanding of the nature and purposes of books, printed words, and letters. Weaknesses in these domains are particularly common among children from low income families, children with limited proficiency in English, and children whose parents report difficulty learning to read.

Although the importance of assessing pre-readers' linguistic awareness and knowledge about print is apparent, and these skills are being measured effectively by researchers, no empirically-based set of developmental milestones or parent/teacher-oriented screening tools measures presently exist. Furthermore, activities that recent research has shown to be effective in advancing these key skills are not yet available in a usable form for implementation by parents, preschool teachers, librarians, child care providers, and other early childhood professionals.

The National Center for Learning Disabilities (NCLD) has launched the *Get Ready to Read!* initiative, a multi-year collaborative endeavor that will enable parents, and early childhood professionals, and other care providers to help young children develop the skills they need to be ready to read. The

initiative is the first coordinated, nationwide effort to provide caregivers with a reliable, research-based method to determine whether 4-5 year old children have the necessary skills to be ready to read, as well as research-based information and resources to help adults provide children with the experiences needed for literacy skill development.

NCLD contracted with Applied Research Partners, Inc. (ARP) to develop a screening instrument that would identify pre-readers at risk of later reading problems because of low levels of emergent literacy. The principal investigator for the project at ARP was Grover Whitehurst, who is a leading researcher in the field of early literacy. Co-principal investigator was Christopher Lonigan, also an internationally visible researcher in the field.

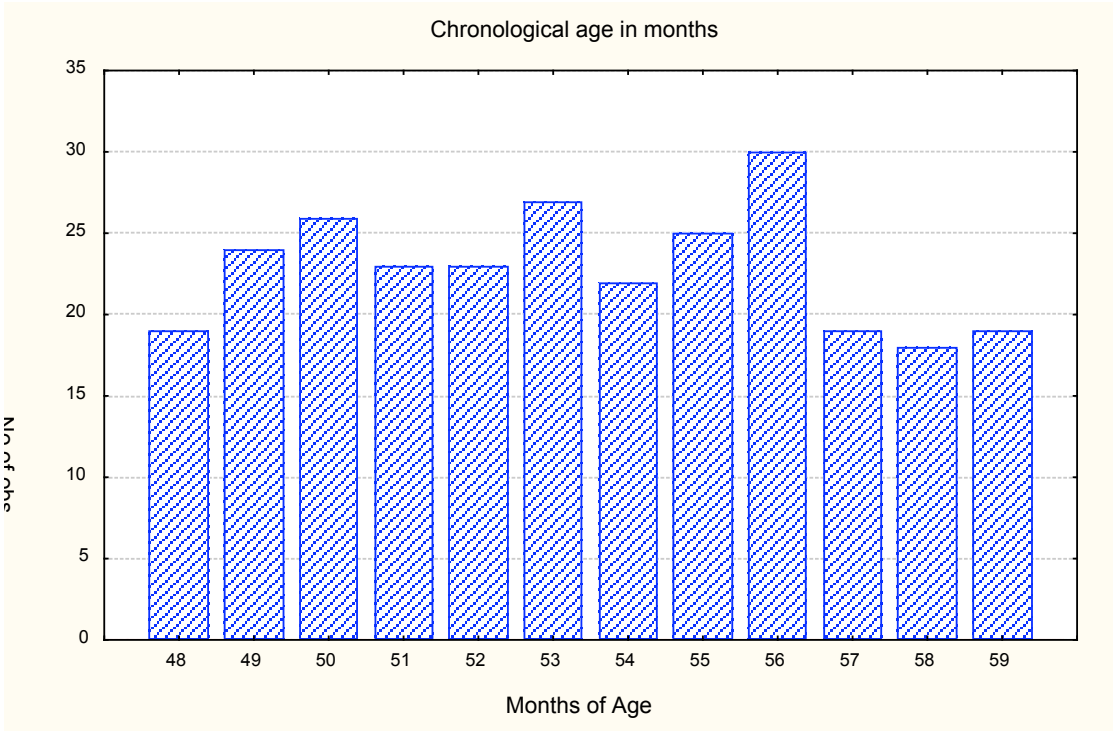
Project Design

The aim of the project was to provide a screening tool that could be administered to children in the pre-kindergarten year by non-professionals, and that would produce a score for individual children that could be linked empirically to the potential for reading success. Achieving this aim within the time span of less than a year that was budgeted for the project depended on a procedure of secondary validation of the screening tool. In this procedure, the screening instrument was linked statistically to results from another professionally administered assessment, the *gold standard*, that had already been validated as a predictor of later reading problems. A strong correlation between the results from the screening instrument and results from the same children on the gold standard would indicate that the screening instrument is also valid.

