Practical Assessment, Research, and Evaluation

2006

Elementary school teachers' perceptions of curriculum-based measures of written expression

Kristin A. Gansle

Donna N. Gilbertson

Amanda M. VanDerHeyden

Follow this and additional works at: https://scholarworks.umass.edu/pare

Recommended Citation

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Elementary School Teachers' Perceptions of Curriculum-Based Measures of Written Expression

Kristin A. Gansle, Louisiana State University
Donna N. Gilbertson, Utah State University
Amanda M. VanDerHeyden, University of California, Santa Barbara

Despite evidence indicating the general acceptability of curriculum-based measures (CBM) to teachers (e.g., Eckert, Shapiro, & Lutz, 1995), there has to date been no empirical evidence demonstrating the relative level of acceptability of measures of written expression that might set the stage for their increased adoption by teachers. Results of an exploratory survey designed to assess teachers' perceptions of various measures of written expression are reported. The relative acceptability of the measures for this sample of 335 teachers from several locations throughout the United States, the relationship of the measures to demographics, and directions for future research are discussed.

For students who experience failure or very limited success in the general education school curriculum, the traditional norm-referenced assessments of achievement commonly used by educators in the special education referral, eligibility determination, and placement process are poorly suited to the development, evaluation, and refinement of interventions designed to remediate those academic problems (Shinn, 1986, 1989). In these times of increasing accountability in education (e.g., No Child Left Behind, 2001), it is critical to evaluate the effectiveness of the procedures used to remediate those specific academic deficits so that they can be continued if successful, and changed or discarded if not. For both of these kinds of service delivery, formal evaluation must proceed on an ongoing basis, and often dictates that data are collected every day or every few days.

This need for repeated, quality assessment data targeted toward identifying specific academic deficits is not one that easily can be filled. The available academic achievement tests are ill-suited to the demands of this kind of assessment. First, published achievement tests are designed to be stable over time; they are deliberately made reliable and are therefore not sensitive to changes that occur on a daily basis. Second, published tests are not designed to be administered multiple times over the weeks-long duration of an intervention or a skill instruction set in special education of similar duration. Without frequent administration, it is impossible to demonstrate ongoing student growth in specific content areas. Third, published tests tend to be given once or twice annually and usually assess students' skills in a number of areas that are relevant to an entire year's growth in given curricula over broad skills area such as Mathematics or Language Arts. However, short-term gains from interventions that focus on specific skill areas will not necessarily be demonstrated using this type of

test. Fourth testing often represents high costs in terms of administration time and personnel resources.

Teacher-made tests have problems as well. Although these tests correspond closely to content that is taught in the classroom prior to their administration, it would be unusual for growth over time in specific skill areas to be easily demonstrated after multiple administrations. For example, even weekly spelling tests, which might seem easily comparable over weeks, would only tell how a student performs on different words each week. There is no way to know the cumulative effect of a spelling intervention on a group of common words. In mathematics, scores on weekly tests tell how the scores change on the different skills and problems that are being taught at that time. Variable difficulty is an additional problem—even reading passages from a single grade level reading text can vary several grade levels when subjected to various reading formula calculations (Witt, Daly, & Noell, 2000).

Curriculum-based measurement (CBM) was designed to address many of those issues. CBM is sensitive to change over short periods of time, can be administered frequently for progress monitoring, and requires relatively little time and resources to administer and score (Jenkins, Deno, & Mirkin, 1979; Marston, 1989). Teachers and classroom paraprofessionals are routinely taught to administer and score the products of CBM, and can do so reliably (Tindal, 1989). One of the best aspects of CBM is that it has been validated for use as "dynamic indicators of basic skills" or DIBS (Shinn & Bamonto, 1998); it provides data from which the effectiveness of an intervention or instructional program can be determined. In other words, it eases and standardizes the process of formative evaluation for improving students' academic outcomes.

CBM differs from teacher-made tests in that it is tied to specific basic skills that students in early grades must master in order to be successful in later curriculum. In reading, one common CBM assessment procedure involves presentation of a grade-level reading passage that has been verified to be on grade level to a student and recording how many words that student read correctly and/or incorrectly over the duration of one minute.

Median scores of fluency and errors over three passages are recorded, and these scores, taken several times in a week, can be used to determine student progress. In reading, fluency is a prerequisite for comprehension (Marston, 1989). This positive relationship between fluency and what is ultimately important, comprehension, supports the validity of CBM (Fuchs, Fuchs, & Maxwell, 1988). For mathematics, assessments are usually tied to a specific skill. For example, for addition facts, a fact "probe" is designed, on which many addition facts problems are presented. This probe or worksheet is administered to the student, who is instructed to complete as many problems as possible over two minutes. The number of digits that the student writes correctly is tracked over time to establish the effectiveness of instructional or intervention procedures. The same can be done for subtraction, mixed facts, or any type of problem, as long as the probes administered are of similar difficulty and are comparable when considering the specific skills that are assessed.

In reading and mathematics, the curriculum-based measures that are collected appear to be tied to the ultimate goals in those areas. However, for assessment in written expression, the connection between the curriculum-based measures used and the ultimate goal, competent writing, is not as clear. A wide variety of curriculum-based measures has been investigated for use and is recommended for assessment and progress monitoring in written expression, including total words written (TWW), words spelled correctly, correct word sequences, correct letter sequences, and long words, in addition to a variety of rate-based measures (see Howell, Fox, & Morehead, 1993; Marston, 1989 for reviews). However, CBM for written expression has not received the same attention in the literature as CBM for reading or mathematics, and anecdotal reports indicate that teachers find some specific curriculum-based measures to be less acceptable than others for measuring student achievement (Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002). In other words, teachers report that knowing how many words a student has written over the course of three minutes in response to a writing prompt (TWW) may only tangentially be related to the quality of that written product. The same has been said about other curriculum-based

Gansle, Gilbertson, & VanDerHeyden, Teachers' Perceptions of Writing Measures

measures of writing; however, to date there is no empirical evidence that describes what teachers think of these measures of written expression, and whether what they think is related to whether they would use these measures. This can be described as the acceptability of assessment procedures.

