Reading Psychology, 26:285–312, 2005

ISSN: 0270-2711 print / 1521-0685 online
DOI: 10.1080/02702710590967834

INSTABILITY OF DOUBLE-DEFICIT SUBTYPES AMONG AT-RISK FIRST GRADE READERS

JANET E. SPECTOR
University of Maine, Orono, Maine, USA

The double-deficit hypothesis provides a framework for identifying students at-risk for persistent reading difficulties. I examined the temporal stability of four double-deficit subtypes (no-deficit, naming-speed deficit, phonological deficit, and double-deficit) in 197 low-performing, first-grade readers. Concurrent analyses in fall and spring revealed that students with poor phonemic awareness and naming speed were the lowest readers in the sample; however, having a double-deficit in fall was not a good predictor of subsequent progress, most likely because subtypes were not stable over the first half of grade 1. Only about half the sample retained the same subtype from fall to spring.

Over the years, there have been numerous attempts to identify reading difficulties subtypes, or groups of poor readers with similar profiles (e.g., Castles & Coltheart, 1993; Cromer, 1970; Fletcher et al., 1994; Manis, Seidenberg, Stallings, et al., 1999; Morris et al., 1998; Stanovich, Siegel & Gottardo, 1997; Wolf & Bowers, 1999). Together, these efforts have increased our understanding of the diversity of factors that are associated with persistent reading difficulties. Ultimately, subtypes may be useful in identifying students for early intervention (Speece & Cooper, 2004). Although single variables have not been effective in screening for future reading difficulties, profiles based on a combination of variables may have greater predictive validity (Scarborough, 2001).

The double-deficit hypothesis of reading disabilities (Bowers & Wolf, 1993; Wolf & Bowers, 1999) offers a conceptual framework that may be effective in identifying children who are at-risk for persistent reading difficulties. The variables that compose the framework—phonemic awareness and naming speed—are two of the best predictors of reading (Wolf, 2003). Furthermore, they can

Address correspondence to Janet E. Spector, 5766 Shibler Hall, College of Education and Human Development, University of Maine, Orono, ME 04469-5766. E-mail: janet.spector@umit.maine.edu

both be assessed in children who do not have formal reading skills. If a particular pattern of performance associated with later reading difficulties can be identified before students have fallen years behind average achieving peers, then instruction can be appropriately aimed to prevent the development of a more intractable reading problem.

A major tenet of the double-deficit hypothesis is that phonemic awareness and naming speed have independent and additive effects on reading (Wolf, Bowers, & Biddle, 2000). The independence of the two types of measures receives support from two lines of research. First, many studies report low or zero-order correlations between naming speed and phonemic awareness (see review in Swanson, Trainin, Necoechea, & Hammill, 2003; cf. Cardoso-Martins & Pennington, 2004; Schatschneider, Carlson, Francis, Foorman, & Fletcher, 2002). That is, how well students do on phonemic awareness tasks is not closely related to how well they do on tasks that require them to name objects, colors, letters, or digits.

Second, phonemic awareness and naming speed tend to correlate with different dimensions of reading. While phonemic awareness correlates more highly with tasks such as pseudoword decoding, naming speed correlates more highly with tasks that require reading fluency (Lovett, Steinbach, & Frijters, 2000). Indeed, a recent meta-analysis of research on the relationship between naming speed and phonemic awareness confirmed that measures of the two load on different factors (Swanson et al., 2003). Specifically, phonemic awareness loaded most highly on a factor that represented ability to read pseudowords and naming speed loaded most highly on a factor that represented reading comprehension. Other investigators have identified the unique contribution of naming speed to prediction of performance on tasks that tap orthographic rather than phonological processing (Bowers, Sunseth & Golden, 1999; Cardoso-Martins & Pennington, 2004; Manis, Doi, & Badha, 2000; Manis, Seidenberg, & Doi, 1999; Wolf & Bowers, 1999; Wolf et al., 2002).

Evidence in support of additivity comes primarily from studies that sort students by subtype based on their performance on phonemic awareness and naming speed and then examine differences between subtypes in reading. If effects of phonemic awareness and naming speed are additive, then students with weak phonemic awareness or slow naming speed will be more impaired in reading than students with adequate abilities on both, and students who are low performers on measures of both phonemic awareness and naming speed will be more impaired than those with a single phonological or naming-speed deficit.

To date, several studies have confirmed predictions of the double-deficit hypothesis, either in part or in full (e.g., Badian, 1997; Bowers et al., 1999; Lovett et al., 2000; Manis et al., 2000; Sunseth & Bowers, 2002; cf. Ackerman, Holloway, Youngdahl, & Dykman, 2001). Application to the practical task of early identification, however, remains largely unexplored. First, most double-deficit studies have examined readers with well-established histories of reading difficulties, most often in grades 3 and above (cf. Kirby, Pfeiffer, & Parilla, 2003). Second, the temporal stability of profiles based on PA and RAN has yet to be established. Although a number of studies have investigated the stability over time of PA and RAN using correlational techniques (e.g., Scarborough, 1998b), these studies have not reported degree of consistency at the level of the individual student (i.e., percentage of students who retained their rank over time) nor have they shed light on consistency of profiles based on the two variables combined.

Historically, efforts to identify reading disabilities subtypes have been hampered by the instability of student profiles from one point in time to the next (Kavale, 2003). That is, students who demonstrate a particular pattern of strengths and weaknesses at one time may not demonstrate the same pattern of performance when tested again (Manis, Seidenberg, Stallings, et al., 1999; White & Wigle, 1986). Although a number of studies have investigated the double-deficit hypothesis using longitudinal designs, I was unable to identify any that examined whether students who are initially identified with a particular pattern of performance on phonemic awareness and naming speed demonstrate the same pattern upon retest. In studies that have been longitudinal, subtyping has either been done at pretest or posttest, not at both points in time (e.g., Kirby et al., 2003; Lovett et al., 2000). Instability may be a particularly significant problem in identifying subtypes during the beginning stages of reading acquisition because performance on all three variables—naming speed, phonemic awareness, and reading—changes rapidly in first grade. The purpose of the present study was to investigate the temporal stability of subtypes generated by the double-deficit hypothesis in a sample of low performing, beginning readers who were school-identified as at-risk for future reading difficulties.

