

Writing productivity development in elementary school: A systematic review

Catherine Martin ^{*,1}, Julie E. Dockrell ^{2,3}

Psychology and Human Development, UCL, IOE, London, England, UK

ARTICLE INFO

Keywords:

Systematic review
Writing productivity
Elementary school
Words written
Correct word sequences
Text genre

ABSTRACT

The ability to produce fluent and coherent written text impacts learning and attainments. Valid and reliable assessments of writing are needed to monitor progression, develop goals for writing and identify struggling writers. In order to inform practice and research a systematic review was conducted to investigate which writing productivity measures captured writing development and identified struggling writers in elementary school. Sixty-seven empirical studies were identified for inclusion, appraised, and their data extracted under the themes of writing genre, duration of writing task, use of priming of topic knowledge prior to the writing assessment, use of planning time, writing modality, gender, age of participants and learning difficulties. Total Number of Words and Correct Word Sequences were the most common means of measuring productivity. Productivity varied significantly between genres and durations of writing tasks and was higher in girls than boys. Students with learning difficulties scored significantly lower in writing productivity when compared to typically developing peers. Insufficient research was available to draw conclusions regarding the effects of priming of topic knowledge, planning and modality on writing productivity. Study limitations, links to the assessment of writing and recommended further research are discussed.

Writing is a higher order skill that develops over time through interactions between the child's skills and cognitive resources, the instructional context, and the demands of the writing task (Kellogg, 2008). Significant advances have been made in our understanding of writing development and the demands placed on the writer. By corollary there has been an increase in research designed to identify effective approaches to the teaching of writing (see special issue of Reading and Writing, Vol 34, No 7: Teaching Writing to Elementary Grade Students) and targeted interventions (Gillespie & Graham, 2014). Conceptual and practical advances are underpinned by reliable and valid assessments of written products and an awareness of their strengths and limitations. Text produced at word and sentence levels are commonly used metrics to evaluate the written products of children in elementary school. Although the amount of text produced is not the ultimate goal of writing, longer texts allow ideas to be articulated, developed and elaborated with sufficient detail (Kim et al., 2014). The current systematic review focuses on metrics used to evaluate text production in elementary school measured in terms of the amount of written text generated (words, utterances or ideas) in a composition (Dockrell et al., 2018; Puranik et al., 2008) and the accuracy of the amount of text produced (correct word sequences) in order for benchmarks to be established. To

* Corresponding author.

E-mail address: catherine.martin.20@alumni.ucl.ac.uk (C. Martin).

¹ Orcid ID 0000-0002-0719-2987

² Postal Address: IOE, UCL, 20 Bedford Way, London, WC1H 0AL.

³ Orcid ID 0000-0003-3595-6064

<https://doi.org/10.1016/j.asw.2024.100834>

Received 2 September 2022; Received in revised form 25 February 2024; Accepted 2 March 2024

Available online 27 March 2024

1075-2935/© 2024 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

our knowledge, no systematic review has yet examined differences in the measures of text production across the elementary school grades and the factors that are associated with the amount produced and text accuracy. The current systematic review aims to address this gap in our understanding of count measures that capture children's written products and, where possible, provide indicative benchmarks for assessments and interventions.

The assessment of writing is dependent on sensitive measures of writing competencies; that is, measures which reveal the student's ability to produce a written text and capture the characteristics of proficient writing competence. Assessment of writing has been identified as problematic, with suggestions that it is the single greatest barrier to both writing instruction and research (Cole, Haley, & Muenz, 1997; Dockrell & Connelly, 2021). Early studies regarding written composition identified quality and productivity as key dimensions for assessment (Berninger & Swanson, 1994; Graham et al., 1997; Olinghouse & Graham, 2009), with the quality of writing products operationalised through the content and organisation of information (Graham et al., 1997; Wechsler, 2009) and productivity operationalised as the number of words, sentences, or ideas produced within the writing products, sometimes described as writing fluency (Abbott & Berninger, 1993; Kim et al., 2011). Assessing quality or productivity alone fails to capture the multidimensional nature of writing, typically excluding macro-structure (text organization, structure and cohesion at the whole-text level) and micro-structure elements (complexity and mechanics at a word, sentence or whole-text level) (Dockrell & Connelly, 2021; Kim, Al Otaiba et al., 2015; Koutsoftas, 2018). Nonetheless count metrics of the amount and the accuracy of text produced continue to feature as a key component in writing models, as an indicator of struggling writers and a component assessed in several writing intervention studies (Puranik et al., 2008; Salas et al., 2021). These measures are important as they are strongly correlated with writing quality (Puranik et al., 2008). Measures of text produced are important as they are, typically, more reliable than assessments of writing quality and as such are an ideal starting point for systematically examining changes over development and establishing developmental benchmarks (Puranik et al., 2008). However, to do so requires an understanding of which measures vary across development and which factors are associated with performance.

To date there has been no consistent approach to identifying which measures of the written product should be considered and developmental benchmarks have yet to be established. A wide range of metrics have been used to evaluate children's written products, variously referred to as productivity, fluency and accuracy (see for example <https://www.interventioncentral.org/assessment/writing-assessment-elementary>; Abbott & Berninger, 1993; Kim, Puranik et al., 2015). These metrics have considered both the amount produced and the accuracy of the text produced. The amount written has been captured by Total Number of Words (TNW), numbers of sentences, ideas and/or punctuation marks (Beers & Nagy, 2011; Dockrell et al., 2018; Kim, Al Otaiba et al., 2015; Kim et al., 2014; Puranik et al., 2008; Ritchey & Coker, 2014; Scott & Windsor, 2000; Dockrell & Connelly, 2020). Accuracy measures, by contrast, reflect the grammatical accuracy and complexity of the produced texts. Metrics have included Correct Word Sequences (CWS; the number of two adjacent writing units, such as words or punctuation, that are grammatically correct and spelled correctly), clauses, and T-units (defined as consisting of a main clause plus all subordinate clauses and non-clausal structures that are attached or embedded to it). CWS, Words Spelt Correctly (WSC), and the difference between the number of CWS in a writing sample and the number of incorrect CWS (CIWS) are often considered the most reliable measures of accurate text production (McMaster & Campbell, 2008) with strong technical adequacy across grades. However, differential patterns of performance across the measures have not been systematically addressed. Both TNW and CWS are considered robust measures of writing development (Kim et al., 2014; Weiss et al., 2019), that are sensitive to age (Wagner et al., 2011; Koutsoftas, 2018; Dockrell & Connelly, 2015), and strongly correlated with teacher ratings (Gansle et al., 2004; Jewell & Malecki, 2005) and standardised writing assessments (Dockrell, Connelly, Walter, & Critten, 2015; Gansle et al., 2006; Jewell & Malecki, 2005). There is growing evidence that text production metrics can be used successfully in research and practice (Piercy & Dockrell, 2023). However, correlations weaken as children become more competent writers (Kim et al., 2014; Weissenburger & Espin, 2005), comparisons between different metrics of written text production are underexplored and longitudinal data are rare. In the current systematic review we examine data on different measures of written text production capturing amount and accuracy to examine the possibility of producing reliable developmental benchmarks.

