

FOCUS ON EXCEPTIONAL children

Throw 'em Out or Make 'em Better? State and District High-Stakes Writing Assessments

Steve Graham, Michael Hebert, and Karen R. Harris

The writing of school-aged children is assessed for many reasons (Graham, Harris, & Hebert, 2011). Teachers assess writing to monitor students' growth as writers, inform instruction, provide feedback, and evaluate the effectiveness of their teaching. Students assess their own writing to appraise growth, identify strengths, and determine areas in need of further development. Peers assess other students' writing to provide them with feedback on what works in a paper and what still needs work. States and school districts assess writing to determine how many students meet local or state performance standards, identify youngsters who need extra help, and evaluate the effectiveness of individual teachers and schools. The national government administers the National Assessment of Educational Progress (NAEP) writing test to measure American students' collective writing success, evaluating students' writing performance across time.

Given the heavy emphasis now placed on assessment and evaluation as a tool for improving and reforming writing and other aspects of education in the United States (Gewertz, & Robelen, 2010; National Commission on Writing, 2003), it is important to ask whether the various forms of assessment, ranging from classroom-based writing assessments to state and district evaluations (the focus of this article) do, in fact, make a difference in improving how well students write? For students with disabilities, such questions are especially important, as so many of these students experience difficulty learning to write. On the 2007 NAEP (Salahu-Din, Persky, & Miller, 2008), just 6% of eighth-grade and 5% of twelfth-grade students with disabilities performed at or above the "proficient" level in writing (defined as solid academic performance). Students scoring below this level are classified as obtaining only partial mastery of the literacy skills needed at their respective grade. Thus, this assessment indicates that 19 of every 20 students with disabilities do not acquire the writing skills needed for success in school.

Although classroom-based assessments are not the focus of this article, evidence shows that such assessments can make a difference in improving how well students write. A recent meta-analysis of experimental and quasi-experimental studies conducted mostly with typically developing students (Graham, Kihara, McKeown, & Harris, 2011) provided empirical evidence that writing assessments that are part of typical classroom practices improve the overall quality of students' writing. When students receive feedback about their writing and learning progress, writing improves. When students evaluate their own writing, writing

Dr. Graham and Dr. Harris share the Currey Ingram Professor of Special Education and Literacy chair at Vanderbilt University. Mr. Hebert is a predoctoral fellow at Vanderbilt University.

improves. When students are partners in writing assessment, giving and receiving peer feedback, writing improves. While research examining the effectiveness of such classroom-based assessment practices specifically with students with disabilities is needed, these findings are encouraging.

STATE AND DISTRICT HIGH-STAKES WRITING ASSESSMENTS

In addition to typical classroom-based assessment practices, most states and school districts annually engage in another form of writing assessment, tallying how many students in specific schools pass standards-based writing tests (Jeffery, 2009). Such tests are considered to be high stakes for students, as they may determine whether a student receives a high school diploma, establish eligibility for merit-based scholarships, or indicate readiness for college (Beck & Jeffrey, 2007). These tests are also high stakes for districts, schools, and teachers; these summative results are used by the public, policy makers, and educators to make judgments about the effectiveness of writing instruction and,

more recently, to evaluate teachers (as part of the annual review process; such evaluations are increasingly emphasized in states that receive Race to the Top funding from the federal government).

Supporters of high-stakes writing tests argue that they are beneficial because they raise expectations for students' writing and make these expectations more explicit, improving the teaching of writing and encouraging students to become better writers. This viewpoint about the value of high-stakes tests is typically shared by the general public (Afflerbach, 2005).

In contrast to classroom-based writing assessment procedures, however, where the impact of such assessments have been subject to stringent scientific study, the value of high-stakes writing tests rests primarily on correlational data and qualitative analyses. Such studies have provided some limited support for the positive effects of high-stakes writing assessments, including improving students' writing (based on correlational data where students' writing scores improved after such assessments were implemented), making writing instruction more central to the mission of schools, and changing teachers' writing practices in positive ways (Callahan, 1999; Dappen, Iserhagen, & Anderson, 2008; Parke, Lane, & Stone, 2006).

Despite these positive outcomes, there is considerable reason to be concerned about the technical adequacy of high-stakes writing assessments and the consequences of these tests for relevant stakeholders: students in general, their classmates with disabilities, and these students' teachers and schools. In this article, we present seven recommendations for improving high-stakes writing assessments. These recommendations are based primarily on findings from empirical and qualitative studies. The data reviewed provides validation for our concerns about high-stakes assessment in writing. While we focus our discussion on the implications of these recommendations for students with disabilities, our concerns and suggestions are applicable to the use of high-stakes writing assessments with all students.

The basic conclusion that we draw from the review presented here is that high-stakes writing assessment as currently conceptualized and conducted is not defensible. This raises the question as to why we made recommendations for improving such tests at all. Why not just recommend that they be thrown away? First, much of the reform effort in the United States is driven by standards-based assessment (National Commission on Writing, 2003). Reform of writing instruction is a necessity in this country, as almost two out of every three youngsters do not write well enough to meet grade-level demands, and writing instruction in too many classrooms is impoverished (see Applebee & Langer, 2006; Gilbert & Graham, 2010; Kihara, Graham, & Hawken, 2009). If writing is to be a player in the reform movement, measuring the writing progress of students individually and collectively is

FOCUS ON EXCEPTIONAL CHILDREN

ISSN 0015-511X

FOCUS ON EXCEPTIONAL CHILDREN (USPS 203-360) is published monthly except June, July, and August as a service to teachers, special educators, curriculum specialists, administrators, and those concerned with the special education of exceptional children. This publication is annotated and indexed by the ERIC Clearinghouse on Handicapped and Gifted Children for publication in the monthly *Current Index to Journals in Education* (CIJE) and the quarterly index, *Exceptional Children Education Resources* (ECER). The full text of *Focus on Exceptional Children* is also available in the electronic versions of the *Education Index*. It is also available in microfilm from Serials Acquisitions, National Archive Publishing Company, P.O. Box 998, Ann Arbor, MI 48106-0998. Subscription rates: individual, \$50 per year; institutions, \$68 per year. Copyright © 2011, Love Publishing Company. All rights reserved. Reproduction in whole or part without written permission is prohibited. Printed in the United States of America. Periodical postage is paid at Denver, Colorado. **POSTMASTER:** Send address changes to:

Love Publishing Company
Executive and Editorial Office
P.O. Box 22353
Denver, Colorado 80222
Telephone (303) 221-7333

CONSULTING EDITORS

Steve Graham
Vanderbilt University

Ron Nelson
University of Nebraska-Lincoln

Eva Horn
University of Kansas

Carrie E. Watterson
Senior Editor

Stanley F. Love
Publisher

part of the entry fee. Nevertheless, as the National Commission on Writing (2003) indicated, we must “ensure that the assessment of writing is fair and authentic” (p. 29).