Method

Participants

The sample for the present study was drawn from a prior investigation of progress in Reading Recovery, a short-term reading intervention that provides one-to-one tutoring for a maximum of around 20 weeks (Spector & Moore, 2001). As part of that study, a volunteer sample of 27 Reading Recovery teachers randomly selected two to four students from their caseloads ($n = 106$) and two to four students who were eligible for Reading Recovery but were on the waiting list for later service ($n = 90$). All were first graders who had been school-identified as the lowest performing readers in their class based on classroom performance and statewide norms on Clay's (1993) Observation Survey, a battery of six early literacy measures. Table 1 depicts sample means and standard deviations on two informal inventories of alphabetic knowledge (letter name

TABLE 1 Means and Standard Deviations on Test Variables in Fall and Spring

	Fall[a]		Spring[b]	
Measure	*M*	*SD*	*M*	*SD*
Phonemic awareness				
Items (% correct)	16.71	18.31	63.77	23.09
Segments (% correct)	29.60	23.58	75.20	18.11
Naming speed				
Raw score (in seconds)	219.70	66.58	185.55	54.77
Standard score ($M = 100$, $SD = 15$)	87.00	15.99	92.59	16.86
Letter sound knowledge (% correct)	44.90	24.92	85.40	14.19
Letter name knowledge (% correct)	76.39	19.29	96.62	5.69
Hearing sounds in words (% correct)	28.49	22.81	88.83	12.49
Text level	.91	.88	9.39	4.84

[a] $n = 197$ for all fall measures except naming speed ($n = 196$). [b] $n = 197$ for all spring measures except letter-sound knowledge ($n = 190$), letter-name knowledge ($n = 196$) and naming speed ($n = 196$).

and letter-sound knowledge) and two Observation Survey measures, Text Level and Hearing and Recording Sounds.

As can be seen, beginning first graders in the sample were not proficient in letter naming. In addition, the majority scored at floor on Text Level, a measure of oral reading accuracy. More specifically, 79% scored at or below the lowest text level and 93% scored below what is commonly identified as an easy preprimer level. From a normative perspective, students in the present sample scored .67 *SD*s below the level of a nationally-constituted, random sample of over 35,000 beginning first graders on Text Level and 1.67 *SD*s below the random sample on Hearing and Recording Sounds (Gomez-Bellengé, Rodgers, & Fullerton, 2003). The more modest effect size associated with Text Level reflects restricted variability on this measure due to a floor effect, both in the current sample and in the random sample.

The mean age of students in the fall was 78.07 months ($SD =$ 4.89) and the sample comprised almost twice as many males as females (65% male, 35% female). All were native English speakers and, consistent with the demographics of the state in which the study was conducted, over 90% of participants were white.

Information regarding socioeconomic status came from lunch-cost statistics, available for about half of the sample. Among students for whom these data were available, 46% received free or reduced-cost lunch, while 54% paid the regular price for lunch. Approximately two thirds of the 31 schools represented in the sample were located in towns with populations less than 10,000 ($n =$ 20). All but one of the remaining schools were in towns with populations ranging from 10,100 to 49,999. Classrooms of participating students varied unsystematically in methods of teaching reading, incorporating the eclectic mix of instructional practices found in public schools in the state (i.e., guided reading, phonics, basal readers).

As I mentioned above, all participants were school-identified as the lowest performing readers in their class. No additional tests of intelligence or verbal ability were administered; however, none of the participants had been identified with cognitive, physical, or sensory disabilities, and teacher reports indicated that none had obvious cognitive, sensory, or physical limitations. A small number ($n = 11$) had received speech and language services prior to grade 1, most typically once or twice a week from a speech and language therapist. I conducted all subsequent analyses with and without these students with no change in results. I elected, therefore, to retain these students in all analyses reported below.

Procedures

All students had been screened in their schools at the beginning of first grade using the Observation Survey. Posttesting occurred about 20 weeks later ($M = 20.37$ weeks, $SD = 2.30$) when early intervention participants had completed a full set of Reading Recovery lessons. Although wait-listed students were retested at the same time as early intervention participants, they were evaluated on only two Observation Survey measures: Text Level and Hearing and Recording Sounds.

Testers also administered measures of letter-name and letter-sound knowledge, phonemic awareness, and naming speed to all students. Testing occurred in the fall (September or early October) and near the beginning of the spring semester (January or February) to coincide with administration of Observation Survey tests.

Measures

PHONEMIC AWARENESS (PA)

Studies of the double-deficit hypothesis have used a wide range of tasks to identify students with weak phonological skills. Most have focused exclusively on phonemic awareness (e.g., Ackerman et al., 2001; Bowers et al., 1999; Schatschneider et al., 2002; Sunseth & Bowers, 2002; Wimmer, Mayringer, & Landerl, 2000), some have relied on phonologically-based reading tasks such as pseudoword decoding (e.g., Badian, 1997; Wolf, 1997), and others have included a combination of the two (e.g., Lovett et al., 2000; Manis, Seidenberg, Stallings, et al., 1999). Present analyses drew on the Yopp-Singer Test of Phoneme Segmentation (Yopp, 1988), an aural measure of PA, to avoid confounding with print skills. The test has high internal consistency ($\alpha = .95$) and correlates positively through Grade 6 with a wide range of literacy measures, including word attack, vocabulary, comprehension, and spelling (Yopp, 1995).

On the Yopp-Singer, examiners asked students to pronounce, in order, each of the sounds in a spoken word. Sounds, not letter names, were the appropriate response. Corrective feedback was given on trial items and after any incorrect responses throughout the 22 item test. Using Yopp's approach, I scored the test for number of words that were segmented correctly. Roughly a third of the sample, however, scored zero at pretest. To alleviate the floor effect, I created a partial credit measure, awarding one point for each sound isolated. For example, on the word *zoo*, the response/ zu/–/u/ received one point (for the phoneme/u/). The maximum score possible using partial credit scoring was 56. Table 1 includes samples means and standard deviations for both conventional scoring (i.e., number of items correct) and partial credit scoring (i.e., number of segments correct).

In addition to the Yopp-Singer, I examined scores on two print-related measures of phonological skills: an informal inventory of letter-sound knowledge and Hearing and Recording Sounds from the Observation Survey. I used these measures to cross-validate judgments regarding strengths and weaknesses in the phonological domain. The informal inventory of letter-sound knowledge was a modified version of the Letter Identification subtest from Clay's (1993) Observation Survey. Specifically, testers asked students to provide a sound for each of the 54 alphabetic characters, the upper and lower case standard letters as well as the print form of *a* and *g*. Scores indicate the percentage of sounds identified correctly.

Hearing and Recording Sounds, a measure from Clay's Observation Survey (1993), taps PA and alphabetic knowledge. On this test, examiners read a short sentence or two and asked children to write the words that they heard. Credit was awarded based on accuracy in representing phonemes rather than on conventional spelling. The measure has high internal consistency ($\alpha = .96$) and correlates positively with word recognition in first grade ($r = .79$; Clay, 1993). Scores indicate the percentage of sounds out of 37 represented appropriately.

