

The Role of Curriculum Based Measures in Assessing Writing Products

Julie Dockrell, Vincent Connelly, Kirsty Walter and Sarah Critten

The Challenge

Literacy, including writing, is a key gateway skill (Buchanan & Flouri, 2001). Failure to produce text quickly, legibly, and accurately results in poor educational achievements, reduced job opportunities, and reduced earning potential. Thus, the ability to produce written text, either manually or electronically, is a key transferable skill. The heightened awareness that “writing today is not a frill for the privileged few, but an essential skill for the many” (The National Commission on Writing, 2003, p. 11), coupled with the numbers of students 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 (Bew, 2011).

These assessments should capture the key components of written text production as children are learning to write (Berninger et al., 2002). Arguably such assessments should be timely and lead to targeted teaching or specific interventions that can be monitored (Saddler & Asaro-Saddler, 2013). 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 chapter considers the assessment of children’s written texts by exploring the use of curriculum-based measures of writing (CBM-W). We draw on data from children in English primary schools (Dockrell, Connelly, Walter & Critten, 2015) to consider the extent to which such measures capture developmental differences and changes in writing skills over time (Fewster & MacMillan, 2002). Using these data, we also assess whether the CBM-W can accurately reflect the performance found on more time-consuming and complex standardized measures of written text, and finally, we consider if CBM-W can identify struggling writers.

Assessing Writing

In England, National Curriculum (NC) assessments have provided teachers with a framework from which to evaluate children's written text products (Qualifications and Curriculum Authority—QCA 1999; Department for Education, 2013). The NC set out both the programme of study that schools needed to follow and the attainment targets that pupils were expected to achieve. However, the use of NC assessments for writing has not gone unchallenged. There is substantial variation in the marks assigned to the same script by trained markers, with evidence of regression to the mean at both ends of the distribution (He, Anwyll, Glanville, & Deavall, 2013). This has implications for teachers' assessment of writing and the extent to which the tests are "fit for purpose" (Bew, 2011). In addition, teachers have found the assessment schemes overly bureaucratic, with English primary teachers devoting, on average, 5 hours a week to assessing and marking pupils' work (Deakin, James, Tickner, & Tidwell, 2010). Teachers comment on the perceived heavy workload and complexity of current writing assessments, mirroring earlier comments on this particular assessment of writing drawn from a large-scale research review (Stanley, MacCann, Gardner, Reynolds, & Wild, 2009). Therefore, there is scope for exploring less complex and less time-consuming alternative forms of assessment for writing and establishing whether adequate levels of reliability, sensitivity, and acceptability can be established for such measures.

Written texts 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; this type of assessment is typical in research studies. By contrast, formative assessment is more likely to involve tracking performance over time to identify both progress and the child's strengths and needs (Mansell et al., 2009). 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.

Various approaches to evaluation of written composition have been used by researchers and teachers, including holistic scoring, analytic scoring, quantitative scoring, and curriculum based measures (Abbott & Berninger, 1993; Dockrell, Ricketts, Charman, & Lindsay, 2014; Lee, Gentile, & Kantor, 2010; Mackie & Dockrell, 2004; McMaster & Espin, 2007; Puranik & Al Otaiba, 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 that are made.

Holistic scoring measures have been used in research, psychometric assessments, and in practice. Global quality of the text is rated on a single ordinal scale (see, for example, Weschler, 2005) rather than on any specific dimensions of the text produced. Holistic measures have the advantage of providing a single score with relatively little time involvement, but they are limited in their ability to reliably differentiate among writing levels, monitor change over time, or capture differential performance on the key components of writing (Espin et al., 2000; Kim et al., 2015). In younger children and those with learning disabilities, the short amount of text often written by these children 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 Language Dimensions of writing or UK writing Key Stage 2 SATS assessment scores (Department for Education, 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 the assessor to have specialized training to reliably identify the target dimensions, and the construct validity of the various analytical dimensions is often lower than for holistic scoring schemes (Espin, De La Paz, Scierka & Roelofs, 2005; Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002).

