School Psychology Review,
2014, Volume 43, No. 1, pp. 30–40

Research Brief

The Relation of Word Reading Fluency Initial Level and Gains With Reading Outcomes

Jean Louise M. Smith, Kelli D. Cummings, Joseph F. T. Nese, Julie Alonzo, Hank Fien, and Scott K. Baker
University of Oregon

Abstract. The purpose of this research brief is to present results from a study investigating the relation between word reading fluency (both initial level and fall–winter gain scores), passage reading fluency, and reading comprehension. Word reading fluency data were collected in the fall and winter; outcome measures were administered in the spring of Grade 1 (N = 150 students). We found that both initial level and gains on the word reading fluency measure predicted important reading outcomes in Grade 1 and that the impact of gains was attenuated for students with strong initial skills. These findings are consistent with previous research on fluency-based measures of early reading skill.

Students with poor word reading skills often do not reach adequate skill levels in reading comprehension and thus struggle to access subsequent levels of the general education curriculum (Adams, 1990; Torgesen, 2000). There are varied theoretical models of how readers construct meaning from print, bridging four decades of research; however, most theories are premised on the importance of effortless, or automatic, word reading (LaBerge & Samuels, 1974). This process of reading words automatically, seemingly effortlessly, has been termed reading words by “sight” (Ehri, 2005, p. 168), but the process is actually much more nuanced that the name implies. Accurate and efficient word reading

Scott K. Baker is now at Southern Methodist University. This research was supported in part by funds from the Institute for Education Sciences (PR/Award No. R324A100014 funded from June 2010–June 2014). The opinions expressed are those of the authors and do not represent the views of the Institute or the US Department of Education. Support for this work was also provided by the Center on Teaching and Learning and Behavioral Research and Teaching, College of Education, University of Oregon. The authors acknowledge their research team and school partners for their support of this project. They thank Drs. Doris L. Baker, Edward J. Kame’enui, and Gerald Tindal, who worked with this team on the design and management of this study.

Correspondence regarding this article should be addressed to Jean Louise M. Smith, University of Oregon Center on Teaching and Learning, 5292 University of Oregon, Eugene, OR 97403-5292; e-mail: jmercier@uoregon.edu

represents a confluence of multiple sublexical and cognitive/memory processes (Ehri, 2005; Perfetti, 1992; Share, 1995). Generally, theory suggests that comprehension requires cognitive resources and that word recognition, as well as the sublexical processes that produce it, is a candidate for automatic processes that would preserve processing resources for comprehension (LaBerge & Samuels, 1974; Perfetti, 2007). Perfetti's lexical quality hypothesis poses that comprehension processes rely on high-quality lexical representations, including a highly specified orthographic representation (i.e., spelling); a phonological representation, through either spoken language or phonemic decoding; and semantic information (i.e., meaning). To be high quality, the representation must be coherent, accurate, complete, and synchronously available (Perfetti, 2007).

Word reading is often the target of screening in the early grades because of the robust relation between word reading and comprehension. Assessing skilled word reading can be firmly grounded within a strong theoretical framework of word reading processes (Olson, Forsberg, Wise, & Rack, 1994). Direct measures of word reading fluency (WRF), such as reading a list of words, are designed to measure efficient and accurate word reading, without the influence of context-related comprehension processes (Olson et al., 1994). Word reading assessments differ from decoding assessments in that word reading assessments include both phonetically regular and irregular words whereas decoding assessments target phonemic decoding skill by including phonetically regular words (or decodable nonwords).

Rationale for Examining WRF Gains

Although the most common use of word reading assessments is to make normative and criterion-referenced comparisons at single points in time, recent research has highlighted the added benefit of evaluating changes between levels (i.e., gain scores) from one measurement occasion to the next (Ardoin & Christ, 2008; Fien et al., 2010). For the purposes of this report, we define *gains* as change in scores between two assessment occasions (e.g., winter scores minus fall scores) whereas *growth* represents a mathematically derived slope estimate that is established across three or more data points (e.g., the estimated slope for five data points that were gathered across 5 weeks). Progress monitoring typically involves the analysis of slope (growth) and includes the use of assessment data collected multiple times per year (Christ, Silberglitt, Yeo, & Cormier, 2010). Within a multitiered approach to instructional support, data teams often use gains (difference scores) to evaluate the efficacy of their tiered instructional support plans and to guide necessary modifications (Fien et al., 2010). Although many basic psychometric properties of a measure are retained when gain scores are examined, additional information is needed to determine the extent to which gain scores, above and beyond skill level, relate to later outcomes (Cummings, Dewey, Latimer, & Good, 2011; Fien et al., 2010; Fuchs, Fuchs, & Compton, 2004; Good, Baker, & Peyton, 2009; Harn, Stoolmiller, & Chard, 2008).

