Assessing children's writing products: the role of curriculum based measures

Dockrell, J.E. , Connelly, V. , Walter, K. and Critten, S.

Assessing Children's writing products: the role of curriculum based measures

Julie E. Dockrell,

Department of Psychology and Human Development

Institute of Education,

London WC1H 0AL

Vincent Connelly

Department of Psychology,

Oxford Brookes University

Kirsty Walter

Department of Psychology and Human Development

Institute of Education,

London WC1H 0AL

Sarah Critten

Psychology Department

Coventry University

Abstract

The assessment of children's writing raises technical and practical challenges. In this paper we examine the potential use of a curriculum based measure for writing (CBM-W) to assess the written texts of pupils in Key Stage 2 (M age 107 months, range 88 to 125). Two hundred and thirty six Year three, five and six pupils completed a standardized assessment of writing quality. Pupils also responded to two writing probes, one expository and one narrative, at baseline and five months later. In response to the probe pupils wrote for five minutes. Transcribed texts were scored for productivity (total words produced, correct word sequences, number of punctuation marks and sentences produced) and accuracy (proportion of words spelled correctly, correct word sequences and punctuation marks). The CBM –W measures were differentially sensitive to development and showed construct validity as evidenced by their association with the norm- referenced test measuring writing quality. Change over time was also evident and significant differences between narrative and expository texts were found. Pupils with special educational needs scored significantly more poorly on the CBM-W. Potential implications for research and practice are considered.

Key words: writing, assessment, curriculum based measures, Key Stage 2

The ability to produce written text, either manually or electronically is a key transferable skill.

Writing is central to academic achievement, to gaining employment, and to communicating more

widely. For students in formal education producing written text remains the primary way to

demonstrate knowledge (Graham & Harris, 2004). Writing is a complex skill which develops

over time through interactions between the child's cognitive resources, the instructional context

and the demands of the writing task. Not surprisingly, given its complexity, many students

struggle with learning to write (Graham & Harris, 2004), teachers often find writing instruction

challenging (Graham et al., 2008) and the assessment of written products raises a unique set of

problems for practitioners and researchers alike (Huot, 1990). Over the primary school years

students develop their writing skills and some, when they enter secondary schools, are highly

competent wordsmiths (Myhill, 2009; Myhill & Jones, 2007). However, the inability to produce

sustained, accurate, and competent writing remains a pervasive weakness for many students;

teachers need to be able to profile pupils developing writing skills so appropriate action can be

taken (Fuchs & Fuchs, 2009). To date, studies of writing development and the ways in which

writing products are assessed have been relatively neglected (Miller & McCardle, 2011). This

paper considers the assessment of children's written texts by exploring the use of curriculum

based measures of writing (CBM-W) with children in English primary schools to examine the

extent to which these measures captured developmental differences and changes in writing

performance overtime (Fewster & MacMillan, 2002). The validity of the CBM-W measure was

evaluated in relation to a standardized measure of written text quality and the potential for CBM-

W measures to identify struggling writers was examined.

Despite concerns about students' ability to produce written text in the UK, USA and

Europe (Department for Education, 2013; Persky, Daane, & Jin, 2003; Torrance et al., 2012),

research on writing has lagged behind research on reading and oral language development. Recent years have witnessed an increased focus on the development of children's writing processes and a recognition of the specific difficulties experienced by some pupils (Connelly, Barnett, & Dockrell, 2009; Graham & Harris, 2004; Myhill & Jones, 2009). This heightened awareness that "writing today is not a frill for the privileged few, but an essential skill for the many" (National Commission on Writing, 2003; 11), coupled with the numbers of students who are challenged by writing and the complexity of the writing process, has called attention to the importance of using reliable and valid assessments of written text production.

Understanding writing development

A prerequisite to the assessment of written text is an understanding of the skills that are developing as children learn to write. The multiple components of the developing writing process have been captured in the 'simple view of writing' (Berninger & Amtmann, 2003; Berninger et al., 2002) and more recently the 'not-so-simple view of writing (Berninger & Winn, 2006). Effectively the model synthesizes diverse traditions in compositional research whereby developing writing can be represented, figuratively, as a triangle in a working memory environment in which transcription skills (handwriting/typing and spelling) and executive functions are the vertices at the base that enables the goal of text generation at the top of the triangle to proceed efficiently (see Berninger & Amtmann, 2003: 350). The applicability of the framework to educational settings is evident through the inclusion of developmental processes, the emphasis on the importance of transcription skills until these become automatic and an acknowledgement of the different phases of writing development. Writing takes time to develop and is not mastered until well into the teenage years and beyond but it is lower level processes such as handwriting and spelling constrain the text production in the beginning writer and for those struggling with writing. Given these constraints

on writing development, the majority of children up to about age 11 rely on a simple linear text production method when writing most texts (Bereiter & Scardamalia, 1987). Productivity and the accuracy of the text produced are the key indicators of writing quality in children of primary school age (Berninger, Nielsen, Abbott, Wijsman, & Raskind, 2008; Graham, Berninger, Abbott, Abbott & Whitaker, 1997; Wagner et al, 2011). Assessments of writing will need to capture the key components of the written text product as children are learning to write if progress is to monitored reliably and appropriate interventions implemented.

Assessment of written texts

Writing can be evaluated in a number of different ways and assessments can be made of single or multiple texts. Often summative assessments are made from single texts and this type of assessment is typical in research studies. By contrast formative assessment is more likely to involve tracking progress over time to identify both progress and the child's strengths and needs (Mansell, et al, 2009; Qualifications and Curriculum Authority – QCA 1999; Department for Education – DfE, 2011). Typically primary school children's writing is evaluated in response to a written, spoken or pictorial probe and these texts are then scored to capture the children's competence in written text production. Irrespective of the approach taken writing assessments are needed to provide information about the pupils' current level of performance and future teaching and learning needs and the teacher or the researcher is required to make explicit decisions about the dimension(s) of the text which are to be evaluated (Saddler & Asaro-Saddler, 2013).

Various approaches to evaluation of written composition have been used by researchers and teachers, including holistic scoring, analytic scoring, quantitative scoring, and curriculum based measurement measures (Abbott & Berninger, 1993; Lee, Gentile, & Kantor, 2010; McMaster & Espin, 2007; Puranik & AlOtaiba, 2012; Scott & Windsor, 2000; Wagner et al., 2011). These various evaluation approaches differ in purposes and in the underlying assumptions about the dimensionality of written composition. It is not clear, however, how these various evaluative approaches are related.

Holistic scoring measures have been used in research, psychometric assessments and in school practice. Global quality of the text is rated on a single ordinal scale (see for example Weschler, 2005) and not on any specific dimensions of the text produced. Holistic measures have the advantage of providing a single score with relatively little time involvement from the assessor, but they are limited in their ability to reliably differentiate among writing levels, monitor change over time and capture differential performance on the key components of writing (Espin et al., 2000). In younger children and those with learning disabilities the short amount of text which is typically produced can also reduce the validity of a holistic approach to evaluation (McMaster & Espin, 2007).