The number of dimensions thought to underpin written text production has been a matter of debate. Earlier studies of composition identified two dimensions in written texts: quality and productivity (Berninger & Swanson, 1994; Graham, Berninger, Abbott, Abbott, & Whitaker, 1997; Olinghouse & Graham, 2009). Recently researchers have refined these dimensions by including 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 (e.g., numbers of words generated), and accuracy (Puranik, Lombardino, & Altmann, 2008; Wagner et al., 2011).

Text genre is an added consideration when evaluating written texts. Most research assessments of pupils' writing rely on single transcripts and single genres and, as such, may not be consistent with children's writing competence. Because children need to learn to write for multiple purposes and multiple audiences, assessments which only examine a single writing product may fail to capture the demands of different types of writing tasks (Scott & Windsor, 2000). For example, narrative and expository writing are common school tasks, but expository texts which involve conveying facts or describing procedures take longer to master (Berman & Verhoeven, 2002), and differences identified

in students' performances are evident in their texts (Beers & Nagy, 2011; Koutsoftas & Gray, 2012). Thus, a tool that can be used for different text types and can differentiate among them has advantages in supporting teaching and learning. Currently, it remains to be established whether measures of text production can reliably differentiate across genres and whether any differences identified should inform the assessment of writing.

Overall, assessing writing is challenging. Arguably assessments of pupils' written text should capture dimensions of productivity and accuracy without placing undue demands on staff training and time. In addition, the ability to distinguish across text types would provide teachers with a flexible approach to the assessment of writing.

Formative Assessment

Timely sensitive assessment of students' writing competencies is a key step to monitoring progression (Nelson, 2014), but students need to be assessed on reliable and valid measures; frequently high stakes national tests do not provide this information (Graham, Hebert, & Harris, 2011). Thus, it becomes important to examine the ways in which formative assessment can drive writing development for students. Formative assessment produces a picture of learners' strengths and needs in terms of their writing skills. It requires professional judgment, something that is often challenging without objective and measurable benchmarks (Marlow et al., 2014).

Writing needs to be evaluated at the word, sentence, and text levels, capturing the key dimensions of both productivity and accuracy for pupils and examining different phases of writing development (Connelly & Dockrell, 2015; Dockrell & Connelly, 2016). Monitoring change is a key component in that activity, one which requires two basic elements to be effective: first students need to be assessed over time, and second, the writing task needs to be tailored to the competencies that are being examined. In addition, where possible, comparisons should be made across different writing genres (Berman, 2008; Olinghouse & Wilson, 2013). Curriculum based measures (CBM) offer one solution to these challenges.

CBMs offer a way of measuring a child's academic progress through direct assessment of specific 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 pupils' productivity and accuracy of 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 assess-

ments 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 7 and 12 (Weissenburger & Espin, 2005). Thus, they provide a potentially quick and reliable assessment of younger school children's writing products. Given their dual focus on productivity and accuracy, CBMs also reflect current models of the writing process.

A variety of different text measures have been used to evaluate productivity, including numbers of words written, correct word sequences (CWS), punctuation marks, and words spelled correctly, and the appropriateness of these measures varies with the pupil's age. (See McMaster & Espin, 2007, for a review of the technical features of the measures). 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 assessment of a complex skill such as writing (McMaster & Espin, 2007; Ritchey & Coker, 2013). The inclusion of correct word sequences and measures of spelling may also provide more face validity for teachers (Coker & Ritchey, 2010; Gansle, Noell, VanDerHeyden, Naquin, & Slider, 2002).

The measures of the writing product vary in their scoring reliability. Inter-rater reliability can be high, with 80–90% agreement (Gansle, VanDerHeyden, Noell, Resetar, & Williams, 2006; Graham, Harris, & Hebert, 2011; Weissenburger & Espin, 2005). There is also evidence of validity where some elements are correlated with standardized assessments (.69 for TOWL) and with teacher ratings (.76) (Parker, Tindal, & Hasbrouck, 1991). For example, Espin and colleagues (Espin et al., 2000) confirmed the validity and reliability of the number of CWS as an indicator of general writing performance. In general, validity is higher when the CBM-W is being examined in relation to a writing task rather than more general performance across the curriculum (McMaster & Espin, 2007). Simpler measures such as total words written have lower criterion-related coefficients than more complex measures, such as correct word sequences or measures reflecting spelling and word choice. Narrative probes have demonstrated the best technical adequacy to date (see McMaster & Campbell, 2008).