There is evidence that initial scores on WRF measures, including word identification fluency, function well for screening purposes in the early grades (Clemens, Shapiro, & Thoemmes, 2011; Compton, Fuchs, Fuchs, & Bryant, 2006; Fuchs et al., 2004; Zumeta, Compton, & Fuchs, 2012). In addition to evidence of the predictive validity of Grade 1 fall score levels, when word reading curriculum-based measures are used to document skill growth for relatively short periods (i.e., < 9 weeks), growth from the fall to winter of Grade 1 predicts reading outcomes at the end of Grade 1 (Fuchs et al., 2004; Zumeta et al., 2012) and Grade 2 (Compton et al., 2006, 2010). In some cases, the inclusion of fall-to-winter growth in classification models improves the accuracy of screening decisions compared with fall scores alone (Clemens et al., 2011; Compton et al., 2006, 2010).

Given that growth, and possibly gains, on measures of word reading can predict later reading outcomes, we are also interested in the extent to which initial skill moderates the relation between fall-to-winter gains and end-of-grade outcomes. For example, there is compelling evidence that relations between gains on a phonemic decoding measure and reading outcomes are moderated by initial skill (Clemens et al., 2011; Cummings et al., 2011; Fien et al., 2010; Good et al., 2009; Harn et al., 2008). If an interaction between initial skill and gains on WRF exists, then—as with other fluency assessments—examining gains in the absence of initial skill may attenuate relations between gains and important outcomes. However, because WRF measures can be theoretically conceptualized as measuring a confluence of skills (Clemens et al., 2011; Ehri, 2005; Share, 1995), it is not clear whether the finding from a measure of phonemic decoding (e.g., nonsense word fluency) will replicate to a WRF measure.

Purpose and Research Questions

The purpose of this study was to examine the extent to which initial skill level and fall-to-winter gains on the easyCBM WRF (Alonzo & Tindal, 2007) relate to later reading outcomes, including end–of–first grade passage reading fluency (PRF) and reading comprehension. This study addressed the following two research questions: (a) What is the relation between WRF initial level, fall–winter WRF gains, and spring PRF and reading comprehension? (b) What are the unique variances in spring reading outcomes that are predicted by both WRF initial level and WRF gain?

Methods

Participants and Setting

Primary data collection activities occurred during the 2009–10 academic year with a sample of 150 Grade 1 students. The sample includes students from four schools and nine classrooms within a single school district located in a small city in the Pacific Northwest. All students in the participating classrooms participated in the data collection, but only students with data for all measures at all time points were included in the analytic sample. School participation in the study was voluntary, and the student sample represented approximately 9% of the district's approximately 1,717 Grade 1 population.

Of the students in this sample, 51% were male; 10% were identified for special education services at the time of the study. Most students (84%) spoke English as their primary language. Of the students in this sample, 64% were identified as White, 17% as Asian/Hawaiian/Pacific Islander, 3% as Black, 2% as American Indian/Alaskan Native, and 5% as multiethnic; 9% were of unknown race. Across all categories of race, 29% of the sample was identified as Hispanic/Latino. Districtwide policy stipulates that schools provide 90 minutes of daily Tier 1 reading instruction in a core program, with additional instructional support for students in Tier 2 and Tier 3.

Measures

Word reading fluency. The easyCBM WRF measure is a standardized, individually administered test of word reading accuracy and fluency (see Alonzo & Tindal, 2007, for a full description of the measure). The reported median alternate form reliability for Grade 1 WRF is 0.95, and the median test–retest reliability is 0.94 (Alonzo & Tindal, 2009). The reported predictive validity for Grade 1 WRF is 0.67–0.69 with the Stanford Achievement Test, 10th edition (SAT10), reading total standard score (Lai et al., 2010). We used the fall score as the first predictor in this study and WRF first-semester gain score (difference between winter and fall WRF) as the second predictor.

Passage reading fluency. We used easyCBM PRF (Alonzo & Tindal, 2007) spring benchmark fluency scores as one outcome variable in our analyses. PRF is a standardized, individually administered test of reading accuracy and fluency in connected text (see Alonzo & Tindal, 2007, for a full description of the measure). The reported alternate form reliability for Grade 1 PRF is between 0.95 and 0.97, and the reported test–retest reliability is between 0.96 and 0.97 (Alonzo & Tindal, 2009). The concurrent correlation between Grade 1 PRF and WRF is 0.75, and the predictive correlation between

PRF and the SAT10 reading total standard score is 0.63–0.69 (Lai et al., 2010).

