CENTER FOR APPLIED RESEARCH
AND EDUCATIONAL IMPROVEMENT
UNIVERSITY OF MINNESOTA

Research Brief
Fountas & Pinnell Benchmark Assessment System
Considerations for Use as a Universal Screener

What is a universal screener?

The screening process is designed to allow for the **efficient and early identification** of needs. Screening involves administering brief, **reliable and valid assessments** to **all students** at multiple points per year. Screening provides a quick way to identify which students are expected to exceed, meet, or fall below grade level standards. Without screening, all students who may need additional support are unlikely to be identified.

There are many options when selecting a universal screener in the area of literacy. The **technical adequacy** of a screener is important to consider. Most educators are familiar with validity and reliability; however, **additional considerations** regarding the **diagnostic accuracy** of the measure(s) **must be considered** (Klingbiel, McComas, Burns, & Helman, 2015).   The terms sensitivity and specificity are terms used to describe the diagnostic accuracy of an instrument.

- Sensitivity: "The proportion of truly at-risk students who were identified as at risk by the screener" (Klingbeil et a.l, 2015, p. 502). Did the screening identify all the at-risk students in the school? If an assessment does not have adequate levels of sensitivity, students who need intervention may not be identified and therefore not receive intervention.
- Specificity: "The proportion of students who were truly not at risk among all students classified as as not at risk" (p. 502).   If an assessment does not have adequate levels of specificity, some student who are not actually at risk may be identified as such and receive intervention.

Given that all assessment involves some error, it is critical that screening measures be as **accurate as possible**. **Reducing** the rate of **false positives (providing interventions to a student who does not need them)** and **false negatives (not providing interventions to a student who needs them)** and **increasing** the rate of **true positives** and **true negatives** is not only important for efficiency and resource allocation purposes, but also to provide students with the **best possible outcomes**.

Review of Research

The Fountas & Pinnell Benchmark Assessment System (FPBAS) is a popular literacy assessment, with a 3rd Edition being recently published. The assessment is administered individually to each student and takes about 20 – 40 minutes. There is an oral reading component and a comprehension component, which together provide an instructional level for each student (independent, instructional, frustrational). The authors recommend administering the FPBAS three times per year. The 3rd edition has replaced the comprehension questions with comprehension conversations.

Independent studies examining the utility of the FPBAS as a universal screener are limited. This review summarizes the findings from three studies published in peer reviewed journals. No studies were found using the 3rd edition. Another limitation is that only 2nd and 3rd grade students were included in the studies.

Key Terms
ORF – Oral Reading Fluency or the number of words read correctly in one minute.
FPBAS – Fountas & Pinnell Benchmark Assessment System
MAP – Measures of Academic Progress

Klingbeil, D. A., McComas, J. J., Burns, M. K., & Helman, L. (2015). Comparison of predictive validity and diagnostic accuracy of screening measures of reading skills. *Psychology in the Schools, 52*(5), 500-514. http://doi.org/10.1002/pits.21839

	FPBAS	ORF	MAP
Minutes per Student	15 – 30	5	N/A
32 Students Total Time (Hours)	8 – 16	3	1 – 1.5

Research Questions & Results

1. **How much variance do two screening measures account for in end of the year reading comprehension?** *The results showed that all measures were significant predictors of spring MAP scores. ORF and FPBAS accounted for similar percentages of variance in MAP (.44 & .41).*

2. **Does using a combination of fluency and comprehension effectively predict end of year comprehension scores?** *Adding the FPBAS to ORF increased variance accounted for from 40% to 54%. All three assessments (FPBAS, ORF, and MAP) accounted for the most variance in EOY MAP (65%). ORF & MAP accounted for more variance than ORF & FPBAS (61%).*

3. **What was the diagnostic accuracy of screening measures using district or national norms?** *Using district cut-scores & a judgement based approach, specificity and sensitivity were analyzed.*

    a. *Sensitivity: Using either the FBAS or ORF to identify students as at-risk resulted in moderate sensitivity, meaning which "would result in unacceptable rates of false negatives" (p.507). Using combinations of two screening measures did not improve sensitivity and neither did using all three assessments. The MAP alone did provide adequate sensitivity.*
    b. *Specificity: When specificity was analyzed for each test individually, the MAP and the ORF were adequate, the FBAS was not. When an at-risk score on two assessments was required, all combinations met adequate levels. When students who need intervention were identified by having two out of three scores in the at-risk range specificity was adequate.*