The use of a CBM has also not gone unchallenged. There are a number of significant criticisms of the use of such measures, criticisms which relate to scoring, the extent to which these measures are valid across different populations, and their sensitivity in capturing text quality. Despite the challenges of some children's handwriting, good inter-rater reliability has been achieved following training for numbers of words and words spelled correctly, but other measures are not so straightforward and require a more subjective judgment. Moreover,

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 more sensitive indicators of early writing. Work by McMaster and colleagues has begun to systematically address these issues with children below the age of eight (McMaster, Ritchey, & Lembke, 2011). Most work using the CBM-W has examined static scores, and evidence examining growth trajectories has been contradictory. McMaster and colleagues found stable and valid growth curves for children between the ages of eight and nine. However, Costa and colleagues (Costa, Hooper, McBee, Anderson, & Yerby, 2012) concluded that only the CBM-W variables of total words written, words spelled correctly, and correct word sequences showed clear developmental trends and argued that CBM-W be used in combination with other forms of assessment. A final limitation rests in the paucity of research on CBM-W 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 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, for example, the current English Key Stage 2 (age 7–11 years) where children are expected to make much progress as young writers. As yet we do not know which, if any, CBM-W elements differentiate pupils' writing performance or whether these are sensitive to change over time (Ritchey & Coker, 2013).

Our Study

To further examine the potential use of a writing CBM (CBM-W), we studied 263 pupils who were 8, 9, and 10 years old (English school Years 3, 4, and 5). We used both a narrative and an expository probe to examine genre effects on performance and examined changes over a five-month period to track developmental trajectories. (For full details of the study, see Dockrell, Connelly, Walter & Critten, 2015.) We scored the texts for productivity (total words produced, correct word sequences (CWS), number of punctuation marks, and sentences produced) and accuracy (proportion of words spelled correctly, CWS, and punctuation marks).

For all productivity measures, except sentences produced, we found significant age trends and genre effects, with narrative genre resulting in increased productivity. In Figure 10.1 the changes for total numbers of words produced are presented. As Figure 10.2 shows, there were also clear and significant differ-

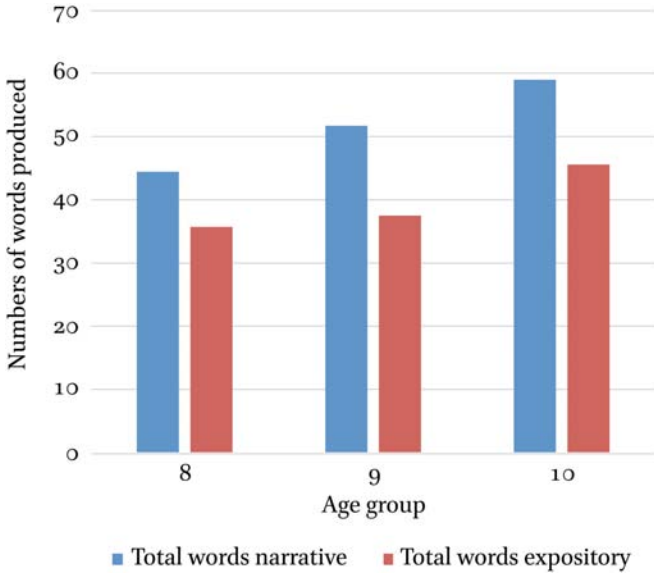


FIGURE 10.1 Mean numbers of words produced in five minutes in typically developing children