Stanford Achievement Test. We used scaled scores from the reading comprehension subtest of the SAT10 (Harcourt Educational Measurement, 2002) as the second outcome measure in this study. The SAT10 is a group-administered, norm-referenced test of overall reading proficiency. The reliability coefficients of the SAT10 reading comprehension subset for Grade 1 are 0.87 (alternate form) and 0.91 (Kuder-Richardson 20); criterion-related validity coefficients range from 0.50–0.82 (Harcourt Assessment, 2004).

Procedures

Data collection occurred at three time points, once in the fall, winter, and spring, and only those students with scores for each occasion were included in the study. District personnel administered the WRF measure in the fall and winter and both the PRF and reading comprehension measures in the spring. The district administered the word reading and passage fluency measures as part of standard practice. Research staff, including the lead author, provided WRF and PRF training to district data collectors in the fall, winter, and spring. Data collectors were required to meet a 90% training reliability criterion on the total score of each measure according to a pre-prepared script.

Interrater Reliability

A shadow-scoring procedure was used for interrater reliability in the fall, winter, and spring with 30% of the district. A district data collector administered and scored the probe, and a project research assistant observed the administration and scored the same probe in a different booklet. Interrater agreement was computed by dividing the total number of words on which both raters agreed by the total number of words, which ranged from 92%–100% for WRF raw scores within each of the fall, winter, and spring data collection periods.

Analysis

We conducted a sequential multiple regression analysis to examine the relation between each of the reading proficiency outcomes (i.e., PRF and reading comprehension) and our two predictors (i.e., WRF initial level and WRF fall–winter gain). Sequential multiple regression is used to build successive linear regression models in which variables are entered into the equation as separate blocks, with each block adding more predictors. The proportion of variance accounted for by the subsequent blocks of predictors can then be examined and tested for statistical significance.

Previous findings have indicated that the relation between gains (or growth) and outcomes is moderated by student initial skill level or, in other words, that estimates of gain scores depend on where students start the year (Baron & Kenny, 1986). This interaction can be captured by including polynomial terms for continuous scores (e.g., Zumeta et al., 2012) or by grouping students according to their risk level and examining the regression coefficients for each group (e.g., Cummings et al., 2011; Harn et al., 2008). In this study, we used a dummy-code approach to create five quintiles that separate the Grade 1 students by initial WRF scores. Quintile 1 contained those students at or below the 20th percentile on the fall WRF (21% of the sample), Quintile 2 contained students between the 21st and 40th percentiles (20%), Quintile 3 contained students between the 41st and 60th percentiles (20%), Quintile 4 contained students between the 61st and 80th percentile (19%), and Quintile 5 contained students at or above the 81st percentile (20%). The fall WRF scores are used as an index of students' WRF initial level.

In the first block, we entered WRF fall quintiles with the exception of Quintile 3, which served as the reference group for all analyses. Quintile 3 corresponds to students scoring between the 41st and 60th percentiles on the fall WRF, or the "average" scoring group. In the second block, we entered the mean-centered WRF gain score. In the third block, we entered WRF initial level–by–gain

Table 1
Mean, Standard Deviation (SD), and Minimum (Min.) and Maximum (Max.) Scores of Word Reading Fluency (WRF) Fall-Winter Gains (uncentered), Passage Reading Fluency (PRF), and Reading Comprehension Scores Associated with the Full Sample and the Fall WRF Quintiles

		WRF gain (uncentered)			PRF			Reading comprehension		
	n	Mean (*SD*)	Min.	Max.	Mean (*SD*)	Min.	Max.	Mean (*SD*)	Min.	Max.
Full sample	150	15.93 (10.36)	−8	61	70.44 (40.96)	4	209	566.82 (45.50)	443	667
Quintile 1	32	13.06 (7.36)	3	39	33.22 (22.57)	4	96	526.87 (49.06)	443	617
Quintile 2	30	12.70 (5.66)	0	22	46.20 (23.24)	10	113	544.60 (29.68)	483	617
Quintile 3	30	13.60 (7.19)	−1	31	61.07 (22.08)	11	105	565.43 (26.81)	514	643
Quintile 4	28	23.54 (13.41)	3	61	89.89 (22.80)	46	131	588.29 (27.63)	534	643
Quintile 5	30	17.47 (12.63)	−8	41	125.60 (28.68)	80	209	613.00 (28.76)	550	667

interaction terms to test for moderation of gain effects based on initial level.