The concept of acceptability was originally applied to intervention, and was described as "judgments by laypersons, clients, and others of whether treatment procedures are appropriate, fair, and reasonable for the problem or client" (Kazdin, 1981, p. 493), and has been applied to assessment, though not to the same extent to which it has been applied to treatment or intervention (Allinder & Oats, 1997; see Elliott, 1988; Elliott et al., 1991, for reviews on treatment acceptability). It is important due to the possibility of a link between acceptability and use of a measure or of an intervention. If a lack of acceptability leads to low intervention integrity, decreasing the probability that positive outcome will be realized (Gresham, 1989), there is little point to designing those interventions at the outset. Similarly, if measures are unacceptable, they may be less likely to be used, and therefore less likely to contribute to treatment-valid intervention development.

The acceptability of assessment has been investigated in the literature. In addition to a study that discussed the acceptability of curriculum-based assessment (CBA) to school psychologists (e.g., Shapiro & Eckert, 1993), Eckert, Shapiro, and Lutz (1995) investigated teachers' ratings of acceptability of CBA methods using the Assessment Rating Profile (ARP, Kratochwill & Van Someren, 1984, as cited in Eckert, et al., 1995). The ARP was designed to measure the acceptability of assessment scales and methods, and was a refinement of a similar scale designed by Witt and Martens (Intervention Rating Profile, 1983) to assess teachers' perceptions of intervention acceptability. Assessment acceptability, as used in this scale, addresses issues concerning assessment such as severity of the problem, effectiveness in identifying problems, time cost, willingness to use, and the perceived benefit to the child, among others. The factors addressed did not include the acceptability of specific measures and their relationship to the constructs of interest, in favor of assessment of the more general aspects of CBA. Eckert and

colleagues (1995) compared the acceptability of CBA to published, norm-referenced tests, and found that general and special education teachers rated CBA procedures as highly acceptable, and more acceptable than published, norm-referenced tests.

Allinder and Oats (1997) investigated the effects of acceptability of CBM on special education teachers' use in mathematics, and determined that those who perceived CBM as more acceptable used more probes than those who did not, and set more ambitious goals for their students than the teachers who rated the acceptability of CBM as low. In addition, the students of teachers who rated CBM as more acceptable made greater gains in mathematics performance than those in the low CBM acceptability group.

Curriculum-based methods of assessment have been shown to be more acceptable to teachers than published, norm-referenced tests, and greater acceptability has been associated with greater use of interventions. However, what remain unknown are teachers' perceptions of the suitability of individual curriculum-based measures to determine specific skills. Research cited above suggests that teachers find CBM, in general, acceptable. Investigators should establish that teachers find specific CBM procedures acceptable over the range of specific curriculum-based measures in academic areas. This is especially true for written language, where anecdotal reports indicate that the acceptability of procedures for assessment appears to be lower than for CBM in other academic areas.

In response to teachers' concerns regarding the common curriculum based measures for written expression such as total words written, correct word sequences, and words spelled correctly, Gansle and colleagues (2002) evaluated a variety of curriculum-based measures of writing, some of which previously had been evaluated in the literature, some of which had not. Based on this evaluation, they found that there is a variety of indices of writing skill that were more closely related to criterion measures such as the Iowa Tests of Basic Skills (ITBS; Hoover, Hieronymus, Frisbie, & Dunbar, 1996) than total words written or words spelled correctly. For example, the number of correct punctuation marks, as well as the number of complete sentences was more strongly related to

ITBS scores than were total words written or words spelled correctly. The measures investigated included total words written, the total numbers of nouns, verbs, and adjectives written for each passage, the number of words that were spelled correctly in isolation and contained 8 or more letters, and the total number of words spelled correctly. In addition, the total number of punctuation marks, the number of correct punctuation marks, correct uses of capital letters, number of complete sentences (started with a capital letter, had recognizable subject, verb, and ending punctuation), total words in all sentences counted as complete sentences, correct word sequences, sentence fragments, and simple sentences were examined.

Further investigation using the measures best related to the criterion measures indicated that while not sensitive to brief intervention, measures such as total punctuation marks and correct word sequences were better correlated with the Writing Samples subtest of the Woodcock-Johnson individual achievement test than was total words written (Gansle et al., 2004).

Despite the promise that the measures showed for evaluating students' skills in written language, there is no evidence to suggest that teachers perceive that any of the measures is more appropriate or acceptable for use in evaluating those skills than is total words written or words spelled correctly. As a result, there is no evidence that even if they can be validated, they will be deemed acceptable and used by teachers in schools. Given that primary interest rests with use of effective technologies in schools toward better student outcomes, this study was designed as an exploratory investigation of teachers' perceived acceptability of a variety of curriculum-based measures of writing skills, and as a point of reference, looked at other measures of written expression with which teachers are likely to have at least passing familiarity. These measures included curriculum-based measures of written language that previously had been researched, newer curriculum-based measures, and more traditional measures of student writing skill.