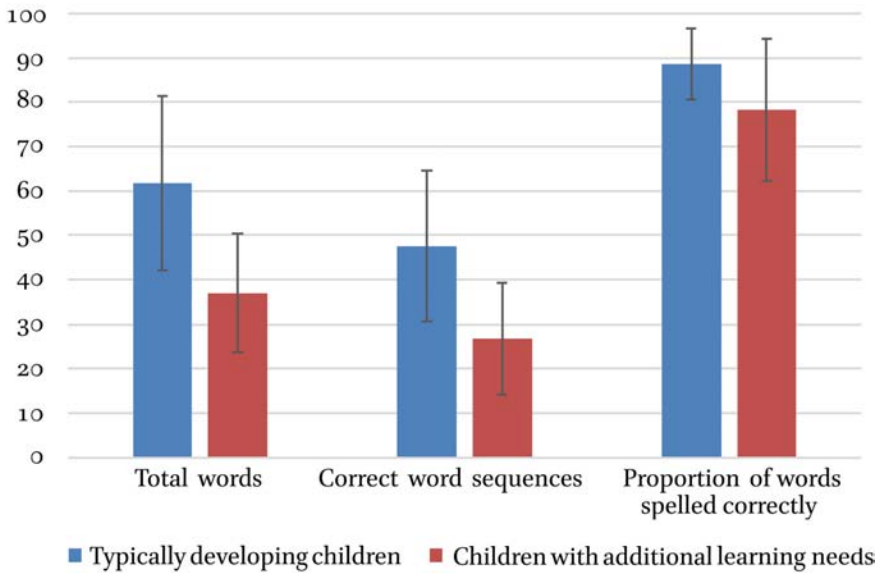


FIGURE 10.2 Mean (SD) on CBM productivity measures for children typically developing children and those with special educational needs

ences between children who had additional learning needs and their chronologically age-matched typically developing peers. The differences were large and significant for all measures of productivity and showed little overlap between the two groups as illustrated by the standard deviations.

Overall the results were consistent with other reports in the literature. The different elements of the CBM-W were differentially sensitive to development with clear age trends. We also found good construct validity as evidenced by their significant association with the norm-referenced test measuring writing quality, the Wechsler Objective Language Dimensions (Rust, 1996), that had been standardized in the UK against an age range from 8 to 16 years of age. All the different elements of the CBM-W (for both narrative and expository texts) correlated significantly with the overall quality scores from the norm-referenced test measuring writing quality. Furthermore, a principal components analysis demonstrated that both the norm-referenced test and the CBM tasks loaded onto two constructs of writing which represented measures of productivity and measures of accuracy. Change over time was also evident, and significant differences between narrative and expository texts were found. Pupils with special educational needs scored significantly less well on the CBM-W, demonstrating the sensitivity of the measure to identify struggling writers.

There was also evidence of more complex patterns across the different measures of accuracy and productivity. For example, while numbers of sentences produced failed to discriminate by age and genre across the sample, it was an important factor for the older children. Developments in writing for older competent writers may be more evident if detailed assessments of sentence structure at the clausal level are used (Berman, 2008). However, such analyses are not compatible with rapid assessment. As yet there are also no data suggesting that such sentence structure measures change reliably with development at this age.

In contrast to some previous work (McMaster & Espin, 2007), the only accuracy measure that discriminated between age groups was the proportion of words spelled correctly. This result replicates that of Costa and colleagues (Costa et al., 2012). Transcription skills, both handwriting and spelling, account for the majority of the variance in writing quality for both children developing typically and those with developmental difficulties learning to write in English (Berninger, Nielsen, Abbott, Wijsman, & Raskind, 2008; Graham et al., 1997; Olinghouse, 2008). Our data suggest that a short (five-minute) text written to a writing probe can effectively capture these differences.

It is important to note that despite the statistically significant differences, and in many cases the large effect sizes, found in our study, there was marked heterogeneity within the age groups. Ninety-five percent confidence inter-

vals revealed little overlap between the year groups, but 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 standardized Wechsler Objective Language Dimensions (Rust, 1996) scores for the sample; the national data on writing tests for England also show similar patterns of heterogeneity within year groups (Department for Education, 2011, 2012). In fact, despite this heterogeneity, there was little overlap between the scores for pupils with Special Educational Needs and those with no recorded special needs, suggesting that CBM-W may be a useful objective measure for children who are struggling to develop writing skills. The variance in children's scores at this stage in writing development can thus serve as a sensitive marker for monitoring progress and identifying pupils struggling with writing.

