How Valid are Decisions Made with Fountas and Pinnell Benchmark Assessment System (F&P) Data?

Matthew K. Burns

University of Minnesota

The F&P is a commonly used reading assessment that is purportedly appropriate for screening and instructional decisions, but little research has examined the utility of the data. The poster presents two studies that examine the validity of decisions made with F&P data regarding screening and estimating a student’s instructional level.

**Study 1: Screening.** The purpose of this study was to determine the diagnostic accuracy of F&P data and oral reading fluency (ORF), which is another commonly used reading screener. Participants included 968 second- and third-grade students from eight elementary schools. The sensitivity and specificity of the data were evaluated by comparing the results to the 25th percentile on the Northwest Evaluation Association Measures of Academic Progress for Reading (MAP-R). All 2nd grade students completed System 1 of the F&P, and scores below an H were considered at risk. The 3rd graders completed System 2 and scores below an L were considered at risk. ORF and F&P data were highly correlated (ρ = .83 for 2nd grade and ρ = .73 for 3rd), and both had high correlations to the MAP-R. ORF data resulted in sensitivity of .86, specificity of .78, and 80% overall correct classification. Estimates for the F&P were much lower, with sensitivity of only .31, specificity of .66, and 54% overall correct classification.

**Study 2: Instructional Level.** The purpose of Study 2 was examine the reading performance of students as they read from texts at their instructional level as determined by the F&P. A total of 64 2nd and 3rd graders read aloud for 1 minute from three F&P leveled texts and the percentage of words read correctly was recorded. The percent accurate from each of the three readings correlated from \( r = .47 \) to \( r = .68 \). Data were then converted to instructional level estimates of 93% to 97% being an instructional level, < 93% being a frustration level, and > 93% being an independent level (Gickling & Havertape, 1981). Percent agreement calculations showed that the categorical scores for the three readings agreed approximately 67% to 70% of the time, which resulted in a kappa estimate of less than .50. There was considerable variability in accuracy for these three readings taken from books rated at the same level of difficulty by the F&P. Finally, 58% of the students who were low readers read at a frustration level from books rated at their instructional level by the F&P, and 71.4% of the middle and 67.7% of the high readers read at an independent level. Thus, only 26% of the students read at an instructional level when reading from books rated to be at their instructional level by the F&P.