

Benjamin J. Lovett  
Lawrence J. Lewandowski  
Cassie Berger  
Rebecca A. Gathje

## *Effects of Response Mode and Time Allotment on College Students' Writing*

*Written essays are a common feature of classroom and high stakes standardized tests at many age levels. However, little is known about how small alterations in the writing task affect students' writing, an issue made more important by the increasing use of task alterations for students with disabilities. In the present study, 140 college students completed a standardized assessment of writing ability under one of two timing conditions (10 minutes, 15 minutes) and with one of two response formats (handwritten, word-processed). Students who used a word processor wrote significantly more than did students who handwrote their essays. However, the extended time allotment only increased students' amount of writing in the word processor condition. Only small differences between the groups' essays were found in spelling, grammar, and related characteristics. Implications of these results for future research and writing assessment policies are discussed.*

**T**raditionally, two types of test items have been used in educational assessment: items in which the student selects the correct response from a set (as seen in multiple-choice, true-false, and matching items) and items in which the student constructs a response on his or her own (as in short answer or essay items). The

latter type of test item has become more common in large-scale standardized testing for several reasons. First, writing samples are thought by many to be the best method of assessing writing ability (Conlan, 1986; Linn & Miller, 2005). Second, computer programs have been developed to score writing samples (Dikli, 2006), reducing the financial and logistical challenges associated with this type of assessment. Third, multiple-choice tests have continued to raise concerns about the ability to measure complex reasoning and problem solving (for a review, see Phelps, 2003).

Although essay tests have certain acknowledged advantages over selected-response tests (e.g., utility in assessing certain higher-order learning objectives), essay tests also have potential limitations. One such limitation is that the response mode of the test may significantly affect examinees' scores; composing an essay using a computerized word processor program may lead to a different score than composing an essay by hand (Russell, 1999). A second limitation is that time limits, which determine the amount of text that students can compose, may significantly affect students' scores, since the amount of text written is a robust correlate of holistic measures of essay quality (Hopkins, 1998; Powers, 2005). In the present study, we explore these issues empirically, asking how time limits and response modes interact to affect students' essay composition.

Before reviewing relevant literature on writing assessment, this study briefly discuss one of the motivations behind the study. An increasing number of students in higher education have diagnoses of common disabilities (e.g., learning disabilities, attention problems, psychiatric disorders, etc.) that may adversely impact their scores on standardized tests (Konur, 2002). In many countries, alterations are made to the administration of the tests (for example, testing accommodations) in the hopes of giving these students a fair chance at showing their skills (Bolt & Thurlow, 2004; Hampton & Gosden, 2004). Extending test time limits and allowing examinees to use computers to write are among the most common accommodations offered, but little is known about how these accommodations affect essay examinations.

An appropriate test accommodation should mitigate performance obstacles of students with disabilities (e.g., large print for a student with visual limitations), while having less of an effect on the performance of non-accommodated students (Fuchs & Fuchs, 2001). However, many testing accommodations provide at least some benefit to students both with and without disabilities (Sireci, Scarpati, & Li, 2005). Indeed, some work shows that students who have poor academic skills but no disability diagnoses benefit more from accommodations than those with

official diagnoses do (e.g., Elliott & Marquart, 2004). The present study, then, was conducted in part for its potential implications for the use of extended time and word processor accommodations, both for students with and without disability diagnoses.

### **Response Mode and Writing Performance**

A number of studies have compared writing produced with and without the aid of a word processor. In a recent meta-analysis, Goldberg, Russell, and Cook (2003) examined 26 of these studies published between 1992 and 2002 using students in K-12 educational settings. These investigators concluded that word processors lead reliably to writing in greater quantities (with a weighted effect size of  $d = .50$ ) and to writing of better quality (judged using a variety of measures, depending on the study; weighted effect size of  $d = .41$ ). These positive effects are consistent with results from earlier research (summarized in Bangert-Drowns, 1993), as well as the few studies that have been published since Goldberg and colleagues' meta-analysis (e.g., Burke & Cizek, 2006).