This procedure of secondary validation can be contrasted with primary validation, which would require that children assessed on the screening instrument be followed into school and tested on their reading skills two or three years later. It would then be possible to determine directly that low scorers on the screening instrument during the preschool period are more likely than high scorers to have reading difficulties once they reach school. While primary validation is desirable, most new assessment instruments in psychology and education are validated against existing instruments (secondary validation), rather than against primary measures of the outcomes of interest. For instance, a new intelligence test would be validated by demonstrating that it correlates well with an existing test, such as the Stanford-Binet (a gold standard), rather than by demonstrating that people who score higher on the new test do better in

school and life. Thus there is nothing unusual or suspect about the secondary validation procedure employed for the *Get Ready to Read!* Screening Tool.

Sample. 342 children participated in the development of the screening instrument. These children were from two locations: Suffolk County, NY, and Tallahassee, FL. The Suffolk County sample consisted of 139 children drawn entirely from Head Start centers: 48% of these children were Latino; 38% were African-American, 9% were Caucasian, and the remaining 5% were from Other racial/ethnic categories. The Tallahassee sample of 203 children was comprised of three groups: A Head Start sample of 84 children was 96% African American and 4% Caucasian; a group of 69 children attending state-sponsored pre-K classrooms was 71% African-American, 25% Caucasian, and 4% Other; 50 children attending private nursery schools were 8% African-American, 86% Caucasian, and 6% Other. Across the sub-samples and the sample as a whole, children were evenly divided on sex. Children were also distributed roughly evenly across the age range of 48 months to 59 months that included the four-year-olds who were the target of the screening instrument, as indicated by the following figure.



The nursery school sub-sample was middle-class. The state pre-K sub-sample in Florida, and the Head Start samples in both New York and Florida were from low-income families. There are two reasons

for the disproportionate representation of low-income children in the sample. First, as will be described in greater detail below, the gold standard assessment used in this project grew out of research on low-income children. Given that the success of the present project requires that the screening instrument correlate highly with the gold standard, it made sense to use children as participants who were similar in background to the children used in previous research to develop the validity of the gold standard. Second, preschoolers from low-income backgrounds have very elevated risks of later reading problems compared with children from middle-class backgrounds. For instance, among African-American and Hispanic students in the U.S. (two groups who experience disproportionate rates of poverty) the percentages of Grade 4 students reading below the basic level are 64% and 60%, respectively, compared with 38% for children in general (National Center for Educational Statistics, 1998). Thus, for a screening instrument designed to identify children at-risk of later reading problems, it made sense to over-sample children in the highest risk group for reading difficulties. Note, however, that the inclusion of a middle-class sub-sample in the project design allows the relevance of the screening instrument for children from different economic and racial/ethnic backgrounds to be determined.

The gold standard. The science demonstrating that preschooler's emergent literacy abilities are robust predictors of later reading achievement is recent. Much of the relevant research has been conducted by the principal investigator and the co-principal investigator for this project (e.g., Lonigan, Burgess, & Anthony, 2000; Whitehurst & Fischel, 2000; Whitehurst & Lonigan, 1998). The gold standard for the present project is derived from longitudinal research by Whitehurst and his colleagues. Beginning in 1993, three successive waves of four-year-olds attending Head Start received extensive assessment on various emergent literacy skills. These children were followed through their elementary school careers, with annual spring assessments of reading ability. Virtually all of these children were at high risk of reading problems due to conditions associated with economic poverty.

For the purposes of the present project, we focused on 260 of these children who were either successful readers who were experiencing serious reading problems at the end of second grade. The successful readers (N = 124) had an average normal curve equivalent (NCE) of 50 or above on four standardized tests of reading ability described below. A child receiving a NCE of 50 or above is performing in the top 50% of children nationally. The unsuccessful readers (N = 124) had an average NCE

of 30 or below. A child receiving an NCE score of 30 or below would be in the lower 18% of children nationally in reading performance.

The second grade tests of reading ability used to define the successful and unsuccessful readers were the Word Reading subscale of the Stanford Achievement Test- Eighth Edition (SAT) (Psychological Corporation, 1989), a test of the ability to match printed words in context with pictures; the Word Study subscale of the SAT, at test of the ability to understanding the relationships between letters and sounds and to understand suffixes, contractions, and compound words; the Word Attack subscale of the Woodcock Reading Mastery Tests-Revised (WRMT-R) (Woodcock, 1987), a test of the ability to sound out printed pseudowords; and the Reading Comprehension subtest from the SAT, a test of the ability to extract meaning from and answer questions about short paragraphs.