Written text production is also influenced by task and child factors. Factors related to the writing task include the length of time children write for, the genre, modality of the text and the preparation children engage in before they write. While child factors including age, gender and learning disabilities, have all been shown to be associated with written text production. Table 1 outlines different tasks used to capture children's written text production. As the table shows tasks include Curriculum Based Measures of Writing (CBM-W), standardised assessments and bespoke tasks designed for specific studies. CBM-W tasks in response to a writing prompt are the shortest tasks, lasting between three and five minutes and can be scored for numbers of words written or grammatical accuracy (Dockrell, Connelly, Walter & Critten, 2015; Kim et al., 2014). These measures are more sensitive with younger writers and with struggling writers. Standardised assessments can involve longer writing times such as the writing fluency subtest of the

Table 1
Different methods of measuring writing productivity.

Writing productivity measures	Descriptor
Curriculum-Based Measures of Writing (CBM-W)	CBM-W tasks in response to a writing prompt are the shortest tasks, lasting between three and five minutes. These written samples can be scored in a number of ways including numbers of words written, spelling accuracy, correct word sequences, incorrect word sequences and grammatical accuracy.
Standardised assessments	Assessments include the 7-minute writing fluency subtest of the Woodcock-Johnson Tests of Achievement-3 (Woodcock et al., 2001), the essay composition subtest of the Wechsler Individual Achievement Test-III (Wechsler, 2009), or the 15-minute story composition subtest of the Test of Written Language-4 (Hamill & Page, 2009).
Unstandardised/ bespoke writing tasks	Studies can develop their own writing tasks which vary in topic, length of time to write, planning time etc.

Woodcock-Johnson Tests of Achievement-3 which allows children to write for seven minutes (Woodcock et al., 2001), the essay composition subtest of the Wechsler Individual Achievement Test-III (Wechsler, 2009), or the story composition subtest of the Test of Written Language-4 (Hammill & Page, 2009) where children write for 15 min (Kim et al., 2014). More complex prompts with longer writing durations are reported to be more valid for older learners (Espin et al., 2008; Espin et al., 2005; McMaster & Campbell, 2008).

Writing tasks also differ in genre including narrative writing genres such as personal narratives and story narratives, or expository writing genres such as informational, compare-contrast and persuasive writing (Beers et al., 2018). Most research studies have focused on narrative writing, with limited data in other genres (Beers et al., 2018; Puranik et al., 2008). For example, the AIMSweb (NCS Pearson, 2012) produces normative data for written expression providing guidelines for TWW, CWS, and WSC, but uses only narrative writing prompts. The AIMSweb draws on two studies (Gansle et al., 2002; Jewell & Malecki, 2005) both included in the current systematic review. The focus on narrative writing alone is problematic, as there is indicative evidence that writing metrics are not stable across all genres (Dockrell et al., 2018; McMaster & Campbell, 2008). More text is produced in narrative writing than in expository writing (Beers & Nagy, 2011; Hall-Mills & Apel, 2015; Puranik et al., 2008), while writing complexity is higher in expository samples (Beers & Nagy, 2011; Hall-Mills, 2010; Hall-Mills & Apel, 2015; Koutsoftas & Gray, 2012; Scott & Windsor, 2000). However, McMaster and Campbell (2008) state that narrative writing appeared to be the most promising in terms of its technical adequacy across grades. Expository writing is known to take longer to master, requires more complex vocabulary and places a greater cognitive load on the writer (Dockrell & Connelly, 2021) and, as such, productivity measures to capture writing competence may differ. Finally, children can produce text using different modalities such as handwriting or typing (Alves et al., 2008) and with different opportunities to plan the content of the written product (Koutsoftas & Gray, 2013; Llauro & Dockrell, 2019). To date how these different task factors are related to metrics of written text production has not been systematically examined.

Child-based factors, such as gender and learning disabilities, also are associated with writing performance. Girls consistently outperform boys on writing assessments across grades (Kim, Al Otaiba et al., 2015; Truckenmiller et al., 2014; Fearing et al., 2014). This gender performance gap widens during the middle school years (Malecki & Jewell, 2003) and has been shown to be greater when the writing time is longer (Berninger et al., 2008). Children with dyslexia (Sumner et al., 2013), dysgraphia (Smits-Engelsman & VanGalen, 1997) and language difficulties (Dockrell et al., 2007) all show reduced amounts of text produced, although the factors which underpin these differences likely differ (Dockrell, Connelly, & Arfe, 2019; Dockrell et al., 2014). These difficulties persist across the writing modalities of handwriting, typing and use of a stylus (Beers et al., 2018; Corkett & Benevides, 2016; Berninger et al., 2009). CWS are reported to differentiate learners with writing disabilities from typically-developing learners (Kim et al., 2014; Weiss et al., 2019) whereas TNW show mixed results (Weiss et al., 2019). However, again, lack of normative data for typical and struggling writers makes identification of delays and differences challenging.

The production of spontaneously produced written text is key to children's academic attainment. To inform our understanding of children's production of written text we synthesised available data reporting production of TNW, CWS, number of different words produced, number of sentences produced, length of T-units, length of sentences and number of ideas produced across the elementary school grades. There were two main research questions: 1. To what extent do current data allow the establishment of reliable and valid benchmarks for text production (amount and accuracy)? 2. To what extent are these writing metrics associated with task- and child-based factors?

We anticipated that more complex measures of writing productivity (CWS and T-Units) would differentiate performance at the upper ends of elementary school while for younger children TNW would be a reliable indicator. We also reasoned that children with developmental difficulties would produce less text at all ages. By contrast the opportunity to plan and priming of topic knowledge should improve text production for all children. To maintain the focus of the review on empirical studies examining metrics of written text production inclusionary and exclusionary criteria were developed.

1. Method

This systematic review was conducted⁴ and reported following guidance from the 2020 PRISMA statement (Page et al., 2021) and the EPPI-centre (Gough et al., 2017).

1.1. Selection criteria

To address the research questions, empirical studies were selected that met the following criteria:

- (1) Children with typical development and/or children with developmental difficulties.
- (2) Elementary school children in Grade 1 to 7, between six and 13 years of age.
- (3) Assessments that involved written expression.
- (4) Studies that included metrics of written text production such as the total number of words, clauses, sentences, or ideas produced within the writing products, irrespective of the task format e.g., CBM-W tasks, standardised assessments and/or unstandardised/bespoke writing tasks.
- (5) Studies in which participants were assessed in their language of instruction.

⁴ Data was extracted by the first author.

- (6) Studies conducted and reported in English. This aids comparability across written products and reduces potential differences across orthographies and morphological complexities (see [Dockrell & Connelly, 2021](#)).
- (7) Studies published between 2000 and 2021. Published literature and grey literature were included.
- (8) Assessment using any modality to produce written text.