Variability in word processing's effects has led researchers to search for potential moderators and mediators. Student age appears to be one moderator: Goldberg et al. (2003) found, in regard to both quantity and quality of writing, that middle and high school students benefited more from the use of word processors than elementary school students. Computer skills, a related variable, may also determine students' degree of benefit; Russell (1999) found that middle school students with above-average keyboarding skills benefited substantially from word processing, whereas students with below-average skills actually showed a decrease in essay performance. Similarly, a recent study by Burke and Cizek (2006) found a set of complex interactions between response mode and self-perceived computer skills on the essays written by sixth graders in which higher computer skills were generally associated with more benefit from word processing.

One issue in measuring the effects of response mode on writing performance has been rater bias in the assessment of writing composed using different modes; that is, it is reasonable to suspect that judges who score essays may be more generous to essays that are either handwritten or word-processed. Research has generally shown that judges are more generous to essays that are handwritten. In a landmark study on the topic, Powers, Fowles, Farnum, and Ramsey (1994) asked college students to write two essays on different topics, one in handwritten form and one in word-processed form. The handwritten essays were then word-processed exactly as originally written, without correcting for misspellings and grammar, and the word-processed essays were written

down on paper. These investigators found that when original handwritten essays were word-processed and scored again, the scores decreased significantly, and when the computer-produced essays were handwritten and scored again, the mean increased. In qualitative assessments, raters reported that word-processed essays appeared to be shorter and claimed to give handwritten essays the “benefit of the doubt,” especially when the handwriting was poor. Other studies on the topic have found similar results (e.g., Arnold et al., 1990; MacCann, Eastment, & Pickering, 2002), suggesting that all essays should be presented in the same way (either typed or handwritten) to scorers to avoid rater bias effects.

### **Time Allotment and Writing Performance**

The substantial relationship between essay length and quality is well known (e.g., Hopkins, 1998; Powers, 2005). Time limits clearly have the potential to serve as ceilings in determining essay length. However, few studies have directly examined the effects of time allotment on writing. Many studies have focused on the speededness of tests—that is, the degree to which scores on tests are the product of working quickly (Lu & Sireci, 2007)—but the tests in these studies are rarely measures of writing skill. In their review of the small extant literature on time allotment and writing quality, Powers and Fowles (1996) noted that time allotment has generally, but not always, been found to increase writing performance. In their own comprehensive study, these investigators assigned college students randomly to 40-minute or 60-minute conditions, finding that scores of the examinees in the 60-minute conditions were approximately one half of a standard deviation higher than scores of examinees in the 40-minute conditions; extended time allotments, then, had a meaningful effect on scores.

In a more recent study, Crawford, Helwig, and Tindal (2004) examined several features of fifth- and eighth-grade students' compositions when the students were given either 30 minutes or three days to complete a composition assignment. Each student completed one assignment under each of the two time conditions. Crawford and colleagues found that although fifth-graders' three-day compositions were superior to their 30-minute compositions, no such effect was present for eighth-graders, suggesting that age may be a moderator of timing effects. In addition, time had a greater effect on the scores of students who were receiving special education services, even though the nondisabled fifth-grade students' essays also showed a (smaller) benefit. Of course, giving examinees three days to complete an assignment raises other concerns, since examinees will clearly not be spending the entire interval working on the assignment. In our study, we felt it was important to keep all time

intervals brief enough that examinees could be expected to be working steadily throughout the intervals.

### **The Present Study**

If response mode and time allotment have been found to affect the writing performance of students of varying ages and education levels, it is reasonable to ask how these factors might interact. That is, does the effect of time depend on the response mode, and vice versa? This question is important because many students with disabilities are given alterations in test administration that include simultaneous changes to both response mode and time allotment. In addition, more generally, high stakes tests (e.g., college and graduate admissions tests) and classroom assessments both vary substantially in time allotment and response mode. The present study investigated this issue in a sample of college students whose writing was evaluated under varying conditions.