METHOD

Participants

The sample consisted of 335 teachers from schools located in eight states in various geographic regions throughout the United States: Louisiana (16%), California (6%), Iowa (1%), Ohio (6%), New York (29%), Utah (5%), Idaho (17%), and Arizona (20%). To obtain survey participants, the investigators enlisted the assistance of colleagues in higher education and K-12 schools to recruit school teachers and administrators in rural, suburban, and urban areas throughout the United States. In one state (Iowa), teachers were recruited during a professional development workshop run by the Area Education Agency for local school personnel. Although schools were not selected randomly, an effort was made to include participants from a variety of U.S. geographic regions.

Participants were recruited via a three-step process. First, school principals were contacted, the study was described, and their consent requested. Verbal permission was obtained from the principals to contact their teachers. Second, depending on the school district, written permission was obtained from any additional boards or personnel required by the principal and/or the school district. Third, all teachers in each school were contacted and included in the study if they agreed to complete the survey and returned it to the research team after they were given information about the study.

Survey Development

The survey was constructed for this study, based on the results of previous studies of alternative curriculum-based measures in writing (Gansle et al., 2002, 2004) and was designed to examine teachers' preferred assessment and scoring methods which best represent student writing abilities when using curriculum based measurement and other assessment procedures. Several phases were followed in the construction of the survey. A literature review was conducted to determine potential scoring methods that have been used to assess writing ability. Based on information obtained from the literature, an initial draft of the survey was developed to obtain teachers' ratings on individual types of assessment and scoring methods. Second, the survey was revised according to feedback from several research and educational

Gansle et al.: Elementary school teachers' perceptions of curriculum-based measu

Practical Assessment Research & Evaluation, Vol 11, No 5 5
Gansle, Gilbertson, & VanDerHeyden, Teachers' Perceptions of Writing Measures

professional experts regarding clarity of content and directions as well as ease of administration.

A cover letter providing a rationale, detailing the purpose of the study, and assurance of confidentiality was attached to each survey. The four-page survey consisted of items that were divided into subsections: scoring methods, assessment methods, and demographics. To introduce teachers to the survey topic, a brief description of CBM writing assessment procedures (as per Shinn, 1989) was presented at the beginning of the survey. Operational definitions of specific scoring methods as used in Gansle et al. (2002) were provided for teachers. These were the same definitions that were provided to scorers in the 2002 study; they are available from the first author on request. The teachers' judgments of individual scoring methods were measured by asking them to rate the extent to which they thought the scoring items were representative of students' writing skills. They indicated that items were "not at all important indicators of student writing skill" by selecting a "1" on the scale to "extremely important indicators of student writing skill" by selecting a "7" on the scale.

In addition to curriculum-based measures, respondents were asked to use the same scale to indicate their perceptions of holistic ratings, group achievement test scores, and individual achievement test scores. Holistic ratings were described: "Read a sample of student writing and score it based on your judgment of the overall quality of the writing." Research and educational experts suggested that most classroom teachers are not conversant in the specific strengths and weaknesses of large numbers of commonly-used standardized tests of achievement; consequently, items describing individual tests were not included in the survey. One item describing group and one item describing individual achievement tests were included. Examples of common tests were provided so that teachers would know which kinds of tests were indicated. For the former, the ITBS and the Metropolitan Achievement test were suggested as examples, and for the latter, the Key Math, the TOWL, and the Woodcock-Johnson were suggested as examples. Teachers were advised to ask questions if they did not understand any of the descriptions on the survey or wanted additional examples of group or individual achievement tests,

and clarifications were provided whenever they were requested.

Respondents were asked to indicate their gender, race, years of experience, grade(s) currently taught, highest level of education, and primary area(s) of teacher certification. Finally, teachers were asked to rate the extent to which they agreed with the following statement on a "1" (strongly disagree) to "7" (strongly agree) Likert scale: "I have received training in the use of Curriculum-based Measurement." This item was intended to provide investigators with a snapshot of whether and how much training the respondent had completed without spending a great deal of time specifying the amount and intensity of whatever training there had been. Copies of the survey and cover letter are available from the first author.

Data Collection

The surveys and cover letter were distributed to teachers at each participating school during faculty meetings. Teachers were provided a brief explanation of the purpose of the survey and how the information would be used, and were given time to read the cover letter that was attached to the survey instrument. If teachers chose to participate, they completed the survey before the end of the meeting and returned it to the person who distributed them. Surveys were distributed only one time to each participant. Individuals who distributed the surveys reported that they distributed 415 surveys during faculty meetings and workshops. Three hundred thirty-eight surveys were returned, for an 81% return rate. The high return rate is likely due to the teachers' superiors asking for their participation during a meeting at which they were present.

Exclusion of Invalid Surveys or Items

Three of the original 338 surveys returned were discarded as the respondents filled out less than one of the four pages of the survey. If a teacher endorsed more than one importance rating on the Likert scale for any of the items rated in that way, or if the answer to a given question was not clearly marked, that item was considered missing and not entered into the database.

RESULTS

Although 335 surveys were returned with usable data, not all teachers chose to answer all of the questions on the survey. Therefore, where applicable, data are reported as a percentage of or a mean for the respondents who completed that particular question.

Demographics

Teachers reported a mean 11.1 years (sd = 9.6) teaching experience. General education was the current teaching assignment for 82.1% of the teachers, special education for 11.3% of the teachers, and "other" for 6.6% of the teachers in the sample. Teachers were asked to indicate the grade levels included in their current assignments. These are not listed as percentages, as it was possible for teachers to list more than one grade level. Fifty-two of the teachers reported that their current teaching assignment included kindergarten or younger students, 92 reported first grade, 92 reported second grade, 93 reported third grade, 95 reported fourth grade, and 76 reported fifth grade. Although the emphasis on progress monitoring for written expression may increase as students move toward high-stakes testing years (often fourth and seventh or eighth grades), no assumptions were made regarding the extent to which written expression is the focus of instruction in lower versus higher grades. The certification structures for most states tend to group elementary certification between kindergarten and (approximately) fifth grade. This suggests that the grade taught by any given teacher may be a historic artifact, and that any of the teachers could be teaching any of the elementary grades as the environment demanded. In other words, just because a teacher is teaching first grade does not mean that he or she is unqualified to teach fifth grade. On the contrary, elementary teachers are commonly trained and certified to teach pre-literacy skills as well as more advanced writing skills.