With reading success and reading failure thus defined, we searched for measures of emergent literacy skills obtained when these same children were four-year-olds that would predict their second grade reading outcomes. Fourteen different measures obtained during the pre-K Head Start year were evaluated. Each of these measures was selected from an extensive battery of standardized tests of school readiness, The Developing Skills Checklist (CTB/McGraw-Hill,1990), that covers many more domains than emergent literacy. The 14 measures utilized were: naming pictured letters, identifying letters and sounds, segmenting sentences, segmenting compound and regular words, rhyming, identifying the correct way to hold a book, identifying people engaged in reading, differentiating print from pictures and letters from numerals, identifying functions of print, identifying components of written communication, printing first name, drawing a person, message-writing mechanics, and message writing quality. Each of these 14 measures consists of multiple questions that are administered to children individually in the typical style of a standardized test of language or intelligence.

A statistical procedure called discriminant function analysis was utilized to determine the degree to which these 14 measures obtained when children were 4-year-olds could predict whether the same children would be successful or unsuccessful readers four years later, when they were in second grade. The results are presented in the table below. The overall classification accuracy is 78%, which is quite good for prediction using tests of emergent literacy to an entirely different set of assessments of reading obtained four years later. Note that the predictive accuracy for poor reading outcomes is remarkably strong.

Specifically, four year olds receiving low scores on the 14 professionally administered measures of emergent literacy have an 85% chance of being reading failures at the end of second grade. These results indicate that we have a strong gold standard against which to determine the effectiveness of the *Get Ready to Read!* Screening Tool.

Classification Matrix			
	Predicted poor readers	Predicted good readers	% correct classification
Observed poor readers	116	20	85%
Observer good readers	38	86	69%

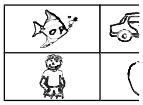
The alpha items for the screening instrument. Potential items for the screening instrument were developed by surveying results from previous research by the Principal Investigator and co-Principal Investigator. Items from the DSC, as described above, and from a large battery of phonological assessments developed by the co-PI, were examined to identify areas of emergent literacy knowledge in four year olds that correlated with subsequent reading outcomes. Where individual items from these preschool batteries correlated well with later reading outcomes on standardized tests, we generated new items that we hoped would capture the same underlying knowledge. The new items, unlike most of the original items, were designed in a multiple choice, receptive format in which an adult reads a question and the child points to one of four pictures as the best answer to that question. Items that require only pointing responses from children are much easier to administer and score than questions that require children to produce an oral response or written response. Ease of administration and scoring is a critical characteristic of an assessment instrument intended for use by nonprofessionals.

A pool of approximately 100 items was initially developed, and was then trimmed to 60 items by eliminating items that were substantially redundant, or items that in the view of the development team were more difficult to understand or graphically more difficult to represent. The surviving pool of 60 items will be referred to as the *alpha* pool. It sampled the following sub-domains of emergent literacy knowledge: book knowledge ("Find the picture that shows the back of the book."); print knowledge ("Find the picture of a cereal box that shows the name of the cereal."); letter knowledge ("Find the letter G."); letter-sound correspondence ("Find the letter that makes the *buh* sound."); emergent writing ("Some children wrote the letter F. Find the one that is written best."); linguistic awareness – initial phonemes ("Find the picture of the thing that starts with the *duh* sound."); linguistic awareness -- rhyming ("Find the one that rhymes with

ball."); linguistic awareness – compound words ("Find what you get when you put SEA and SHELL together.").

Graphic representations of a correct choice and three foils were developed for each of the 60 items. A testing form was developed for children using these graphics, with four picture choices for each question arranged on a single 8 X 10 page in landscape mode. The resulting 60 pages were arranged in a loose leaf binder, which constituted the testing material seen by each child.

A corresponding scoring form was developed for the examiner/tester. It consisted of small representations of the pictures seen by the children, with the questions to be asked written next to the pictures. The adult's task was to turn the child's testing booklet pages to each question in turn, ask the question on the scoring form, and mark the picture the child pointed to by making a corresponding mark on the picture on the scoring form. The initial sample question for the alpha items is presented below as it appeared on the scoring form.



Sample: These pictures are: fish, car, boy, apple. Which one is car? Find car. *If child answers incorrectly:* That was a good try, but this is car. Let's try it again. Which one is car?