More recent work has attempted to identify specific dimensions of children's written text products providing guidelines of where and how to evaluate children's written compositions. Sometimes these dimensions are considered together to create a single score such as in the Wechsler Objective dimensions of writing or current UK writing Key Stage 2 SATs assessment scores (DfE, 2013; Rust, 1996). Other times the hypothesized dimensions are scored separately and profiles of writing are produced in terms of analytic or quantitative scoring schemes (Huot, 1990). However, all these approaches require specialized training to reliably identify the target

dimensions and even with training the reliability of the analytic scales are too low to guide

decision making and planning (Graham, Harris & Hebert, 2011; He, Anwyll, Glanville &

Deavall, 2013). Moreover, the construct validity of the various analytical dimensions is often

lower than holistic scoring schemes (Espin, De La Paz, Scierka, & Roelofs, 2005; Gansle,

VanDerHeyden, Noell, Resetar, & Williams, 2006). Both holistic and analytic assessments of

writing provide information about writing performance but their use in evaluating pupil's writing

is challenging as the measures are hard to define objectively.

The number of dimensions which are thought to underpin written text production have

also been a matter of debate. Earlier studies of composition identified two dimensions in

children's written texts: quality and productivity (Berninger & Swanson, 1994; Graham, et al.,

1997; Olinghouse & Graham, 2009). Recently researchers have refined these dimensions and

included factors related to text complexity and organization (Wagner et al., 2011). Although

these dimensions vary by age and population tested, they all capture dimensions of productivity

and accuracy (Puranik, Lombardino, & Altmann, 2008; Wagner et al., 2011).

Timely assessment of students' writing competencies that are sensitive to measures of

productivity and accuracy is a key step to monitoring progression (Nelson & Tattersall, 2014).

Students need to be assessed on reliable and valid measures and frequently high stakes national

tests do not provide this information (Graham, et al, 2011). Thus, it becomes important to

examine the ways in which formative assessment can support writing development for students.

Monitoring change is a key component in this activity. Two basic elements are required if this is

to be effective: students need to be assessed over time and the writing task needs to be tailored to

the competencies that are being examined.

Text genre is an added consideration when evaluating written texts and comparisons should be made across different writing genres (Berman, 2008; Olinghouse & Wilson, 2013). Assessments of single writing products may fail to capture the demands of different types of writing tasks (Scott & Windsor, 2000). Narrative and expository writing are common school tasks. Narrative writing involves telling a story, while expository texts involve conveying facts or describing procedures. Expository texts take longer to master (Berman & Verhoeven, 2002) and differences are evident in the student's text (Beers & Nagy, 2011; Koutsoftas & Gray, 2012). Expository writing is argued to be more cognitively demanding for children in comparison to the production of narrative texts. Typically children produce shorter summaries which are more error prone when they produce expository texts, but more complex text structures may be used (Scott & Windsor, 2000).

Accurate measurement of pupils' writing is necessarily premised on the identification of core components or dimensions in a developmentally sensitive manner. This includes an evaluation of both the quality of the text and the quantity of text produced**.** These measures must also be sensitive to the performance level of struggling writers so that appropriate remediation and targets can be implemented. Struggling writers produce texts that are generally shorter, less interesting and poorly organized at the sentence and paragraph level (Hooper, Swartz, Wakely, de Kruif, & Montgomery, 2002). The children's texts are marred by inordinate numbers of mechanical, spelling and grammatical errors (MacArthur, Graham, & Fitzgerald, 2006). To ensure that assessments should encapsulate performance at word, sentence and text level capturing the key dimensions of both productivity and accuracy. These assessments should be

sensitive to small changes over time and involve uncomplicated scoring techniques to aid

regular, reliable administration.

Curriculum based measures of writing

Curriculum based measures (CBM) of assessment are ways of measuring a child's

academic progress through direct assessment of academic skills and have been well established

for reading and numeracy (Stecker, Fuchs, & Fuchs, 2005). They are also argued to be a

sensitive index of pupil's written text production (Espin et al., 2000) and have been successfully

used to examine the skills of English language learners (Campbell, Espin, & McMaster, 2013).

These assessments involve pupils writing for short periods (between three and seven minutes) in

response to a probe and have been shown to be valid and reliable measures of writing proficiency

for students aged between seven and 12 (Weissenburger & Espin, 2005). Thus, they provide a

potentially quick and reliable assessment of primary school children's writing products. Students

are typically assessed on two dimensions: productivity and accuracy. CBM-Ws have also been

promoted as a potential screening tool and as a means of monitoring progress (McMaster, Parker

& Jung, 2012).

A variety of different text measures have been used to evaluate productivity in CBM-Ws

including numbers of words written, correct word sequences (CWS), punctuation marks and

words spelled correctly. The appropriateness of these measures vary with pupil's age (see

McMaster & Espin, 2007 for a review of the technical features of the measure). Although

productivity measures such as total words written has often been considered the hallmark

measure of CBM-W tasks there is increasing evidence that inclusion of other quantitative

measures in combination with qualitative measures provides a more comprehensive measure of a

complex skill such as writing (McMaster & Espin, 2007; Ritchey & Coker, 2013). The inclusion

of CWS and measures of spelling may also provide more face validity for teachers (Coker &

Ritchey, 2010; Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002).

　　　　CBM-W measures vary in their scoring reliability. Inter rater reliability can be high,

with 80-90% agreement (Gansle et al., 2006; Graham, et al., 2011; Weissenburger & Espin,

2005) but this varies in relation to the types of texts produced, the age group of the pupils studied

and the text measure that is used (Campbell, et al., 2013; McMaster et al., 2012). Validity of the

CBM-W has also been examined and some measures are correlated with standardized

assessments (.69 for Test of Written Language, Hammill & Larsen, 1996)) and with teacher

ratings (.76, Parker, Tindal, & Hasbrouck, 1991). Simpler CBM measures such as total words

written have lower criterion related coefficients than more complex measures such as CWS or

measures reflecting spelling and word choice. Narrative probes have demonstrated the best

technical adequacy to date (see McMaster & Campbell, 2008).

　　　　The use of CBM-W has not gone unchallenged. Concerns relate to scoring, the extent to

which these measures are valid across different populations and their sensitivity in capturing text

quality. While numbers of words and words spelled correctly show good interater reliability

other measures are not so straightforward to score (CWS) or require a more subjective judgment.

Only modest criterion validity coefficients have been achieved, although this may be a more

general problem of writing measures (Huot, 1990). Of particular concern has been the

identification of sensitive indicators of early writing (but see McMaster, Ritchey, & Lembke,

2011). Most work using the CBM-W has examined static scores and evidence examining growth

trajectories has been contradictory. McMaster et al (2011) found stable and valid growth curves

between the ages of eight and nine. However, other studies have indicated that only CBM-W

variables of total words written, words spelled correctly and CWSs showed clear developmental

trends (Costa, Hooper, McBee, Anderson, & Yerby, 2012). A final limitation rests in the paucity

of research on CBM-W measures in educational contexts outside North America. Given the

interaction between teaching and learning and the different pedagogical approaches used outside

North America the utility of CBM-W as a measure to evaluate writing performance and progress

for children at similar stages of learning to write remains limited. It therefore becomes important

to consider whether CBM- W can be used to complement current writing assessments in KS2. As

yet we do not know which, if any measures, differentiate pupils writing performance and

whether these are sensitive to change over time (Ritchey & Coker, 2013).