The teachers who completed the survey were from public schools in eight states in the United States, from the northeastern, midwestern, western, and southern sections of the country. In this initial look at teachers' perceptions, efforts were made to get responses from a distribution of teachers throughout the country and within different population density areas. Forty-three percent described their districts as suburban, 35% as rural, and 22% described their school districts as urban. The majority (89%) reported holding a bachelor's degree in elementary education. Thirty-nine percent reported holding a master's degree. Ninety percent reported elementary education as their primary teacher certification area, and 100% were certified in either elementary or special education. Ninety percent of the teachers who responded to the survey were female. Eighty-six percent of the teachers were Caucasian, 10% of the teachers were African-American, 2.2% were Hispanic, and 1.3% were of Asian descent.

Survey Ratings

The means, standard deviations, and 95% confidence intervals for all of the 17 writing variables were calculated. These data can be found in Table 1. A visual description of the overlap of the confidence intervals can be found in Figure 1.

Elementary school teachers in this sample scored holistic ratings as the most important indicator of student writing skill (M = 5.89), followed by the number of complete sentences (M = 5.63) and the number of correct capitals (M = 5.10). The 95% confidence intervals for holistic ratings, complete sentences, and correct capitals did not overlap with any of the other variables rated. Teachers judged their own holistic ratings as more representative of student writing skills than any other measure evaluated in this study. This is especially noteworthy given the fact that many of the measures rated have been investigated in the literature, and have a large body of data to support their validity and reliability. These teachers appear to trust their own ratings of student skills better than any of the measures presented here.

Number of complete sentences and correct capitals followed holistic ratings and were distinct from each other. These teachers appear to believe that the number of complete sentences a student can write within three minutes is indicative of their writing skills, as is the number of correct capitalizations they write within the same time period.

Gansle, Gilbertson, & VanDerHeyden, Teachers' Perceptions of Writing Measures

Several variables seemed to cluster together in importance following the initial three. These included nouns, verbs, adjectives, correct punctuation marks, correct word sequences, and simple sentences ($4.40 \le M \le 4.67$). These variables appear to be grammar-related. Parts of speech (nouns, verbs, and adjectives), correct punctuation marks, correct word sequences, and simple sentences all focus on grammar and syntax.

Figure 1: Overlap of 95% confidence intervals of writing variables ratings
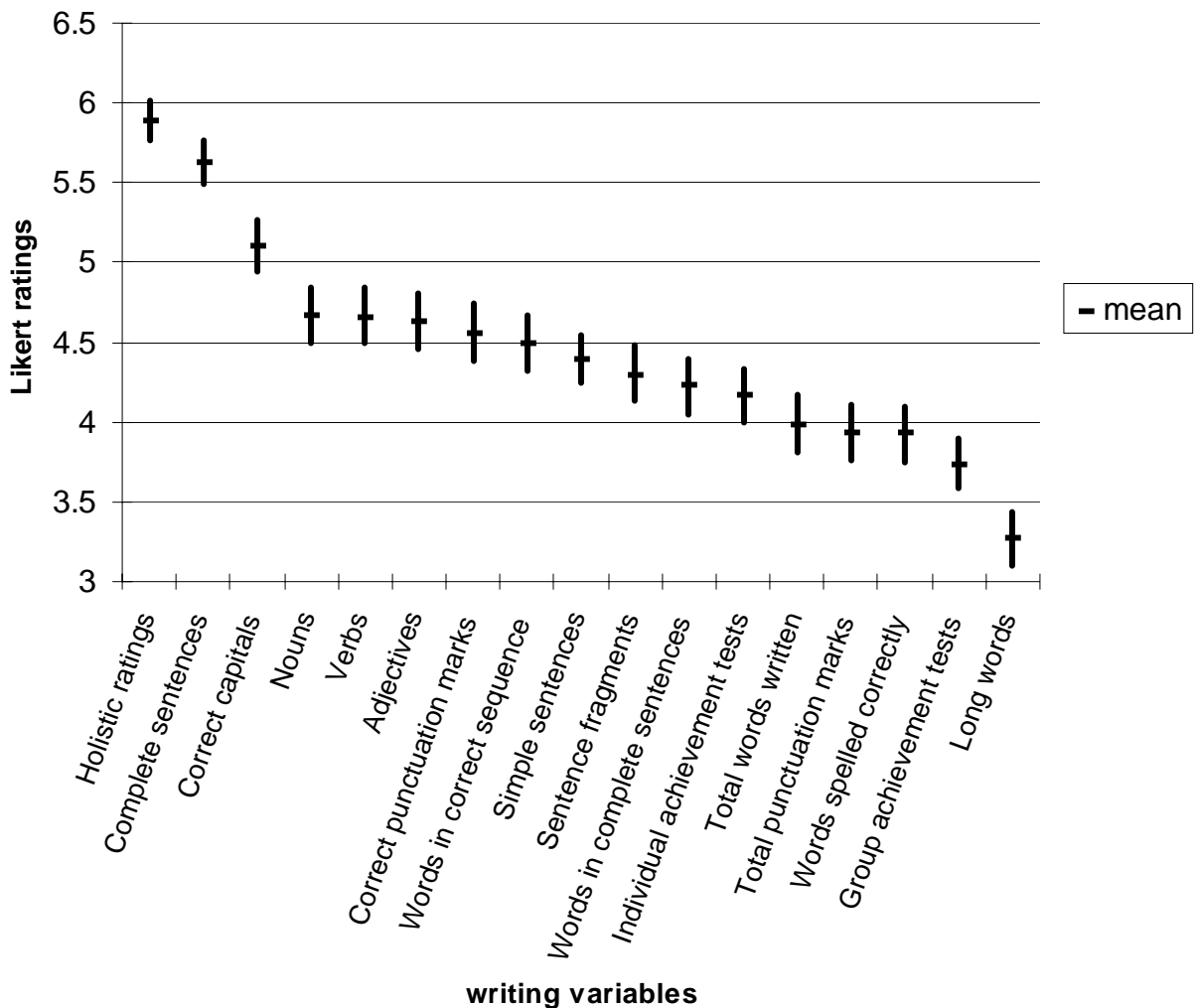
The lower bound of the 95% confidence interval for the remainder of the variables rated all fell below the upper bound for the standard CBM writing measure, total words written (TWW). The means for sentence fragments and words in complete sentences exceeded the upper bound for TWW, and the rating for individual achievement tests equaled the upper bound for TWW. The two lowest-rated variables were group achievement tests (M = 3.74) and long words (M = 3.27).