Testing procedures. Between early October and early December of 2000, the 342 children who were participants in the study were administered the 60 item alpha pool. On another occasion within two weeks, the 14 gold standard measures from the DSC, described previously, were also administered. Children in the Florida sample, who were part of another ongoing study, were administered various other tests, including a test of oral language vocabulary, a test of letter knowledge, and a collection of tests of phonological awareness. The testing occurred at the child's Head Start, preschool, or nursery school center. Children were taken individually to a quiet location in the school. Examiners were graduate students in psychology.

Selection of the *Get Ready to Read!* Items. The goal of item selection was to reduce the alpha item pool to a selection of 20 items that would form the screening instrument. The target of 20 items was pragmatic: A longer test such as the full 60 item alpha pool, though desirable from a measurement perspective, would be impractical as a screening instrument because of the time it would take to administer.

On the other hand, a very short test, e.g., 10 items, would be unlikely to have sufficient reliability to provide valid screening. Note that the reliability of a test is strongly influenced by the number of items on the test; further a test that is too short to be reliable cannot be valid.

A number of statistical procedures are available for test construction. One group of procedures is designed to select items that relate maximally to each other so that one can say that all the items on the resulting test measure the same thing. Examples of such statistical procedures are factor analysis and item analysis. Items that relate maximally to each other, i.e. are internally consistent, also produce tests that have the highest reliability, e.g., if the test is split into two halves, the halves will correlate highly with each other. A second group of statistical procedures is designed to select items that relate maximally to some external measure or criterion, such as the gold standard items employed in the current project. Examples of such procedures are canonical correlation and multiple regression. Sometimes these two types of procedures, one focused on internal consistency and the other focused on external validity, do not offer identical answers to the question of which items to include or exclude from a test. In a screening test intended to cover several sub-domains of knowledge one does not necessary expect extremely high internal consistency. After all, it is not only conceivable but likely that there are children who have book knowledge, but not linguistic awareness, or letter knowledge but not emergent writing. For that reason, we chose to favor external validity over internal reliability when choices were necessary on item selection, with the aim of finding a set of 20 items that would relate as strongly as possible to the gold standard, while still having acceptable internal reliability.

Other considerations in item selection included item difficulty, with the aim being to eliminate items that were either quite easy or quite difficult for most children, and item coverage, with the aim being to sample from each of the sub-domains of emergent literacy knowledge that had formed the conceptual basis for the alpha item pool.

The following steps were followed in generating the 20 item screening instrument: 1) A *difficulty* set was selected from the alpha pool by identifying the 20 items whose mean correct response was closest to .50, i.e., about half the children answered the item correctly. Such items are ideal from a measurement perspective because they provide the maximal discrimination among people taking the test. In contrast, an item that everyone answers correctly or an item that everyone fails is useless in detecting differences

among people. As it turned out each of the 20 items thus selected fell between a mean correct of .4 and .6; in other words, about half the children got each of these questions correct. 2) A *reliability* set was selected by identifying the 20 items from the alpha pool that generated the highest coefficient alpha (a standard measure of internal consistency). As it turned out, coefficient alpha was .81 for the 20 items thus selected. This is a respectably high alpha for a psychological/educational assessment, particularly one with just 20 multiple choice items. 3) A *validity* set was selected by identifying the 20 items that generated the highest statistical relationship with each child's score on the DSC, the gold standard. Using the DSC score as a dependent variable, we employed a statistical technique called forward stepwise regression to select the 20 items from the alpha pool that most strongly correlated with the DSC. The multiple R between these 20 items and the DSC was .78, indicating a very strong relationship between the validity set and the DSC.

The steps outlined above generated three different sets of 20 items, the difficulty set, the reliability set, and the validity set. In order to select a final set of 20 items, we first eliminated any items on any of the three sets that were either particularly easy (mean proportion of correct responses of .80 or higher) or particularly hard (mean correct proportion of correct responses of .20 or below). We then selected any remaining item that appeared on each of the three item sets: 10 items were thus selected. We then selected any remaining item that appeared on two of the three sets: 4 items were thus selected. We then selected 6 additional items that appeared on one or another of the three item sets with the aim of generating as wide a coverage as possible of the subdomains covered by the 60 item alpha pool, and favoring items from the validity set, where possible. The scores generated by summing correct responses over the final 20 item set as well as each of the three 20 item preliminary sets were highly intercorrelated, ranging from .89 to .96. This indicates that the three different methods of initial item selection and the procedures used to select the final 20 items generated roughly equivalent scales.