Aims of the study

The primary purpose of the study was to examine the potential uses of CBM-W as a

means of evaluating writing products in a cohort of English primary school pupils. It has been

argued that three stages are essential in establishing the usefulness of CBMs for monitoring

progress: examining the technical features of the static score, examining the sensitivity of the

measure to growth and finally examining the use of CBM in monitoring teaching effectiveness

(Fuchs, 2004). In this paper we examine the first two stages using both narrative and expository

probes for pupils in between the ages of eight and 11.

To examine the technical features of the CBM-W we first established that the children's

written texts could be scored reliably for measures of productivity (numbers of words produced,

numbers of CWS and numbers of punctuation marks) and text accuracy (proportion of CWS,

proportion of words spelled correctly and proportion of correct punctuation marks). We then

examined whether the measures used differentiated across age groups and between genres and

the extent to which the measures correlated with a UK standardized test of writing quality

(Wechsler Objective Language Dimensions). This allowed us to examine the concurrent validity

of the CBM-W measures. Further to this we carried out a principal components analysis on the

CBM to establish the extent to which the measures reflected productivity and accuracy described

in dimensional measures of children's writing for this age range (Puranik et al., 2008).

To establish sensitivity to growth, for measures that captured age related differences, we

considered changes in pupil's performance over a five-month period during the same academic

year. Finally given the overall poorer performance of pupils with special educational needs

(SEN) on writing tasks (Graham & Harris, 2005) and evidence that CBM-W can profile levels of

writing skills for pupils with SEN (Dockrell, Ricketts, Charman & Lindsay, 2014) we explored

whether the measures were sensitive to children's SEN status. To our knowledge this is the first

study to consider CBM-W measures in the UK and, as such, to provide an evidence base about

its potential use in schools.

Method

Participants

Two hundred and thirty six pupils from two primary schools in two shire counties participated in

the study. Schools were representative of the local authority in terms of national data on

percentages of free school meals, ethnicity and numbers of children with special educational

needs.

All Year 3 ($n = 71$), Year 4 ($n = 83$) and Year 5 ($n = 82$) pupils participated in the study

(n boys = 130; n girls = 95). Data on free school meal eligibility, presence of SEN (school action,

school action plus or statement) and English as an additional language was available for 82

pupils. Of these 82 pupils nine were eligible for free school meals, three had English as an

additional language and 19 were reported to have special educational needs.

Both schools scored above the national average in Key Stage 2 writing assessments

(Writing level 4 or above 93% and 79%). However, Mean standard scores for the WOLD were

at the lower margins of the average range (Mean Year 3 = 85.27, SD = 6.81; Mean Year 4 =

85.51, SD = 7.90; Mean Year 5 = 87.17, SD = 10.68) and there were no significant differences

between the three year groups in the pupils standard scores (F (2, 315) = 1.507, ns).

Design

This was a repeated measures design where pupils completed two CBM-W tasks

(narrative and expository) and a standardized measure of writing at two different time points

during one academic year, five months apart. Presentation of narrative and expository tasks was

balanced across year groups and time of assessment. The standardized writing task (WOLD) was

always completed last.

Materials

Curriculum based writing measures

Pupils responded to two probes. In the narrative condition children responded to a probe

"One day I had the best/worst day ever at school". In the expository condition children

responded to the probe "There are many things that make a day at my school very

interesting/boring".

Each probe was presented at the top of a sheet of lined paper, with additional sheets

available if required. Participants were asked to write the best story/description they could within

the time limit. After explaining the task, students were given 30 seconds to think about what they

wanted to write and five minutes to write it.

The CBM-W was scored for both productivity and text accuracy. Productivity was

examined using four measures: total number of words produced, number of CWS, total

punctuation marks and total number of complete sentences produced. Total words produced was

calculated by counting the number of words produced in the five minute period, excluding

numerals and crossed-out words irrespective of whether the words were correctly spelled. CWS

was defined as any two adjacent words that were acceptable within the context of the sample to a

native English speaker (Espin et al., 2005). End punctuation and beginning capitalization were

also taken into account in the scoring (Tindal & Parker, 1989). To distinguish transcription errors

and text generation skills spelling errors were not penalized. All punctuation marks were counted

and total sentences were defined by the use of a full stop.

Accuracy scores were proportion scores created for words spelled correctly, CWS,

correct punctuation marks and correct sentences. Proportion scores were calculated using the

correct use of the textual feature taking into account the total production of that specific feature.

For example proportion of words spelled correctly was calculated from the total number of

words produced.

Standardized measure of writing

The Wechsler Objective Language Dimensions (WOLD): writing expression (Rust,

1996). The child is asked to write a letter outlining his or her ideal house. Children are allowed

15 minutes to complete the task. The written output can either be scored holistically or

analytically: reliability .89, correlation with Woodcock-Johnson Psycho-Educational Battery-

Revised, Dictation = 0.72. The analytic scale was used to assess the children's written text in

relation to UK standardized norms. This comprises six dimensions, each rated on a four point scale, which are scored independently of each other: Ideas and development; Organization, Unity and coherence; Vocabulary; Sentence structure and variety; Grammar and usage; Capitalization and punctuation. These scores are then combined to produce a total score with a maximum of 24 points. This score is standardized against a set of UK normative data.

Procedure

Ethical approval was given by Oxford Brookes University, which follows British Psychological Society guidelines. Parents were provided with information sheets explaining the objectives of the study.

All testing took part in class groups and children were informed they were participating in a research study. Students were free to withdraw at any point during the assessments. No additional assistance was provided.

Coding and reliability

CBM measures in school from the handwritten texts by the teachers. Since the raters were unfamiliar with the pupils before scoring the scripts, each script was transcribed onto the computer. This removed the potential impact of unfamiliar handwriting and provided a clear text for the assessors to evaluate. All scripts were scored by graduate psychologists or one of the project directors. Prior to scoring training was provided where 12 example scripts randomly selected from the sample were scored by all raters. A minimum of 80 per cent agreement with the principal marker was achieved for all trainers. Inter-rater reliability was checked for a 10 per cent sample of all scripts. Chronbach's Alpha for the CBM-W scoring criteria was above .8 showing a high rate of reliability across markers.

Results

The results are presented in three sections. In each section we examine CBM-W measures of accuracy and productivity separately and consider the effect of genre type (narrative versus expository). Section 1 addresses the first research questions and examines whether the CBM-W measures were sensitive to developmental differences. Section 2 examines the concurrent validity of CBM-W measures in relation to a norm referenced writing assessment. Section 3 examines, for those measures where developmental differences were evident, whether these CBM-W measures are sensitive to change over time. Finally, for those pupils where we were provided with information of their SEN status we report the extent to which being identified as experiencing an SEN impacted on their performance on the CBM-W scores.