We also examined the unique variance explained by each predictor in the reading outcomes (i.e., squared semipartial correlations, SC^2) to determine which effect accounted for the most unique variance. Squared semipartial correlations represent the unique relation between the predictor and the criterion variables when other predictor variables are held constant.

Results

Of the students in participating classrooms, only those with scores for each occasion and measure were included in the study ($N = 150$); there were no missing data in our analyses. For the sample in our study, the average fall WRF score was 18.03 words correct per minute (WCPM) ($SD = 19.14$), the average winter WRF score was 33.97 WCPM ($SD = 23.06$), and the average, standardized reading comprehension score was 566.82 ($SD = 45.50$). The mean WRF gain score was 15.93 WCPM ($SD = 10.37$), and the mean PRF spring score was 70.44 WCPM ($SD = 40.96$). Table 1 displays all descriptive statistics. Table 2 displays the correlations between the predictor and outcome variables; the top of the table displays the correlations among fall WRF, uncentered WRF gain, PRF, and reading comprehension for the full sample, and the bottom of the table displays conditional correlations for fall WRF with uncentered WRF gain, PRF, and reading comprehension separate for students within each quintile.

To reduce multicollinearity in the models, the WRF gain score and the interaction terms were mean centered. The final model results yielded only three predictors flagged for multicollinearity (tolerance < 0.3 and variance inflation factor > 1.5), whereas in the model with WRF gain uncentered, all predictors were flagged for multicollinearity. We determined that multicollinearity was not a significant issue for our data but was, rather, an artifact of the interaction model.

Relation Between WRF and PRF and Comprehension

In this section, we present the results for our first research question regarding the relation between fall WRF level and WRF fall-to-winter gain and PRF and reading comprehension.

Passage reading fluency. The PRF regression models were statistically significant for all blocks. The sequential regression analyses yielded R^2 values of 0.67 for Block 1, 0.72 for Block 2, and 0.78 for Block 3. The

Table 2
Correlations between the Full Sample and Fall Word Reading Fluency (WRF) Quintiles, WRF Fall-Winter Gain (uncentered), Passage Reading Fluency (PRF), and Reading Comprehension

	WRF gain (uncentered)	PRF	Reading comprehension
Full sample fall WRF	.15	.82	.64
Full sample WRF gain (uncentered)		.42	.36
Full sample PRF			.74
Quintile 1 fall WRF	−.15	−.48	−.46
Quintile 2 fall WRF	−.16	−.30	−.25
Quintile 3 fall WRF	−.11	−.12	−.02
Quintile 4 fall WRF	.35	.23	.23
Quintile 5 fall WRF	.07	.68	.51

Note. All correlation coefficients greater than .20 are significant at the .01 level.

addition of the WRF fall–winter gain scores (Block 2) yielded a significantly improved fit over the initial fall WRF scores (F change $= 28.48$; $df = 1, 144$; $p < .001$), and the addition of the interaction terms (Block 3) yielded a significantly improved fit over the main effects (F change $= 10.09$; $df = 4, 140$; $p < .001$).

For every unit above average WRF gain, students in Quintile 2 read 2.64 more WCPM on the spring PRF measure; students in Quintile 3 read 2.29 more WCPM on the spring PRF measure; students in Quintile 1 read 2.02 more WCPM on the spring PRF measure; students in Quintile 4 read 1.25 more WCPM on the spring PRF measure; and students in Quintile 5 read 0.36 fewer WCPM on the spring PRF measure. In other words, there was a small, negative relation between WRF gain and spring PRF for students who started the year with the highest scores on the WRF measure (i.e., Quintile 5). However, on the basis of the results presented in Table 3, gains by Quintiles 1, 2, and 4 were not significantly different from those by Quintile 3. To understand the relation among the quintiles, we ran the same regression model separately so that each quintile was the reference group; we found that the effects of gains on the PRF measure were not statistically significantly different for Quintiles 1–4 and the effects of all four quintiles were significantly greater than for Quintile 5 (it should be noted that we have only presented the results for the model in which Quintile 3 is the reference group). In addition, the effects of Quintiles 1–4 were statistically significantly greater than zero; however, the effect for Quintile 5 was not. The predicted PRF scores with quintile trend lines are shown in Figure 1a.