Psychometric properties of the *Get Ready to Read!* Screening Tool

The final 20 item pool, the *Get Ready to Read!* Screening Tool, was evaluated for its reliability, factor structure, relationship with the gold standard, and consistency across low-income and middle class children. The first table below presents reliability and validity data for the entire sample of children who were tested. With an alpha of .78 and split-half reliability of .80, the screening test has respectable internal

consistency. The range of correct responses across children was 1 correct answer to a perfect 20. With a mean of about 9 correct and a standard deviation of about 4, we know that 68% of children taking the test scored between roughly 5 to 13 correct. This is in keeping with the design goal of choosing items that were neither too easy or too hard for most children taking the test.

The validity coefficients indicate that the *Get Ready to Read!* Screening Tool correlates highly with the total score from the DSC and with a test of letter naming. It also correlates well with a measure of language development, and with a professionally administered battery of phonological awareness tests. These robust correlations between the Screening Tool and professionally administered tests of emergent literacy are impressive, given that the overall reliability of the Screening Tool is .78 (which sets the upper limit to correlations that can be obtained between the Screening Tool and other measures). In the context of the secondary validation procedure employed in the present project, there is good evidence for the validity of the *Get Ready to Read!* Screening Tool.

Psychometrics for Total Sample (N = 342)	
Reliability	
Number of items	20
Coefficient alpha	.78
Split half reliability	.80
Mean correct answers	9.14
Range	1 - 20
Standard deviation	4.31
Validity	
Correlation with DSC total	.69
Correlation with PPVT language age (Florida sample)	.58
Correlation with letter knowledge (Florida sample)	.66
Correlation with phonological awareness (Florida sample)	.58

Because the Screening Tool is intended for children from a wide range of socio-economic backgrounds, the information in the table above is presented below for children in the middle-income nursery school subsample. As can be seen by comparing the two tables, the reliability and validity coefficients for the middle-income sample are in the same range as for the sample as a whole.

Psychometrics for Middle-Income Sample (N = 50)	
Reliability	
Number of items	20
Coefficient alpha	.78
Split half reliability	.78
Mean correct answers	12.52
Range	3 - 20
Standard deviation	4.16
Validity	
Correlation with DSC total	.79
Correlation with PPVT language age (Florida sample)	.52
Correlation with letter knowledge (Florida sample)	.65
Correlation with phonological awareness (Florida sample)	.45

The same question can be asked about the Head Start children in the sample, who are from low-income families. Comparison of the coefficients in the table below, for Head Start children, with the coefficients in the table above, for middle-class children, suggests few differences. In fact, the only statistically significant difference between the corresponding entries in the two tables are for the mean correct answers, $p < .001$. A significantly higher mean for the middle-class sample is to be expected give the known role of the home environment in the development of emergent literacy skills, and adds to the inference that the Screening Tool has validity.

Psychometrics for Head Start Sample (N = 223)	
Reliability	
Number of items	20
Coefficient alpha	.75
Split half reliability	.80
Mean correct answers	8.52
Range	1 - 20
Standard deviation	4.16
Validity	
Correlation with DSC total	.70
Correlation with PPVT language age (Florida sample)	.60
Correlation with letter knowledge (Florida sample)	.66
Correlation with phonological awareness (Florida sample)	.67

We can ask whether the Screening Tool has reasonable psychometric properties across White, Black, and Hispanic children. The table below indicates that coefficient alpha is reduced for Hispanic

children, 20% of whom were identified by their Head Start teachers as having limited English proficiency. It is relevant to report that for the sample of children who were involved in the gold standard study, the ability to predict reading success and failure in second grade from DSC scores in Head Start was actually higher for Hispanic children than for Blacks or Whites, suggesting that measures similar to those used in the Screening Tool, administered in English, can still be valid for children from Hispanic homes.