Correlations

Significance levels for correlations are not reported as the magnitude of the largest of correlations was low, and the large number of correlations calculated would increase the Type I error rate of inferences regarding significance. Correlations between the variable ratings and teachers' descriptions of their levels of training in CBM were calculated. The only correlation that had a magnitude of greater than .15 was the relationship

Practical Assessment, Research, and Evaluation, Vol. 11 [2006], Art. 5

Practical Assessment Research & Evaluation, Vol 11, No 5   8
Gansle, Gilbertson, & VanDerHeyden, Teachers' Perceptions of Writing Measures

between training in CBM and holistic ratings (r = -.157). In other words, the more CBM training a teacher reported, the lower he or she rated holistic ratings as indicative of student writing skill.

The correlations between years experience and the ratings of the variables were also calculated. Only one correlation with a magnitude approaching .15 was the relationship between correct word sequences and years experience (r = -.148). The

Table 1: Ratings for Writing Variables (n=335)

Variable	M	SD	95% Confidence Interval	
			Lower	Upper
Holistic ratings*	5.89	1.13	5.77	6.02
Complete sentences	5.63	1.27	5.49	5.77
Correct capitals	5.10	1.55	4.94	5.27
Nouns	4.67	1.61	4.49	4.84
Verbs	4.66	1.65	4.49	4.84
Adjectives	4.63	1.60	4.46	4.81
Correct punctuation marks	4.56	1.65	4.38	4.74
Correct word sequences	4.50	1.61	4.32	4.67
Simple sentences	4.40	1.40	4.25	4.55
Sentence fragments	4.30	1.62	4.13	4.48
Words in complete sentences	4.23	1.63	4.05	4.40
Individual achievement tests*	4.17	1.49	4.00	4.33
Total words written	3.99	1.68	3.81	4.17
Total punctuation marks	3.94	1.64	3.76	4.11
Words spelled correctly	3.93	1.64	3.75	4.10
Group achievement tests*	3.74	1.46	3.58	3.90
Long words	3.27	1.53	3.10	3.43

Note: * indicates variables that are not CBM-type variables.

more years experience a teacher reported, the lower he or she rated correct word sequences as indicative of student writing skill. Overall, years experience and CBM training were unrelated to measures of writing skill in this study.

Comparisons by Groups

In order to determine whether training in CBM or training for and working with students in special education programs had an effect on the ratings of the variables, t-tests were completed.

High CBM training vs. low CBM training. Due to the 7-choice response possibilities for indicating level of training in CBM, teachers were put into one of three groups based on their answer to this question. Teachers who indicated no or low training, who answered 1 or 2 to the question, were put into the low training group (n = 188). Teachers who indicated high levels of training by answering the question with a 6 or a 7, were put into a high

training group (n = 38). All teachers who answered between 3 and 5, inclusive, were put into a third group. The low training group was compared to the high training group. None of the variables rated in the survey was significantly different according to training group.

General education vs. special education. Teachers who reported that their current teaching assignment was special education (n = 38) were compared to those teachers who reported that their current teaching assignment was general education (n = 275). Only one of the measures' p values fell below .05: individual achievement tests (t = 2.24, df = 299, p = .026). Special education teachers rated individual achievement tests higher than general education teachers (M = 4.05, sd = 1.48 for general education teachers, M = 4.63, sd = 1.55 for special education teachers). However, with the large number of writing variables measured, a corrected

9

Gansle, Gilbertson, & VanDerHeyden, Teachers' Perceptions of Writing Measures

alpha does not allow a positive determination of statistical significance to be made.

Special education and training in CBM. The number of teachers who reported high training in CBM is equal to the number of teachers who reported that their current teaching assignment is special education. These groups are not the same teachers. Of those teachers who reported high CBM training, 22 teach general education classes, 10 teach special education classes, and 6 reported "other" as their current assignment.

DISCUSSION

Although they have been shown reliable and valid for screening and progress monitoring in written expression (Marston, 1989), writing measures such as words spelled correctly and total words written do not seem to capture important aspects writing such as content or syntax (Gansle et al., 2002). Despite the amount of anecdotal evidence that psychologists may report from the schools and teachers they serve, there has not yet been a study of the acceptability of specific curriculum-based measures of written language. The survey described above provided an exploratory analysis of and some empirical support for the statements made by teachers and psychologists.