1.1.1. Exclusion Criteria

Studies were excluded from analysis if:

- (1) Samples were from Kindergarten (USA), Reception and Year 1 (UK) to exclude floor effects.
- (2) Studies that used a single case experimental design were excluded due to sample size as their baseline data would not be appropriate for generalisation.
- (3) Studies that had insufficient data to code task dimensions (e.g., genre and/or writing task duration) as the data could not be categorised and analysed.
- (4) Studies involving only copying or dictation tasks as they are not generative writing tasks. If a study included data on both copying/dictation and generative writing tasks, the independent writing data were included in this systematic review.
- (5) Samples including children with acquired disabilities such as traumatic brain injury, were excluded in order to maintain the focus on writing development.

1.2. Location and selection of studies

1.2.1. Databases

Educational, psychological and general databases were used as well as library services. Databases included the British Education Index, Australian Education Index, Open Grey, ProQuest Central, ERIC, PubMed, PsychArticles, Scopus and Web of Science.

1.2.2. Boolean search terms

Search terms were identified following a review of the relevant literature, discussion with writing researchers and support from the library review team. A Boolean search was then used in each database to identify relevant records (see Appendix A). The Boolean search included terms associated with writing productivity, elementary school years and assessment. Terms relating to preschool education, secondary school years and adults were excluded. The database search was conducted on 16 February 2021.

1.2.3. Selection of studies

Inclusion and exclusion criteria were applied in two stages: during the title and abstract screening and then during the full text screening by the first author and discussed with the second author. Data were unavailable in the public domain for two theses from two university online libraries and were requested by the first author. The reference lists of each of the included studies were also searched. The selection process is summarised in [Fig. 1](#). Sixty-seven studies met the inclusion criteria and these are marked * in the reference list.

1.3. Data extraction and coding

1.3.1. Data extraction

Following guidance from the PRISMA statement ([Page et al., 2021](#)), information from the 67 included studies was recorded on a spreadsheet including information related to the output (authors, date of publication, country of study), study aims, characteristics of the participants recruited, methodological features of the writing assessment (writing task type, writing duration, writing modality, priming of topic knowledge, opportunity to plan prior to writing) and a summary of results. In experimental studies, only pre-test scores were extracted, with the exclusion of post-test scores and intervention effects (see Appendix B).

1.4. Quality appraisal

In line with guidance from the PRISMA statement ([Page et al., 2021](#)) and the EPPI-centre ([Gough et al., 2017](#)), a quality appraisal of the included studies was conducted. Most quality appraisal checklists target a specific type of research design ([Critical Appraisal Skills Programme, 2018](#); [Joanna Briggs Institute, 2020](#)), yet the included studies in this paper held multiple study designs. As such a modified bespoke critical appraisal checklist was used across all study designs to assess the rigour of the included studies. The checklist focused on evaluating the studies' sampling methods, assessment methods, coding and scoring of data. This checklist was developed through combining the Critical Appraisal Skills Programme Checklists ([Critical Appraisal Skills Programme, 2018](#)), JBI Critical Appraisal Tool Checklists ([Joanna Briggs Institute, 2020](#)) and the Cochrane Risk of Bias Assessment tool ([Higgins & Thomas, 2021](#)). The checklist can be found in Appendix C with the results from the studies' appraisals in Appendix D.

1.5. Data analysis

Extracted data were grouped by variables of writing genre (Appendix E), gender (Appendix F), growth by term (Appendix G), growth by week (Appendix H) and use of planning (Appendix I) to investigate the research questions.

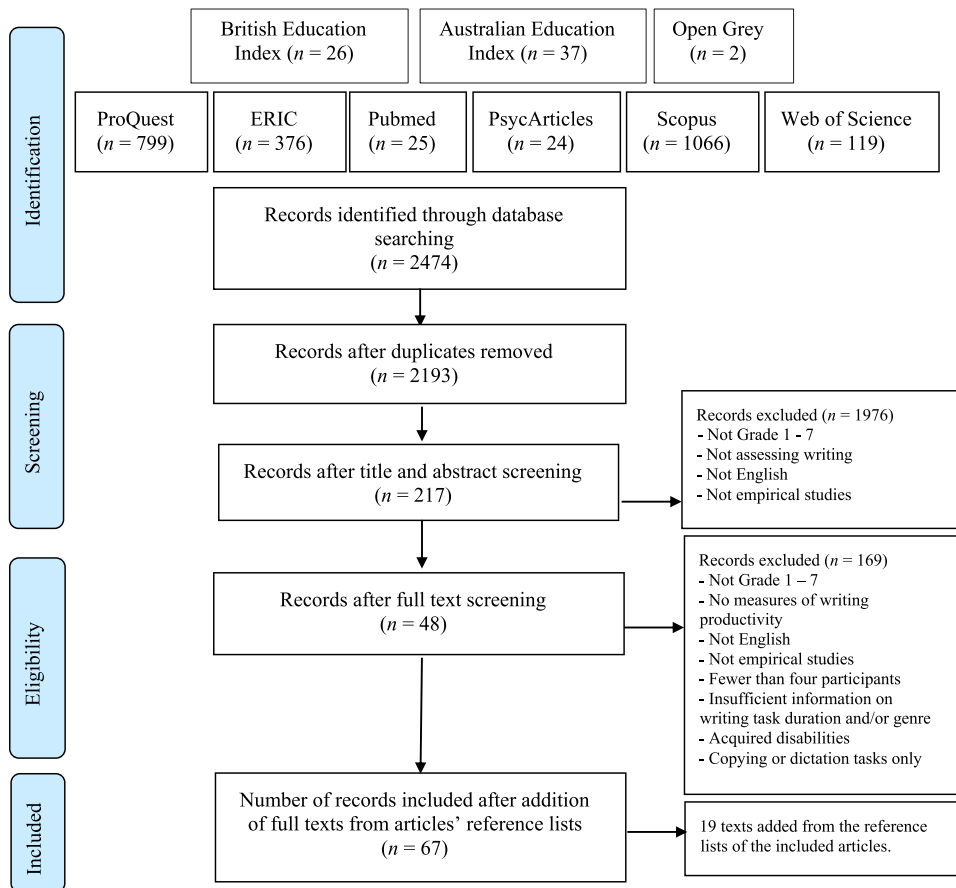


Fig. 1. Flow chart of the selection process.

2. Results

The results of the systematic review are presented in two sections. First, a summary of the 67 included studies' characteristics is reported. Secondly, data were gathered in themes to answer each research question where sufficient studies using the target metrics were available.

2.1. Study characteristics

2.1.1. Study location and sample size

The majority of studies were conducted in the United States ($n = 58$, 86.6%) with the remaining studies being in the United Kingdom ($n = 5$, 7.4%), Canada ($n = 3$, 4.4%) and Slovenia⁵ ($n = 1$, 1.5%). The sample sizes ranged from five to 1240 children ($M = 201$, $SD = 227.4$).