CBM-W measures and sensitivity to age group and genre type differences

One hundred and ninety-two pupils contributed both narrative and expository texts. Means and standard deviations for productivity measures are presented in Table 1 and accuracy measures in Table 2. A series of repeated measures ANOVAs were conducted for each measure, with genre type as the within child measure and Year group as the between group measure.

INSERT TABLE 1 AND 2 ABOUT HERE

There were significant effects of genre type and Year group differences for the majority of the productivity measures. Older children wrote significantly more words (F (2, 190) = 7.50, p = .001, ηp^2 = .07) and there was a significant effect of genre type where children produced significantly more words with the narrative rather than the expository probe. (F (1, 190) = 72.29,

$p < .001$, $\eta p^2 = .28$). There was no interaction between Year group and genre type (F (2, 190) = 2.01, ns). Older children also produced significantly more CWS (F (2, 189) = 4.34, p = 01, ηp^2 = .04) and there was a significant effect of genre type where children produced significantly more CWS with the narrative probe rather than the expository probe (F (1, 189) = 41.81, $p < .001$, ηp^2 = .18). There was no interaction between Year group and genre type (F (2, 189) = 2.96, ns).

Older children also produced more punctuation marks (F (2, 190) = 5.38, $p = .005$, $\eta p^2 = .05$) and again there was an effect for genre type, but for punctuation pupils produced more punctuation marks to the expository probe rather than the narrative probe (F (1, 190) = 7.25, p = .008, $\eta p^2 = .04$). There was no interaction between Year group and genre type (F (2, 190) = .82, ns). Total sentences produced did not differ by year group (F (2, 190) = .94 ns) or genre type (F (1, 190) = .97, ns). There was, however, a significant interaction by genre type and Year group F (2, 190) = 4.89, $p = .009$, $\eta p^2 = .05$. Follow-up univariate ANOVAs with post hoc tests revealed no significant differences between Year groups for expository texts (F (2, 200) = .25, ns). By contrast in pupils in Year 5 in the narrative condition produced significantly more sentences than those in Year 4, but there were no significant differences between Year 4 and Year 3 (F (2, 190) = 3.66, $p = .03$, $\eta p^2 = .05$). In sum productivity measures of total words produced, CWS and numbers of punctuation marks were sensitive to Year group differences and genre type.

Measures of accuracy were proportion scores calculated by considering correct production over total production. There was no significant Year group difference for proportion of CWS (F (2, 189) = .82, ns), nor was there a significant effect of genre type (F (1, 189) = .003, ns). However there was an interaction between Year group and genre type (F (2, 198) = 4.89, p = .1   , $\eta p^2 = .08$). Follow-up univariate ANOVAs with post hoc tests revealed no significant differences for narrative texts (F (2, 200) = 2.83, ns). In contrast there was a marginally

significant effect for expository texts where Year 4 pupils had a significantly greater proportion of CWS than Year 5 pupils (F (2, 190) = 3.06, p = .05, ηp^2 = .03). For the proportion of words spelled correctly there was a significant of Year group with older pupils performing better than younger pupils, and an effect of genre type with performance to the narrative probe being significantly better (F (1, 190) = 11.20, p < .001, ηp^2 = .07) but no interaction between genre type or Age group (F (2, 190) = 1.76, ns). For proportion of correct punctuation marks there was no significant effect of Year group (F (2, 153) = 2.07, ns) but there was a significant effect of genre type with performance to the expository probe being significantly better than performance to the narrative probe (F (1, 153) = 6.26, p = .013, ηp^2 = .04) but no interaction between genre type or Age group (F (2, 153) = 2.24, ns).

In sum, accuracy measures indicated that proportion of words spelled correctly was sensitive to both Year group differences and genre type. Neither proportion of CWS nor proportion of correct punctuation marks were sensitive age group differences but in both cases accuracy patterns differed across genre type. For the subsequent analyses we only considered those measures that have shown statistically reliable age group differences.

CBM-W and standardized measures of writing

We considered the relationship between the CBM-W measures and the pupil's raw score on the WOLD measures. All correlations controlled for participant's age. These correlations are presented in Table 3; correlations for the expository text are presented above the diagonal and for the narrative text below the diagonal. As the table shows there were significant correlations between all the measures and the WOLD total score. Both narrative and expository texts CWS correlations with the WOLD were the largest. We further explored the data using a principal components analysis (PCA) where all the measures in Table 3 were used. As shown in Table 4

the PCA resulted in two components with eigenvalues >1, accounting for 65% of the variance. The two components represented two constructs of writing where productivity measures all loaded on the first factor and the accuracy measures loaded on the second factor. Loadings were similar for the two types of text and the WOLD loaded on both factors, as would be predicted given the aggregate nature of the score.

<p align="center">INSERT TABLE 3 AND 4 ABOUT HERE</p>

CBM-W measures and sensitivity to change over time

We considered change over time in those measures, which were sensitive to age group differences, that is productivity measures of total words written, CWS and punctuation and the accuracy measure of proportion of words spelled correctly. Gains were calculated by subtracting performance at the first test point from assessment at the second test point. Means and standard deviations of gain scores across the five month period are presented in Table 5. Four repeated measures ANOVAs were conducted with Time (2) and Genre type (2) as the repeated measures and Year group (3) as the between group measure.

<p align="center">INSERT TABLE 5 ABOUT HERE</p>

There was a significant effect of time for total words produced for both narrative (F (1, 262) = 209.75, $p < .001$, $\eta p^2 = .45$) and expository texts (F (1, 262) = 23.14, $p < .001$, $\eta p^2 = .09$). There was no interaction by genre type (F (1, 262) = 2.19, ns). As expected from the first assessment point there was also a significant effect of Year group (F (2,261) = 16.87, p < .001) where Year 5 pupils performed better than Years 3 pupils ($p < .001$) and 4 ($p < .001$) and Year 4 pupils performed better than Year 3 pupils (p =.02). There were no other statistically significant differences.

There was also a significant effect of time for CWS for both narrative (F (1, 262) = 166.24, p < .001, ηp^2 = .39) and expository texts (F (1, 262) = 35.62, p < .001, ηp^2 = .12). As expected there was also a significant effect of Year group (F (2,261) = 15.55, p < .001) where Year 5 pupils performed better than Year 3 pupils (p < .001) and 4 (p < .001) and Year 4 pupils produced more words than Year 3 pupils (p =.02). There were no other significant effects.

Punctuation marks improved over time for both narrative (F (1, 262) = 5.29, p= .02, ηp^2 = .02) and expository texts (F (1, 262) = 4.58, p = .03, ηp^2 = .02), but the effect size was very small. As expected there was also a significant effect of Year group (F (2,261) = 14.14, p < .001) where Year 5 pupils performed better than Year 3 pupils (p < .001) and 4 (p < .008) and Year 4 pupils performed better than Year 3 pupils (p =.008).There were no other significant effects.