RAPID AUTOMATIZED NAMING (RAN)

Naming speed tasks require students to label as quickly as possible highly familiar stimuli in randomly ordered arrays. Most commonly, tasks present colors, objects, letters, and/or digits, either in blocks of a single stimulus type (e.g., colors only, digits only) or mixed blocks. At the beginning of first grade, participants had yet to master letter names. Teachers also reported that digits remained problematic for many children in the sample. I opted, therefore, to assess naming speed using a two-part test taken from the Comprehensive Test of Phonological Processes (CTOPP; Wagner, Torgesen, & Rashotte, 1999): Rapid Automatized Naming of Colors and Objects (RAN). First, children named colored squares arranged in rows while the examiner timed them. After that, they named a set of common objects (e.g., key, fish, boat) arranged in rows. According to the test manual, the measure has adequate internal consistency at the ages at which students were tested ($\alpha = .89$ and .87, for six and seven year olds, respectively), acceptable test-retest reliability ($r = .70$) and strong interscorer reliability ($r = .99$). Analyses were conducted using age-based standard scores ($M = 100$, $SD = 15$; Wagner et al., 1999).

READING

Text Level, a measure from Clay's (1993) Observation Survey, assesses oral reading accuracy. On this test, children read aloud from a series of increasingly more difficult texts, drawn from established basal systems (Askew, Fountas, Lyons, Pinnell, & Schmitt, 1998). Texts used for testing were not used in instruction, either in the classroom or in the intervention program. Testers provided a minimal, scripted introduction and recorded word recognition accuracy. Previous reliability studies utilizing Rasch scaling methods yielded a person separation reliability of .83 and an item separation reliability of .93 (Pinnell, Lyons, DeFord, Bryk, & Seltzer, 1994). Scores indicate the highest text level at which students read aloud with at least 90% word-recognition accuracy.

Because so many students scored at floor on text level in the fall, I included a letter-naming test to assess print exposure. I used this measure rather than fall text level as a covariate in analyses of spring reading to control for pre-existing differences in print knowledge. Examiners asked students to name 54 characters, the upper and lower case standard letters as well as the print form of *a* and *g*. Scores indicate the percentage of names produced correctly.

Results

Analyses investigated the temporal stability of double-deficit subtypes in a sample of low performing, first-grade readers. First, I sorted students by subtype based on performance on PA and RAN, once using fall scores and once using spring scores. I then compared subtypes on PA, RAN, and reading. The purpose of these analyses was to investigate whether subtypes based on PA and RAN differed in ways that were predicted by the double-deficit hypothesis. That is, were PA and RAN independent and were low performing beginning readers with a double deficit more impaired in reading than other low performing readers? Next, to estimate the stability of subtypes, I examined the overlap in subtype membership from fall to spring. In doing so, I also investigated possible differences in stability between two groups within the larger sample: students who participated in early intervention during the first half of first grade and those who were wait-listed for service during the same time period. The purpose of this analysis was to assess generalizability across two instructional contexts. Alpha was set at .05 for all tests of statistical significance.

Definition of Double-Deficit Subtypes

Any definition of deficient performance based on a continuous measurement scale is arbitrary. Some double-deficit studies have relied on standard scores on nationally-normed tests to identify students who are below-average (e.g., Lovett et al., 2000), while others have applied standards gleaned from previous research (e.g., Ackerman et al., 2001). Yet another method is within-sample standing, classifying students as deficient based on rank within a particular sample (e.g., Sunseth & Bowers, 2002; Wimmer et al., 2000).

In the present study, I used a combination of approaches to divide students into groups based on their scores on RAN and PA. RAN was assessed using a nationally-normed test (i.e., CTOPP), so I applied a cut-off used in many other studies of the double-deficit hypothesis. Specifically, I considered students to be low performers if they scored more than 1 *SD* below the norm group mean on a composite measure of object and color naming. I used the same standard score criterion in fall and spring although raw scores associated with the standard score cut-off changed due to the developmental nature of performance on RAN (see Table 1).

To identify students with low PA, I used the Yopp-Singer Test of Phoneme Segmentation (Yopp, 1988, 1985). In setting cut-offs, I considered normative data as well as developmental benchmarks.

Although the Yopp-Singer does not provide norms, I examined previous studies that have reported scores on the Yopp-Singer in samples selected to be representative of the range of abilities in a general education classroom. Beginning first graders in the present study who segmented fewer than four items correct on the Yopp-Singer scored more than 1 *SD* below the mean for end-of-year kindergarteners in two independent studies (Yopp, 1988; Spector, 1992). I identified these students as low on the task and those who segmented four or more items correctly as not low.

Next, I cross-validated this cut-off using partial-credit scoring (i.e., number of segments correct). Syntheses of research on PA have concluded that by the end of kindergarten, most students can isolate and pronounce the beginning sound of a word, and by mid-first grade, they can isolate and pronounce all the sounds in two- and three-phoneme words (Snow, Burns, & Griffin, 1998; Torgesen & Mathes, 2000). I considered students who failed to identify at least one sound on at least 75% of the Yopp-Singer items to fall short of the end of kindergarten standard and all others to meet the standard. Examination of the joint distribution of scores confirmed that students who were identified as low on traditional scoring (i.e., items correct) were also identified as low on partial credit scoring.

In the spring, I relied again on the Yopp-Singer to identify students who were low on PA. I was unable to locate any studies that reported mid-first-grade means on the Yopp-Singer in heterogeneous samples. As I mentioned above, however, syntheses of research on PA have concluded that by the middle of grade 1, most students can segment two and three phoneme words (Torgesen & Mathes, 2000). I classified students who segmented less than 75% of the 22 two- and three-phoneme items on the test as low and those who segmented more than 75% of the items as not low. Although arbitrary, this distinguished students who demonstrated mastery of the task from those with emergent skills.

In presenting results, I identify students with low performance on just the Yopp-Singer as having a single phonological deficit (PD); those with low performance on just RAN as having a naming speed deficit (NSD); those with low performance on both RAN and the Yopp-Singer as having a double deficit (DD); and those with low performance on neither RAN nor the Yopp-Singer as having no deficit (ND). To distinguish subtypes by point in time, I add the letter F for beginning of fall semester and S for beginning of spring semester (e.g., DD-F vs. DD-S).

Independence of PA and RAN

In a recent study of the relationship between phonological awareness and naming speed, Schatschneider et al. (2002) identified a potential threat to the validity of studies on the double-deficit hypothesis when there is a positive correlation between PA and RAN. If PA and RAN are positively correlated, then DD may have poorer phonemic awareness than PD. As a result, differences between the two subtypes could be the result of degree of phonological impairment, rather than a double-deficit. Following a similar logic, DD may have weaker naming speed than NSD. If this is the case, then differences between the two subtypes could be the result of quality of performance on naming speed, rather than a double deficit. To rule out each of these possibilities, I examined correlations between PA and NS.