Second, this moment in time presents a unique opportunity for revising and improving high-stakes assessments that states and school districts use to evaluate students’ writing. Two common-assessment consortia (Gewertz, 2011), which collectively involve 44 states and the District of Columbia, are currently developing assessments for measuring the new grade-level expectations for writing specified in the Common Core State Standards Initiative (<http://www.corestandards.org/the-standards/english-language-arts-standards>). This includes the development of new summative high-stakes writing assessments. We hope that the two consortia (Smarter and PARCC) and others address the concerns about high-stakes assessment raised in this article. Our recommendations comprise the following sections.

BASE HIGH-STAKES WRITING ASSESSMENTS ON STUDENTS’ ACTUAL WRITING

A long-standing controversy in the field of writing involves whether writing should be assessed directly, by grading students’ actual writing, or indirectly, through tests that measure knowledge, skills, or processes through the use of multiple choice questions, fill-in-the-blank exercises, and other decontextualized writing activities. The basic assumption underlying the use of indirect measures on a high-stakes assessment is that scores from these measures are highly correlated with scores from assessments of students’ actual writing. More specifically, if the same students are given an indirect and direct measure of writing performance, the rank order of the resulting scores for the two assessments will be highly similar.

Available evidence indicates that this assumption is not valid. We located nine studies examining the relationship between indirect and direct measures of writing performance (Table 1). The indirect measures included a wide range of mostly norm-referenced standardized tests (primarily composed of multiple-choice items). The direct measures involved assessing the quality of students’ writing in a variety of genres: narrative, expository, descriptive, and persuasive. Most of the reported correlations between indirect measures and the quality of students’ writing were in the low to moderate range (.08 to .69). Indirect measures were not a good proxy for the direct assessment of students’ writing in these studies, as they accounted for less than one half of the variability in writing quality scores.

While indirect measures may be less expensive and easier to administer, they do not provide an adequate assessment of students’ actual writing on high-stakes assessments. As the National Commission on Writing (2003) noted, “Assessment of student writing must go beyond multiple-choice, machine-

scorable items” (p. 4). Students with disabilities need to produce text when they are being assessed. This is not to say that indirect measures provide no useful information at all. Instead, a high-stakes assessment that is not largely based on students’ actual writing is insufficient and inadequate.

BASE HIGH-STAKES WRITING ASSESSMENTS ON MULTIPLE SAMPLES OF STUDENTS’ WRITING

On high-stakes assessments, we often infer how well students write by examining their performance on a single writing task (Beck & Jeffrey, 2007). Because writing serves many purposes and different forms of writing require specialized knowledge (e.g., narrative vs. persuasive vs. expository), it seems unlikely that a single composition provides enough evidence to make reliable and valid judgments about a student’s writing capabilities. Many state and district assessments, however, proceed as if a single sample of writing is enough (Jeffrey, 2009).

If a single composition provides an adequate measure of writing ability, students’ scores on different types of compositions (e.g., narrative, persuasive, expository, descriptive) should yield similar outcomes when scored in the same way. Unfortunately, this was not the case in six studies involving high-stakes assessments that we reviewed (Carlman, 1985; Englehard, Gordon, Gabrielson, 1991; Englehard, Gordon, Gabrielson, & Walker, 1994; Gearhart, Herman, Baker, & Whitaker, 1992; Popp, Ryan, Thompson, & Behrens, 2003; Veal & Tillman, 1971). In each study, the writing quality scores for two or more genres of writing differed statistically for the tested students. Consequently, students’ performance on one writing task is not identical to their performance on a different one.

Furthermore, if a single paper produced in a high-stakes assessment provides a sufficient indicator of students’ general writing abilities, performance on one composition in a specific genre should be highly correlated with performance in another genre. This assumption was not supported in five studies involving high-stakes assessments (Hunter, Jones, & Randhawa, 1996; Lehman, 1990; Moss, Cole, & Khampalikit, 1982; Popp et al., 2003; Purves, 1992). In these studies, correlations between writing quality scores for different types of compositions (e.g., persuasive, descriptive, expository, narrative) were small to moderate (.10 to .61). Thus, a single piece of writing on a high-stakes test is not an adequate measure of students’ writing capabilities.

Basing a high-stakes assessment on a single piece of writing is like administering a one-item test. Such tests are likely to be unreliable. For example, Coffman (1966) reported that five writing tasks were needed to reliably assess the writing of high school students (when scored by two raters). While Huang (2008) reported that three writing tasks (scored by two

TABLE 1.
Correlations between Direct and Indirect Measures of Writing

Study	Task	Grade	Quality Measure	Indirect Measures	Correlations
Shermis & Long, 2009	P	4, 8, 10	holistic	FCAT Writing - multiple choice items	.45 (gr 4) .58 (gr 8) .59 (gr 10)
Bright, 1996	Open	3 - 5	analytic	Stanford Achievement Test	.53 (gr 3) .50 (gr 4) .57 (gr 5)
Welch & Miller, 1995	N, EXP	8	holistic	multiple choice	.66-.77
Hogan & Mishler, 1980	N	3 & 8	holistic	Metropolitan Achievement Test	story 1 (gr 3): .52-.66; (gr 8): .50-.57; story 2 (gr 3): .56-.77; (gr 8): .72-.73
Swartz & Whitney, 1985	P, D, E	12	holistic	GED Writing Skills	.55-.69
Veal & Hudson, 1983	NR	10	holistic	ITBS CAT TAP WP	ITBS: .48-.65 CAT: -.20-.37 TAP: .23-.45 WP: .23-.51
Moss, Cole, & Khampalikit, 1982	I, PUR	4, 7, 10	holistic	3Rs Achievement Test	.12 (gr 4) .39 (gr 7) .47 (gr 10)
Goldshalk, Swineford, & Coffman, 1966	D, P, E, N	11 & 12	holistic	English Comp Test	.47-.71
Edmiston & Gingerich, 1942	NR	4-12	holistic	English Usage Test	.45-.68 (gr 4) .30-.54 (gr 5) .20-.40 (gr 6) .23-.56 (gr 7) .28-.60 (gr 8) .23-.44 (gr 9) .18-.55 (gr 10) .24-.44 (gr 11) .08-.55 (gr 12)