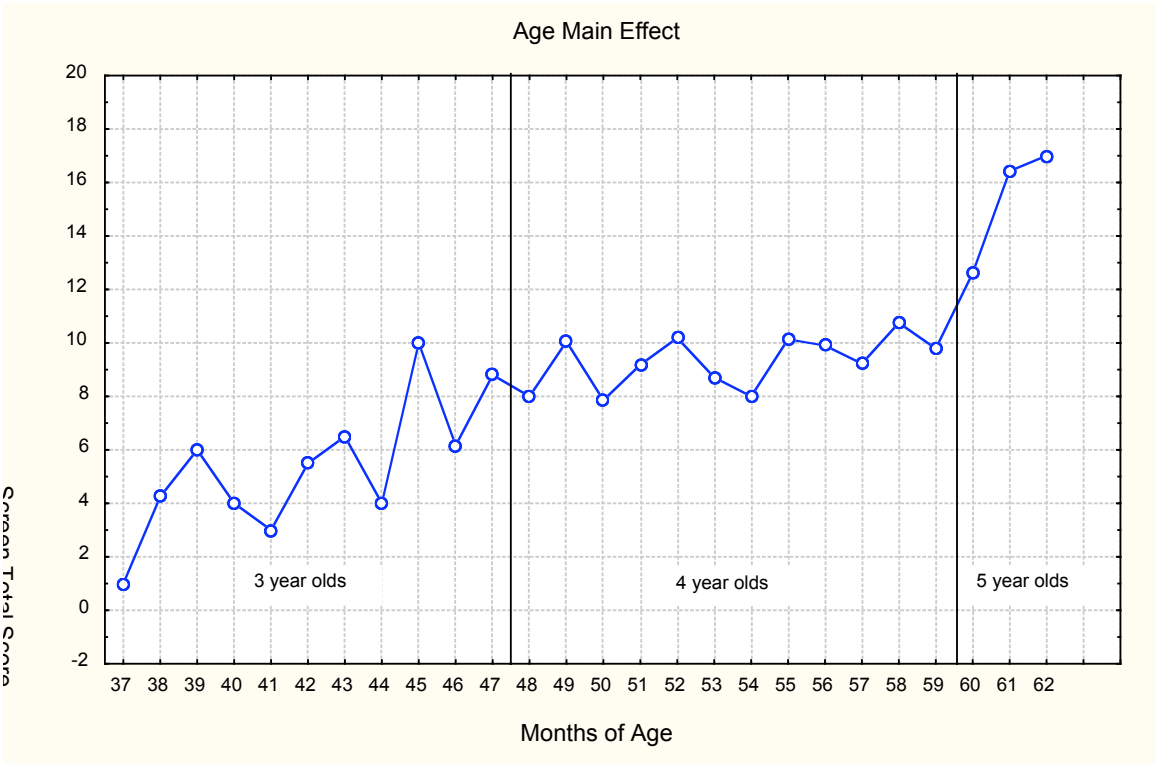
Psychometrics for Different Ethnic/Racial Groups			
Group	*Black*	*Hispanic*	*White*
Coefficient alpha	.78	.55	.76
Split half reliability	.81	.67	.73
Mean correct answers	9.03	6.86	11.23
Range	1 - 19	2-14	3 - 20
Standard deviation	4.29	3.01	4.13
Correlation with DSC total	.62	.60	.75
Correlation with PPVT language age	.58	*	.44
Correlation with letter knowledge	.63	*	.66
Correlation with phonological awareness	.53	*	.59

* Not enough Hispanic children in the Florida sample to compute. Children in the NY sample were not given tests of language age, letter knowledge, or phonological awareness.

A principal components analysis was conducted on the 20 items of the Screening Tool to determine whether different subdomains of emergent literacy could be identified. The first factor extracted accounted for a disproportionate amount of the total variance, suggesting that one interpretation of the Screening Tool is that it constitutes a single scale. The scree plot showed a smooth decrease of eigenvalues after the initial factor, again arguing for a one factor solution. However, five factors after the initial factor had eigenvalues greater than one, hinting at the presence of other underlying factors that would probably manifest themselves more clearly in a longer scale. In any case, the factor structure of the 20 item Screening Tool, with a dominant single factor, does not permit the use of a child's responses to subsets of items on the Screening Tool to indicate weakness or strength in particular subdomains of emergent literacy, such as letter knowledge, or phonological awareness. A longer assessment designed explicitly to identify profiles of emergent literacy skills will be necessary to differentiate specific areas of emergent literacy in which a child needs the most help. The Screening Tool itself only allows inferences to the child's general level of emergent literacy readiness.

The effect of age is important to determine in a screening test for children. Even thought the *Get Ready to Read!* Screening Tool is intended only for children in the pre-K year, if children closer to age

five score significantly higher than children closer to age four, then a child's age would need to be taken into account in scoring the Screening Tool. The present sample of children included 26 five-year-olds, who had just passed their fourth birthdays, and 43 three-year-olds, who were not quite four, as well as the 273 four-year-olds who were the primary focus of the Screening Tool. Analyses of the data represented in the figure below indicate that age is not a statistically significant influence on performance on the Screening Tool, *as long as children are four-year-olds*. However, as illustrated in the figure, scores start to rise after children turn five, and scores decrease in children younger than four. Based on these data, the Screening



Tool should not be used with five-year-olds or three-year-olds without the development of an interpretative scheme that takes age into account. However, the Screening Tool can be used for children between 48 and 59 months of age without concern for age.