4. **How does diagnostic accuracy differ using a regression-based approach?** *This question considers methods for decision making when more than one screening measure is used. Multiple screening assessments can sometimes be confusing, especially when results are contradictory. As a result, there are two commonly used approaches to decision-making: professional judgement and regression. A professional judgement approach involves providing interventions to students based on professional judgement or collective input from a team. A regression-based approach uses a statistical analysis of the combination of test scores to predict which students need intervention. Using a regression based approach was more diagnostically accurate than a judgement based approach.*

Burns, M. K., Pulles, S. M., Maki, K. E., Kanive, R., Hodgson, J., Helman, L. A., Preast, J. L., (2015). Accuracy of student performance while reading leveled books rated at their instructional level by a reading inventory. *Journal of School Psychology, 53*, 437-445.

Research Questions and Results

1. **How consistent are the instructional level estimates based on accuracy while reading three books rated to be at the same difficulty level by an IRI?** *The mean number of words read across three reading assessments was somewhat inconsistent. When the accuracy scores were converted to instructional levels agreement between the measures based on categorical levels was not considered to be strong. Kappa coefficient of .70 is considered strong and the study kappa coefficient was .50.*

2. **To what extent do students accurately read books that are rated at their instructional level using IRI data?** *Only 28% of the median scores fell within the instructional level range (93-97%) when reading from books at the assessed instructional level as indicated by performance on FPBAS. Overall, 15.6% of the students were reading from books that were too difficult and 56.3% were reading from books that were too easy. Additional analyses (e.g., ANOVA) indicated that students at the lower end of reading levels tended to read with less accuracy despite being in books at their instructional level as measured by the FPBAS.*

3. **How do reading skills affect the accuracy with which students read books identified to be at their instructional level with IRI data?** Overall, the FPBAS overestimated the reading levels of students at or below the 25th percentile and underestimated reading levels for students above the 25th percentile. *Students below the 25th read accurately at their instructional level 41.7% of the time. This means that students who were at or below the 25th percentile were reading from books that were at their frustrational level about 58% of the time. Students within the 26th-75th%ile were reading materials that were too easy for them 71% of the time. And students above the 75th%il were reading materials that were too easy 67.8% of the time. These results call into question the accuracy of the initial reading level provided by the FPBAS.*

Parker, D. C., Zaslofsky, A. F., Burns, M. K., Kanive, R., Hodgson, J., Scholin, S. E., & Klingbeil, D. A, (2015). A brief report of the diagnostic accuracy of oral reading fluency and reading inventory levels for reading failure risk among second- and third-grade students. *Reading & Writing Quarterly, 31*, 56-67.

Research Questions and Results

1. **What is the relationship between ORF data and scores on a district-administered criterion assessment of reading (MAP)?** *Correlations between MAP & ORF were high (.84 for 2nd grade & .74 for 3rd grade). Correlations between ORF and FPBAS were high .83 2nd grade & .73 3rd grade.*

2. **What is the relationship between reading inventory level and scores on a district-administered criterion assessment of reading?** *Correlations between MAP and FPBAS were also high (.76 2nd grade; .69 3rd grade).*

3. **What is the diagnostic accuracy of ORF for scoring proficiently on a district-administered criterion assessment of reading? What is the diagnostic accuracy of reading inventory level scores for scoring proficiently on a district-administered criterion assessment of reading?** *Using the MAP 25th percentile as the criterion, ORF was more accurate at identifying the correct overall classification (.80) than the FPBAS (.54). ORF also had higher levels of sensitivity (.86) and specificity (.78) than did the FPBAS (.31 & .66).*

   "In a hypothetical school with 100 students needing intervention, 86 of the students who actually need an intervention based on MAP performance would be correctly identified using ORF criteria. Only 31 of those students would be accurately identified using the IRI [FPBAS] screening data" (p. 64).

Conclusions

The results of the three studies underscore the important issues school districts must consider when selecting a universal screener. Time, resources, and diagnostic accuracy must all be factored into the final decision. The FPBAS falls short when compared to other measures considered in these articles (CBM and MAP).

Additional Information on Universal Screeners

More information regarding other reading screeners may be found at the following links.

Center on Response to Intervention: Screening Tools Chart

National Center on Intensive Intervention : Click on the **General Outcome Measures** tab.