Sensitivity to growth was evident for two productivity measures (CWS and total words written) and for proportion of words spelled correctly, a measure of accuracy. Effect sizes for both CWS and total words written were large. This is a promising finding, as teachers may be able to track progress using these more objective measures, which also provide for the detailed monitoring of children, especially those who struggle with writing. 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 pupils' writing.

Performance on CBM-W also reliably differentiated between narrative and expository texts. This confirms previous work examining these genre differences (Koutsoftas & Gray, 2012; Scott & Windsor, 2000) and provides a further source of information about the validity of the CBM-W. Pupils produced less text and less accurate text 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. In order to differentiate between pupils 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 we used for children with this range of ages and educational needs. Good reliability of the scoring was established and there was validity with a UK nationally-standardized measure. The CBM-W differentiated across year groups and for pupils with and without special educational needs. It was also sensitive to change over the five-month period of 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 age 7 to age 11 for both typical and atypical

writers and can chart change over periods of time within those years. There is more work required to see if CBM 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 common (McMaster & Espin, 2007).

There is still much to investigate around CBM-W and more specific research is required to specify the utility of CBM-W at various ages and to identify which measures best reflect the underlying strengths or weaknesses in the children's written products. For example, the utility of CBM-W may change depending on the macro-structural dimensions of the text that are assessed. This future work will be important for demonstrating if CBM-W can be used to support directly the development of specific skills in writing through interventions. Similarly, it may be that a different measure of text complexity at the 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 ages and their writing skills.

In our work we examined children up to age 11 but there is currently limited data on CBM-W in older children's writing and in advanced older writers in post-school education. More detailed validity profiles, for example, examining informational validity for instructional actions are also lacking. It may be the case that CBM-W may be more useful in older writers for tracking struggling pupils as the higher-level components of writing such as ideation or narrativity may be more difficult to measure with a CBM approach. However, some recent research has shown that linguistic-based automated text analysis can detect complex differences in narrativity in written texts between children at U.S. High School and this may provide a fruitful avenue for research on older writers' more complex texts (Allen, Snow, & McNamara, 2016).

In terms of our own data presented here, there are also limitations. For example, we were unable to control for potential teacher effects across the study and we did not have detailed data on the nature of the children's learning difficulties. Trained graduate assistants coded the texts, so it is not yet 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 only provide partial information on writing as they assess the product but not the process of writing. However, further research tackling these issues may, in turn, help enhance the face value of CBM-W. Some teachers and researchers remain to be convinced that the complex set of processes represented in writing can be adequately measured by seemingly simple measures such as the number of words and spelling errors.

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. 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 the relatively straightforward 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, but we hope that CBM-W can be a useful tool to assist them in this process.

References

- Abbott, R.D., & Berninger, V.W. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary-grade and intermediate grade writers. *Journal of Educational Psychology, 85*(3), 478–508. doi: 10.1037/0022-0663.85.3.478
- Allen, L.K., Snow, E.L., & McNamara, D.S. (2016). The narrative waltz: The role of flexibility in writing proficiency. *Journal of Educational Psychology, 108*(7), 911–924. <http://dx.doi.org/10.1037/edu0000109>
- 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.A. (2008). The psycholinguistics of developing text construction. *Journal of Child Language, 35*(4), 735–771. doi: 10.1017/S0305000908008787
- Berman, R., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text production abilities: Speech and writing. *Written Language and Literacy, 5*(1), 1–43.
- 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., 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), 1–21. doi: 10.1016/j.jsp.2006.11.008