Scoring

The intent of the *Get Ready to Read!* Screening Tool is to identify preschoolers who can profit from extra help in acquiring the emergent literacy skills that are the foundation for learning to read. The data

and analyses presented above indicate that the Screening Tool correlates well with a gold standard, the DSC, that has been demonstrated to predict reading failure vs. reading success in second grade. In other words, children who get higher scores on the Screening Tool are likely to do better at the task of learning to read when they enter school than children who get lower scores on the Screening Tool. The challenge addressed here is how to translate a child's specific score on the Screening Tool into information on whether the child needs extra help. Yes, a score of 18 is better than a score of 8, but what is the safety zone – the score at or above which a child has an acceptable likelihood of eventual reading success?

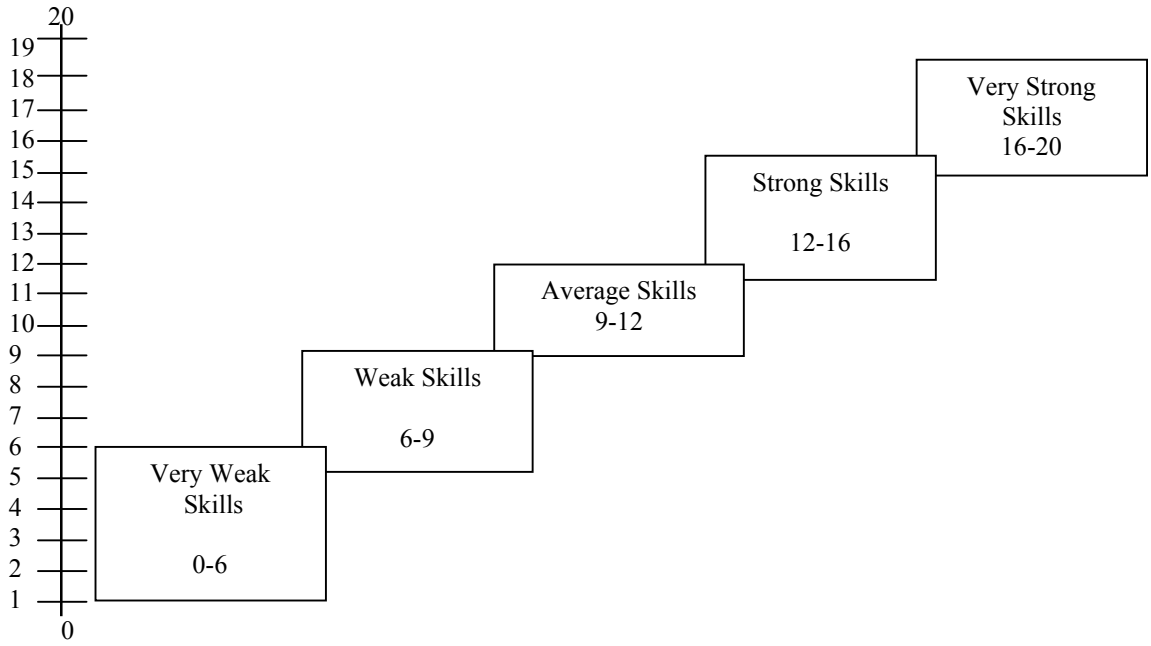
We addressed this question by again using data from the earlier study of Head Start children from which the gold standard measures were selected. Recall that we classified children in that study as successful readers if they scored within the upper 50% of children nationally on standardized tests of reading at the end of second grade. In contrast, we classified children as unsuccessful readers if they had NCE scores of 30 or below, corresponding to the 18th percentile in reading. We determined the mean number of correct answers for these two groups of children on the DSC given during Head Start. Children who were successful readers in second grade had a mean score of 25.86 correct on the DSC. In contrast, children who were unsuccessful readers in second grade had a mean score of 17.77 correct on the DSC. We then returned to the data collected from the present sample of children and asked the question: "What score on the Screening Tool is predicted from a score of 25.86 vs. a score of 17.77 on the DSC?" Because the Screening Tool and the DSC total score are strongly correlated (r = .69) in the present sample, that question can be asked and answered using a statistical technique called linear regression. A score on the Screening Tool of 11.14 correct is predicted for children with a 25.86 score on the DSC. In contrast, a Screening Tool score of 8.56 is predicted for children with the lower DSC score of 17.77.