### **Method**

#### **Participants**

The participants were 140 college students (73 females, 67 males) taking an introductory psychology course at a large, private university in the northeastern United States. They ranged in age from 18 to 22 years old ( $M = 19.1$ ), and most (75%) were either first- or second-year students. The participants' college grade point averages (GPAs) ranged from 2.0 to 4.0 ( $M = 3.2$ ). They completed testing in small group sessions. Each session was randomly assigned, using one of four conditions: hand writing for 10 minutes (HW-10), hand writing for 15 minutes (HW-15), word-processed writing for 10 minutes (WP-10), and word-processed writing for 15 minutes (WP-15). This assignment resulted in four groups of participants ( $n$ s ranging from 29 to 42) that did not differ significantly with respect to gender composition ( $\chi^2 = 1.04, p = .80$ ), age ( $F = 1.38, p = .25$ ), year in school ( $F = .30, p = .83$ ), or college GPA ( $F = .84, p = .48$ ; for more details on these data, see Table 1).

#### **Measures**

The study included three primary measurement tools: one assessed participants' motor speed in either typing or hand writing, a second tool assessed general essay writing skill, and a third tool assessed participants' ability to compose brief sentences quickly. To measure each of these three qualities, we sought tasks that could be scored objectively, and, where relevant, tasks that had evidence of validity showing that the tasks measured the desired qualities.

**Table 1**  
*Group Means on Demographic and Performance Test Variables*

Measure	Handwritten		Word-processed		<i>F</i>	$\eta^2$
	10 min. ( <i>n</i> = 29)	15 min. ( <i>n</i> = 30)	10 min. ( <i>n</i> = 39)	15 min. ( <i>n</i> = 42)		
Age (years)	19.5	19.0	19.0	19.0	1.39	.03
College GPA	3.3	3.1	3.2	3.3	.838	.02
WJ-III Fluency	29.5	27.4	29.1	25.7	3.61*	.07
Writing Speed Task	136.6	137.7	199.7	181.2	28.58***	.39
Words Written	215	216	261	365	22.04***	.33
Context. Conven- tions	11.1	10.5	11.5	9.8	1.96	.04
Context. Language	20.2	19.5	21.0	21.3	2.23	.05
Story Construction	9.9	10.3	11.1	11.7	1.49	.03
TOWL Total	41.2	40.3	43.5	41.8	1.17	.03

\*  $p < .05$

\*\*  $p < .01$

\*\*\* $p < .001$

**Writing/typing speed task.** Students were asked to type or hand write (depending on the experimental condition) the word forms of numbers (i.e., “one, two, three...”) for one minute as quickly as possible. Each character correctly typed/written was counted to form a total score used to measure motor speed.

**Assessment of spontaneous writing skills.** The spontaneous writing materials from Form B of the Test of Written Language, Third Edition (TOWL-3; Hammill & Larsen, 1996) were used to assess participants’ writing quality. Students were shown a picture and given a set length of time to write a story about the picture. The TOWL-3 essay rating procedures yield three separate scores: one for “contextual conventions” (spelling and mechanics), one for “contextual language” (vocabulary, grammar, and syntax), and one for “story construction” (prose, plot, and organization); these three scores are typically summed to generate a “spontaneous writing composite” score. Additionally, the number of words that participants wrote were counted to obtain a writing quantity measure.

The TOWL-3 is one of the most commonly used standardized measures of writing skill (Salvia & Ysseldyke, 2007), and although it is typi-

cally administered to students at the high school level and below, most of the dimensions of writing that it assesses are widely recognized as being important at the college level and beyond (e.g., Jones, 1994). In any case, the present study did not use the TOWL-3 norm-referenced scores to make absolute judgments of students' writing skill levels, but rather used the scores to compare groups that completed the test under different conditions. In addition, although the essay task is brief (10-15 minutes), this is a common length of time allotted for individual essays on both high-stakes and classroom assessments.