RAN did not correlate with performance on the Yopp-Singer in fall, ($r = .03$, $p = .69$) or spring ($r = .09$, $p = .21$). In addition, t tests indicated that single- and double-deficit subtypes did not differ on PA or RAN in fall or spring. The two groups with a phonological deficit (PD and DD) did not differ on the Yopp-Singer in fall, $t(107) = .43$, $p = .67$, or spring, $t(75) = .30$, $p = 77$. In addition, results indicated no significant differences between NSD and DD in either fall, $t(75) = 1.49$, $p = .14$, or spring, $t(62) = 1.45$, $p = .15$. Together, these results indicated that in the present sample, number of deficits was not confounded with level of impairment on PA or RAN.

Subtype Performance on PA, RAN, and Reading

As can be seen in Table 2, the majority of participants (73%) fell into one of the three deficit categories in both fall and spring, NSD ($F = 17\%$; $S = 12\%$), PD ($F = 33\%$, $S = 40\%$), DD ($F = 23\%$, $S = 21\%$). Table 2 also depicts observed subtype means and standard deviations on RAN and PA and several early literacy measures. To determine whether subtypes differed in ways predicted by the double-deficit hypothesis, I compared subtypes on RAN, PA, and reading. The purpose of comparisons on RAN and PA was to

TABLE 2 Observed Means and Standard Deviations on Test Variables in Fall and Spring by Double-Deficit Subtype

	Fall subtype				Spring subtype			
Measure	ND $n = 54$	NSD $n = 33$	PD $n = 64$	DD $n = 45$	ND $n = 53$	NSD $n = 23$	PD $n = 79$	DD $n = 42$
Phonemic Awareness								
Items correct								
M	32.15	35.40	3.69	3.33	86.88	84.58	50.11	48.92
SD	15.49	14.50	4.98	4.98	7.85	6.53	16.79	19.85
Segments correct								
M	49.27	51.57	13.89	12.86	91.24	89.75	66.82	62.75
SD	17.66	17.69	12.35	12.36	6.11	4.70	14.85	18.52
Naming speed								
M	97.85	73.18	97.00	70.18	101.58	76.00	101.90	72.34
SD	10.47	7.65	9.88	9.96	9.78	8.24	11.52	10.39
Letter sound knowledge (% correct)								
M	52.92	49.11	41.98	37.27	91.58	92.12	80.45	83.31
SD	22.81	25.50	25.64	22.88	8.68	4.37	17.31	12.66
Letter name knowledge (% correct)								
M	81.24	79.07	75.35	70.95	98.32	97.75	96.23	94.53
SD	16.83	20.31	19.71	18.60	2.19	2.08	7.36	5.84
Observation survey								
Hearing and Recording Sounds in Words (% correct)								
M	37.79	37.02	22.42	20.36	94.34	95.18	87.10	81.66
SD	24.64	24.32	19.47	17.69	5.71	6.15	12.21	16.73
Tqext Level								
M	1.24	.92	.83	.64	11.64	11.17	8.94	6.40
SD	1.04	.99	.73	.66	4.20	5.31	4.62	4.00

Note. Fall subtypes were based on fall phonemic awareness and naming speed scores. Spring subtypes were based on spring phonemic awareness and naming speed scores. DD = double-deficit; ND = no deficit; NSD = naming speed deficit; PD = phonological deficit.

confirm that the cut-offs described above resulted in groups that were significantly different on RAN and PA. The objective of comparisons on reading was to determine whether students with a double deficit in fall or spring were more impaired in reading than other low-performing readers without a double deficit.

NAMING SPEED

As can be seen, at both points in time, NSD and DD had mean standard scores on RAN that approached 2 *SD*s below the mean of same age peers in the national norm group. In contrast, ND and PD had mean scores on RAN in the average range. *T* tests confirmed the significance of differences on RAN between students with and without a naming speed deficit, $t(193) = -18.50$, $p < .001$.

PHONEMIC AWARENESS

Fall and spring scores on PA revealed the degree of separation between students with and without a phonological deficit. As can be seen in Table 2, students with a phonological deficit in fall (PD-F and DD-F) were generally unable to segment items on the Yopp-Singer. In contrast, students without a phonological deficit (ND-F and NSD-F) demonstrated emergent skill, averaging about a third of the items correct. *T* tests confirmed that the two groups differed on percentage of items correct, $t(101.26) = -18.17$, $p < .001$, as well as on percentage of segments correct, $t(148.90) = -16.69$, $p < .001$.

Furthermore, low scorers on the Yopp-Singer also tended to be low scorers on Hearing and Recording Sounds and letter-sound knowledge, two measures that tap print-related phonological skills (see Table 2). *T* tests confirmed the significance of the difference between students with relatively strong versus weak segmentation skills on Hearing and Recording Sounds, $t(157.87) = -5.09$, $p < .001$, and letter-sound knowledge, $t(195) = -3.38$, $p < .001$. This finding cross-validates judgments regarding the weak phonological skills of PD-F and DD-F.

At the beginning of the spring semester, students who I identified with a phonological deficit (PD-S and DD-S) segmented correctly only about half of the items on the Yopp-Singer. In contrast, students without a phonological deficit averaged over 85% accuracy. *T* tests also confirmed that students identified as low based on spring scores on the Yopp-Singer scored lower than those identified as not low on two spring measures of print-related phonological skills: Hearing and Recording Sounds, $t(173.43) = -6.48$, $p < .001$, and letter-sound knowledge, $t(178.41) = 6.00$, $p < .001$.

READING

As can be seen in Table 2, the direction of differences between subtypes on text level was generally consistent with predictions of the double-deficit hypothesis. Descriptively, students with a double deficit were the lowest performing readers both in fall and spring. I did not test the statistical significance of differences between subtypes in fall due to a floor effect on text level; however, students with a double deficit were over-represented in the group of students who were targeted by their teachers for intervention in the fall, $\chi^2(3, N = 196) = 14.56$, $p = .002$. This finding confirms their low standing relative to other poor readers in the sample. Specifically, almost 80% of DD-F was assigned to immediate intervention compared to about 50% for each of the other three subtypes. When all of the low-performing students in a first-grade cohort can not be assigned to one-to-one tutoring immediately, it is Reading Recovery policy to serve the lowest-performing first graders first and wait list remaining students who meet eligibility criteria. Decisions about which students will receive priority are based on teacher recommendations and performance on the Observation Survey. Selection decisions were made before the study began, so teacher judgments were not influenced by measures that I used to define a double-deficit.