School psychologists and educational consultants who use CBM have been trying to "sell" measures such as correct word sequences, words spelled correctly, and total words written to their teachers for well over 10 years; there is still work to be done in this area. In fact, even teachers who reported high levels of training in CBM find total words written to be no better related to student writing skills than teachers who reported low levels of training. CBM training may not have had an effect on teachers' perceptions of the importance of a variety of CBM-type measures to assessing student writing skills. According to these teachers, correct word sequences is perceived as more representative of student writing skill than total words written or words spelled correctly; however, it is not perceived as favorably as complete sentences.

This suggests one of two possibilities. Either we are not making the case well enough, or those who actually have had the opportunity to use the measure have discovered evidence in their own practice that contradicts that which has been published. It may be a combination of the two. When a teacher has the opportunity to look at writing samples that have the same number of words but which are grossly different in terms of diction, spelling, syntax, and grammar, it can be obvious that no matter what the research says, the same number for two students does not mean that they have demonstrated comparable levels of writing skill. They may find that total words written and words spelled correctly do not adequately represent the growth that they believe their students are making in their writing skills when exposed to instruction or practice with feedback.

Evidence now exists that teachers perceive variables such as complete sentences, number of nouns or verbs, and correct punctuation marks to represent student writing skill better than total words written, words spelled correctly, or correct word sequences. When coupled with evidence that suggests that there may be curriculum-based measures of written expression that are better related to criterion measures like norm-referenced measures or teacher rank of student writing skills (e.g., Gansle et al., 2002; Gansle et al., 2004), measures that have been shown here to be well accepted by teachers as representative of student writing skills deserve to be subjected to greater empirical scrutiny to place more confidence in their utility and validity for this purpose.

Teachers believe that their holistic ratings of student writing skill outperform all other measures investigated in this survey. Despite their pre-service preparation, the in-service training they may have received during their time on the job, and the experience they have working with psychologists and teachers who are responsible for completing and explaining assessments for students with special needs, they still believe that their own professional judgment is a better tool for assessing students' skills than any other measure rated here. Unfortunately, the increasing emphasis on accountability is incongruent with basing decisions on vaguely defined, data-free judgments. This is a fact of which psychologists and educational consultants should be aware when discussing progress monitoring with teachers, so that they will provide good evidence for the superiority of

Gansle, Gilbertson, & VanDerHeyden, Teachers' Perceptions of Writing Measures

whatever measures they suggest to professional judgment.

Number of complete sentences was rated second to holistic ratings. This is a measure that shows promise for the evaluation of writing skill. Although it did not appear in the regression equations that predicted criterion measures in the two Gansle et al. studies (2002, 2004), the inter-observer reliability of that measure was low compared to others in the studies, which may have attenuated its relationship to the criterion variables. Efforts should be made to strengthen the scoring criteria for the variable so as to increase the probability that if there is a relationship to criterion measures, it can be discerned.

There was a variety of grammar, syntax, and parts-of-speech variables that clustered together after holistic ratings and complete sentences in the ratings for writing variables. Teachers appear to think that these are decent, and better than TWW for assessing student writing skill. Given that they are very similar in terms of ranking, and that many of them were significantly related to criterion measures in the two Gansle et al. studies (2002, 2004), it might be possible to use a rubric that sums these variables and determine if the resulting measure is related to criterion measures. If this is possible, a measure that captures the multi-faceted mechanics of writing is likely to have a good relationship to criterion measures and to be rated as very acceptable to teachers, which may be the best of all possible worlds.

Given the omnipresence of high stakes group testing, it is noteworthy that group achievement tests are rated lower than even TWW in the assessment of students' writing. It is unfortunate that teachers must spend a large portion of their time preparing their students for an event that they do not feel bears much relationship to the focus of their instructional time.

Limitations and Future Directions

Sampling procedures constitute a limitation of the study. Although the sample of teachers surveyed included teachers from schools in many locations throughout the United States, it is difficult to say whether the results would generalize to all teachers in the US. Future studies should use a stratified random sample of teachers to increase the

generalizability of the results. In addition, given the nature of the consent procedures used, it was impossible to report a return rate on the survey. Teachers who chose not to complete the survey within the faculty meetings at which they were distributed did not leave a paper trail that allowed for a response rate determination.

This study focused on elementary school teachers, and did not address writing with teachers of more advanced students. It was designed that way due to the apparent relationship between basic writing skills and the curriculum-based measures investigated. It is possible that teachers who address more sophisticated aspects of writing in their lessons might find different measures of writing more suited for assessment in their classrooms.

Although there was no difference between special and general education teachers for curriculum-based measures, it is possible that the relatively small number of special education teachers attenuated the results of these comparisons. The data may represent the distribution of special and general education teachers in public schools throughout the United States. There are many more general educators than special educators. Future research might attempt to secure more balanced groups of teachers.

This study addresses multiple aspects of evaluating written language. It is not, however, an exhaustive selection of the possible measures of writing, but an assessment of some that had been investigated for validity and reliability in previous studies. There are likely a plethora of good curriculum-based measures of written language that have yet to be investigated. Future research should focus on some that would bear a good relationship to criterion measures, be sensitive to the effects of instruction and practice with feedback, and be acceptable to teachers.

Although expert opinion and previous literature were used to construct the survey instrument described herein, there are no data to support the technical adequacy of the instrument used in terms of reliability or validity. There are currently no data to indicate how much of a difference in the teachers' perceptions' score is important or meaningful. In addition to using stratified random

Gansle, Gilbertson, & VanDerHeyden, Teachers' Perceptions of Writing Measures

sampling, future research should focus on determining the technical adequacy of the instrument used to survey teachers.