**Writing fluency measure.** The writing fluency subtest from the Woodcock-Johnson Tests of Achievement, Third Edition, Form A (WJ-III; Woodcock, McGrew, & Mather, 2001) was used to assess students' ability to write simple expository text quickly. For each item on this test, students were asked to write a sentence that contained all three of the words shown for that item (e.g., dog, big, is); students were given seven minutes to complete as many items as they could. This test has shown adequate reliability ( $\rho = .88$ ) and there is substantial evidence of construct validity, based on correlations with other WJ-III subtests as well as with other measures of writing.

### **Procedures**

We conducted all sessions in the same small, quiet classroom, with no more than 10 students being tested at a time. After completing informed consent forms, we gave participants one minute to complete the motor speed task, and when the minute was over, we passed out the TOWL-3 stimulus picture. We gave participants in the hand written (HW) sessions pencils and lined paper to write with, whereas participants in the word processor (WP) sessions used a word processing program (Microsoft WordPad Version 5.1) to type their compositions on laptop computers (Dell Inspiron 1405 models with standard keyboards and a 14-inch monitor) that we provided. This program does not have spell-checking or similar aids. After completing the TOWL-3 essay assignment, all students completed the WJ-III Writing Fluency subtest, using a hand written response format. We then asked students to complete a brief questionnaire, including demographic information as well as questions about computer usage.

For scoring, we double-spaced and printed all WP essays, and we typed, double-spaced, and printed all HW essays, so that judges would not allow the typed format to influence their ratings of writing quality. We trained two judges to score TOWL-3 essays, and each of the essays was scored by one of the two judges, with 20% of the essays double-scored to ensure reliability across judges. The inter-rater reliability (using a

Pearson correlation) for number of words written was 1.0 (100%); the coefficients for the contextual conventions, contextual language, and story construction subtests from the TOWL-3 were .90, .88, and .76, respectively, and the coefficient for the TOWL total score was .88.

## Results

We present mean scores in Table 1 for demographic and performance variables of all four groups (WP-10 minutes, WP-15 minutes, HW-10 minutes, and HW-15 minutes). There were no significant differences on demographic variables. The differences on performance variables were generally unsurprising; students in the WP conditions wrote more words in their essays and wrote more words during the speed task than students in the HW conditions did, although no significant differences in quality (i.e., TOWL scores) were present in these analyses.

**Table 2**  
*Performance Task Intercorrelations*

Variable	1	2	3	4	5	6	7
1. Speed Task	—	.31*	.05	.11	.09	.13	.28*
2. WJ-III Fluency	.25*	—	.18	.19	-.24	.00	.25
3. Context. Conventions	.52**	.33**	—	.19	.16	.69**	-.04
4. Context. Language	.46**	.15	.52**	—	-.05	.45**	.29*
5. Story Construction	.46**	.24*	.59**	.63**	—	.72**	-.09
6. TOWL Total	.57**	.28*	.83**	.85**	.88**	—	.03
7. Words Written	.34**	.18	.40**	.54**	.60**	.61**	—

*Note.* Correlations above the diagonal were observed in students in the hand written conditions ( $n = 59$ ); those below the diagonal were observed in the word-processed condition ( $n = 81$ ).

\*  $p < .05$

\*\*  $p < .01$

A second group of analyses examined correlations among writing measures for the HW groups separately from the WP groups. Because time allotment had very little effect on correlations, data was collapsed across time conditions. Table 2 displays the correlation matrices for both response modes, and in some ways the correlations are quite different. For example, the writing/typing speed score correlated significantly with all writing measures for the WP conditions, but did not correlate



significantly with any quality measures for the HW groups. Similarly, the number of words written by the WP groups correlated significantly with all writing measures, yet much lower correlations were found between these measures for the HW groups. Writing speed and fluency, then, were better predictors of quality (and vice versa) when essays were completed using a word processor.