Reading comprehension. The reading comprehension regression models were also statistically significant for all blocks. The sequential regression analyses yielded R^2 values of 0.46 for Block 1, 0.50 for Block 2, and 0.58 for Block 3. The addition of the WRF fall–winter gain scores (Block 2) yielded a significantly improved fit over the initial fall WRF scores (F change $= 10.40$; $df = 1, 144$; $p < .01$), and the addition of the interaction terms (Block 3) yielded a significantly improved fit over the main effects (F change $= 8.89$; $df = 4, 140$; $p < .001$).

For every unit above average WRF gain, students in Quintile 1 scored 4.06 points higher on the spring comprehension measure; students in Quintile 2 scored 3.04 points higher on the spring comprehension measure; students in Quintile 3 scored 2.04 points higher on the spring comprehension measure; students in Quintile 5 scored 0.04 points higher on the spring comprehension measure;

Table 3
Multiple Regression Results for the Final Model (all blocks entered)

Variable	*b*	*SE*	*p*	SC^2	f^2
Passage Reading Fluency (PRF)					
Intercept	66.41	3.78	<.001		–
Fall WRF quintile 1	−27.40	5.32	<.001	.04	.11
Fall WRF quintile 2	−11.66	5.61	.040	.01	.02
Fall WRF quintile 4	13.97	5.72	.016	.01	.03
Fall WRF quintile 5	59.74	5.23	<.001	.20	.21
WRF fall-winter gain	2.29	0.51	<.001	.03	.05
WRF gain x quintile 1	−0.27	0.70	.696	.00	.03
WRF gain x quintile 2	0.35	0.82	.667	.00	.00
WRF gain x quintile 4	−1.04	0.58	.076	.00	.04
WRF gain x quintile 5	−2.65	0.58	<.001	.03	.04
Reading Comprehension					
Intercept	570.19	5.70	<.001	–	
Fall WRF quintile 1	−31.68	8.02	<.001	.04	.21
Fall WRF quintile 2	−15.76	8.47	.065	.01	.05
Fall WRF quintile 4	19.06	8.63	.029	.01	.06
Fall WRF quintile 5	42.73	7.89	<.001	.08	.39
WRF fall-winter gain	2.04	0.77	.009	.02	.09
WRF gain x quintile 1	2.02	1.05	.058	.01	.05
WRF gain x quintile 2	1.00	1.24	.420	.00	.01
WRF gain x quintile 4	−2.16	0.88	.015	.02	.08
WRF gain x quintile 5	−1.98	0.88	.026	.01	.07

Note. WRF fall-winter gain scores were mean-centered and consequently so were the interaction terms. SC^2 = squared semipartial correlation. $f^2 = SC^2 / 1 - R^2$.

and students in Quintile 4 scored 0.12 points lower on the spring comprehension measure. The effects of WRF fall–winter gains on reading comprehension were not statistically significantly different for Quintiles 1–3, and the effects of all three quintiles were significantly greater than for Quintiles 4 and 5. In addition, the effects of Quintiles 1–3 were statistically significantly greater than zero; however, the effects for Quintiles 4 and 5 were not. The predicted reading comprehension scores with quintile trend lines are shown in Figure 1b.

Variance in PRF and Comprehension Explained by WRF

In this section, we present results for our second research question regarding the unique variance in our outcome measures explained by WRF initial level and WRF fall-to-winter gain, as well as their interactions.

Passage reading fluency. The SC^2 values in Table 3 represent the unique variance accounted for in the spring PRF measure by each predictor, after we controlled for prior predictors. Given the multicollinearity among the predictors and thus the large amount of shared variance in the outcome, the SC^2 value for each predictor tends to downplay its overall contribution to the model. For example, the final model explained 78% of the total variance in spring PRF scores. Approximately 32% of the total variance in the PRF measure can be uniquely attributed to the predictors (i.e., the sum of the total of the PRF SC^2 values in Table 3), meaning that 46% of the total variance is shared between the predictors. However, we can use the SC^2 values to describe the relative contribution of fall WRF scores and fall–winter WRF gain values. Fall WRF scores exclusively accounted for 26% of