In the spring, results of analyses of covariance (ANCOVA) confirmed the statistical significance of the effect of spring subtype, $F(3, 192) = 11.54$, $p < .001$, $\eta^2 = .16$, on spring text level, after adjusting for differences on the covariate, fall letter-name knowledge, $F(1, 192) = 8.18$, $p = .005$, $\eta^2 = .05$. Post-hoc tests using Bonferroni adjustment for multiple comparisons confirmed that DD-S (adjusted $M = 6.43$, $SD = 4.41$) scored lower on text level than each of the other three subtypes: ND-S, adjusted $M = 11.51$, $SD = 4.38$, $d = -1.05$; NSD-S, adjusted $M = 11.21$, $SD = 4.42$, $d = -.99$; PD-S adjusted $M = 9.0$, $SD = 4.40$, $d = .53$). PD-S scored lower than ND-S ($d = -.52$) but was not significantly different from NSD-S. NSD-S was not significantly more impaired in reading than ND-S, a finding that is consistent with previous research indicating that students with a single deficit in naming speed are more selectively impaired and so may not score lower than the no-deficit subtype on accuracy-based measures of word recognition (Lovett et al., 2000). Indeed, some students with a single naming speed deficit may not be identified by their teachers as poor readers until grade 3 or later when problems with fluency interfere with reading comprehension (Wolf & Katzir-Cohen, 2001).

Finally, I tested for differences in spring text level by fall subtype. That is, do students with a double deficit in the fall continue to stand out as the most impaired readers in the spring? Overall, ANCOVA results indicated a small but significant effect of fall subtype on spring text level, $F(3, 191) = 3.68$, $p = .01$, $\eta^2 = .06$, after adjusting for the covariate, fall letter-name knowledge, $F(1, 191) = 5.12$, $p = .02$, $\eta^2 = .03$. At the same time, however, post-hoc tests using Bonferroni adjustment for multiple comparisons failed to detect significant differences between DD-F (adjusted $M = 8.93$, $SD = 4.70$) and the other three subtypes: ND-F (adjusted $M = 10.96$, $SD = 4.70$), NSD-F (adjusted $M = 9.93$, $SD = 4.64$), and PD-F (adjusted $M = 8.17$, $SD = 4.64$). Contrary to predictions of the double-deficit hypothesis, the only two subtypes that differed were PD-F and ND-F ($d = -.58$). Implications of this finding are discussed in a later section.

Stability of Double-Deficit Subtypes

Analyses of the concurrent relationship between double-deficit subtypes and reading indicated that students with a double deficit were the lowest-performing readers at each point in time; however, predictive analyses revealed that students with a double deficit in the fall were not the lowest-performing readers in the spring. A likely explanation for poor predictive validity was that fall estimates of PA and RAN were not accurate measures of student standing on spring PA and RAN. To test this hypothesis, I examined the stability of subtypes across the two points in time.

First, I examined the correlations of PA with itself and of RAN with itself from fall to spring. For the Yopp-Singer, I conducted analyses using scores based on both partial credit (segments correct) and traditional scoring (items correct) due to a floor effect on traditional scoring in fall. The correlation between fall and spring scores was small, but significant, $r = .21$, $p < .00$, both for segments correct and items correct. The correlation approximated the coefficient that I obtained in an earlier study of growth in PA from mid to end of kindergarten (i.e., $r = .30$); however, it was lower than stability estimates that have been reported for measures of phonological processing that have not been limited to phonemic awareness (e.g., Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993) and for composite measures of PA (e.g., Cardoso-Martins & Pennington, 2004). For RAN, the correlation ($r = .76$) was consistent with the test–retest reliability coefficient reported in the test manual ($r = .70$) and more stable than coefficients reported by Cardoso-Martins and Pennington (2004) in a study of the relationship between RAN and PA that, like the present study, included first graders ($r = .62$ for RAN for colors and objects; $r = .50$ for RAN for letters and digits). Correlations, however, do not indicate the percentage of individuals who retained the same profile on a combination of PA and RAN from fall to spring. I turn, therefore, to analyses of the distribution of subtypes in fall and spring.

TOTAL SAMPLE

The distribution of subtypes in fall and spring is depicted in Table 3. About 80% of the students who presented with a deficit profile in fall (i.e., NSD-F, PD-F, or DD-F) also presented with a deficit profile in spring. Fewer than half, however, demonstrated the same profile at both points in time. Descriptively, NSD-F was the least stable. Only 30% of NSD-F was classified the same in spring. PD-F was the most stable; almost 2/3 of the students with this profile demonstrated the same pattern in spring.

I also examined the overall stability of phonological and naming speed deficits by collapsing across the two groups with a phonological deficit (PD and DD) and the two groups with a naming speed deficit (NSD and DD), respectively. Overall, the stability for phonological deficits was 61%. Over 70% of students with a phonological deficit in the fall (PD-F and DD-F) continued to score low on PA in spring, and about 50% of students who were not low in fall continued to demonstrate adequate PA in spring. The overall stability for naming speed deficits (NSD and DD) was 74%. Almost 90% who began with adequate scores on RAN (ND-F and PD-F) remained in the average range in spring; however, 43% of the students who began with low scores on RAN did not score low

TABLE 3 Correspondence Between Double-Deficit Subtypes in Fall and Spring Showing Number of Students with Same Versus Different Profiles over Time

	Spring subtypes														
	Total sample					Intervention participants					Wait-listed students				
Fall subtypes	ND	NSD	PD	DD	Total	ND	NSD	PD	DD	Total	ND	NSD	PD	DD	Total
ND	*26*	5	20	3	54	*15*	3	3	1	22	*11*	2	17	2	32
NSD	3	*10*	7	13	33	2	*7*	3	5	17	1	*3*	4	8	16
PD	14	1	*42*	7	64	5	0	*23*	4	32	9	1	*19*	3	32
DD	10	7	10	*18*	45	9	5	7	*14*	35	1	2	3	*4*	10
Total	53	23	79	41	196	31	15	36	24	106	22	8	43	17	90

Note. Italicized values represent students with the same subtype in both fall and spring. DD = double-deficit; ND = no deficit; NSD = naming speed deficit; PD = phonological deficit.

in spring. Interestingly, shifts between single deficit profiles were infrequent; only seven students changed from NSD-F to PD-S and only one student changed from PD-F to NSD-S.

STABILITY FOR INTERVENTION PARTICIPANTS VERSUS WAIT-LISTED STUDENTS

To what extent was subtype instability a function of participation in early intervention? To address this question, I computed the percentage of early intervention participants and wait-listed students who were classified the same in both fall and spring. Overall, results of chi-square analyses indicated that the stability rate did not vary by group, $\chi^2(1, N = 197) = 2.08$, $p = .15$. Among early intervention participants, 56% were classified the same at both times and among students who were waited-listed for services, 41% were classified the same at both times (also see Table 3). This finding indicates that the magnitude of instability was not accounted for by participation in early intervention.