Accuracy in terms of the number of words spelled correctly improved over time for both narrative F (1, 262) = 26.96, p < .001, ηp^2 = .09) and expository texts (F (1, 262) = 5.49, p = .2 , ηp^2 = .02, ηp^2 = .09), but as for punctuation the effect sizes were modest. As expected there was also a significant effect of Year group (F (2,261) = 12.18, p < .001) where Year 5 pupils performed more accurately than Year 3 pupils (p < .001) but not Year 4, but Year 4 pupils produced more words than Year 3 pupils (p< .001).There were no other significant effects.

The fluency and the accuracy measure were sensitive to change. Although change was evident for both narrative and expository texts the effect sizes were larger for narrative texts for total words produced, CWS and words spelled correctly. Age trends were significant where Year 5 performed better than Year 4 and Year 4 performed better than Year 3 for all measures as found at the first assessment point.

CBM-W and pupils with special educational needs

For 80 pupils we had information about their SEN status. There was no significant difference in age between those with SENs ($n = 20$, M age = 8;5) and those without SEN ($n = 60$, M age = 9;0) t(79) =1.8, *ns.*) We examined differences between the two groups at the first assessment point on the four measures that differentiated between Year groups. As Table 6 shows on all variables, apart from punctuation, participants with SEN scored significantly lower. Where the comparisons were significant the effect sizes were large, indicating that pupils with SENs were performing one standard deviation lower than those without reported SENs on these measures. These differences were corroborated when the pupil's performance was compared on the standardized writing measure (SEN $M = 81.30$ $SD = 6.53$, No SEN $M = 90.70$, $SD = 10.07$; $t(50.54) = 4.83$ $p < .001$ $d = 1.40$).

INSERT TABLE 6 ABOUT HERE

Discussion

The primary purpose of the study was to examine the potential uses of CBM-W as a means of

evaluating writing products in a cohort of English primary school pupils both concurrently and

over time. We established that some CBM-W measures differentiated across age groups and

correlated significantly with the standardized measure of writing quality, demonstrating validity

and confirming previous. A principal components demonstrated that productivity and accuracy

measures loaded on different factors. Over all sensitivity to growth was evident for some of the

text measures. These measures reflected both productivity and accuracy, confirming previous

research in this area. The CBM-W was also sensitive to the pupil's special educational needs

status, demonstrating significant and large differences between pupils with designated SENs and

those without.

Three measures of productivity and one measure of text accuracy reliably distinguished

between the year groups (see Costa et. al, 2012 for similar results). Additionally the children's

productivity in terms of amount of words produced aligns with those reported in other studies for

similar age groups (McMaster & Espin, 2007). As with previous research we found total number

of words written and CWS discriminated between year groups (Gansle et al., 2002; Gansle et al.,

2006; McMaster & Espin, 2007; Weissenburger & Espin, 2005). Few studies have considered

punctuation marks and, although they were relatively infrequent in the pupils written texts,

number of punctuation marks also discriminated between year groups (Gansle et al., 2006). Total

sentences produced did not vary significantly across year groups, although there was a

significant interaction effect for the Year 5 pupils in terms of sentences produced in the narrative

task. These results suggest that children may need to have reached a level of competence in the

genre before considering sentences as a text measure. Assessments of the texts of more

competent writers are likely to require more detailed assessments of sentence structure at the

clausal level were used (Berman, 2008). However, such analyses pose significant challenges for

reliable scoring (Graham et al., 2011) and as yet there are no data suggesting that such measures

change reliably with development in the primary school period.

In contrast to previous work (McMaster & Espin, 2007) the only accuracy measure which

discriminated between age groups was the proportion of correct spelling (see also Costa et. al,

2012). Transcription skills, both handwriting and spelling, account for the majority of the

variance in writing quality for both typically developing children and those with developmental

difficulties learning to write in English (Berninger, et al., 2008; Graham et al., 1997; Olinghouse,

2008). Our data suggest that a short five minute written text to a writing probe which assesses

spelling errors in relation to the total text produced can capture these differences.

Despite these statistically significant differences and, in many cases, large effect sizes

there was marked heterogeneity within the age groups. Although 95 per cent confidence intervals

revealed little overlap between the year groups, variation within year groups was often large.

This variation was also evident in gains that pupils made in each measure over the five month

period. While this variation deserves further investigation, similar heterogeneity was found in the

WOLD scores for the sample and the national data on writing tests for England also show similar

patterns of heterogeneity within year groups (DfE, 2011, 2012). Of particular note is the fact that

there was no overlap between the scores for pupils with SEN and those with no recorded special

needs suggesting that such measures may be a useful to identify objectively children who are

struggling to develop writing skills and monitor progress (see also McMaster, et al., 2013).

Importantly the CBM-W measures correlated significantly with a standardized measure

of text quality. This provides indicative evidence of the validity of CBM-W measures as an

indicator of text quality. Moreover, the principle component analyses supported the apriori distinction between measures of productivity and measures of accuracy. These measures have become important indicators of the microstructural aspects of children's texts (Puranik et al., 2008; Wagner et al., 2011) and the data suggest that different CBM-W measures may tap these dimensions.

Sensitivity to growth was evident for two productivity measures (CWS and total words written) and proportion of words spelled correctly, a measure of accuracy. Effect sizes for the both CWS and total words written were large. This is a promising finding as teachers may be able to track progress using these objective measures which can be quickly and accurately scored. Information can then be used to inform decision making about the need for further support and by corollary the subsequent effect of that support on the pupil's writing.

The CBM-W measures reliably differentiated between narrative and expository texts. This confirms previous work using other forms of writing assessments examining these genre differences (Apel & Apel, 2011; Koutsoftas & Gray, 2012; Scott & Windsor, 2000), and provides a further source of information about the potential validity of the CBM-W measure. Pupils produced less text and were less accurate in response to the expository probe. In contrast more punctuation marks were used than in the narrative texts, perhaps indicating the more list like nature of narrative texts at this point in development. There were large effect sizes for these differences, as would be expected when children are new to writing in a genre. This raises an important caveat in using these assessments to differentiate between pupils and across time, comparisons need to be made using similar types of probes.

Overall we were able to identify a number of strengths in the CBM-W measure at this point in development. Good reliability of the scoring was established and there was validity with

a nationally standardized measure of writing quality. Both accuracy of scoring and validity are reported to be problematic for other forms of writing assessment (Graham et al., 2011). Some of the CBM-W measures differentiated across year groups and pupils with and without special educational needs. It was also sensitive to change over the five month period used in the current study, providing a sound basis for formative assessment. Together these data suggest that the CBM-W can be used across the primary years from Year 3 to Year 6 and can chart change over periods of time within those years if the appropriate measures are chosen. There is more work required to see if CBM-W can be reliably administered more frequently than the current 5 month period validated here. However, other work on CBM-W would suggest that more fine grained administration periods with weekly administration are possible (McMaster & Espin, 2007). Our current results with a 5 month gap between tests show CBM-W can, at least, be used in a broadly similar way to the Assessing Pupils Progress (APP) writing assessment in UK schools which is currently recommended to be used no more than 2-3 times a year (DCSF, 2009; QCA, 2008).