Note: P = persuasive; Open = student choice; N = narrative; EXP = expressive; D = descriptive; NR = not reported; I = Informative; PUR = a letter serving a specific purpose; E = expository; FCAT = Florida California Achievement Test; ITBS = Iowa Test of Basic Skills; CAT = California Achievement Test; TAP = Test of Achievement and Proficiency; WP = Writing Proficiency Test; COMP = composition.

raters) provided a reliable estimate of the quality of high school seniors' writing, he found that five writing tasks were needed to reliably estimate ($r > .80$) the writing of at-risk students. While more research is needed, we find it reasonable that more writing samples may be needed to reliably assess the writing of students with disabilities. The higher the stakes in these assessments, the more important it becomes to collect multiple samples of students' writing. Overall evaluation of students, teachers, and schools provided by high-stakes assessments should not be based on a single throw of the dice.

ENSURE THAT HIGH-STAKES WRITING ASSESSMENTS ARE RELIABLE

A basic assumption underlying a valid high-stakes writing assessment is that it is reliable. The validity of such an

assessment rests upon the following assumptions: Different raters will provide the same score, the score obtained today will be obtained a few days later, and a similar assessment produces similar results. We located nine studies that examined the reliability of holistic quality measures used to score high-stakes writing tests (see Table 2). All of these studies involved more than two raters (similar to how high-stakes writing tests are scored on state assessments). Twelve similar studies were located where an analytic scale was used to score such assessments (see Table 3). With a holistic scale, a single rating of the general quality of the composition is made, whereas an analytic scale produces separate ratings for specific attributes such as ideation, organization, style, and so forth. These are the two basic approaches to scoring writing samples on high-stakes assessments.

TABLE 2.
Reliability and Validity Information for Holistic Scoring on High-Stakes Assessments

Study	Task	Rubric Range	Gr (Age)	Raters	Trained	Interrater Agreement			Validity and Reliability
						%	<i>r</i>	<i>G</i>	
Sevigny, Savard, who & Beaudoin, 2009	E	5	(13-16)	NR	Y	63%– 64%			construct validity: students received higher holistic scores also received higher scores on analytic measures, suggesting that each level of the holistic measure represents a different level of writing ability
Huang, 2008	I, INT, US	6	12	NR	NR			.70–.80 (ELL); .80–.86 (non-ELL)	construct validity: greatest source of variation in students' writing scores was due to differences among ELL students and non-ELL students
Hunter, Jones, & Randhawa, 1996	NR	5	5, 8, 11	NR	Y		.82		concurrent validity: correlation between holistic score and total analytic score was .69; relation between holistic score and analytic traits (I, O, SYN, VOC, C) ranged from .47 to .55; correlation between high stakes holistic score and holistic score on students' best writing ranged from .27 to .34
Linn, Kiplinger, Chapman, & LeMahieu, 1992	NR	4–100	3–5; 7–8; 11–12	EXP	Y				construct validity: papers scored using different state guidelines were rank-ordered in a similar fashion across states, suggesting that the holistic scoring systems of the 6 states define writing quality in a similar way
De Ayala, Dodd, & Koch, 1991	N, EXPRES	4 or 8	HS	NR	Y	76%			reliability: when 4-point scale was lengthened by averaging the scores of 2 raters, it was more reliable validity: the range of scores for the 8 point scale was restricted
Lehmann, 1990	LET, N, P, E, PAR	5	11	T	Y	73%	.84	.65 .89 (inter) .94 (intra)	
Goldberg & Walker-Bartnick, 1988	N,	E	7	NR	NR	NR			reliability: when writing samples were rescored using the same rubric 3 years later, scores were lower
Swartz & Whitney, 1985	P,D, E	6	12	EXP	Y		.82–.87		reliability: corrected by Spearman Brown formula
Moss, Cole, & Khampalikit, 1982	I, PUR	4	4, 7, 10	NR	NR		.86–.94	.41–.50	
Stewart & Grobe, 1979	I, P, E	4	5, 8, 11	T	Y		.90		
Godshalk, Swineford, & Coffman, 1966	D, P, E, N	3	11 & 12	Some Exp	N			.92	

Note: E = Expository, I = Informative, INT= Interpretation, NR = Not Reported, N = Narrative, Express = Expressive, LET = Letter, P = Persuasive, PAR =

TABLE 3.
Reliability and Validity Information for Analytic Scoring on High-Stakes Assessments

Study	Task	Rubric Range	Gr (Age)	Skills Scored	Trained	Interrater Agreement			Validity and Reliability	
						%	<i>r</i>	<i>G</i>		
Crawford & Smolkowski, 2008	N	10	5 & 8	C/O, ST, L	T (TR)	49%	.61-.96		reliability: was calculated for each skill assessed	
Crawford, Helwig, & Tindal, 2004	N, E, P, I	6	5 & 8	I, O, SF, C	T (TR)	49%	.64-.79			
Brown, Glasswell, & Harland, 2004	I, D	11	5-8	A, PUR, I, O, C, U	T (TR)			.77	.75	construct validity: the scale was sensitive to students' growth in writing in a short study involving persuasive writing instruction with high school students
Popp & Ryan, 2003	N	6	3	I, O, V, VOC, SF, C	EXP					validity: magnitude and range of scores varied depending upon whether the anchor points were based on the performance of grade 3 students versus a broader range of grade 3, 5, & 8 students
Haladyna & Hess, 1999-2000	EXP, I, N, P	6	8 & 10	I/O, O, V, VOC, SF, C	T-TR		.61 total .50 each skill	.94-.96		reliability: alpha is for total score reliability; construct validity: factor analysis indicated only a single dimension measured (halo effect); more students failed test if they had to have a passing score on each skill versus total score summed for all skills
Hollenbeck, Tindal, & Almond, 1999	SC	8	6	I, O, VOC, W, SF, C	T (TR)	47%	.57-.97			reliability: most reliabilities calculated for each trait; one fourth of students were classified as proficient by one rater and not proficient by another rater
Schoonen, Vergeer, & Eiting, 1997 (study 1)	P, D, & INST	5	6	C, U	T-EXP; LAY			.88-1.00		reliability: When using a scale with benchmarks g for experienced teachers ranged from .88 to 1.00 for C and from .91 to .96 for U; g for laypersons ranged from .70 to .91 for C and .30 to .70 for U
Schoonen, Vergeer, & Eiting, 1997 (study 2)	P, D, & INST	5	6	C, U	T-EXP; LAY			.70-.94		reliability: when using a scale with a scoring guide, g for experienced teachers ranged from .70 to .95 for C and .86 to .94 for U; g for laypersons ranged from .78 to .99 for C and from .65 to .78 for U
Kuhlemeier & van den Bergh, 1997	LET	5	9	I, ST, O, E	T (TR)		.83-.99			
Gearhart, Herman, Novak, & Wolf, 1995	N	6	1-6	F/O, D, C	T-EXP (TR)	28% - 37%	.60-.63	.63-.68		construct validity: students' scores increased by grade; halo effect observed as scores for each skill were moderately to highly correlated (.74-.92)