Discussion

Previous research has yielded some support for the double-deficit hypothesis of reading disabilities; however, practical application to the task of early identification has yet to be fully explored. One issue that is particularly important for purposes of early identification is temporal stability of profiles based on phonemic awareness and naming speed. In the present study, I investigated continuity and changes in double-deficit subtypes within a school-identified sample of low-performing, first-grade readers.

Two key findings contribute to the literature on early identification and the double-deficit hypothesis. First, consistent with the double-deficit hypothesis, concurrent analyses revealed that first graders with weaknesses on both PA and RAN were the weakest readers in the sample. Second, although students with a double deficit in fall were more impaired in reading in the fall and those with a double-deficit in spring were more impaired in reading in spring, subtypes were not stable over the first half of first grade. In the sections below I discuss these findings. In doing so, I consider limitations of the present study and identify questions for future research.

Support for the Double-Deficit Hypothesis

The present study extends the generalizability of previous tests of the double-deficit hypothesis. Most previous double-deficit studies have targeted students with well-established reading disabilities, most typically in grades 3 or above. The present study suggests that significant effects of a double deficit can be identified by mid first grade on conventional measures of reading such as oral reading of connected text. Before that, however, it may be difficult to discern between-subtype differences in skill because so many students score at floor on formal tests of reading.

At the same time, although test scores were insensitive to differences among the lowest-performing beginning first-grade readers, school-based judgments regarding degree of need for intervention were sensitive to qualitative differences between subtypes, even at the beginning of grade 1. Students with a double deficit at pretest were significantly more likely than those with a single phonological or naming-speed deficit to be targeted by their teachers for immediate intervention. When these decisions were made, measures used in the study to identify double-deficit subtypes had not yet been administered, so it is unlikely that teacher judgments were influenced by the double-deficit framework. Unfortunately, I did not have access to scores of wait-listed students on all Observation Survey measures or on other standardized measures of reading to explore more fully factors that contributed to these judgments.

It remains possible that I might have found more powerful effects of RAN had I been able to assess speed of naming letters or digits rather than colors and objects. Unfortunately, many students in the present study were not yet proficient in naming letters and numbers at the beginning of first grade so they could not be assessed using alphanumeric stimuli. Previous research on RAN has yielded less consistent results regarding the predictive strength of naming speed for nonsymbolic stimuli, like colors and objects, than has been the case for symbolic stimuli like letters and digits. On the one hand, some studies, particularly those that have targeted the poorest readers and those in preschool or kindergarten, have documented significant correlations between RAN for objects or colors and future reading (Cornwall, 1992; Fawcett & Nicholson, 1994; Korhonen, 1995;

Meyer, Wood, Hart, & Felton, 1998a, 1998b; Scarborough, 1998a, 1998b).

On the other hand, a number of investigators have concluded that naming speed for non-symbolic stimuli is an inconsistent or weak predictor of reading (Badian, 1998; Cardoso-Martins & Pennington, 2004; Manis et al., 2000; Wimmer, 1993; Wolf, 1991; Wolf, Bally, & Morris, 1986). Indeed, recent research using functional magnetic resonance imaging suggests that among adults without a history a reading difficulties, different neurological substrates appear to underlie performance on RAN for symbolic and nonsymbolic stimuli (Misra, Katzir, Wolf, & Poldrack, 2004). Clearly, the results reported in the present paper are most relevant to identification of double-deficit subtypes among students who are at the earliest stages of reading acquisition and who can not be assessed on RAN for letters and numbers.

Temporal Stability of Double-Deficit Subtypes

As I mentioned above, concurrent analyses in fall and spring revealed that students with a double-deficit were the lowest performing readers on a least one indicator of reading skill. At the same time, however, predictive analyses indicated that students with a double-deficit in the fall were not the lowest performing readers in the spring. The most likely explanation for this finding is that profiles on PA and RAN were not stable over a 20 week period from the beginning to the middle of first grade. In other words, the group of students with a double deficit in the fall was not the same group of students with a double deficit in the spring. To the contrary, the overlap in membership was only about 40%.

Changes in subtype reflected improvement in standing on PA and/or RAN for some students and decline in status for others. Stability rate, however, did not vary across two sets of students who were school-identified as eligible for early intervention: those who participated in immediate intervention and those who were wait listed for services during the 20 weeks between fall and spring testing. This finding suggests that the magnitude of subtype instability was not accounted for by participation in the early intervention program. That is, instability was just as likely in early intervention participants as it was in students in a first grade classroom reading program. Below I discuss possible explanations for subtype instability.

Explanations for Subtype Instability

A variety of factors may contribute to subtype instability. First, cut-off scores like those used in the present study are arbitrary, inviting misclassification of students with scores just above or below the cut-off. When a continuous variable is dichotomized, students who score close to the cut-off are particularly vulnerable to changes in status due to test imprecision rather than true change on the ability that is tested. For example, among the 52 students who differed in naming speed status at the two points in time, the majority had fall or spring scores that were within 1 SEM of the cut-off score for a naming speed deficit based on test-retest reliability statistics reported in the test manual (i.e., standard score = 85, $+/-8$). Although some shifts could have been due to true change in proficiency, some, no doubt, were due to measurement error. Similarly, because subtypes were based on low standing on each variable, regression to the mean also may have contributed to subtype shifts. Students who were the lowest scorers in fall might have scored higher in spring as a result of regression to the mean rather than as a result of greater proficiency, particularly students with fall scores close to the cut-off for a deficit.

Some double-deficit studies have achieved wider separation between groups by including only students with extreme scores on PA or RAN and excluding those with average scores. This approach may reduce instability due to measurement error and regression to the mean, as well as misclassification that results when a continuous variable is dichotomized. Analyses based on extreme groups may be quite useful for purposes of theory building. For purposes of practical application, however, exclusion of a large percentage of students results in uncertain generalizability to the broader population of poor readers that school must serve. It is for this reason that I chose not to conduct my analyses using extreme groups.

As an afterthought, however, I used within-sample standing to identify students who fell within the top or bottom third on PA and RAN. I then sorted these students into groups: those who scored in the top third on both PA and RAN, those who scored in the bottom third on RAN but the top third on PA, those who scored in the bottom third on PA but the top third on RAN, and those who scored in the bottom third on both PA and RAN. Applying these criteria, 86 out of 197 met criteria for one of the four subtypes at the beginning of grade 1. I then examined the percentage of students who continued to meet criteria for the same subtype in the spring. Overall, only 20 (23%) were classified the same way in spring. Similarly, I computed stability based on students who fell in either the top and bottom 40% on each measure (i.e., eliminating only the middle 40th–60th percentiles). Among the 131 students who met criteria for one of the subtypes in fall, only 45 (34%) met criteria for the same subtype in spring. Together, these results indicated that stability was not improved by wider separation of groups.