There were limitations with both the current study and the CBM-W. We were not able to control for teacher effects on the children's progress nor did we have detailed information about the nature of the children's special educational needs or their current Key Stage levels. There is growing evidence to indicate that the nature of the child's SEN impacts on the types of difficulties they experience with written text production (Connelly, Campbell, MacLean, & Barnes, 2006; Dockrell et al., 2014). The texts were coded by trained graduate assistants, so it is not possible to generalize the findings to other assessors. Research in other domains has indicated that generalizing from research studies to conventional practice in schools raises additional challenges (McCartney, Boyle, Ellis, Bannatyne, & Turnbull, 2011). CBM-W

measures assess the product not the process of writing. As such it only provides partial

information about the writer.

The micro-structural coding of the current CBM-W texts raises the question of whether

better differentiation between writers and genres would be evident if other macro structural

dimensions were considered (Wagner et al., 2011). Similarly it may be that a different measure

of text complexity and word and sentence level would provide more sensitive indicators of

change. It is likely that the nature of analysis will need to consider both the children's age and

their level of writing skill. CBM-W measures may only be useful for assessing lower level skills

in contrast to higher level skills like ideation which become progressively more important as

children become more competent writers (Juel, 1988).

IMPLICATIONS AND FUTURE RESEARCH

It is clear from the above limitations that there is more work to be done to establish the

potential uses of the CBM-W measures. Nonetheless the current study suggests that the CBM-W

is a useful tool among a repertoire of methods of assessing pupils writing. It has the potential to

be used for targeting intervention goals and as a screening tool to identify those children

struggling to write. Furthermore given the high levels of reliability and relatively straight-

forward scoring system it is likely to be appealing to researchers and educational practitioners

alike. Of course the availability of CBM-W data alone does not lead to changes in instruction or

better outcomes for struggling writers (McMaster et al., 2011). Professionals using such

measures will need to ensure that pupils are supported with effective, targeted teaching to

develop their writing skills.

References

Abbott, R. D., & Berninger, V. W. (1993). Structural equation modelling of relationships among

developmental skills and writing skills in primary-grade and intermediate-grade writers.

Journal of Educational Psychology, 85(3). doi: 10.1037/0022-0663.85.3.478

Apel, K., & Apel, L. (2011). Identifying Intraindividual Differences in Students' Written

Language Abilities. *Topics in Language Disorders, 31*(1), 54-72. doi:

10.1097/TLD.0b013e31820a22b4

Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to

seven: syntactic complexity and genre differentiation. *Reading and Writing, 24*(2), 183-

202. doi: 10.1007/s11145-010-9264-9

Berman, R., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text

production abilities:Speech and writing. *Written Language and Literacy, 5*(1), 43.

Berman, R. (2008). The psycholinguistics of developing text construction. *Journal of Child

Language, 35*(4). doi: 10.1017/s0305000908008787

Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008). Writing

problems in developmental dyslexia: Under-recognized and under-treated. *Journal of

School Psychology, 46*(1). doi: 10.1016/j.jsp.2006.11.008

Berninger, V. W., & Swanson, H. L. (1994). Modifying Hayes and Flowers' model of

skilled writing to explain developing writing. In E. C. Butterfield (Ed.), *Advances in

cognition and educational practice. Children's writing: Toward a process theory of the

development of skilled writing* (Vol. 2, pp. 1-30). Greenwich: JAI Press.

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., . . .

Graham, S. (2002). Teaching spelling and composition alone and together: Implications

for the simple view of writing. *Journal of Educational Psychology, 94*(2). doi:

10.1037//0022-0663.94.2.291

Campbell, H., Espin, C. A., & McMaster, K. (2013). The technical adequacy of curriculum-

    based writing measures with English learners. *Reading and Writing, 26*(3), 431-452. doi:

    10.1007/s11145-012-9375-6

Coker, D. L., & Ritchey, K. D. (2010). Curriculum-Based Measurement of Writing in

    Kindergarten and First Grade: An Investigation of Production aid Qualitative Scores.

    Exceptional Children, 76(2), 175-193.

Connelly, V., Barnett, A., & Dockrell, J. E. (2009). *The teaching of writing: Lessons from

    research*. UK: BPS Blackwell.

Connelly, V., Campbell, S., MacLean, M., & Barnes, J. (2006). Contribution of lower order

    skills to the written composition of college students with and without dyslexia.

    Developmental Neuropsychology, 29(1). doi: 10.1207/s15326942dn2901_9

Costa, L. J. C., Hooper, S. R., McBee, M., Anderson, K. L., & Yerby, D. C. (2012). The Use of

    Curriculum-Based Measures in Young At-Risk Writers: Measuring Change Over Time

    and Potential Moderators of Change. *Exceptionality, 20*(4), 199-217. doi:

    10.1080/09362835.2012.724623

DCSF. (2009). Getting to Grips with Assessing Pupils' Progress.*Report for Department for

    Children, Schools and Families*. Nottingham.

Department for Education. (2013). Key Stage 2 Writing exemplification, levels 2–6. Retrieved

    from

    http://www.education.gov.uk/schools/teachingandlearning/assessment/keystage2/b00200

    837/ks2_writing_moderation/eng-mod-ks2 website:

Dockrell, J. E., Ricketts, J., Charman, T., & Lindsay, G. (2014). Exploring writing products in

    students with language impairments and autism spectrum disorders. *Learning and

    Instruction, 32*(0), 81-90. doi: http://dx.doi.org/10.1016/j.learninstruc.2014.01.008

Espin, C., De La Paz, S., Scierka, B. J., & Roelofs, L. (2005). The relationship between

curriculum-based measures in written expression and quality and completeness of

expository writing for middle school students. *Journal of Special Education, 38*(4), 208-

217.

Espin, C., Shin, J., Deno, S. L., Skare, S., Robinson, S., & Benner, B. (2000). Identifying

indicators of written expression proficiency for middle school students. *Journal of

Special Education, 34*(3), 140-153.

Fewster, S., & MacMillan, P. D. (2002). School-based evidence for the validity of curriculum-

based measurement of reading and writing. *Remedial and Special Education, 23*(3), 149-

156.

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research.

School Psychology Review, 33(2), 188-192.

Fuchs, L. S., & Fuchs, D. (2009). On the Importance of a Unified Model of Responsiveness to

Intervention. *Child Development Perspectives, 3*(1), 41-43.

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*(4), 477-497.

Gansle, K. A., VanDerHeyden, A. M., Noell, G. H., Resetar, J. L., & Williams, K. L. (2006).

The technical adequacy of curriculum-based and rating-based measures of written

expression for elementary school students. *School Psychology Review, 35*(3), 435-450.

Graham, S., Berninger, V. W., Abbott, R. D., Abbott, S. P., & Whitaker, D. (1997). Role of

mechanics in composing of elementary school students. A new methodological approach.