REFERENCES

Allinder, R. M., & Oats, R. G. (1997). Effects of acceptability on teachers' implementation of Curriculum-based Measurement and student achievement in mathematics computation. *Remedial and Special Education*, 18, 113-120.

Eckert, T. L., Shapiro, E. S., & Lutz, J. G. (1995). Teachers' ratings of the acceptability of curriculum-based assessment methods. *School Psychology Review,* 24, 497-511.

Elliott, S. N. (1988). Acceptability of behavioral treatments in educational settings. In J. C. Witt, S. N. Elliott, & F. M. Gresham (Eds.), *Handbook of behavior therapy in education* (pp. 121-150). New York: Plenum Press.

Elliott, S. N., Witt, J. C., & Kratochwill, T. R. (1991). Selecting, implementing, and evaluating classroom interventions. In G. Stoner, M. R. Shinn, & H. M. Walker (Eds.), *Interventions for achievement and behavior problems* (pp. 99-135). Silver Spring, MD: National Association of School Psychologists.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education*, 9, 20-28.

Gansle, K. A., Noell, G. H., VanDerHeyden, A. M., Naquin, G. M., & Slider, N. J. (2002). Moving beyond total words written: The reliability, criterion validity, and time cost of alternate measures for curriculum-based measurement in writing. *School Psychology Review*, 31, 477-497.

Gansle, K. A., Noell, G. H., VanDerHeyden, A. M., Slider, N. J., Naquin, G. M., Hoffpauir, L. D., & Whitmarsh, E. L. (2004). An examination of the criterion validity and sensitivity of alternate curriculum-based measures of writing skill. *Psychology in the Schools*, 41, 291-300.

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review*, 18, 37-50.

Hoover, H. D., Hieronymus, A. N., Frisbie, D. A., & Dunbar, S. B. (1996). Iowa tests of basic skills, Form M. Itasca, IL: Riverside Publishing.

Howell, K. W., Fox, S. L., & Morehead, M. K. (1993). *Curriculum-based evaluation: Teaching and decision making* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Jenkins, J. R., Deno, S. L., & Mirkin, P. K. (1979). Measuring pupil progress toward the least restrictive environment. *Learning Disability Quarterly,* 2, 81-92.

Kazdin, A. E. (1981). Acceptability of child treatment techniques: The influence of treatment efficacy and adverse side effects. *Behavior Therapy*, 12, 493-506.

Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford Press.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 20 U.S.C. § 6301 et seq.

Shapiro, E. S., & Eckert, T. L. (1993). Curriculum-based Assessment among school psychologists: Knowledge, use, and attitudes. *Journal of School Psychology,* 31, 375-383.

Shinn, M. R. (1986). Does anyone care what happens after the refer-test-place sequence: The systematic evaluation of special education program effectiveness. *School Psychology Review*, 15, 49-58.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children.* New York: Guilford.

Shinn, M. R., & Bamonto, S. (1998). Advanced applications of curriculum-based measurement: "Big ideas" and avoiding confusion. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement.* New York: Guilford.

Tindal, G. (1989). Evaluating the effectiveness of educational programs at the systems level using curriculum-based measurement. In M. Shinn (Ed.), *Curriculum-based assessment: Assessing Special Children* (pp 202-238). New York: Guilford Press.

Witt, J. C., Daly, E., & Noell, G. H. (2000). Functional assessments: A step-by-step guide to solving academic and behavior problems. Longmont, CO: Sopris West.

Witt, J. C., & Martens, B. K. (1983). Assessing the acceptability of behavioral interventions used in classrooms. *Psychology in the Schools,* 20, 510-517..

Citation

Author Note

Correspondence concerning this article should be addressed to Kristin A. Gansle, College of Education, Department of Curriculum & Instruction, Louisiana State University, 223 Peabody Hall, Baton Rouge, Louisiana 70803, (225) 578-7213. E-mail: kgansle@lsu.edu.

Appendix A
Cover Letter & Survey Materials

Dear Teacher:

You may be familiar with Curriculum-based Measurement (CBM) methods for reading, mathematics, and written language assessment. They use a one-, two-, or three-minute assessment of the area of interest, and provide scores that can be used for frequent assessment in those areas. For example, in reading, children are asked to read a passage at a particular grade level for one minute. The number of words read and the number of errors made are the scores that are recorded for reading.

For CBM in written language, students are given a half-sentence story starter and asked to write for three minutes. These writing samples are usually scored for the number of words written. Researchers have developed other methods of scoring these writing samples, and I would like your help in determining your ideas about some writing sample measures and whether these are representative of students' writing skills.

Behind this page, you will find a three-page survey about writing assessment and Curriculum-based measures of writing. If you decide to participate in this survey, please return the form to the person who gave it to you. Your responses are confidential, and neither your name nor any code to identify you is on this form. All surveys will be returned to the investigator, Kristin Gansle, at Louisiana State University, and will be maintained in a locked cabinet. No publication of this data will mention individual participants. Your consent to participate is your completion of the form. You may choose to decline your participation at any time. It is expected that it will take approximately 15 minutes to complete this survey.

Thank you for your help. Please feel free to contact me (information is below) if you should have any questions or concerns about the project.

Sincerely,

Kristin A Gansle
Assistant Professor

e-mail: kgansle@lsu.edu

Kristin A Gansle, Ph.D.
School of Social Work
Louisiana State University
311 Huey P. Long Field House
Baton Rouge, Louisiana 70803

Writing samples are collected according to standard CBM instructions (see Shinn, 1989, or contact Kristin Gansle for more information). Students are given a story starter and told to think about what to write for one minute and then to write for three minutes. After three minutes they are told to put down their pencils. The written products can be scored in a variety of ways. Please read the description of the scoring methods below and rate the extent to which you think they are representative of students' writing skills. A rating of 1 indicates that you think the item not an important indicator of student writing skill, and a 7 indicates that you think the item is an important indicator of student writing skill.