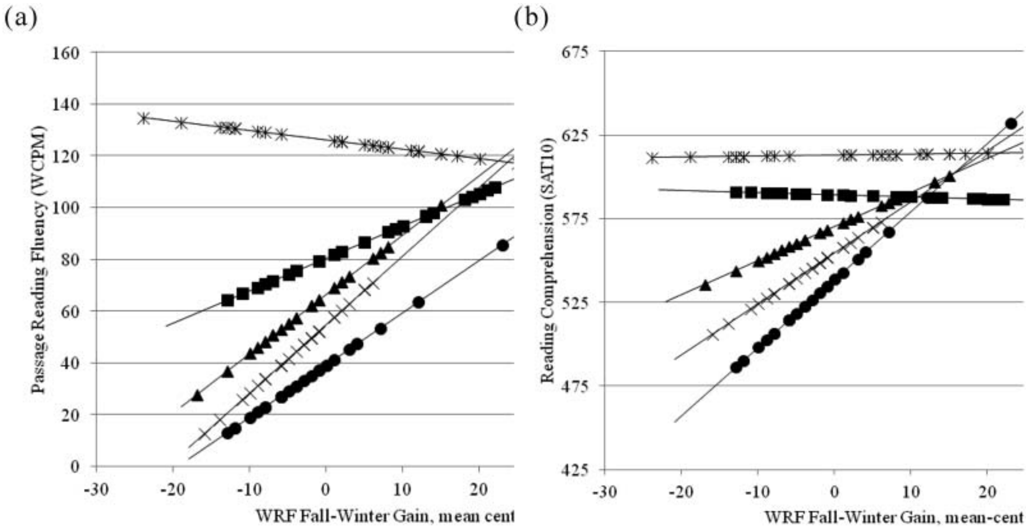

Figure 1. Predicted passage reading fluency scores (a) and predicted reading comprehension (Stanford Achievement Test, 10th edition; SAT10) scores (b) with fall–winter word reading fluency (WRF) gains (mean centered), by quintile and with quintile trend lines. mean-cent = mean centered; Quintile 1 (circles) = students at or below 20th percentile on fall WRF, quintile 2(x's) = students between the 21st and 40th percentiles, quintile 3 (triangles) = students between the 41st and 60th percentiles, quintile 4 (squares) = students between the 61st and 80th percentile, and quintile 5 (asterisks) = students above the 81st percentile; WCPM = words correct per minute.

the variance in spring PRF scores, whereas fall–winter WRF gain exclusively accounted for 3%. The fall-gain interaction terms also exclusively accounted for 3% of the variance in the spring PRF measure.

To help frame the meaningfulness of these estimations of unique variance, we can use Cohen's (1992) effect size (ES) index. The values associated with a small, medium, and large ES for a multiple partial correlation (f^2) are 0.02, 0.15, and 0.35, respectively. Thus, the ES for fall WRF on spring PRF (0.37) was large, whereas the ES for the fall–winter gain (0.05) and the ES for the interaction effects (0.11) were both small (Table 3).

Reading comprehension. The final model explained 58% of the variance in spring reading comprehension scores, and approximately 38% of the total variance explained was shared between the predictors. With partitioning of the unique variance between initial WRF level and fall–winter WRF gain, fall WRF scores exclusively accounted for 14% of the variance in spring reading comprehension scores, whereas fall–winter WRF gain exclusively accounted for 2%, and the fall-gain interaction terms exclusively accounted for 4% of the variance in spring reading comprehension scores. The ES for fall WRF on spring reading comprehension (0.71) was large, the ES for the fall–winter gain (0.09) was small, and the ES for the interaction effects (0.21) was medium.

Discussion

Our results indicate that, for all but the highest initially scoring students (i.e., the upper 20th percentile of all students when predicting PRF and the upper 40th percentile of all students when predicting reading comprehension), fall–winter gains in Grade 1 were predictive of both PRF and reading comprehension measured in the spring. As indicated in Table 2, students who began the school year with a low score (i.e., at or below the third quintile) tended to have higher growth. This phenomenon has been documented in other studies (see Clemens et al., 2011; Cummings et al., 2011; Fien et al., 2010; Good et al., 2009; Harn et al., 2008). Students in all other quintiles except Quintile 5 also were predicted to have a WCPM boost if they made WRF gains that outperformed the average. When predicting spring reading comprehension, we observed a similar pattern for this interaction effect. WRF fall–winter gain had a significant impact on spring reading comprehension for students in Quintiles 1, 2, and 3. WRF gains were not efficient predictors of reading comprehension outcomes for the students in our sample with the highest initial WRF scores.