Journal of Educational Psychology, 89(1). doi: 10.1037/0022-0663.89.1.170

Graham, S., & Harris, K. (2004). Writing instruction. In B. Wong (Ed.), *Learning about learning

disabilities* (3rd ed., pp. 281-313). San Diego, California: Elsevier.

Graham, S., & Harris, K. R. (2005). Improving the writing performance of young struggling writers: Theoretical and programmatic research from the center on accelerating student learning. *Journal of Special Education, 39*(1). doi: 10.1177/00224669050390010301

Graham, S., Harris, K., & Hebert, M. (2011). *Informing writing: The benefits of formative assessment. A Carnegie Corporation Time to Act report*. Washington, D.C.: Alliance for Excellent Education.

Graham, S., Hebert, M., & Harris, K. R. (2011). Throw 'em Out or Make 'em Better? State and District High-Stakes Writing Assessments. *Focus on Exceptional Children, 44*(1).

Graham, S., Morphy, P., Harris, K., Fink-Chorzempa, B., Saddler, B., Moran, S., & Mason, L. (2008). Teaching spelling in the primary grades: A national survey of instructional practices and adaptations. *American Educational Research Journal, 45*(3), 796-825. doi: 10.3102/0002831208319722

He, Q., Anwyll, S., Glanville, M. & Deavall, A. (2013). An investigation of the reliability of marking of the Key Stage 2 National Curriculum writing tests in England. *Educational Research, 55*(4), 393-410.

Hooper, S. R., Swartz, C. W., Wakely, M. B., de Kruif, R. E. L., & Montgomery, J. W. (2002). Executive functions in elementary school children with and without problems in written expression. *Journal of Learning Disabilities, 35*(1). doi: 10.1177/002221940203500105

Huot, B. (1990). The literature of direct writing assessment – major concerns and prevailing trends. *Review of Educational Research, 60*(2), 237-263. doi: 10.3102/00346543060002237

Juel, C. (1988). Learning to read and write- a longitudinal study of 54 children form 1st through 4th grades. *Journal of Educational Psychology, 80*(4), 437-447. doi: 10.1037/0022-0663.80.4.437

Koutsoftas, A. D., & Gray, S. (2012). Comparison of narrative and expository writing in students
        with and without language-learning disabilities. *Language, Speech, and Hearing Services
        in Schools, 43*(4). doi: 10.1044/0161-1461(2012/11-0018)

Lee, Y. W., Gentile, C., & Kantor, R. (2010). Toward Automated Multi-trait Scoring of Essays:
        Investigating Links among Holistic, Analytic, and Text Feature Scores. *Applied
        Linguistics, 31*(3), 391-417. doi: 10.1093/applin/amp040

MacArthur, C., Graham, S., & Fitzgerald, J. (2006). *Handbook of research on writing*. New
        York: Guildford.

Mansell, W., James, M., & Advisory Group. (2009). Assessment in schools. Fit for purpose? A
        Commentary by the Teaching and Learning Research Programme *Economic and Social
        Research Council, Teaching and Learning Research Programme*. London.

McCartney, E., Boyle, J., Ellis, S., Bannatyne, S., & Turnbull, M. (2011). Indirect language
        therapy for children with persistent language impairment in mainstream primary schools:
        outcomes from a cohort intervention. *International Journal of Language &
        Communication Disorders, 46*(1), 74-82. doi: 10.3109/13682820903560302

McMaster, K., & Espin, C. (2007). Technical features of curriculum-based measurement in
        writing - A literature review. *Journal of Special Education, 41*(2). doi:
        10.1177/00224669070410020301

McMaster, K., Ritchey, K., & Lembke, E. (2011). Curriculum-based mesurement for beginning
        writers:recent developments and future directions. In T. Scruggs & M. Mastropieri (Eds.),
        Assessment and Intervention (Vol. 24). Bingley,UK: Emerald Group Publishing.

McMaster, K. L., & Campbell, H. (2008). New and Existing Curriculum-Based Writing
        Measures: Technical Features Within and Across Grades. *School Psychology Review,
        37*(4), 550-566.

McMaster, K. L., Parker, D., & Jung, P.-G. (2012). Using Curriculum-Based Measurement for
        Beginning Writers Within a Response to Intervention Framework. *Reading Psychology,
        33*(1-2), 190-216. doi: 10.1080/02702711.2012.631867

Miller, B., & McCardle, P. (2011). Reflections on the need for continued research on writing. *Reading and Writing, 24*(2), 121-132. doi: 10.1007/s11145-010-9267-6

Myhill, D. (2009). Children's patterns of composition and their reflections on their composing processes. *British Educational Research Journal, 35*(1), 47-64. doi: 10.1080/01411920802042978

Myhill, D., & Jones, S. (2007). More than just error correction - Students' perspectives on their revision processes during writing. *Written Communication, 24*(4), 323-343. doi: 10.1177/0741088307305976

Myhill, D., & Jones, S. (2009). How talk becomes text: investigating the concept of oral rehearsal in early years' classrooms. *British Journal of Educational Studies, 57*(3), 265-284. doi: 10.1111/j.1467-8527.2009.00438.x

Nelson, N. K., & Tattersall, P. (2014). Assessing Integrated oral Language and Writing and Linking Assessment Results to Intervention. In B. Arfé, J. E. Dockrell & V. W. Berninger (Eds.), *Writing Development in Children with Hearing Loss, Dyslexia or Oral Language Problems: Implications for Assessment and Instruction.* New York: Oxford University Press.

Olinghouse, N. G. (2008). Student- and instruction-level predictors of narrative writing in third-grade students. *Reading and Writing, 21*(1-2), 3-26. doi: 10.1007/s11145-007-9062-1

Olinghouse, N. G., & Graham, S. (2009). The Relationship Between the Discourse Knowledge and the Writing Performance of Elementary-Grade Students. *Journal of Educational Psychology, 101*(1), 37-50. doi: 10.1037/a0013248

Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing, 26*(1), 45-65. doi: 10.1007/s11145-012-9392-5

Parker, R. I., Tindal, G., & Hasbrouck, J. (1991). Progress monitoring with objective measures of writing performances for students with mild disabilities. *Exceptional Children, 58*(1), 61-73.

Persky, H. R., Daane, M. C., & Jin, Y. (2003). *The Nation's Report Card: Writing 2002*, *U.S. Department of Education. Institute of Education Sciences. National Center for Education Statistics*. Washington, D.C.

Puranik, C. S., & AlOtaiba, S. (2012). Examining the contribution of handwriting and spelling to written expression in kindergarten children. *Reading and Writing, 25*(7). doi: 10.1007/s11145-011-9331-x

Puranik, C. S., Lombardino, L. J., & Altmann, L. J. P. (2008). Assessing the microstructure of written language using a retelling paradigm. *American Journal of Speech-Language Pathology, 17*(2), 107-120.

QCA. (2008). Assessing pupils' progress: Assessment at the heart of learning *Report for Qualifications and Curriculum Authority*.