In this mathematical context, it could be argued that children with scores of 11 or above could be considered good bets for reading success in second grade, while children with scores of 8 or below could be considered at risk of reading difficulties. However, we think that the number correct that represents the safety zone should be higher than 11 for four reasons. First, a mathematical consequence of linear regression is that predicted scores have a compressed range compared to observed scores; the amount of this compression increases as the correlation between the two measures decreases. In other words, with a less than perfect correlation of about .70 between DSC scores and Screening Tool scores, the predicted

success score of 11.14 on the Screening Tool is likely to be somewhat low. Second, the present data were collected in the fall of the pre-K year. We know from prior research that emergent literacy abilities increase significantly over the pre-K year. Thus a score of 11 correct on the Screening Tool might be a good sign for a child just entering a pre-K program, but the same score would be less rosy for a child just a few weeks away from entering kindergarten. Third, the help that we anticipate being made available to children as a result of problematic scores on the Screening Tool is in the form of educational activities that should be useful to any child who has not yet mastered all of the emergent literacy domains tapped by the Screening Tool. We would not want to see parents or teachers assume that emergent literacy activities are unimportant for children who received a score of 11 or above on the Screening Tool. Fourth, the average number of correct answers for the middle-class sample was 12.5. While this was clearly not a nationally representative sample, the PPVT language age for this subsample was within 2 months of their chronological age (i.e., the children were about average as a group), suggesting that their performance on the Screening Tool might also be close to what one would find with a nationally representative sample. If a score of 12-13 is average on the Screening Tool, then we would not want to indicate that a lower score is safe.

In the context of these analyses, we recommend the following interpretation key for scores on the *Get Ready to Read!* Screening Tool:

Score Interpretation Key for the *Get Ready to Read!* Screening Tool



Next Steps

The following are among the activities that should follow the acceptance of this document and the Screening Tool:

1) The *Get Ready to Read!* Screening Tool is intended for use by nonprofessionals who may know very little about emergent literacy and reading readiness. The Screening Tool should be distributed with a companion primer on the nature of emergent literacy and its consequences for later reading.

2) The graphics and format of the existing Screening Tool should be updated by a professional design team. It is important to keep the pictures in the child's testing form very close to the current graphics. Seemingly unimportant changes in the depiction of objects may alter the probability of children responding correctly.

3) A web version of the screen and primer need to be developed.

4) Educational activities that promote the skills assessed by the Screening Tool should be developed and made available in both print and web-versions.

5) Links to sources of professional help should be made available in the case of children receiving low scores on the Screening Tool.

6) The development of norms for the Screening Tool is advisable as a buttress to score interpretation suggested herein. A nationally representative sample of 250 children should be sufficient to accomplish norming. Because performance on the Screening Tool will be affected by the passage of the pre-K school year, norms should be developed for the beginning, middle, and end of the school year.

7) A full emergent literacy diagnostic battery, to be administered by professionals, would be an appropriate companion to the Screening Tool.

8) A Spanish version of the Screening Tool would be desirable.

9) Studies are warranted to determine whether the use of the Screening Tool changes the likelihood that parents and teachers will engage in more or different teaching activities related to emergent literacy.

10) A longitudinal primary validation study is warranted in which scores on the screen are related to long-term literacy outcomes.

References

CTB/McGraw-Hill. (1990). *Developing Skills Checklist*. Monterey, CA: CTB/McGraw-Hill.

Lonigan, C.J., Burgess, S.R., & Anthony, J.L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology, 36*, 596-613.

National Center for Educational Statistics (1998). *National Assessment of Educational Progress*. Washington, D.C.: U.S. Department of Education.

Psychological Corporation. (1989). *The Stanford Achievement Test: Eighth edition*. Orlando, FL: Harcourt Brace Jovanovich, Inc.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998*). Preventing reading difficulties in young children.* Washington, D.C.: National Academy Press.

Whitehurst, G.J., & Fischel, J.E. (2000). A developmental model of reading and language impairments arising in conditions of economic poverty. In D. Bishop & L. Leonard (Eds.), *Speech and language impairments in children: Causes, characteristics, intervention and outcome* (pp 53-71). East Sussex: Psychology Press.

Whitehurst, G.J., & Lonigan, C.J. (1998). Child development and emergent literacy. *Child Development, 69*, 848-872.

Woodcock, R.W. (1987). *Woodcock reading mastery tests - Revised.* Circle Pines, MN: American Guidance Service.

Appendices

1) Draft of Scoring Form

2) Draft of Child's Test Form