2.1.2. Frequency of writing assessment

The majority of studies measured written text production at a single time point for each grade ($n = 58$, 86.6%), with the remaining ($n = 9$, 13.4%) measuring growth over time within an academic year: every term in fall, winter, and spring ($n = 3$, 4.4%), twice a year ($n = 4$, 6%) and every week for 12 weeks ($n = 2$, 3%).

2.1.3. Metrics of written text production

Typically, studies used more than one indicator of writing productivity ($n = 46$, 68.7%). As Table 2 shows the most common indicators used were TNW ($n = 60$, 89.6%) and CWS ($n = 29$, 43.3%) with larger textual units (number/length of sentences, length of T-units) being markedly less frequent.

⁵ The Slovenian study met the inclusion criteria as the learners were bilingual, being taught in both English and Slovenian. Data from only the English writing assessment was extracted.

Table 2
Frequency of writing productivity indicators from selected studies $N = 67$.

Indicator	Number of studies
Total number of words (TNW)	60
Correct Word Sequences (CWS)	29
Number of different words	12
Number of sentences	10
Length of T-units	8
Length of sentences	8
Number of ideas	5
Number of T-units	5
Words per minute	4
Correct punctuation marks	4
Number of punctuation marks	1
Words per clause	1
T-units per minute	1
Total number of letters	1
Words spelt correctly	1

2.1.4. Frequency of grades being assessed

Table 3 presents the frequency of grades assessed within the studies. As the table shows most of the studies focused on Grades 2 to 4, although there were data available across the elementary school years. Two studies (3%) combined data for Grade 7 and 8, yet were retained due to the Grade 7 data meeting the inclusion criteria. In four studies (6%), the mean age of cohorts was given without providing grades. For the purposes of this study, the data from these cohorts were matched to the closest grade or, if between two grade brackets, represented as a combination of both grades.

2.1.5. Genre across grades

Table 4 provides details of all the writing genres presented in the studies. The majority of studies ($n = 53$, 79.1%) assessed only one genre, with story narrative being the most frequently assessed ($n = 42$, 62.7%), followed by personal narrative ($n = 10$, 14.9%).

2.1.6. Samples assessed

Table 5 provides details of the approach to writing assessment that were used in the identified studies. Fifty-five studies (82.1%) investigated only one group and typically data were collected as a whole class activity ($n = 41$, 61.2%), which included both learners with typical development and learners with learning difficulties present in the mainstream class.

2.1.7. Duration of writing tasks

Table 6 provides details of the duration of the writing tasks used. Typically, these involved a short writing task of three-minutes ($n = 22$, 32.8%) or five-minutes ($n = 14$, 20.9%). Comparisons of the effect of writing duration were rare, with only four studies making a comparison (6%).

2.2. Priming of topic knowledge

Most studies ($n = 61$, 91%) did not use any form of priming of topic knowledge. The six studies that introduced the topic prior to writing used reading comprehension ($n = 2$), listening comprehension ($n = 1$), class discussion to brainstorm ideas ($n = 1$), use of a model text ($n = 1$), watching a movie ($n = 1$) and reading comprehension combined with retell ($n = 1$). Five of these six studies used expository writing genres.

2.2.1. Planning time

The provision of planning time for writing was infrequent, with only 30% ($n = 20$) providing time to plan. When planning time was provided this ranged from 30 s ($n=6$) to 45 min ($n=6$), with the modal time provided being one minute ($n=13$).

2.2.2. Modality of writing

Ninety-five percent ($n = 64$) assessed writing using pen/pencil and paper while five studies (7.5%) assessed writing using typing and one study a stylus. Three of these studies used typing and/or a stylus to compare learner performance across modalities.

2.3. Factors associated with writing metrics

The descriptive data generated from the studies indicated variability in populations sampled and in study measures. TNW and CWS were used to examine relationships associated with writing metrics as these two measures provided sufficient data for comparison; while other metrics were used infrequently. For example, t-units was only used in five studies and words spelt correctly in one (see Table 2). We explored the extent to which it was possible to answer the research questions related to genre, writing duration, priming and topic knowledge, gender and learning disabilities for TNW and CWS.

Table 3
Frequency of grades sampled from selected studies
 $N = 67^a$.

Grade	<i>N</i>
Grade 1	15
Grade 1 & 2 combined	3
Grade 2	23
Grade 2 & 3 combined	2
Grade 3	26
Grade 3 & 4 combined	1
Grade 4	25
Grade 4 & 5 combined	1
Grade 5	18
Grade 5 & 6 combined	2
Grade 6	12
Grade 7	5
Grade 7 & 8	2

^a For ease of comparison UK year levels were translated into USA grades

Table 4
Frequency of genres ($N = 67$).

Genre	Number of studies
Story narrative	42
Personal narrative	10
Informational writing	8
Persuasive writing	7
Sentence writing	6
Unspecified expository writing	4
Descriptive writing	4
Summary	2
Compare-Contrast writing	2
Argumentative writing	1
Opinion writing	1
Letter writing	1

Table 5
Frequency of study design ($N = 67$).

Population type	Sample	Number of studies
Whole class		41
Targeted participants	Learners with typical development	19
	Learners with language difficulties/ Specific Language Impairment	6
	Learners with writing difficulties	6
	Learners with learning difficulties	4
	Learners with reading difficulties and/or dyslexia	2
	Learners with English as an additional language	2
	Learners with dysgraphia	1
	Learners with spelling difficulties	1
	High achievers	1
	Bilinguals	1
	Recipients of free/ reduced lunch (lower socioeconomic status)	1
	Not recipients of free/ reduced lunch (higher socioeconomic status)	1

2.3.1. Genre

To allow comparisons across studies that allowed different writing times TNW and CWS per minute were used as the units of comparison across studies, only writing tasks spanning three to 15 min for each genre were included and studies with data extending across two or more grades were excluded. In total, seven studies were excluded (10.4%). Argumentative, persuasive and opinion genres were combined given the limited data for each genre individually and the rationale that they share the core writing characteristic of taking a point of view and supporting it with reasoning (Crowhurst, 1990). The TNW per minute and CWS per minute for each grade and genre can be seen in Figs. 2 and 3 respectively. When more than one study was considered, *SD* was calculated and presented in the figures.

As Figs. 2 and 3 illustrate there were both developmental trends and genre effects in TNW and CWS. Only data for the story narrative genre had TNW and/or CWS data for all grades, making establishing benchmarks difficult. There were fewer data points

Table 6
Frequency of writing task durations ($N = 67$).

Writing task duration in minutes	Number of studies
3	22
5	14
6	1
7	2
10	7
15	11
20	3
25	1
30	5
35	1
45	4
60	2

available for CWS than TNW and, as a result, insufficient CWS data were available for the analysis of genres such as personal narrative, informational writing, compare-contrast and summaries. For both measures there was significant variability at each age point, with the greatest variability evident in Grade 5.