**Table 3**  
*Two-Way Analyses of Variance for TOWL-3 Essay Scores and Numbers of Words Written*

Variable and Source	MS	F	$\eta_p^2$
Contextual Conventions			
Response Mode	.44	.04	.00
Time	45.69	4.25*	.03
Response Mode x Time	9.30	.87	.01
Contextual Language			
Response Mode	52.0	5.57*	.04
Time	.78	.09	.00
Response Mode x Time	9.23	.99	.01
Story Construction			
Response Mode	55.08	3.77	.03
Time	9.27	.63	.00
Response Mode x Time	.24	.02	.00
TOWL Total Score			
Response Mode	195.12	3.17	.02
Time	21.16	.34	.00
Response Mode x Time	.23	.01	.00
Number of Words Written			
Response Mode	325338.44	38.67***	.22
Time	93458.25	11.11**	.08
Response Mode x Time	90251.87	10.73**	.07

\* $p < .05$

\*\* $p < .01$

\*\*\* $p < .001$

Table 3 presents analysis of variance (ANOVA) models that examined the interaction between response mode and time allotment. Of most interest were the Total TOWL score (quality) and essay length (quantity)

measures. With regard to quality, there were no differences found for writing format/response mode or time condition and no interaction ( $p$ -values ranging from .08 to .95). However, when considering quantity, we found an interaction of time and response mode conditions,  $F(1, 139) = 10.73$ ,  $p < .01$ . The main effects for both time and response mode were also significant ( $p < .01$ ). Further analysis revealed that the WP groups wrote significantly longer essays than the HW groups, and additional time increased essay length only in the WP conditions. The WP-15 group ( $M = 365$  words) significantly outperformed the WP-10 group ( $M = 261$  words), yet there was no difference in essay length between HW-10 ( $M = 215$  words) and HW-15 ( $M = 216$  words) groups.

The three remaining TOWL scores were of less interest in that they are constituents of the Total TOWL score already analyzed for writing quality. However, analyses of these variables yielded a small effect for time on the Contextual Conventions subtest, with slightly better scores for the groups receiving 10 minutes to write. We also found a small effect of response mode on the Contextual Language subtest score, with those in the WP condition outperforming the HW groups. These analyses are also presented in Table 3.

## Discussion

We examined the effects of two common test administration variables, time allotment and response mode, on the length and quality of written essays in college students. The results in this study were mixed. No effect of test conditions was found for writing quality, while the word processor proved to be better than hand written response mode for essay length (quantity). The effect was larger when students had more time to write. Increasing the time allotment had no effect when students were hand writing essays, but the same increase in time allotment had a greater effect when essays were completed using a word processor. Moreover, there was no relationship between essay length and quality ( $r = .03$ ) for hand written essays, yet this relationship was quite strong ( $r = .61$ ) for word-processed essays. These and other correlations among measures for both response modes suggest that the relationship between writing speed and writing quality is sensitive to whether compositions are typed or hand written.

The findings in this study are consistent with many other studies examining the effects of administration conditions on writing performance. For instance, Russell and Haney (1997) found, as did we, that longer essays were composed on a computer rather than by hand. Our finding that essay length correlated significantly with quality is also congruent with past work (e.g., Espin, Scierka, Skare, & Halverson, 1999;

Mattern, Camara, & Kobrin, 2007), although only for the word processor response mode. Finally, our finding that word-processed essays were longer but not of higher quality recalls the results of Truell, Alexander, and Davis (2004), who found that marketing students who used word processors completed a classroom test more quickly but did not obtain higher scores. In sum, our study coincides with others in the literature, showing that word processing affords an opportunity to be more time efficient and/or more generative in a fixed time period, and that it may or may not improve writing quality, depending on the task, scoring procedures, sample under study, and other features.

