



**Wisconsin Reading Coalition**

April 25, 2012

## **Implementation of Wisconsin's Statutory Screening Requirement**

The selection of an early reading screener for Wisconsin is a decision of critical importance. Selecting the best screener will move reading instruction forward statewide. Selecting a lesser screener will be a missed opportunity at best, and could do lasting harm to reading instruction if the choice is mediocre or worse.

After apparently operating for some time under the misunderstanding that the Read to Lead Task Force had mandated the Phonological Assessment and Literacy Screen (PALS), the Department of Public Instruction is now faced with some time pressure to set up and move through a screener evaluation process. Regardless of the late start, there is still more than enough time to evaluate screeners and have the best option in place for the beginning of the 2012-13 school year, which by definition is the time when annual screeners are administered.

The list of possible screeners is fairly short, and the law provides certain criteria for selection that help limit the options. Furthermore, by using accepted standards for assessment and understanding the statistical properties of the assessments (psychometrics), it is possible to quickly reduce the list of candidates further.

### **Is One Screener Clearly the Best?**

One screener does seem to separate itself from the rest. The Predictive Assessment of Reading (PAR) is consistently the best, or among the best, in all relevant criteria. This comment is not a comparison of PAR to all known screeners, but comparing PAR to PALS does reveal many of its superior benefits.

Both PAR and PALS assess *letter/sound knowledge and phonemic awareness*, as required by the statute.

In addition, PAR assesses the important areas of *rapid naming and oral vocabulary*. To the best of our knowledge, PAR is the only assessment that includes these skills in a comprehensive screening package. That extra data contributes unique information to identify children at risk, including those from low-language home environments, and consequently improves the validity of the assessment, as discussed below.

Both PAR and PALS have high *reliability* scores that meet the statutory requirement. PAR (grades K-3) scores .92, PALS-K (kindergarten) scores .99, and PALS (grades 1-3) scores .92. Reliability simply refers to the expected uniformity of results on repeated administrations of an assessment. A perfectly reliable measurement might still have the problem of being consistently *inaccurate*, but an unreliable measurement *always* has problems. Reliability is necessary, but not sufficient, for a quality screener. To be of value, a screener must be valid.

In the critical area of *validity*, PAR outscores PALS by a considerable margin. Validity, which is also required by the statute, is a measure of how well a given scale measures what it actually intends to measure; leaving nothing out and including nothing extra. In the case of a reading screener, it is validity that indicates how completely and accurately the assessment captures the reading performance of all students who take it. Validity is both much harder to achieve than reliability, and far more important.

On a scale of 0-1, the validity coefficient (r-value) of PAR is .92, compared to validity coefficients of .75 for PALS-K and .68 for PALS. It is evident that PAR outscores PALS-K and PALS, but the validity coefficients by themselves do not reveal the full extent of the difference. Because the scale is not linear, the best way to compare validity coefficients is to square them, creating r-squared values. You can think of this number as the percentage of success in achieving accurate measurement. Measuring human traits and skills is very hard, so there is always some error, or noise. Sometimes, there is quite a lot.

When we calculate r-squared values, we get .85 for PAR, .56 for PALS-K, and .46 for PALS. This means that PAR samples 51 to 84 percent more of early reading ability than the PALS assessments. The PALS assessments measure about as much random variance (noise) as actual early reading ability. Validity is not an absolute concept, but must always be judged relative to the other options available in the current marketplace. Compared to some other less predictive assessments, we might conclude that PALS has valid performance. However, compared to PAR, it is difficult to claim that PALS is valid, as required by law.

PAR is able to achieve this superior validity in large part because it has used 20 years of data from a National Institutes of Health database to determine exactly which sub-tests best predict reading struggles. As a consequence, PAR includes rapid naming and oral vocabulary, while excluding pseudo-word reading and extensive timing of sub-tests.

PAR is *norm-referenced* on a diverse, national sample of over 14,000 children. That allows teachers to compare PAR scores to other norm-referenced formative and summative assessments, and to track individual students' PAR performance from year to year in a useful way. Norm referencing is not required by the statute, but should always be preferred if an assessment is otherwise equal or superior to the available options. The PALS assessments are not norm-

referenced, and can only classify children as at-risk or not. Even at that limited task of sorting children into two general groups, PAR is superior, accurately classifying children 96% of the time, compared to 93% for PALS-K, and only 73% for PALS.

PAR provides the unique service of an individualized report on each child that includes specific *recommendations for differentiated instruction* for classroom teachers. Because of the norm-referencing and the data base on which it was built, PAR can construct simple but useful recommendations as to what specific area is the greatest priority for intervention, the intensity and duration of instruction which will be necessary to achieve results, and which students may be grouped for instruction. PAR also provides similar guidance for advanced students. With its norm-referencing, PAR can accurately gauge how far individual children may be beyond their classmates, and suggest enriched instruction for students who might benefit. Because they are not norm-referenced, the PALS assessments can not differentiate between gray-area and gifted students if they both perform above the cut score.

PAR *costs* about the same as PALS. With bulk discounts for statewide implementation, it will be possible to implement PAR (like many other screeners) at K5, 1<sup>st</sup> grade, 2<sup>nd</sup> grade, and possibly 3<sup>rd</sup> grade with the funds allocated by statute for 2012-13. While the law only requires kindergarten screening at this time, the goal is to screen other grades as funds allow. The greatest value to screening with a norm-referenced instrument comes when we screen in several consecutive years, so the sooner the upper grades are included, the better.

PAR takes less *time to administer* than PALS (an average of 12-16 minutes versus 23-43).

The *procurement procedure* for PALS apparently can be simplified because it would be a direct purchase from the State of Virginia. However, PAR is unique enough to easily justify a single-source procurement request. Salient, essential features of PAR that would be likely to eliminate or withstand a challenge from any other vendor include demonstrated empirical validity above .85, norm-referencing on a broad national sample, the inclusion of rapid naming and oral vocabulary in a single, comprehensive package, empirically valid recommendations for differentiated intervention, guidance on identifying children who may be gifted, and useful recommendations on grouping students for differentiated instruction.

## **Conclusions**

The selection of a screener will be carefully scrutinized from many perspectives. It is our position that a single, superior choice is fairly obvious based on the facts. While it is possible that another individual or team may come to a different conclusion, such a decision should be supported by factual details that explain the choice. Any selection will have to be justified to the public as well as specific stakeholders. Some choices will be easier to justify than others, and

explanations based on sound criteria will be the most widely accepted. Simple statements of opinion or personal choice, or decisions based on issues of convenience, such as ease of procurement, would not be convincing or legitimate arguments for selecting a screener. On the other hand, the same criteria that separate PAR from other screeners and may facilitate single-source procurement also explain the choice to the public and various stakeholder groups. We urge DPI to move forward reasonably, deliberately, and expeditiously to have the best possible screener in place for the largest possible number of students in September.