TABLE 3. (continued)

Study	Task	Rubric Range	Gr (Age)	Skills Scored	Trained	Interrater Agreement			Validity and Reliability
						%	<i>r</i>	G	
Gabrielson, Gordon, & Engelhard, 1995	P	4	11	C/O, ST, SF, C	NR (TR)		.87		
Engelhard, Gordon, Gabrielson, & Walker, 1994	N, D, E	4	8	C/O, ST, SF, U, C	T & Other (TR)		.82		
Engelhard, 1994	N, D, E	4	8	C/O, ST, SF, U, C	EXP (TR)				validity: the following rater errors in scoring were observed: severity, halo effect, central tendency, and restricted range of scores
Engelhard, Gordon, & Gabrielson, 1991	N, D, E	4	8	C/O, ST, SF, U, C	NR (TR)		.82		reliability: based on the Spearman-Brown formula
Engelhard, 1992	N, D, E	4	8	C/O, ST, SF, U, C	EXP (TR)				reliability: there were problems with rater severity

Note: N = Narrative, E = Expository, I = Informational, P = Persuasive, D = Descriptive, EXP = Expressive, SC = Student Choice, INST = Instructional, LET = Letters, C/O = Content/Organization, ST = Style, L = Language Use, I = Ideation, SF = Sentence Fluency, I/O = Idea/Organization, A = Audience, PUR = Purpose, U = Usage, V = Voice, VOC = Vocabulary, W = Words, F/O = Focus/Organization, D = Detail, T = Teachers, TR = Trained, EXP = Experienced, NR = Not Reported, LAY = Layperson

Some of the located studies employed a single measure of reliability, whereas others applied multiple measures to judge consensus, consistency, and generalizability. We judged that a scale was reliable when it met criteria for each computed measure (e.g., 70% agreement or greater for consensus, and correlations of .80 or greater for consistency and generalizability). Holistic and analytic scoring methods were reliable in less than one half of the studies. Even if we adopted a less stringent criteria (i.e., only one type of reliability measure per study had to meet criteria), holistic scales were reliable in 78% of studies and analytic scales in just 50%.

The findings from these studies raise serious concerns about the validity and use of current high-stakes writing assessments. If such assessments are to be used to assess the writing of students with disabilities (especially given the consequences for students, teachers, and schools), the reliability of these scoring procedures must be improved. Evidence-based practices for improving reliability are presented in Figure 1.

MAKE HIGH-STAKES WRITING ASSESSMENTS FAIR

It is not enough to make high-stakes writing assessments reliable; they need to be fair, too. Nationwide, high-stakes

writing assessments involve an “elastic yardstick” for measuring student achievement and progress—important differences in state testing conditions create unequal, and thus unreliable, measurements. This is reflected in the variability in students’ passing rates from one state to the next on high stakes writing assessments (Jeffrey, 2009). The percent of students identified as proficient writers ranged from a low of 24% to a high of 94%. As these data demonstrate, place of residence may well be the best predictor of students’ writing scores on high-stakes assessments.

One reason for this variability in students’ performance involves differences in how students are tested (see Jeffrey, 2009). For instance, tests are not timed in some states, but other states impose specific time limits, ranging from 30 minutes for writing to 2 hours. Further, writing tasks vary considerably. A few states allow students to choose the genre assessed, but most states assign them in various ways, including testing a specific genre once, a specific genre twice, one of several possible genres a single time, or multiple genres a single time each. To add to this variety, there is no consistency in terms of the types of writing that are assessed, with at least 15 different types of writing assessed across states (e.g., personal narrative, story, persuasive, expository, expressive, informative). The grades in which students are assessed also varies, and some states make a

-
- Provide training on how to score compositions (Shohamy, Gordon, & Kraemer, 1992)
 - Use experienced raters to score papers (Schoonen, Verger, & Eiting, 1997; Underwood & Murphy, 1998; Wolfe, 1997; Wolfe, Kao, Ranney, 1998; Wolfe & Ranney, 1996)
 - Have multiple raters score each paper (Burgin & Hughes, 2009; Coffman, 1966; Gearhart, Herman, Novak, & Wolf, 1995; Godshalk, Swineford, & Coffman, 1966; Swartz et al., 1999).
 - Base students' writing score on multiple writing tasks (Burgin & Hughes, 2009; Finlayson, 1951; Godshalk Swineford, & Coffman, 1966; Lehman, 1990).
 - Increase the range of scores on the writing test (De Ayala, Dodd, & Koch, 1991; Godshalk et al., 1966).
 - Provide raters with benchmarks (descriptions or examples) for each point on the scale (Hwang, 1930; Kan, 2007).
 - Apply a two-step scoring process where the rater matches the composition to the closest benchmark and then scores it again if it does not match this benchmark perfectly by adding a plus or minus to the first score (Penny, Johnson, Gordon, 2000a, 2000b; Johnson, Penny, Fisher, & Kuhs, 2003).
 - Have raters discuss and resolve differences in their scores (Johnson, Penny, Gordon, Shumate, & Fisher, 2005).
 - Combine the scores of two disagreeing raters with a third score provided by a more experienced rater (Johnson, Penny, & Gordon, 2000, 2001).
-

FIGURE 1.

Evidence-Based Methods for Improving Reliability of Scoring High-Stakes Writing Assessments

passing test score a requirement for graduation, while others do not impose this condition.