However, two discrepancies between our findings and previous work raise interesting questions. First, why did increased time help only students who word-processed their essays? Composition is generally understood to be a high-effort academic task (for a review, see Torrance & Galbraith, 2006), and if students experienced hand writing essays as requiring more effort than word processing them (given that our students were used to writing using computers), students may have been more motivated to expand and revise their essays with the extra time when using a word processor, whereas students in the hand written conditions may have been satisfied with a minimally adequate essay product.

A related question is raised by our correlations between essay length and quality: why was a positive correlation between these variables only found in the word processor conditions? Part of the answer may lie in the fact that the variables exhibited higher variability in the word-processor conditions than in the hand written conditions; the standard deviation for essay length was smaller for all hand written essays than for word-processed essays, and lack of variability restricts the possible values of the correlation coefficient (Nunnally & Bernstein, 1994). Another possible explanation is that word processing makes it easier to revise text by deleting text that is repetitive, irrelevant, or that otherwise detracts from the essay (MacArthur, 2006), and so word-processed essays may be less likely to contain additional text that does not contribute to the essays' purposes. If hand written essays had more irrelevant text (owing in part to the higher effort required to revise), any relationship between length and quality would be attenuated significantly.

### **Implications for Testing Accommodations**

Two of our findings appear to have implications for the provision of extended time and word processor testing accommodations. First, when combined, the longer time allotment and word processor use led to students writing substantially longer essays, and so if length of composition is important on a writing test, test users and administrators should

know that students with disabilities who are given extended time *and* word processor accommodations are given a very substantial benefit over their classmates, at least in terms of essay length. Since the present study showed that the effects of extended time and word processor are found in students without disabilities, there may be little rationale for restricting these accommodations to students with disabilities (see Zuriff, 2000, for more on the logic of restricting accommodations).

A second implication of our findings for testing accommodations comes from the interaction between time allotment and response mode. Our finding that extended time benefited only the fluency of students who word-processed their essays suggests that providing extended time alone as an accommodation may not be beneficial to students on writing tests, unless a word processor is also provided. Since word processor accommodations may lower the requirements for examinee effort and self-motivation, more students may take advantage of any extra time that is provided.

A new trend in the testing accommodations arena is the design of tests that all examinees can access, so that accommodations are unnecessary, a concept called universal test design (Thompson, Johnstone, Anderson, & Miller, 2005). A writing test with universal design might be taken using a word processor and under conditions of generous time limits, equivalent to providing the accommodations to all examinees. With regard to word processors, computer-based writing is becoming so ubiquitous that some test administrators (e.g., those of the GRE) have moved to completely computer-based testing. Alternatively, word processors may be given as an option to any examinees that request them. However, the "optional" approach should be used cautiously, as the present study has shown that correlations between different essay scores are quite different for essays that are word-processed.

### Limitations and Directions for Future Research

The present study has several limitations, each of which suggests future research that would be beneficial. First, although we were interested in the effects of common testing accommodations, our lack of a group of participants with disabilities precluded a complete test of appropriateness. Although some scholars (e.g., Phillips, 1994; Zuriff, 2000) would consider our finding that nondisabled students benefited from accommodations to be sufficient evidence against their being restricted to students with disabilities, other scholars (e.g., Fuchs & Fuchs, 2001; Sireci, Scarpati, & Li, 2005) argue that as long as students with disabilities benefit *more* from an accommodation than nondisabled students, the accommodation is appropriate, and as such, future work should include

students with disabilities.

A second extension would involve manipulating the test administration variables as within-subject factors, to determine whether individuals' scores actually improved as a function of time allotment and response mode. Examinees could complete one essay under each of several conditions, and for time allotment, they could even be stopped after a certain length of time, and then given an extension to see how the same essay changes. When extended time has been studied in the context of reading and mathematics performance (e.g., Lewandowski, Lovett, Parolin, Gordon, & Coddington, 2007), this procedure has been useful, although writing is a somewhat different task, and changes to an essay may be difficult to record.