From the data, sentence writing tasks yielded the highest TNW, although it was only used as an assessment for younger writers. For sentence writing there was an average words per minute of 5.75 ($SD = 1.11$) in Grade 1, 10.26 ($SD = 2.26$) in Grade 2 and 11.25 ($SD = 1.71$) in Grade 3. The story narrative genre demonstrated consistently higher writing TNW rates than other genres. However, in some studies the TNW elicited from story narrative writing prompts was lower than that elicited by descriptive writing prompts (Beers et al., 2018; Coker et al., 2018) and expository writing prompts (Hall-Mills & Apel, 2015; Edman, 2012). In addition, as Fig. 2 shows, in Grade 7, both informational ($M = 16.5$) and compare-contrast ($M = 17.01$) writing prompts elicited higher TNW than the story narrative writing prompts ($M = 15.09$). However, given the data it was not possible to establish whether this difference was statistically significant.

2.3.2. Writing task duration

There were more data for the narrative writing genre across ages and samples, and these data were used to examine the influence of different writing durations. Texts from studies that involved typing ($n = 1$, 1.5%), and/or studies in which results spanned two or more grades ($n = 2$, 3%) were excluded to reduce confounding variables. Figs. 4 and 5 represent the total TNW and CWS produced by learners across the grades by writing times. The data suggest that longer writing durations resulted in longer total length of texts. However, the TNW produced in three-minute and five- to seven-minute tasks was comparable. The provision of longer times (above 15 min) did not appear to increase TNW between Grades 2 – 4. It should be noted that there was no CWS data for writing times longer than 15 min, as evidenced in Figs. 5 and 7.

Speed of text production (per minute) provides an alternative measure for capturing the production of written text across different writing durations. Figs. 6 and 7 represent the TNW and CWS produced per minute by different writing times. The shorter writing time of 3 min increased both TNW and CWS across the grades. With shorter writing times, learners in Grade 7 were producing up to 15.9 words per minute ($SD = 1.67$) and 14.9 CWS per minute ($SD = 1.52$). This difference in the speed of output elicited by short and longer writing tasks increased with age. At Grade 1 there was a difference of 5.07 words per minute when comparing writing output on three-minute writing tasks ($M = 5.9$, $SD = 2.77$) and 20 – 60-minute writing tasks ($M = 0.83$, $SD = 0.84$), suggesting that these extended tasks

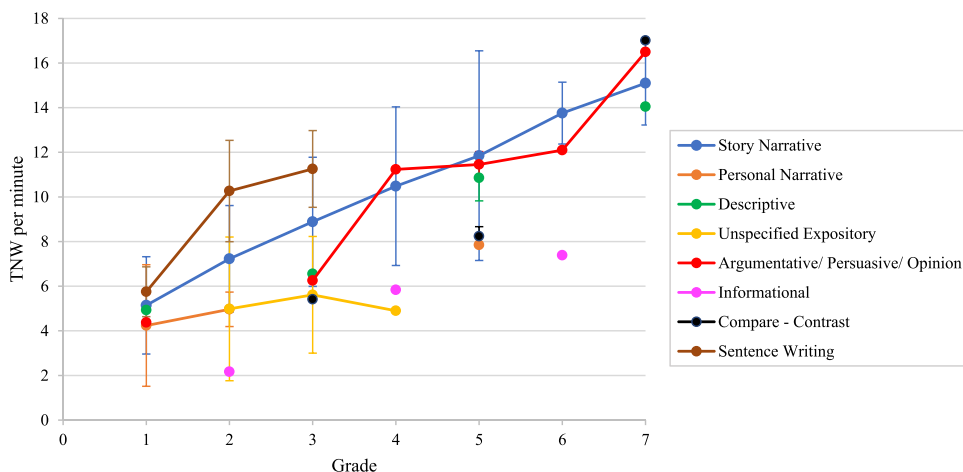


Fig. 2. TNW produced per minute (SD) by learners in Grade 1 - 7 across the different genres.

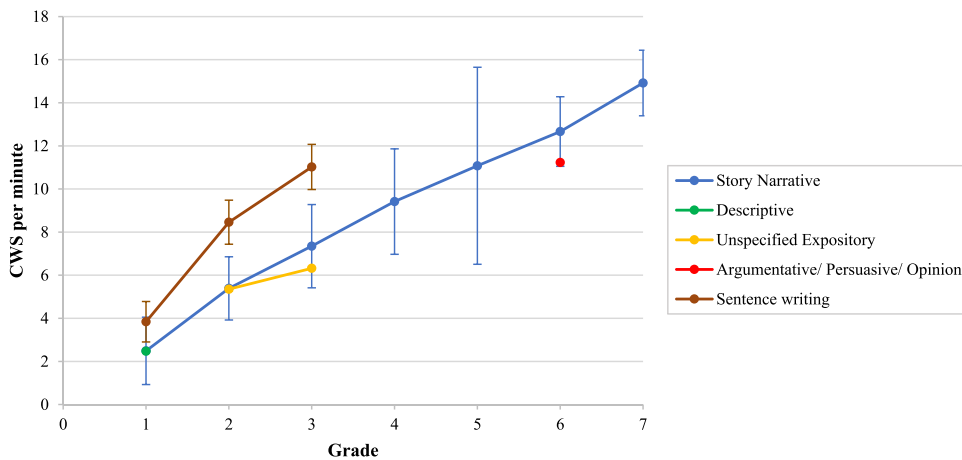


Fig. 3. CWS produced per minute (SD) by learners in Grade 1–7 across the different genres.

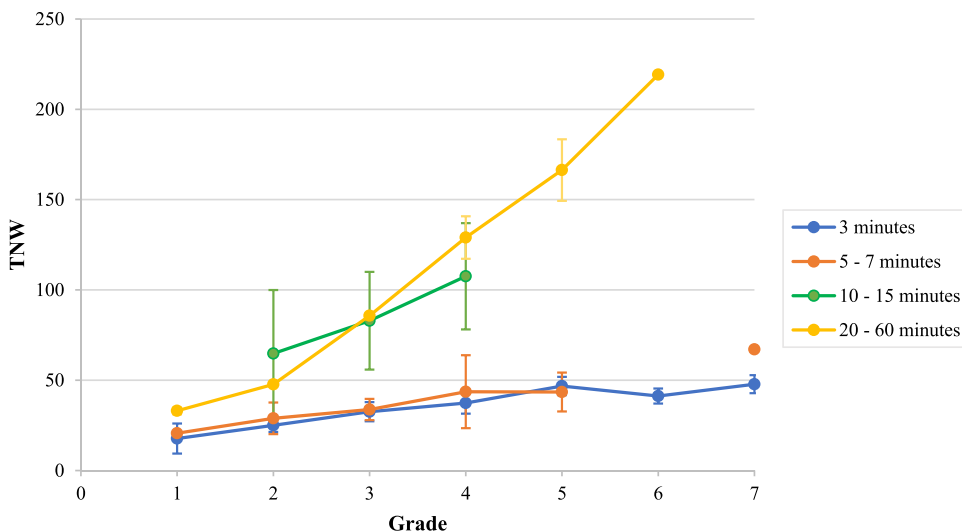


Fig. 4. Average TNW (SD) produced across the grades according to the length of writing task.

resulted in lower text production rates for the younger children. By Grade 5 the difference in average words per minute was 11.9 when comparing writing output on three-minute tasks ($M = 15.9, SD = 1.69$) and 20 – 60-minute writing tasks ($M = 3.7, SD = 2.38$), likely reflecting the longer texts produced at this age and their ability to engage with the tasks.