I also adjusted the cut-off for a deficit on PA or RAN upward and downward at both pretest and posttest to determine whether application of more stringent or more liberal standards for each subtype would result in greater stability. I was unable to find any combination of cut-off scores that resulted in stability for more than about half of the total sample. Interestingly, RAN deficits were not significantly more stable than PA deficits, despite the more robust correlations between fall and spring performance for RAN ($r = .76$) versus PA ($r = .21$). This finding underscores the importance of looking at profile agreement, not just score consistency, when evaluating stability.

Clearly, it remains possible that I would have found greater stability with different measures of RAN and PA, or with multiple measures. In a study that investigated stability of measures of phonemic awareness from end of kindergarten to end of grade 1 and from end of grade 1 to end of grade 2, Cardoso-Martins and Pennington (2004) reported test-retest correlations in the range of .60 to .70 for a composite measure of PA. In contrast, test-retest correlations for five individual measures that were first administered at the end of kindergarten ranged from .29 to .53. Although I was unable to evaluate stability of double-deficit subtypes using multiple measures of PA, I did have access to PA and RAN scores for a small sample of students who participated in Reading Recovery in a subsequent year ($n = 29$). I used the same measure of RAN as I used in analyses reported above, but a different measure of PA: Phoneme Elision, a norm-referenced subtest from the

CTOPP (Wagner et al., 1999). For both PA and RAN, I defined a deficit as performance that was 1 or more *SD*s below the norm group mean. I then identified subtypes using the same procedures as I described in a previous section. Out of nine students with a double deficit in fall, only one continued to demonstrate a double deficit in spring. Overall, the stability of phonological deficits (PD and DD) was 66% and the stability of naming speed deficits was 69%, values that were comparable to those obtained in the present study.

Furthermore, temporal stability was not improved by requiring that students demonstrate low performance on both the Yopp-Singer and Phoneme Elision to be classified with a phonological deficit, and low performance on both RAN objects and RAN colors to be classified with a naming-speed deficit. The convergence of findings suggests that the present results may not be atypical in school-identified samples of low-performing, beginning first-grade readers.

Indeed, a plausible explanation for low temporal stability is substantive, rather than methodological. The first half of grade 1 is a time of rapid growth on a wide range of reading-related skills, including PA and RAN. The most sizeable improvements in naming speed occur during the first two years of school (Wolf & Bowers, 1999). After that, score increases are much more gradual. Similarly, much of the growth on measures of PA, like those administered in the present study, occurs from the beginning of kindergarten to the first half of grade 1. Not all students accelerate at the same time, so standing on PA and RAN can change drastically over the first semester of first grade. It is for this reason that profiles based on fall performance on PA and RAN may not be good predictors of progress over the first semester of first grade. As reported above, students with a double deficit in the fall had the highest rate of assignment to immediate intervention and students with a double-deficit in the spring had the lowest text levels in the spring. Students who were identified with a double deficit in the fall (i.e., DD-F), however, were not more impaired than students without a double deficit on spring text level, perhaps because fall estimates of phonemic awareness and naming speed were not accurate measures of spring standing on phonemic awareness and naming speed.

Conclusions

Having a double deficit in fall or spring of grade 1 may be a red flag for concurrent reading difficulties; however, fall subtypes do not appear to be stable predictors of progress over the first half of grade 1. If scores on PA and RAN had been used in the present sample to identify students with the poorest prognosis in reading, a different set of students would have been targeted depending on whether PA and RAN were measured at the beginning or middle of the year. Furthermore, students with a double deficit in the fall were not more impaired readers in the spring than any of the other three subtypes of low-performing readers.

Finally, it is important to note that results obtained in the present study may not generalize to older students with well-established histories of reading disabilities. Indeed, one of the hallmarks of the population of students with the most severe reading disabilities is intractability, particularly for fluency-rated difficulties (see reviews in Torgesen, 2004; Wolf & Katzir-Cohen, 2001). It is possible that double-deficit subtypes are quite stable over time in students with a longer track record of reading failure. Furthermore, results may not generalize to double-deficit subtypes that are based on measures of phonological processing other than phonemic awareness or measures of naming speed other than RAN for colors and objects. Future research should examine the temporal stability of double-deficit subtypes at different ages and grades to determine at what point a double deficit may be a reliable indicator of future risk. In addition, investigations are needed to determine the reliability and validity of different approaches to identifying double-deficit subtypes.

References

Ackerman, P. T., Holloway, C. A., Youngdahl, P. L., & Dykman, R. A. (2001). The double-deficit theory of reading disability does not fit all. *Learning Disabilities Research & Practice, 16*, 152–160.

Askew, B. J., Fountas, I. C., Lyons, C. A., Pinnell, G. S., & Schmitt, M. C. (1998). *Reading Recovery review: Understanding outcomes and implications.* Columbus, OH: Reading Recovery Council of North America.

Badian (1997). Dyslexia and the double-deficit hypothesis. *Annals of Dyslexia, 47*, 69–87.

Badian (1998). A validation of the role of preschool phonological and orthographic skills in the prediction of reading. *Journal of Learning Disabilities, 31,* 472–481.

Bowers, P. G., Sunseth, K., & Golden, J. (1999). The route between rapid naming and reading progress. *Scientific Studies of Reading, 3,* 31–53.

Bowers, P. G., & Wolf, M. (1993). Theoretical links among naming speed, precise timing mechanisms and orthographic skill in dyslexia. *Reading and Writing: An Interdisciplinary Journal, 5,* 69–85.

Cardoso-Martins, C., & Pennington, B. (2004). The relationship between phoneme awareness and rapid serial naming skills and literacy acquisition: The role of developmental period and reading ability. *Scientific Studies of Reading, 8,* 27–52.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition, 47,* 149–180.

Clay, M. M. (1993). *An observation survey of early literacy achievement.* Portsmouth, NH: Heinemann.

Cornwall, A. (1992). The relationship of phonological awareness, rapid naming, and verbal memory to severe reading and spelling disability. *Journal of Learning Disabilities, 25,* 532–538.

Cromer, W. (1970). The difference model: A new explanation of some reading difficulties. *Journal of Educational Psychology, 61,* 471–183.

Fawcett, A. J., & Nicholson, R. I. (1994). Naming speed in children with dyslexia. *Journal of Learning Disabilities, 27,* 641–646.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Fowler, A., Francis, D. J., Stuebing, K. K., & Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 85,* 1–18.

Goméz-Bellengé, F. X., Rodgers, E., & Fullerton, S. K. (2003). Reading Recovery and Descubriendo la Lectura national report 2001–2002. Columbus, OH: National Data Evaluation Center.

Kavale, K. A. (2003). Discrepancy models in the identification of learning disabilities. In R. Bradley, L. Danielson & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 369–426). Mahwah, NJ: Erlbaum.