- 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), 291–304. doi: 10.1037//0022-0663.94.2.291
- Bew, P. (2011). *Independent review of Key Stage 2 testing, assessment and accountability—Final Report DFE-00068-2011*. London: Department for Education.
- Buchanan, A., & Flouri, E. (2001). Recovery after age 7 from externalizing behavior problems: The role of risk and protective factors. *Children and Youth Services Review*, 23, 899–914.
- 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., & Dockrell, J.E. (2015). Writing development and instruction for students with learning disabilities: Using diagnostic categories to study writing difficulties. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research, 2nd Edition*. (pp. 349–363). New York: Guilford Publications.
- 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
- Deakin, G., James, N., Tickner, M., & Tidswell, J. (2010). *Teachers' workload diary survey 2010. Research report. Department for Education*. UK. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/181773/DFE-RR057-WEB.pdf
- Department for Education (2011). *The framework for the National Curriculum: A report by the expert panel for the National Curriculum review*. London: DfE [online]. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/175439/NCR-Expert_Panel_Report.pdf
- Department for Education (2012). *National Curriculum assessments*. Retrieved from <https://www.gov.uk/government/collections/statistics-key-stage-2>
- Department for Education (2013). *Key Stage 2 Writing exemplification, levels 2–6*. Retrieved from http://www.education.gov.uk/schools/teachingandlearning/assessment/keystage2/b00200837/ks2_writing_moderation/eng-mod-ks2
- Dockrell, J.E., & Connelly, V. (2016). The relationships between oral and written sentence generation in English speaking children: The role of language and literacy skills. In J. Perrara, M. Aparici, E. Rosado, & N. Salas. (Eds.), *Written and spoken language development across the lifespan* (pp. 161–177). New York: Springer International Publishing.

- Dockrell, J.E., Connelly, V., Walter, K., & Critten, S. (2015). Assessing children's writing products: The role of curriculum based measures. *British Educational Research Journal*, 41(4), 575–595.
- 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, 81–90. doi: 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., & 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), 170–182. doi: 10.1037/0022-0663.89.1.170
- 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), 1–12.
- He, Q., Anwyll, S., Glanville, M., & Deavall, A. (2013). An investigation of the reliability of marking of the Key Stage 2 National Curriculum English writing tests in England. *Educational Research*, 55(4), 393–410.
- 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
- Kim, Y.-S.G., Park, C., & Park, Y. (2015). Dimensions of discourse level oral language skills and their relation to reading comprehension and written composition: An exploratory study. *Reading and Writing*, 28(5), 633–654. doi:10.1007/s11145-015-9542-7
- 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), 395–409. 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.
- Mackie, C., & Dockrell, J.E. (2004). The nature of written language deficits in children with SLI. *Journal of Speech, Language, and Hearing Research*, 47(6), 1469–1483.
- Mansell, W., James, M., & Advisory Group (2009). *Assessment in schools. Fit for purpose? A commentary by the Teaching and Learning Research Programme*. London: Economic and Social Research Council, Teaching and Learning Research Programme.
- Marlow, R., Norwich, B., Ukoumunne, O.C., Hansford, L., Sharkey, S. & Ford, T. (2014). A comparison of teacher assessment (APP) with standardised tests in primary literacy and numeracy (WIAT-II). *Assessment in Education: Principles, Policy & Practice*, 21, 412–426.
- 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.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., & Espin, C. (2007). Technical features of curriculum-based measurement in writing—A literature review. *Journal of Special Education*, 41(2), 68–84. doi: 10.1177/00224669070410020301
- McMaster, K., Ritchey, K., & Lembke, E. (2011). Curriculum-based measurement for beginning writers: Recent developments and future directions. In T. Scruggs & M. Mastropieri (Eds.), *Assessment and Intervention* (Vol. 24, pp. 111–148). Bingley, UK: Emerald Group Publishing.
- 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
- Nelson, N. (2014). Integrating language assessment, instruction and intervention in an inclusive writing lab approach. In B. Arfé, J. Dockrell, & V. Berninger (Eds.),

- Writing development in children with hearing loss, dyslexia or oral language problems* (pp. 273–300). New York, NY: 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.
- Puranik, C.S., & Al Otaiba, S. (2012). Examining the contribution of handwriting and spelling to written expression in kindergarten children. *Reading and Writing, 25*(7), 1523–1546. 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.
- Qualification and Curriculum Authority—QCA (1999). *The national curriculum handbook for primary teachers*. London: HMSO.
- 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.
- Stanley, G., MacCann, R., Gardner, J., Reynolds, L., & Wild, I. (2009). Review of teacher assessment: Evidence of what works best and issues for development (R. o. Q.C. 2686, Trans.).
- 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

- The National Commission on Writing (2003). *The neglected R: The need for a writing revolution*. Washington, DC: The College Board. Retrieved from: http://www.writingcommission.org/prod_downloads/writingcom/neglectedr.pdf
- 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.