2.3.3. Priming of topic knowledge

Seven (10.4%) studies involved some level of priming of content prior to the writing task through class discussion, comprehensions, use of a model text and/or watching a movie. However, these studies did not provide detailed descriptions of the priming method such as the teaching style and teaching methods, duration of priming, or level of engagement of learners. Additionally, the studies did not use control groups to compare the influence of the priming activity on TNW/CWS. This significantly affects the ability to reproduce, compare, or draw conclusions from these studies.

2.3.4. Planning

Twenty-five studies (37.3%) included an element of planning prior to the writing phase of the assessment. However, none of the studies compared the written output of learners who utilised planning with those who did not, nor compared the effects of different types of planning (graphic organisers, drafts, drawings) at different ages. Additionally, none of the studies provided a rationale for the length of planning time chosen for the assessment. The data were not sufficient to evaluate the relationship between planning types and planning lengths for text productivity throughout the different grades.

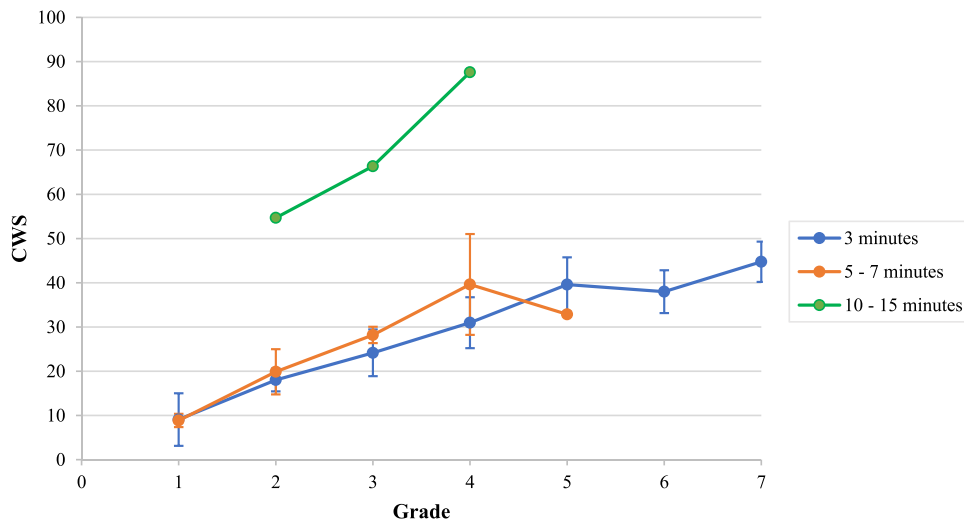


Fig. 5. Average CWS (SD) produced across the grades according to the length of writing task.

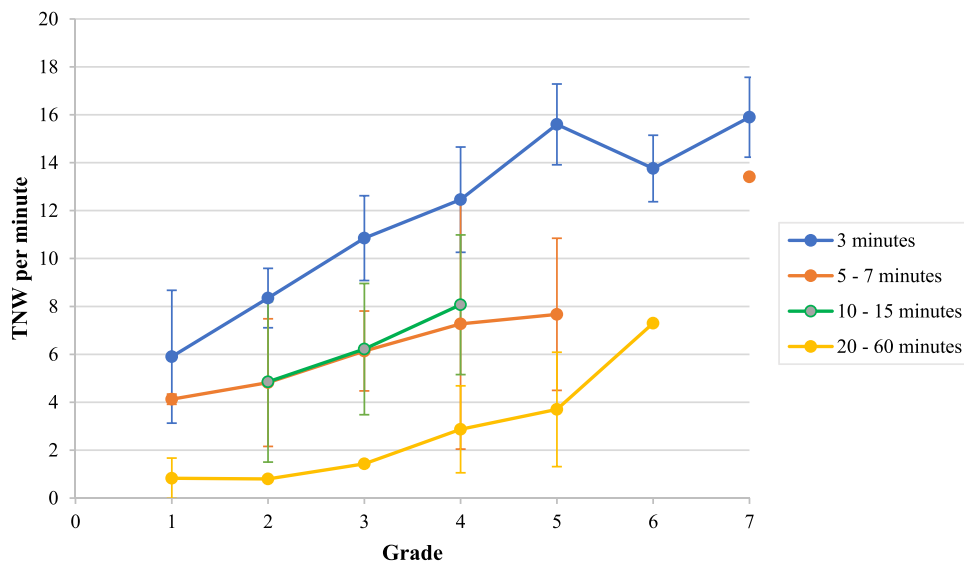


Fig. 6. TNW (SD) produced per minute across the grades according to the length of writing tasks.

2.3.5. Modality

The three studies (4.5%) that compared TNW produced across modalities presented mixed results. [Corkett and Benevides \(2016\)](#) assessed informational writing in Grade 6 learners with learning difficulty and found writing on an iPad facilitated production of more words ($M = 94.67$, $SD = 27.67$) than with pen and paper ($M = 74.56$, $SD = 12.68$). In contrast, [Berninger et al. \(2009\)](#) assessed the informational writing of learners using pen and typing in Grade 2, 4, and 6 and found all grades to produce more words when writing by pen than typing. [Berninger et al. \(2009\)](#) found this to be consistent with typically-developing learners and learners with writing difficulties. [Beers et al. \(2018\)](#) assessed persuasive writing using typing and a stylus in typical, dyslexic and dysgraphic cohorts across Grade 5 - 6 and found all cohorts to produce more words using typing than a stylus. Both [Berninger et al. \(2009\)](#) and [Beers et al. \(2018\)](#) reported that the difference between modalities was greater in the typically-developing cohorts when compared to those with writing difficulties. There were insufficient data to draw conclusions about the effect of modality on productivity.