A third extension relates to our writing stimulus (the TOWL-3 stimulus), which was not normed on college samples or designed specifically for college students, and was only a single, brief essay prompt. Given the importance of interest in motivating students to write, various writing prompts with a range of time demands should be tested to determine which stimuli lead to the greatest motivation, including allowing students choice in selecting a topic to write about. Of course, it should be noted that our own task was of a similar length to those found on many standardized writing tests, suggesting a degree of generalizability to our findings.

The present study, then, should open the door to future research that examines the effects of task characteristics on the writing performance of students with and without disabilities, and with a range of academic skills. On a more practical note, test administrators and users might take away two messages from the present study's results. First, use of a word processor may be necessary for extended time accommodations to exert their full effects. Second, many college students appear to use word processing as their default mode of composition, suggesting that writing tests that require hand written responses may put these students in an unfamiliar situation, compromising the validity of the assessment. Therefore, in high-stakes testing situations, word processors should be considered as an element of universal test design (Thompson et al., 2005), furthering the goal of appropriate assessment for *all* students.

## References

- Arnold, V., Legas, J., Obler, S., Pacheco, M. A., Russell, C., & Umbdenstock, L. (1990). *Do students get higher scores on their word-processed papers? A study of bias in scoring hand-written vs. word-processed papers*. Whittier, CA: Rio Hondo College. Retrieved from ERIC database (ED 345818).
- Bangert-Drowns, R. L. (1993). The word processor as an instructional tool: A meta-analysis of word processing in writing instruction. *Review of Educational Research, 63*, 69-93.

- Bolt, S. E., & Thurlow, M. L. (2004). Five of the most frequently allowed testing accommodations in state policy. *Remedial and Special Education, 25*, 141-152.
- Burke, J. N., & Cizek, G. J. (2006). Effects of composition mode and self-perceived computer skills on essay scores of sixth graders. *Assessing Writing, 11*, 148-166.
- Conlan, G. (1986). "Objective" measures of writing ability. In K. L. Greenberg, H. S. Wiener, & R. A. Donovan (Eds.), *Writing assessment: Issues and strategies* (pp. 109-125). New York: Longman.
- Crawford, L., Helwig, R., & Tindal, G. (2004). Writing performance assessments: How important is extended time? *Journal of Learning Disabilities, 37*, 132-142.
- Dikli, S. (2006). An overview of automated scoring of essays. *Journal of Technology, Learning, and Assessment, 5*(1), 1-36.
- Elliott, S. N., & Marquart, A. M. (2004). Extended time as a testing accommodation: Its effects and perceived consequences. *Exceptional Children, 70*, 349-367.
- Espin, C. A., Scierka, B. J., Skare, S., & Halverson, N. (1999). Criterion-related validity of curriculum-based measures in writing for secondary school students. *Reading & Writing Quarterly, 14*, 5-27.
- Fuchs, L. S., & Fuchs, D. (2001). Helping teachers formulate sound testing accommodation decisions for students with learning disabilities. *Learning Disabilities Research and Practice, 16*, 174-181.
- Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *Journal of Technology, Learning, and Assessment, 2*(1), 1-51.
- Hammill, D. D., & Larsen, S. C. (1996). *Test of Written Language*. (3rd ed.). Austin, TX: Pro-Ed.
- Hampton, G., & Gosden, R. (2004). Fair play for students with disability. *Journal of Higher Education Policy and Management, 26*, 225-238.
- Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation*. (8th ed.). Boston, MA: Allyn & Bacon.
- Jones, E. A. (1994). Defining essential writing skills for college graduates. *Innovative Higher Education, 19*, 67-78.
- Konur, O. (2002). Assessment of disabled students in higher education: Current public policy issues. *Assessment & Evaluation in Higher Education, 27*, 131-152.
- Lewandowski, L., Lovett, B. J., Parolin, R., Gordon, M., & Coddling, R. S. (2007). Extended time accommodations and the mathematics performance of students with and without ADHD. *Journal of Psychoeducational Assessment, 25*, 17-28.
- Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching*. (9th ed.). Upper Saddle River, NJ: Merrill.
- Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice, 26*(4), 29-37.
- MacArthur, C. A. (2006). The effects of new technologies on writing and writing processes. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 248-262). New York: Guilford.