A second reason for variability in students' scores involves how high-stakes writing assessments are scored by different states. We provide three examples to illustrate this. One, if students are required to have a passing score on each element (e.g., ideation, organization, sentence construction) of a writing test versus an overall passing score for the whole test, more students fail the test (Haladyna & Hess, 1999–2000). Two, if raters are trained to score multiple types of writing versus a single type of writing, more students fail the test (Moon & Hughes, 2002). Three, if the benchmarks for scores on a writing assessment for a specific grade are based on writing samples across several grades versus a single grade, more students fail the test (Popp & Ryan, 2003). These findings and the evidence underlying many of the other recommendations presented in this article make a strong case for greater uniformity in how writing is assessed and scored across states and districts if high-stakes writing assessments

in the United States are to be fair assessments for students in general and students with disabilities in particular.

PROVIDE TESTING ACCOMMODATIONS FOR STUDENTS WITH DISABILITIES

Writing is important enough that all students, including students with disabilities, should participate in state and district writing assessments. If students with disabilities do not participate in such assessments, teachers and schools are unlikely to provide these youngsters with the writing instruction they need. The message that teachers and schools receive when students with disabilities are excluded from such assessments is that writing is not important for these youngsters.

Test accommodations provide one means for ensuring that students with disabilities participate in writing assessments. A variety of test accommodations can be provided to students with disabilities (Royer & Randall, in press), including changes in time allowed (e.g., extended time), presentation (e.g., Braille/large print, computer presentation, oral reading of directions), response (e.g., oral vs. written), and setting (e.g., separate room). The purpose of testing accommodations in writing is twofold. One, they allow students to participate in assessments they would not otherwise be able to complete. Two, accommodations are aimed at minimizing the impact of student characteristics that are not relevant to the construct being measured. For example, the writing assignment might be read to a student who is dyslexic.

To be valid, a test accommodation should assess students' writing performance without changing the construct tested (Royer & Randall, in press). If an accommodation is provided to students with and without disabilities, and it only boosts the writing performance of students with disabilities (or boost it more than the performance of students without disabilities), it is assumed that the accommodation reduces the impact of the disability, allowing for a valid measure of writing (Fuchs & Fuchs, 2001). Even though test accommodations for students with disabilities are legally required through the Individuals with Disabilities Act of 2004, we located only one research study assessing the impact of a writing accommodation with students with disabilities. MacArthur and Cavalier (2004) found that dictation and speech recognition (basically a human and electronic scribe, respectively) versus writing by hand improved the writing performance of high schools students with learning disabilities (these youngsters experience difficulty with text transcription skills), but not the performance of high school students without a learning disability.

While additional research is needed, this investigation provides preliminary evidence that writing accommodations can provide a more valid assessment of students' writing without changing the construct measured. Of course, it cannot be

assumed that this is the case with all accommodations (see Royer & Randall, in press; Sireci, Scarpati, & Li, 2005). Likewise, just providing an accommodation periodically (when students are administered high-stakes tests) may minimize its usefulness, as accommodations such as speech synthesis programs require time and practice to master.

For some students with disabilities, test accommodations will not be enough to allow them to participate in writing assessments. Students with significant cognitive disabilities (about 1% of the population), for example, will require alternative assessments (Elliott, Kurz, & Neergaard, in press). Alternative assessments typically involve the use of an altogether different assessment (e.g., assessing earlier grade level content and performance standards), as opposed to making an accommodation to an existing assessment, and often take the form of portfolios (showing examples of students' work), performance tasks/events, or teacher rating scales of achievement. We were unable to locate any studies that examined such assessments in writing.

MAKE HIGH-STAKES WRITING ASSESSMENTS AUTHENTIC

A common criticism of high-stakes writing assessments is that they are not authentic (National Commission on Writing, 2003). Critics contend that such assessments do not capture the types of writing students typically do or will need to do in the future. For example, Herman, Gearhard, and Baker (1993) reported little correspondence between students' writing performance in the classroom and their scores on high-stakes type tests (correlations ranged from .02 to .31). Others have voiced concerns that the types of writing activities common to high-stakes assessments (e.g., writing a personal narrative in grades 4 or 5) are not representative of the type of writing students do in or out of school (Shultz, 2002). For example, some states refer to their assessments as personal narrative and then ask students to write a creative story with themselves as the main character. One such elementary grade prompt reads, "Pretend that for one day you are a tiny raindrop in a storm. For that day you get to go where raindrops go. Now write a story about your day as a raindrop." First, we note that this is not personal narrative. Second, writing a creative story with yourself as the main character is not a writing task typically required of students in elementary grades, is not highly authentic, and is a very demanding writing task that may not be appropriate for most children. In addition, Smagorinsky (1995) questioned the decision to focus just on written text, as composing can now take many forms and modes, involving written, auditory, and visual text.

These criticisms have merit. If high-stakes writing assessments are to provide educators with a credible picture of the

writing capabilities of students with disabilities, they must assess writing in authentic ways. As an example, assessments where students use writing as a tool for learning establish a condition where students write for a genuine purpose. Such an assessment also has the added potential of increasing writing in content classes, as science, social studies, and other teachers are more likely to emphasize writing if it plays a prominent role in the high-stakes assessment of their students. Currently, students in such classes write very little (Applebee & Langer, 2006; Kiuahara et al., 2009).

MINIMIZE NEGATIVE UNINTENDED CONSEQUENCES OF HIGH STAKES WRITING ASSESSMENT

Most states assess the writing performance of students in selected grade levels once a year (Beck & Jeffery, 2007). It was hoped that the implementation of such tests would have a positive effect on students, teachers, and schools. Nevertheless, high-stakes writing assessments have produced a number of unintended negative consequences. One unintended consequence is that these tests may encourage teachers to teach and students to use a formulaic approach to writing, reducing students' originality and voice. In a study by Albertson (2007), for example, 78% of grade 8 students and 66% of grade 10 students used the same basic formula (five-paragraph theme) or a variation of it on the state's direct writing assessment.

A second possible negative consequence of high-stakes writing assessments in middle and high school is that they send the unintended message to teachers that writing is the job of the language arts teachers. This was the case in the study of a high school where English teachers were expected to implement the State's writing assessment framework (Callahan, 1999). Many content teachers in the school came to view writing as something done by the English department. Likewise, statewide assessments of writing are not typically designed to assess writing in specific content areas. Instead, such assessments are more generic in nature, making it more likely that teachers will view writing as the responsibility of the language arts teachers.