Ritchey, K. D., & Coker, D. L. (2013). An Investigation of the Validity and Utility of Two Curriculum-Based Measurement Writing Tasks. *Reading & Writing Quarterly, 29*(1), 89-119. doi: 10.1080/10573569.2013.741957

Rust, J. (1996). The manual of the Wechsler Objective Language Dimensions (WOLD): UK Edition. London, England: Psychological Corporation.

Saddler, B., & Asaro-Saddler, K. (2013). Response to Intervention in Writing: A Suggested Framework for Screening, Intervention, and Progress Monitoring. *Reading & Writing Quarterly, 29*(1), 20-43. doi: 10.1080/10573569.2013.741945

Scott, C. M., & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech Language and Hearing Research, 43*(2), 324-339.

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42*(8), 795-819. doi: 10.1002/pits.20113

Tindal, G., & Parker, R. (1989). Assessment of written expression for students in compensatory and special education programs. *Journal of Special Education, 23*(2).

Torrance, M., Alamargot, D., Castelló, M., Ganier, F., Kruse, O., Mangen, A., . . . van Waes, L.

  (2012). *Learning to Write Effectively: Current Trends in European Research*: Emerald.

Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., Wilson, L. G., Tschinkel, E., & Kantor,

  P. T. (2011). Modeling the development of written language. *Reading and Writing, 24*(2),

  203-220. doi: 10.1007/s11145-010-9266-7

Weissenburger, J. W., & Espin, C. A. (2005). Curriculum-based measures of writing across

  grade levels. *Journal of School Psychology, 43*(2), 153-169. doi:

  10.1016/j.jsp.2005.03.002

Weschler, D. (2005). Weschler Individual Achievement Test (WIAT-II UK). London: Pearson;

  Harcourt Assessments.

Table 1. Mean (standard deviation) productivity scores for narrative and expository

probes for years 3, 4 and 5

Year Group		3	4	5
CBM-W Measure	Genre type	Mean *(SD)*	Mean *(SD)*	Mean *(SD)*
Total words	Narrative	44.35 (18.00)	51.71 (18.28)	58.93 (26.92)
	Expository	35.71 (16.33)	37.54 (15.23)	45.56 (17.15)
Correct word sequences	Narrative	37.86 (17.85)	41.80 (16.76)	49.97 (22.93)
	Expository	31.18 (15.93)	32.98 (15.23)	38.70 (17.03)

Number of punctuation marks	Narrative	3.15 (3.25)	3.04 (2.66)	4.97 (4.32)
	Expository	3.32 (3.07)	3.77 (3.44)	5.29 (4.31)
Total sentences	Narrative	2.40 (2.06)	1.88 (1.71)	3.11 (2.46)
	Expository	2.14 (1.94)	2.12 (2.07)	2.46 (2.12)

Table 2. Mean (standard deviation) accuracy scores for narrative and expository probes

for years 3, 4 and 5

Year group		3	4	5
CBM-W Measure	Genre type	Mean *(SD)*	Mean *(SD)*	Mean *(SD)*
Proportion of words spelled correctly	Narrative	.85 (.10)	.88 (.12)	.90 (.08)
	Expository	.82 (.14)	.85 (.11)	.89 (.11)
Proportion of correct word sequences	Narrative	.81 (.12)	.78 (.14)	.84. (14)
	Expository	.82 (.15)	.83 (.16)	.80 (.15)

Proportion of accurate punctuation marks	Narrative	.90 (.19)	.86 (.25)	.95 (.13)
	Expository	.93 (.16)	.91 (.20)	.92 (.17)
Proportion of correct complete sentences	Narrative	.57 (1.0)	.62 (.42)	.76 (.33)
	Expository	.69 (.41)	.67 (.40)	.70 (.37)

Table 3. Bivariate correlations between CBM-W narrative performance (above the diagonal) and expository performance (below the diagonal) and WOLD raw scores controlling for age

	1	2	3	4	5
1. WOLD Raw score		$.42^{***}$	$.45^{***}$	$.32^{***}$	$.30^{***}$
2. Total Words produced	$.44^{***}$		$.94^{***}$	$.32^{***}$	$.23^{***}$
3. Correct word sequences	$.51^{***}$	$.88^{***}$		$.36^{***}$	$.32^{***}$
4.Number of punctuation marks	$.48^{***}$	$.47^{***}$	$.60^{***}$		$.18^{**}$
5. Proportion of words spelled correctly	$.33^{***}$	$.16^{**}$	$.31^{***}$	$.29^{***}$	

Significance ** p =.005, *** p ≤ .001

Table 4. Principal component analysis

Writing measures	Component scores	
	Productivity	Accuracy
Narrative total words produced	**.85**	.12
Expository total words produced	**.89**	.05
Narrative CWS	**.85**	.26
Expository CWS	**.87**	.15
Narrative total punctuation	**.60**	.40
Expository total punctuation	.44	.35
Narrative proportion of words spelled correctly	.08	**.87**
Narrative proportion of words spelled correctly	.18	**.78**
WOLD raw score	**.58**	**.47**

Table 5. Gain scores for CBM measures

Year group		3	4	5
CBM-W Measure	Genre type	Mean *(SD)*	Mean *(SD)*	Mean *(SD)*
Total number of words	Narrative	5.76 (18.81)	2.25 (22.40)	7.74 (23.52)
	Expository	4.38 (17.85)	9.37 (22.79)	6.88 (23.61)
Correct word sequences	Narrative	5.5 (17.80)	6.85 (22.79)	10.49 (25.36)
	Expository	2.90 (18.15)	6.60 (22.57)	8.52 (23.92)
Correct punctuation marks	Narrative	.14 (4.02)	1.02 (4.66)	.36 (5.29)
	Expository	-.31 (3.44)	.85 (4.66)	.39 (5.16)
Proportion words spelled correctly	Narrative	.0003 (.10)	.02 (.11)	.02 (.09)
	Expository	.02 (.14)	.02 (.13)	-.0003 (.11)

**Table 6. Comparison of pupils with and without Special Educational Needs (SENs) on the
CBM-W**[1]

CBM-W Measure		Total number of words		Correct word sequences		Correct punctuation marks	
	Genre type	Narrative	Expository	Narrative	Expository	Narrative	Expository
SENs	Mean	37.00	33.40	26.70	26.65	2.20	2.80
	(SD)	(13.31)	(12.67)	(12.50)	(10.20)	(2.1)	(2.80)
No SENs	Mean	61.78	47.00	47.63	38.78	3.85	4.10
	(SD)	(19.7)	(17.33)	(16.8)	(14.88)	(3.67)	(3.59)
Significance Effect size		$t(77) = 5.22$ $p < .001$ $d = 1.19$	$t(78) = 3.23$ $p = .002$ $d = .73$	$t(77) = 5.10$ $p < .001$ $d = 1.16$	$t(47.74) = 4.07$ $p < .001$ $d = 1.18$	$t(77) = 1.90$ *ns.*	$t(77) = 1.48$ *ns.*

---

[1] Where Levene's test for equality of variances was significant df for equal variances not

assumed were used