Kirby, J. R., Pfeiffer, S. L., & Parrila, R. K. (2003). Naming speed and phonological awareness as predictors of reading development. *Journal of Educational Psychology, 95,* 453–464.

Korhonen, T. T. (1995). The persistence of rapid naming problems in children with reading disabilities: A nine-year follow-up. *Journal of Learning Disabilities, 28,* 232–239.

Lovett, M. W., Steinbach, K., & Frijters, J. C. (2000). Remediating the core deficits of developmental reading disability: A double-deficit perspective. *Journal of Learning Disabilities, 33,* 334–358.

Manis, F. R., Doi, L. M., & Bhadha, B. (2000). Naming speed, phonological awareness, and orthographic knowledge in second graders. *Journal of Learning Disabilities, 33,* 325–333, 374.

Manis, F. R., Seidenberg, M. S., & Doi, L. M. (1999). See Dick RAN: Rapid naming and the longitudinal prediction of reading subskills in first and second graders. *Scientific Studies of Reading, 3*, 129–157.

Manis, F. R., Seidenberg, M. S., Stallings, L., Joanisse, M., Bailey, C., Freedman, L., & Curtin, S., & Keating, P. (1999). Development of dyslexia subgroups: A one-year follow-up. *Annals of Dyslexia, 49*, 105–134.

Meyer, M. S., Wood, F. B., Hart, L. A., & Felton, R. H. (1998a). Selective predictive value of rapid automatized naming in poor readers. *Journal of Learning Disabilities, 31*, 106–117.

Meyer, M. S., Wood, F. B., Hart, L. A., & Felton, R. H. (1998b). Longitudinal course of rapid naming in disabled and nondisabled readers. *Annals of Dyslexia, 48*, 91–114.

Misra, M., Katzir, T., Wolf, M., & Poldrack, R. (2004). Neural systems for rapid Automatized naming in skilled readers: Unraveling the RAN-reading relationship. *Scientific Studies of Reading, 8*, 241–256.

Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., Katz, L., Francis, D. J., & Shaywitz, B. A. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology, 90*, 347–373.

Pinnell, G. S., Lyons, C. A., DeFord, D. E., Bryk, A., & Seltzer, N. (1994). Comparing instructional models for the literacy education of high risk first graders. *Reading Research Quarterly, 29*, 8–39.

Scarborough, H. S. (1998a). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J. Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75–119). Timonium, MD: York Press.

Scarborough, H. S. (1998b). Predicting the future achievement of second graders with reading disabilities: Contributions of phonemic awareness, verbal memory, rapid naming, and IQ. *Annals of Dyslexia, 48*, 115–137.

Scarborough, H. S. (2001). Connecting early language to later reading (dis)abilities. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy* (pp. 97–110). New York: Guilford Press.

Schatschneider, C., Carlson, C. D., Francis, D. J., Foorman, B., & Fletcher, J. (2002). Relationship of rapid automatized naming and phonological awareness in early reading development: Implications for the double-deficit hypothesis. *Journal of Learning Disabilities, 35*, 245–256.

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

Spector, J. E. (1992). Predicting process in beginning reading: Dynamic assessment of phonemic awareness. *Journal of Educational Psychology, 84*, 353–363.

Spector, J. E., & Moore, P. (2001). Predicting success in Reading Recovery: The contribution of phonological processes. Paper presented at the annual meeting of the *American Educational Research Association*. Seattle, WA.

Speece, D. L., & Cooper, D. H. (2004). Methodological issues in research on language and early literacy from the perspective of early identification and instruction. In C. A. Stone, E. R. Silliman, B. J. Ehren, & K. Appel (Eds.), *Handbook of language and literacy: Development and disorders* (pp. 83–94). New York: Guilford.

Stanovich, K. E., Siegel, L. S., & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability. *Journal of Educational Psychology, 89,* 114–128.

Sunseth, K., & Bowers, P. (2002). Rapid naming and phonemic awareness: Contributions to reading, spelling, and orthographic knowledge. *Scientific Studies of Reading, 6,* 401–429.

Swanson, H. L., Trainin, G., Necoechea, D. M., & Hammill, D. D. (2003). Rapid naming, phonological awareness, and reading: A meta-analysis of the correlation evidence. *Review of Educational Research, 73,* 407–440.

Torgesen, J. (2004). Lessons learned from research on interventions for students who have difficulty learning to read. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 355–382). Baltimore: Paul Brookes.

Torgesen, J. K., & Mathes, P. G. (2000). *A basic guide to understanding, assessing, and teaching phonological awareness.* Austin, TX: Pro-ed.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processes.* Austin, TX: Pro-Ed.

Wagner, R. K., Torgesen, J. K., Laughon, P., Simmons, K., & Rashotte, C. (1993). The development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85,* 83–103.

White, W. J., & Wigle, S. E. (1986). Patterns of discrepancy over time as revealed by a standard-score comparison formula. *Learning Disabilities Research, 2,* 14–20.

Wimmer, H. (1993). Characteristics of developmental dyslexia in a regular writing system. *Applied Psycholinguistics, 14,* 1–33.

Wimmer, H., Mayringer, H., & Landerl, K. (2000). The double-deficit hypothesis and difficulties in learning to read a regular orthography. *Journal of Educational Psychology, 92,* 668–680.

Wolf, M. (1991). Naming speed and reading: The contribution of the cognitive neurosciences. *Reading Research Quarterly, 26,* 123–141.

Wolf, M. (1997). A provisional, integrative account of phonological and naming-speed deficits in dyslexia: Implications for diagnosis and intervention. In B. Blachman (Ed.), *Foundations of reading acquisition* (pp. 67–92). Hillsdale, NJ: Erlbaum.

Wolf, M. (2003). Response to "clinical judgments in identifying and teaching children with language-based reading difficulties." In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 725–736). Mahwah, NJ: Erlbaum.

Wolf, M., Bally, H., & Morris, R. (1986). Automaticity, retrieval processes, and reading: A longitudinal study in average and impaired readers. *Child Development, 57,* 988–1000.

Wolf, M., & Bowers, P. (1999). The "double-deficit" hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91,* 1–24.

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming-speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities, 33,* 387–407.

Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading, 5,* 211–239.

Wolf, M., O'Rourke, A. G., Gidney, C., Lovett, M., Cirino, P., & Morris, R. (2002). The second deficit: An investigation of the independence of phonological and naming-speed deficits in developmental dyslexia. *Reading and Writing: An Interdisciplinary Journal*, *5*, 43–72.

Yopp, H. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly*, *23*, 159–177.

Yopp, H. (1995). A test for assessing phonemic awareness in young children. *The Reading Teacher*, *49*, 20–29.