A third, and potentially the most negative, consequence of high-stakes writing assessments is that they narrow the writing curriculum. Hillocks (2002), for instance, found that when states specify which genres of writing are to be assessed, teachers place greater emphasis on teaching those genres, excluding the teaching of other genres or aspects of writing. Possibly the worst example of this we have witnessed involves states where a single genre of writing is tested at selected grade levels. Some schools seek to maximize their writing test performance by teaching only that genre for the first half of the year, before the test is administered. During the rest of the year, and years when writing

is not tested, very little writing instruction takes place, as it is not tested and other subjects are.

While these are not the only negative unintended consequences of high-stakes writing assessments, it is important that policymakers at the district and state level carefully consider how teachers and schools will seek to “game” the system to improve the test performance of their students, including students with disabilities, and what messages their writing assessment system sends to teachers, students, and parents.

CONCLUDING COMMENTS

High-stakes writing assessments are evaluative and have consequences for students, teachers, schools, and school districts. As a result, it is critical that they are reliable, valid, and fair. The evidence reviewed in this article demonstrated good reason to be concerned about such tests in the United States. Concerns ranged from unintended consequences of these assessments to issues involving the very accuracy and validity of these tests. Available scientific evidence suggests that scores on typical state and district high-stakes writing assessments are not consistent enough to make reliable decisions about groups of students, much less an individual youngster. Moreover, many states collect just a single piece of writing (see Jeffery, 2009). This is not enough to provide a valid estimate of writing ability. Test scores are influenced by where students live. What is tested, how it is tested, and how it is scored differs from one state to the next. As a result, a passing score on a high-stakes test might be granted to a student living in one state but denied to a youngster with equivalent writing skills living in the bordering state.

If states and school districts are to continue to use high-stakes assessments to assess students in general and students with disabilities in particular, these tests must be improved. At the present time, it is not defensible to use these tests to determine which youngsters graduate from high school. They provide a shaky tool for monitoring the writing progress of students in a specific school, district, or state. Principals and teachers have every right to be concerned if such assessments are used to evaluate teachers' competence. Frankly put, such assessments do not provide an accurate gauge of students' writing competence—they are too narrow in their conception, varied in their execution, and uncertain in their findings.

We recommend that states and school districts examine their current high-stakes writing assessments and redesign them (as needed) so that they address the recommendations presented here. This point in time provides a unique opportunity for states to design better writing assessments for summative purposes, as 44 states and the District of Columbia are involved in developing new English/language arts tests through funding from the Race to the Top initiative (Gewertz & Robelen, 2010). We think that it is especially important

that these development efforts carefully consider the issues presented in this article. In addition, we encourage those who develop high-stakes assessments in content areas such as science and social studies to use writing tasks as part of their assessments. This can include performance assessments where students construct written responses, apply writing as a learning tool, or both. Such assessments increase the likelihood that writing will become a common staple across the curriculum and that students learn to take advantage of the potential power of writing to enhance content learning.

REFERENCES

- Albertson, B. (2007). Organization and development features of grade 8 and grade 10 writers: A descriptive study of Delaware student testing program (DSTP) essays. *Research in the Teaching of English, 41*, 435–465.
- Afflerbach, P. (2005). National Reading Conference policy brief: High stakes testing and reading assessment. *Reading Research Quarterly, 37*, 151–162.
- Applebee, A., & Langer, J. (2006). *The state of writing instruction: What existing data tell us*. Albany, NY: Center on English Learning and Achievement.
- Beck, S. W., & Jeffery, J. V. (2007). Genres of high-stakes writing assessments and the construct of writing competence. *Assessing Writing, 12*, 60–79.
- Bright, E. R. (1996). *Portfolio assessment and standardized achievement measures as outcomes in Title I evaluation at the school-district level*. Unpublished doctoral dissertation. University of Florida, Gainesville, FL.
- Brown, G. T. L., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing, 9*, 105–121.
- Burgin, J., & Hughes, G. D. (2009). Credibly assessing reading and writing abilities for both elementary student and program assessment. *Assessing Writing, 14*, 25–37.
- Callahan, S. (1999). All done with the best of intentions: One Kentucky high school after six years of state portfolio tests. *Assessing Writing, 6*, 5–40.
- Carlman, N. (1985). *Variations in the writing performance of grade 12 students: differences by mode and topic*. ERIC Document Reproduction Services No. ED269766.
- Coffman, W. (1966). On the validity of essay tests of achievement. *Journal of Educational Measurement, 3*, 151–156.
- Crawford, L., Helwig, R., & Tindal, G. (2004). Writing performance assessments: How important is extended time? *Journal of Learning Disabilities, 37*, 132–142.
- Crawford, L., & Smolkowski, K. (2008). When a “sloppy copy” is good enough: Results of a state writing assessment. *Assessing Writing, 13*, 61–77.
- Dappen, L., Isernhagen, J., & Anderson, S. (2008). A statewide writing assessment model: Student proficiency and future implications. *Assessing Writing, 13*, 45–60.
- De Ayala, R. J., Dodd, B. G., & Koch, W. R. (1991). Partial credit analysis of writing ability. *Educational and psychological measurement, 51*, 103–114.
- Edmiston, R. W., & Gingerich, C. N. (1942). The relation of factors of English usage to composition. *Journal of Educational Research, 36*, 269–271.
- Elliott, S., Kurz, A., & Neergaard, L. (in press). Large-scale assessment for educational accountability. In K. Harris, S. Graham, & T. Urdan (Eds.), *American Psychological Association Educational Psychology Handbook*. Washington, DC: American Psychological Association.
- Engelhard, G. (1992). The measurement of writing ability with a many-faceted Rasch model. *Applied Measurement in Education, 5*, 171–191.