2.3.6. Gender

Eight studies (11.9%) provided data to evaluate the relationships between gender on TNW and CWS and each of these studies assessed whole class samples. Of these, only five had gender data for individual grades, while the remaining three studies present gender data across multiple grades. Only one study examined both expository and narrative writing, while the others assessed narrative writing only. [Figs. 8 and 9](#) present productivity data by gender and grade. The data are displayed as TNW and CWS per

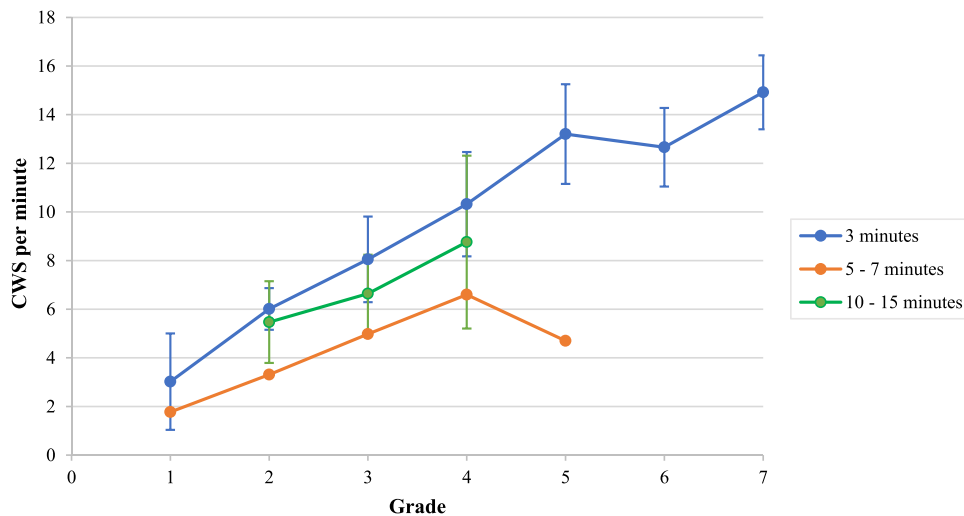


Fig. 7. CWS (SD) produced per minute across the grades according to the length of writing task.

minute of writing, to enable comparison between studies with different writing times. Most of the studies ($n = 5$) that analysed gender differences in writing focused on assessment in younger Grades 2 and 3. Only [Ferrarington et al. \(2014\)](#) provided data for Grades 4 – 7 and as a result, the data in [Figs. 8 and 9](#) should be interpreted with caution. Nonetheless the data indicated a trend, suggesting that girls produce more text than boys throughout Grades 2 – 7, with a suggestion that the performance gap might widen with age.

2.3.7. Learning disabilities

A total of 10 studies (14.9%) directly compared typically developing learners and those with learning disabilities (LD). Most studies ($n = 8$) focused on assessing a single LD while the remaining studies ($n = 2$) assessed the writing productivity of two groups of learners with different diagnoses: general learning difficulty ($n = 2$), language difficulty ($n = 3$), writing difficulty ($n = 5$) and reading difficulty ($n = 1$). Genres varied: story narrative ($n = 6$), story narrative and summary ($n = 1$), informational writing ($n = 2$) and persuasive writing ($n = 1$). The TNW data were extracted as the measure due to a lack of CWS data across the studies. [Fig. 10](#) shows that learners with LDs all experienced significantly reduced word production when compared to typically developing peers (in some cases TNW was equivalent to learners 30 – 48 months younger) ([Dockrell & Connelly, 2015](#); [Mackie & Dockrell, 2004](#); [Scott & Windsor, 2000](#)).

There was indicative evidence that the gap between typical learners in Grade 1 and those with LD widened with age ([Kim, Puranik et al., 2015](#)). The typically-developing cohort's baseline TNW in the fall term was 20.99 words ($SD = 10.48$). This increased by 12.33 words in the spring term ($M = 33.32$, $SD = 13.76$). Comparatively, the language difficulty cohort started with a lower baseline of 13.61 words ($SD = 10.22$) in the fall term and made a smaller increase of 11.87 words by the spring term ($M = 25.48$, $SD = 14.6$).

3. Discussion

This systematic review aimed to evaluate whether benchmarks for metrics of written text production in elementary (primary) school could be established and to capture the extent to which these measures were associated with task- and child-based factors. Previous research has established that metrics of text production are strongly correlated with writing quality ([Puranik et al., 2008](#)), an effective tool to identify struggling writers ([Puranik et al., 2008](#); [Salas et al., 2021](#)), and are typically more reliable than assessments of writing quality making them ideal for establishing benchmarks and systematically examining changes in writing development ([Puranik et al., 2008](#)). To achieve this aim, we examined which measures varied across development and which task and child factors were associated with children's performance. We anticipated that more complex measures of writing productivity (CWS and T-Units) would differentiate performance at the upper ends of elementary school while for younger children TNW would be a reliable indicator. We also expected that children with learning disabilities would produce less text at all ages. By contrast the opportunity to plan and priming of topic knowledge should improve productivity. Overall, we identified significant gaps in the evidence making the establishment of benchmarks for different measures of text production across writing tasks challenging. By contrast data comparing performance of children with learning difficulties and their peers produced robust differences.

Sixty-seven studies met the inclusionary criteria and across the studies a range of different metrics were used (see [Table 2](#)), although many of these were only used in a small number of studies. The most commonly-used indicators of writing productivity were TNW and CWS. To present sufficient comparative data these measures were used to examine productivity for the task- and child-based factor comparisons. Overall, the included studies indicated that TNW and CWS were an effective, age-sensitive method to assess writing development with the ability to differentiate typically-developing learners from those with difficulties. Data were reported across the elementary school range, but the focus was mainly on grades two to four and studies generally used narrative tasks. More

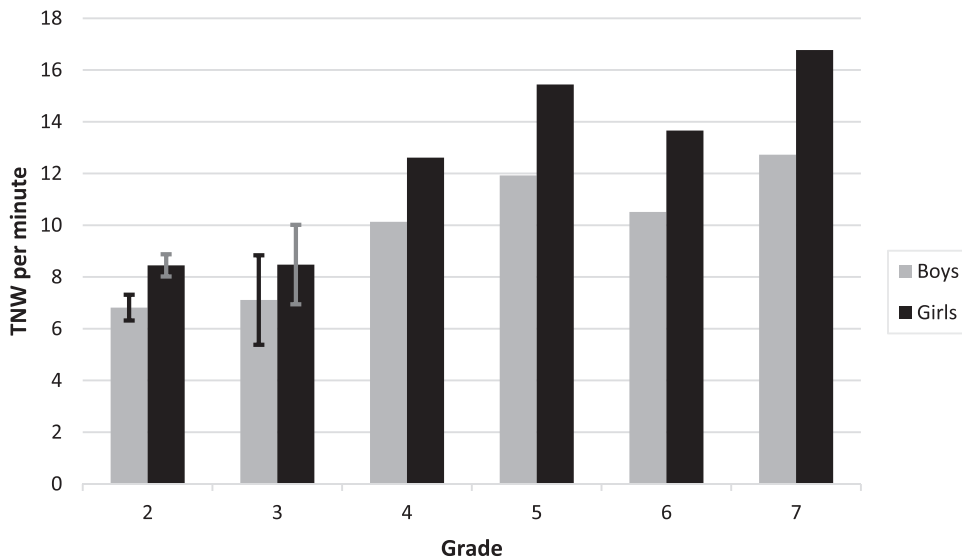


Fig. 8. Average TNW (SD when available) per minute produced by boys and girls across the grades.