- MacCann, R., Eastment, B., & Pickering, S. (2002). Responding to free examination questions: computer versus pen and paper. *British Journal of Educational Technology, 33*, 173-188.
- Mattern, K., Camara, W., & Kobrin, J. L. (2007). *SAT Writing: An overview of research and psychometrics to date* (Research Note RN-32). New York: College Board.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. (3rd ed.). New York: McGraw-Hill.
- Phelps, R. P. (2003). *Kill the messenger: The war on standardized testing*. Edison, NJ: Transaction.
- Phillips, S. E. (1994). High-stakes testing accommodations: Validity versus disabled rights. *Applied Measurement in Education, 7*, 93-120.
- Powers, D. E. (2005). "Wordiness": A selective review of its influence, and suggestions for investigating its relevance in tests requiring extended written responses (ETS Research Memorandum RM-04-08). Princeton, NJ: Educational Testing Service.
- Powers, D. E., & Fowles, M. E. (1996). Effects of applying different time limits to a proposed GRE writing test. *Journal of Educational Measurement, 33*, 433-452.
- Powers, D. E., Fowles, M. E., Farnum, M., & Ramsey, P. (1994). Will they think less of my handwritten essay if others word process theirs? Effects on essay scores of intermingling handwritten and word-processed essays. *Journal of Educational Measurement, 31*, 220-233.
- Russell, M. (1999). Testing on computers: A follow-up study comparing performance on computer and on paper. *Education Policy Analysis Archives, 7*(20). Retrieved January 28, 2008 from <http://epaa.asu.edu/epaa/v7n20>.
- Russell, M., & Haney, W. (1997). Testing writing on computers: An experiment comparing student performance on tests conducted via computer and via paper-and-pencil. *Education Policy Analysis Archives, 5*(3). Retrieved January 28, 2008 from <http://epaa.asu.edu/epaa/v5n3.html>.
- Salvia, J., & Ysseldyke, J. E. (2007). *Assessment in special and inclusive education*. (10th ed.). Boston, MA: Houghton Mifflin.
- Sireci, S. G., Scarpati, S., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research, 75*, 457-490.
- Thompson, S. J., Johnstone, C. J., Anderson, M. E., & Miller, N. A. (2005). *Considerations for the development and review of universally designed assessments* (NCEO Technical Report No. 42). Minneapolis, MN: National Center on Educational Outcomes.
- Torrance, M., & Galbraith, D. (2006). The processing demands of writing. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 67-80). New York: Guilford.
- Truell, A. D., Alexander, M. W., & Davis, R. E. (2004). Comparing postsecondary marketing student performance on computer-based and handwritten essay tests. *Journal of Career and Technical Education, 20*, 69-78.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson Tests of Achievement*. (3rd ed.). Itasca, IL: Riverside Publishing.

Zuriff, G. E. (2000). Extra examination time for students with learning disabilities: An examination of the maximum potential thesis. *Applied Measurement in Education*, 13 (1), 99-117.

---

**Benjamin J. Lovett**, Ph.D., is an assistant professor of psychology at Elmira College, where his research focuses on conceptual and psychometric issues in psychoeducational assessment and psychiatric diagnosis. **Lawrence J. Lewandowski**, Ph.D., is Meredith Professor of Teaching Excellence in the psychology department at Syracuse University. **Cassie Berger**, M.S., is a Ph.D. candidate in school psychology at Syracuse University. **Rebecca A. Gathje**, Ph.D., is a postdoctoral fellow at the Munroe-Meyer Institute and a clinician working out of the Grand Island Behavioral Health Clinic in Grand Island, Nebraska.