- Engelhard, G. (1994). Examining rater errors in the assessment of written compositions with a many-faceted Rasch model. *Journal of Educational Measurement*, 31, 93-112.
- Engelhard, G., Gordon, B., & Gabrielson, S. (1991). The influences of mode of discourse, experiential demand, and gender on the quality of student writing. *Research in the Teaching of English*, 26, 315-335.
- Engelhard, G., Gordon, B., Gabrielson, S., & Walker, E. V. S. (1994). Writing tasks and gender: Influences on writing quality of black and white students. *Journal of Educational Research*, 87, 197-209.
- Finlayson, D. S. (1951). The reliability of the marking of essays. *British Journal of Educational Psychology*, 21, 126-134.
- Fuchs, L. S., & Fuchs, D. (2001). Helping students formulate sound test accommodation decisions for students with learning disabilities. *Learning Disabilities Research and Practice*, 16, 174-181.
- Gabrielson, S., Gordon, B., & Engelhard, G. (1995). The effects of task choice on the quality of writing obtained in a statewide assessment. *Applied Measurement in Education*, 8, 273-290.
- Gearhart, M., Herman, J. L., Baker, E. L., & Whitaker, A. K. (1992). *Writing portfolios at the elementary level: A study of methods for writing assessment* (CSE Tech. Rep. No. 337). Los Angeles: University of California, Center for Research on Evaluation, Standards, and Student Testing.
- Gearhart, M., Herman, J. L., Novak, J. R., & Wolf, S. A. (1995). Toward the instructional utility of large-scale writing assessment: validation of a new narrative rubric. *Assessing Writing*, 2, 207-242.
- Gewertz, C. (2011, February 23). Common-assessment consortia add resources to plans. *Education Week*, 30, 8.
- Gewertz, C., & Robelen, E. (2010, September 15). US tests awaiting big shifts: Most states part of groups winning federal grants. *Education Week*, 30(3), 1, 18-19.
- Gilbert, J., & Graham, S. (2010). Teaching writing to elementary students in grades 4 to 6: A national survey. *Elementary School Journal*, 110, 494-518.
- Godshalk, F. I., Swineford, F., & Coffman, W. E., (1966). *The measurement of writing ability*. New York: College Entrance Examination Board.
- Goldberg, G. L., & Walker-Bartnick, L. (1988, April). *Maintaining scoring standards over a rubric transition process*. Paper presented at the American Educational Research Association, New Orleans, LA.
- Graham, S., Harris, K. R., & Hebert, M. (2011). *Assessing Writing*. Report prepared for the Carnegie Corp. of New York, NY.
- Graham, S., Kiuahara, S., McKeown, D., & Harris, K. (2011, February). *Meta-analysis of writing interventions for elementary school children*. Presentation at Writing Research Across Borders II Conference. George Mason University, VA.
- Haladyna, T., & Hess, R. (1999-2000). An evaluation of conjunctive and compensatory standard-setting strategies for test decisions. *Educational Assessment*, 6, 129-153.
- Herman, J. L., Gearhart, M., & Baker, E. L. (1993). Assessing writing portfolios: Issues in the validity and meaning of scores. *Educational Assessment*, 1, 201-224.
- Hillocks, G. (2002). *The testing trap: How state writing assessments control learning*. New York: Teachers College Press.
- Hogan, T. P., & Mishler, C. (1980). Relationships between essay tests and objective tests of language skills for elementary school students. *Journal of Educational Measurement*, 17, 219-227.
- Hollenbeck, K., Tindal, G., & Almond, P. (1999). Reliability and decision consistency: An analysis of writing mode at two times on a statewide test. *Educational Assessment*, 6, 23-40.
- Huang, J. (2008). How accurate are ESL students' holistic writing scores on large-scale assessments? A generalizability theory approach. *Assessing Writing*, 13, 201-218.
- Hunter, D. M., Jones, R. M., & Randhawa, B. S. (1996). The use of holistic versus analytic scoring for large-scale assessment of writing. *Canadian Journal of Program Evaluation*, 11, 61-85.
- Hwang, P. (1930). Errors and improvement in rating English compositions by means of a composition scale. *Contributions to Education*, 417, 1-67.
- Jeffery, J. V. (2009). Constructs of writing proficiency in US state and national writing assessments: Exploring variability. *Assessing Writing*, 14, 3-24.
- Johnson, R. L., Penny, J., & Gordon, B. (2000). The relation between score resolution methods and interrater reliability: An empirical study of an analytic scoring rubric. *Applied Measurement in Education*, 13, 121-138.
- Johnson, R. L., Penny, J., & Gordon, B. (2001). Score resolution and the interrater reliability of holistic scores in rating essays. *Written Communication*, 18, 229-249.
- Johnson, R. L., Penny, J., Fisher, S., & Kuhs, T. (2003). Score resolution: An investigation of the reliability and validity of resolved scores. *Applied Measurement in Education*, 16, 299-322.
- Johnson, R. L., Penny, J., Gordon, B., Shumate, S. R., & Fisher, S. P. (2005). *Language Assessment Quarterly*, 2, 117-146.
- Kan, A. (2007). Effects of using a scoring guide on essay scores: Generalizability theory. *Perceptual and Motor Skills*, 105, 891-905.
- Kiuahara, S., Graham, S., & Hawken, L. (2009). Teaching writing to high school students: A national survey. *Journal of Educational Psychology*, 101, 136-160.
- Kuhlemeier, H., & van den Bergh, H. (1997). Effects of writing instruction and assessment on functional composition performance. *Assessing Writing*, 4, 203-223.
- Lehmann, R. H. (1990). Reliability and generalizability of ratings of compositions. *Studies in Educational Evaluation*, 16, 501-512.
- Linn, R. L., Kiplinger, V. L., Chapman, C. W., & LeMahieu, P. G. (1992). Cross-state comparability of judgments of student writing: Results from the new standards project. *Applied Measurement in Education*, 5, 89-110.
- MacArthur, C., & Cavalier, A. (2004). Dictation and speech recognition technology as test accommodation. *Exceptional Children*, 71, 43-58.
- Moon, T., & Hughes, K. (2002). Training and scoring issues involved in large-scale writing assessments. *Educational Measurement: Issues and Practice*, 21, 15-19.
- Moss, P. A., Cole, N. S., & Khampalikit, C. (1982). A comparison of procedures to assess written language skills at grades 4, 7, and 10. *Journal of Educational Measurement*, 19, 37-47.
- National Commission on Writing. (2003, April). *The neglected R: The need for a writing revolution*. Retrieved from www.collegeboard.com
- Parke, C., Lane, S., & Stone, C. (2006). Impact of a state performance assessment program in reading and writing. *Educational Research and Evaluation*, 12, 239-269.
- Penny, J., Johnson, R. L., & Gordon, B. (2000a). The effect of rating augmentation on inter-rater reliability: An empirical study of a holistic rubric. *Assessing Writing*, 7, 143-164.
- Penny, J., Johnson, R. J., & Gordon, B. (2000b). Using rating augmentation to expand the scale of an analytic rubric. *Journal of Experimental Education*, 68, 269-287.
- Popp, S., & Ryan, J. (2003, April). *The effect of benchmark selection on the assessed quality of writing*. Paper presented at the State and Regional Educational Research Association Annual Meeting, Chicago, IL.
- Popp, S. E. O., Ryan, J. M., Thompson, M. S., & Behrens, J. T. (2003, April). *Operationalizing the rubric: The effect of benchmark selection on the assessed quality of writing*. Paper presented at the American Educational Research Organization, Chicago, IL.
- Purves, A. C. (1992). A comparative perspective on the performance of students in written composition. In A. C. Purves (Ed.), *The IEA study of written composition II: Education and performance in fourteen countries* (pp. 129-152). Oxford: Pergamon Press.
- Royer, J., & Randall, J. (in press). Testing accommodations for students with learning disabilities. In K. Harris, S. Graham, & T. Urdan (Eds.), *American Psychological Association Educational Psychology Handbook*. Washington, DC: American Psychological Association.
- Salahu-Din, D., Persky, H., & Miller, J. (2008). *The nation's report card: Writing 2007* (NCES 2008-468). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Schoonen, R., Vergeer, M., & Eiting, M. (1997). The assessment of writing ability: Expert readers versus lay readers. *Language Testing*, 14, 157-184.
- Sevigny, S., Savard, D., & Beaudoin, I. (2009). Comparability of writing assessment scores across languages: Searching for evidence of valid interpretations. *International Journal of Testing*, 9, 134-150.