Score & Description	Not at all						Extremely well
Total Words Written Count the total number of words written during the 3-minute period, including the words that are spelled incorrectly. Do not count numbers that are not spelled out (1987, 3, 29) as words. Be sure to count the title if written and proper names and nouns as words. If the student writes the story starter as part of the story, be sure to include those words in the count.	1	2	3	4	5	6	7
Words Spelled Correctly Count words spelled correctly. Context is not relevant; if a word can stand alone and be recognized as one spelled correctly, it is correct.	1	2	3	4	5	6	7
Long Words Count number of <u>words spelled correctly</u> that are longer than 7 letters (8 letters or more). If it is not spelled correctly, it is not a long word, regardless of its length.	1	2	3	4	5	6	7
Total Punctuation Marks Count all punctuation marks, correct or incorrect. For quotation marks, please count open and close quotes as individual marks. For example, "cow," would have 3 punctuation marks.	1	2	3	4	5	6	7
Correct Punctuation Marks Count those punctuation marks correctly applied. Punctuation is supposed to be included <u>inside</u> quotation marks.	1	2	3	4	5	6	7
Correct Capitalization Count correct use of capitalization. These can be at the beginning of the sentence, or in the middle of the sentence if they're proper nouns or in quotation marks.	1	2	3	4	5	6	7

For the parts of speech that follow, please consider the part of speech as the word would be in isolation. If there is a word that could be multiple parts of speech, please consider the word in context. For example, "race" can be a noun or a verb. You must consider context when there is a question as to part of speech.

Score & Description	Not at all						Extremely well
Parts Of Speech—Verbs Count recognizable verbs (action words, something that exists, auxiliary, linking, etc.)	1	2	3	4	5	6	7
Parts Of Speech—Adjectives Count recognizable adjectives.	1	2	3	4	5	6	7
Complete Sentences Count complete sentences: 1) starts with capital letter, 2) has a subject, 3) has a verb, and 4) has ending punctuation.	1	2	3	4	5	6	7
Words in Complete Sentences Count the number of words that are in the complete sentences that were counted above.	1	2	3	4	5	6	7
Words In Correct Sequence Count as a correct word sequence the joining of two words together that are spelled correctly and are grammatically correct. Do not count numbers next to words in the total.	1	2	3	4	5	6	7
Sentence Fragments Count sentence fragments. A sentence fragment is incomplete and cannot stand alone. (You might expect that a higher number of fragments be related to a lower level of writing skill.)	1	2	3	4	5	6	7
Simple Sentences Count the number of complete sentences (that were counted before) that are simple sentences. A simple sentence is one independent clause. Each contains only one subject and one main verb.	1	2	3	4	5	6	7
Holistic Ratings Score the writing sample based on your judgment of the quality of the sample.	1	2	3	4	5	6	7

Gansle, Gilbertson, & VanDerHeyden, Teachers' Perceptions of Writing Measures

Please read the description of the assessment methods below and rate the extent to which you think they are representative of students' writing skills. The scale is the same as above: a rating of 1 indicates that you think the item not an important indicator of student writing skill, and a 7 indicates that you think the item is an important indicator of student writing skill.

Score & Description	Not at all Extremely well
Group Standardized Achievement Tests (e.g., ITBS, Metropolitan Achievement Test)	1 2 3 4 5 6 7
Individual Standardized Achievement Tests (e.g., Key Math, TOWL, Woodcock-Johnson)	1 2 3 4 5 6 7

Please provide some information about yourself, your education, and your job. I am interested in whether there is any relationship between your ideas about writing assessment and your background.

Current teaching assignment:
☐regular education   ☐special education   ☐other _____
                      (☐resource ☐self-contained ☐other _____)

What grades do you currently teach? Please check all that apply.
     ☐ECSE   ☐1   ☐2   ☐3   ☐4   ☐alternative HS
     ☐pre-K   ☐5   ☐6   ☐7   ☐8
     ☐K   ☐9   ☐10 ☐11 ☐12

In which state are you located? _____

My teaching assignment is primarily (check one):
         ☐rural.   ☐suburban.   ☐urban.

Please describe your education. Check all that apply.
     Bachelor's degree
         ☐elementary education   ☐secondary education   ☐other _____

     Master's degree
         ☐no   ☐yes (subject area: _____)

Primary Certification Area
     ☐Elementary Education ☐Secondary Education ☐Special Education ☐Other_____

Secondary Certification Area
     ☐Elementary Education ☐Secondary Education ☐Special Education ☐Other_____

Practical Assessment Research & Evaluation, Vol 11, No 5   17
Gansle, Gilbertson, & VanDerHeyden, Teachers' Perceptions of Writing Measures

Please indicate

     your gender: ☐female   ☐male

     your race: ☐African-American ☐Asian-American   ☐Hispanic   ☐Other _____
                        ☐Native American   ☐Pacific Islander   ☐Caucasian

Please write the number of years experience you have teaching as of January 1. _____

Please rate the extent to which you agree with the statements in the left column by circling the corresponding number in the right column.

Statement	Strongly Disagree						Strongly Agree
I have received training in the use of Curriculum-based Measurement (CBM).	1	2	3	4	5	6	7

I sincerely thank you for the time you have taken to complete this survey. If you would like to have access to the results of the survey when they are finished, please provide an e-mail address (below) and I will notify you when they are finished. I will post them on a website and will provide you with the link to the web page. Please PRINT your e-mail address carefully.

e-mail address: _____