As a screening tool, WRF appears to work well across skill levels: both fall Grade 1 WRF score and fall–winter WRF gain score explained significant and meaningful variation in spring PRF and spring reading comprehension measures. Moreover, both WRF scores (i.e., gain and initial status) explained more variance in PRF measures than in reading comprehension measures (78% versus 58%), and initial scores explained more unique variance than gain scores (26% versus 6% for PRF and 14% versus 6% for reading comprehension). Similar to the results of other research (e.g., Fuchs et al., 2004), we found that WRF fall scores, by themselves, were the strongest predictor for initially high-performing students only. Specifically, the percent of unique variance in both of our outcome measures that was explained by initial WRF status alone was always greatest for students in the upper quintile of our sample. One possible interpretation of this result is that we can expect distinctly higher end-of-year PRF and reading comprehension scores for students who are high initial performers on the WRF measure. The year-end performance for students who initially perform lower on the WRF measure is more difficult to predict from their initial scores alone. Gain scores are most meaningful among students who initially perform in the lower quintiles.

Taken together, the results of our study suggest that WRF gains across Grade 1 may be important to consider and that such gains should be contextualized based on the initial skill level of the student. The results of our study also highlight what is becoming a well-documented phenomenon in assessment research: The relation between a screener and an outcome measure is not one-dimensional. How well a screening tool predicts an important later outcome depends on students' skills at the beginning of the school year (e.g., Fien et al., 2010; Zumeta et al., 2012); the scope, sequence, and type of instruction they receive (e.g., Cummings et al., 2011; Harn et al., 2008); and the extent to which formative assessment data are used to tailor instruction to students' rates of progress throughout the year (e.g., Johnson, Jenkins, Petscher, & Catts, 2009).

Limitations

This study represents results from a relatively small sample of students from a single school district in the Pacific Northwest. In addition, approximately one quarter of the students in participating classrooms were not included in the analysis because of missing data from one or more measure, and approximately 3% were not included because of data entry errors. Future research should include a larger, nationally representative sample. A larger sample would allow additional power to detect possible significant interaction effects across quintile groups. It is possible that the significant effects that were found for Quintile 5 in this study could be found in the other quintile groups as well, but because of the small sample size in this study, we were unable to detect these effects. In addition, because of our small sample, we used ordinary least squares regression rather than heirarchical linear modeling. As a result, we cannot account for shared variance at the classroom or school levels nor can we explore covariates at these levels. This limitation should be addressed in future research because we believe that the value of screening tools in early literacy should be evaluated across different instructional contexts. It is reasonable to believe that the value of a screening tool that assesses sight word reading may differ in a school in which those sight words are taught early in the school year. We also think it is worth noting that this study took place in schools in which WRF data were used to screen students, monitor their progress, and adjust instruction. Predictive relationships may be attenuated when effective instruction is put in place to improve student progress (e.g., Johnson et al., 2009). In addition, because school personnel collected data as part of their typical school procedures, we lack information on the procedural integrity and interrater reliability for all but 30% of the fall, winter, and spring data collection activities for curriculum-based measures.

Conclusions

Although the findings are in need of replication, this study provides support for the use of gains in WRF for decision making. However, similar to findings related to gains in phonemic decoding, gains in word reading should be considered within the context of a student's initial skill. The results of this study suggest that gains are most meaningful for students who begin the year with low initial word reading skills.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Alonzo, J., & Tindal, G. (2007). *The development of fifth-grade passage reading fluency measures for use in a progress monitoring assessment system*. Technical report No. 40. Eugene, OR: Behavioral Research and Teaching. Retrieved from http://www.brtprojects.org/publications/dl/54

Alonzo, J., & Tindal, G. (2009). *Alternate form and test-retest reliability of easycbm reading measures.* Technical report No. 0906. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Ardoin, S. P., & Christ, T. (2008). Evaluating curriculum-based measurement slope estimates using data from triannual universal screenings. *School Psychology Review, 37,* 109–125.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(2), 1173–1182.

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based measurement of oral reading: An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review, 39,* 447–462.

Clemens, N. H., Shapiro, E. S., & Thoemmes, F. (2011). Improving the efficacy of first grade reading screening: An investigation of word identification fluency with other early literacy indicators. *School Psychology Quarterly, 26,* 231–244. doi:10.1037/a0025173

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159.

Compton, D. L., Fuchs, D., Fuchs, L., & Bryant, J. D. (2006). Selecting at-risk readers in first grade for early intervention: A two-year longitudinal study of decision rules and procedures *Journal of Educational Psychology, 98,* 394–409. doi:10.1037/0022-0663.98.2.394

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., . . . Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology, 102,* 327–340. doi:10.1037/a0018448

Cummings, K. D., Dewey, E. N., Latimer, R. J., & Good, R. H., III. (2011). Pathways to word reading and decoding: The roles of automaticity and accuracy. *School Psychology Review, 40,* 284–295.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading, 9,* 167–188. doi:10.1207/s1532799xssr0902_4

Fien, H., Park, Y., Baker, S. K., Smith, J. L. M., Stoolmiller, M., & Kame'enui, E. J. (2010). An examination of the relation of nonsense word fluency initial status and gains to reading outcomes for beginning readers. *School Psychology Review, 39,* 631–653.