FOCUS ON EXCEPTIONAL children

- Shermis, M. D., & Long, S. K. (2009). Multitrait-multimethod analysis of FCAT reading and writing. *Journal of Psychoeducational Assessment, 27*, 296-311.
- Shohamy, E., Gordon, C., & Kraemer, R. (1992). The effect of rater background and training on the reliability of direct writing tests. *Modern Language Journal, 76*, 28-33.
- Shultz, K. (2002). Looking across space and time: Reconceptualizing literacy learning in and out of school. *Research in the Teaching of English, 36*, 356-390.
- Sireci, S. G., Scarpata, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interactions hypothesis. *Review of Educational Research, 75*, 457-490.
- Smagorinsky, P. (1995). Constructing meaning in the disciplines: Reconceptualizing writing across the curriculum as composing across the curriculum. *American Journal of Education, 103*, 160-184.
- Stewart, M. F. & Grobe, G. H. (1979). Syntactic maturity, mechanics of writing, and teachers' quality ratings. *Research in the Teaching of English, 13*, 207-215.
- Swartz, C., Hooper, S., Montgomery, J., Wakely, M., Kruif, R., Reed, M., ... White, K. (1999). Using generalizability theory to estimate the reliability of writing scores derived from holistic and analytic scoring methods. *Educational and Psychological Measurement, 59*, 492-506.
- Swartz, R. & Whitney, D. R. (1985). The relationship between scores of the GED writing skills test and on direct measures of writing. *GED Testing Service Research Studies, 6*, 5-12.
- Underwood, T., & Murphy, S. (1998). Interrater reliability in a California middle school English/language arts portfolio assessment program. *Assessing Writing, 5*, 201-230.
- Veal, L. R., & Hudson, S. A. (1983). Direct and indirect measures for large-scale evaluation of writing. *Research in the Teaching of English, 17*, 290-296.
- Veal, L. R., & Tillman, M., (1971). Mode of discourse variation in the evaluation of children's writing. *Research in the Teaching of English, 5*, 37-42.
- Welch, C. J., & Miller, T. R. (1995). Assessing differential item functioning in direct writing assessments: Problems and an example. *Journal of Educational Measurement, 32*, 163-178.
- Wolfe, E. W. (1997). The relationship between essay reading style and scoring proficiency in a psychometric scoring system. *Assessing Writing, 4*, 83-106.
- Wolfe, E. W., Kao, C., & Ranney, M. (1998). Cognitive differences in proficient and nonproficient essay scorers. *Written communication, 15*, 465-492.
- Wolfe, E. W., & Ranney, M. (1996). Expertise in essay scoring. In D. C. Edelson & E. A. Domeshek (Eds.), *Proceedings of ICLS 96* (pp. 545-550). Charlottesville, VA: Association for the Advancement of Computing in Education.

INDEX • Volume 43

Author Index	Chronological Index of Titles
Al Otaiba, Stephanie (March 2011)	Vocabulary and Reading Comprehension: Direct, Indirect, and Reciprocal Influences (September 2011)
Allor, Jill H. (January 2011)	
Aspy, Ruth (November 2010)	The U.S. Supreme Court and Parental Rights under the Individuals with Disabilities Education Act (October 2010)
Bradfield, Tracy (May 2011)	
Carta, Judith J. (May 2011)	Sustainable Change in Quality of Life for Individuals with ASD: Using a Comprehensive Planning Process (November 2010)
Chard, David J. (January 2011)	
Collins, Terri S. (October 2010)	Assessing the Writing of Struggling Learners (December 2010)
Conroy, Terrye (October 2010)	
Denton, Carolyn A. (March 2011)	A Comprehensive Approach to Improving Reading Fluency for Students with Disabilities (January 2011)
Gajria, Meenakshi (April 2011)	
Greenwood, Charles R. (May 2011)	RTI School-Based Practices and Evidence-Based Models (February 2011)
Grossman, Barry G. (November 2010)	
Henry, Shawn A. (November 2010)	Teaching Word Identification to Students with Reading Difficulties and Disabilities (March 2011)
Jitendra, Asha K. (April 2011)	
Kaminski, Ruth (May 2011)	Reading Comprehension Instruction for Students with Learning Disabilities (April 2011)
Katsiyannis, Antonis (October 2010)	
Linias, Maura (May 2011)	The Response to Intervention (RTI) Approach in Early Childhood (May 2011)
Mellard, Daryl F. (February 2011)	
Meros, Delia (September 2010)	
Myles, Brenda Smith (November 2010)	
Nylander, Donna (May 2011)	
Olinghouse, Natalie G. (December 2010)	
Santangelo, Tanya (December 2010)	
Smith, Sheila M. (November 2010)	
Stern, Amelia (February 2011)	
Wagner, Richard K. (September 2010)	
Woods, Kari (February 2011)	
Yell, Mitchell L. (October 2010)	