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: Word identification fluency versus nonsense word fluency. *Exceptional Children, 71,* 7–21.

Good, R., Baker, S. K., & Peyton, J. (2009). Making sense of nonsense word fluency: Determining adequate progress in early first grade reading. *Reading & Writing Quarterly, 25,* 33–56. doi:10.1080/10573560802491224

Harcourt Assessment. (2004). *Stanford achievement test series* (10th ed.). Technical data report. San Antonio, TX: Author.

Harcourt Educational Measurement. (2002). *Stanford achievement test* (10th ed.). San Antonio, TX: Author.

Harn, B. A., Stoolmiller, M., & Chard, D. J. (2008). Measuring the dimensions of alphabetic principle on the reading development of first graders: The role of automaticity and unitization. *Journal of Learning Disabilities, 41,* 143–157. doi:10.1177/0022219407313585

Johnson, E. S., Jenkins, J. R., Petscher, Y., & Catts, H. W. (2009). How can we improve the accuracy of screening instruments? *Learning Disabilities Research & Practice, 24,* 174–185. doi:10.1111/j.1540-5826.2009.00291.x

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6,* 293–323. doi:10.1016/0010-0285(74)90015-2

Lai, C.-F., Nese, J., Jamgochian, E. M., Kamata, A., Anderson, D., Park, B. J., . . . Tindal, G. (2010). *Technical adequacy of the easycbm mathematics measures: Grades k-1, 2009–2010 version.* Technical report No. 1003. Eugene, OR: Behavioral Research and Teaching, University of Oregon.

Olson, R. K., Forsberg, H., Wise, B., & Rack, J. (1994). Measurement of word recognition, orthographic, and phonological skills. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (pp. 243–277). Baltimore, MD: Paul H. Brookes.

Perfetti, C. A. (1992). Cognitive and linguistic factors in learning to read. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Hillsdale, NJ: L. Erlbaum Associates.

Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11,* 357–383. doi:10.1080/10888430701530730

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55*(2), 151–218. doi:10.1016/0010-0277(94)00645-2

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice, 15,* 55–64. doi:10.1207/SLDRP1501_6

Zumeta, R., Compton, D., & Fuchs, L. (2012). Using word identification fluency to monitor first-grade reading development. *Exceptional Children, 78,* 201–220.

Date Received: May 15, 2012
Date Accepted: May 23, 2013
Associate Editor: Joseph Betts ■

Jean Louise M. Smith, PhD, NCSP, Research Assistant Professor and Associate Director at the Center on Teaching and Learning at the University of Oregon. Her research interests are in the area of reading intervention and assessment and the prevention of academic difficulty.

Kelli D. Cummings, PhD, NCSP, is a Faculty Research Associate at the Center on Teaching and Learning at the University of Oregon. Her research focuses on projects that link assessment and intervention technologies to improve student success. She has provided technical assistance and training on problem solving, response to intervention, and the use of curriculum-based measurement for data-based decision making in schools across the United States, Canada, and Great Britain.

Joseph F. T. Nese, PhD, is a Research Associate with Behavioral Research and Teaching at the University of Oregon. He conducts research in educational assessment and applied measurement, focusing on developing and improving systems that support data-based decision making, as well as using advanced statistical methods to measure and monitor student growth.

Julie Alonzo, PhD, is a Research Associate at the University of Oregon, where she co-directs the Behavioral Research and Teaching research institute. Her primary research interests include teacher professional development, response to intervention, and the meaningful inclusion of students with diverse learning needs.

Hank Fien, PhD, is a Research Associate Professor and Co-Director at the Center on Teaching and Learning at the University of Oregon. His research interests include developing and rigorously testing tiered mathematics and literacy instructional approaches in the early grades.

Scott K. Baker, PhD, is the Executive Director of the Center on Research and Evaluation at Southern Methodist University. He has been the principal investigator on a number of grants from the Institute of Education Sciences. Dr. Baker's current interests focus on the impact of interventions on child outcomes. He is interested in the mechanisms that underlie effective interventions, the long-term impact of evidence-based interventions on a variety of childhood indicators of adjustment and well-being, and how intervention impact varies by factors intrinsic and extrinsic to the child. Dr. Baker is also interested in understanding the impact of programs, policies, and practices delivered in various combinations over extended periods of time.