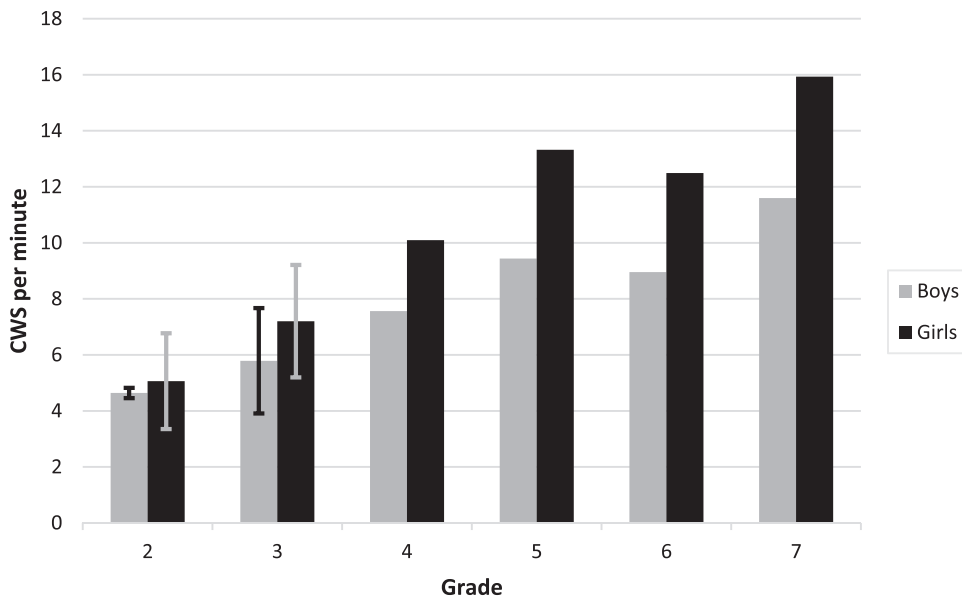


Fig. 9. Average CWS (SD when available) per minute produced by boys and girls across the grades.

than half of the studies used tasks of 10 min duration or less. Given the lack of data across grades for different measures and the inconsistency in task types, establishing benchmarks was not possible but task and child factors on writing productivity were evident.

Analysis of task-based factors associated with TNW and CWS was limited due to the lack of comparative data across tasks and implementation factors. For some grades and genre, data points represented single studies and, as such, may be influenced by differences across samples, administration and writing durations. Nonetheless the review suggests that genre was significantly associated with these writing metrics. Overall studies reported results for the narrative genre and while there was marked variability at each grade it was greatest in Grade 5. This variability is of interest and, potentially, indicates a break point for identifying struggling writers (Dockrell et al., 2019); as it is in intermediate grades when writing tasks become more complex and students can struggle (Allen et al., 2019). There was also indicative evidence of differences across genre with age. Younger children produced the highest numbers of words per minute in sentence writing tasks whereas for older children there was indicative evidence that narrative genres resulted in greater productivity. These data require corroboration as typically children in the lowest grades were given sentence writing tasks and there were few studies with children in the higher grades, none using sentence writing tasks. The variation across genres revealed the importance of assessing and using normative data for each genre, rather than the generalising across genres. More data comparing the

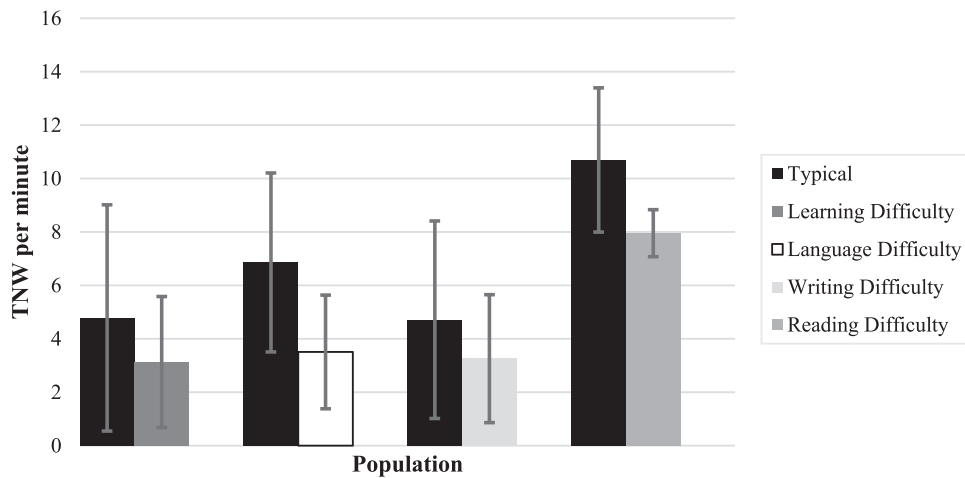


Fig. 10. TNW per minute (SD when available) produced by typically developing learners in comparison to learners with special educational needs and disabilities (SEND) profiles.

performance across genres are needed, using the same administration procedures in order for assessment, monitoring and identification of learners with writing difficulty to be effective.

The data on narrative tasks afforded the analysis of the relationship between writing task duration and TNW and CWS. Overall, longer writing times resulted in a longer total length of text but text produced in the shorter time frames (below five minutes) was similar when comparable data were available across grades (grades 2, 3, and 4). There were few differences when writing time was greater than 10 min. By corollary shorter writing times increased fluency, measured in both TNW and CWS per minute, with higher grades showing the most marked effect. More data are needed but the current results suggest that, for these measures, it is only above Grade 4 that providing time above 15 min made a difference to production rates. The reduced production rates for TNW and CWS in the longer writing tasks raises important questions about the interaction between online planning and writing productivity (see for example [Sénéchal et al., 2018](#)). How time is associated with writing quality remains to be examined systematically by age and task. This is particularly important given the limited data on planning, priming and text modality, and has implications for summative assessments.

Data examining TNW and CWS and child-based factors were also sparse. Five studies produced sufficient data to examine the differences by gender. The data suggested that both that girls produced more text and the gap between girls and boys widened with age. However, this remains to be corroborated with more systematic longitudinal studies which explicitly examine these differences. Children with learning disabilities also produced significantly less text than their peers without learning disabilities, but gender was not always controlled for or analysed. Given the higher numbers of boys experiencing learning disabilities ([American Psychiatric Association, 2013](#)), conclusions need to be interpreted carefully. Research on the metrics of text production of learners with LD remained sparse. However, in all included studies, those with language, learning and literacy difficulties produced less text than typically developing peers. TNW was used consistently in the included studies and was able to differentiate between typically developing writers and those with disabilities (see [Weiss et al., 2019](#) for an argument for the use of CWS). One study suggested that the gap between those with LD and those without widened with age ([Kim, Puranik et al., 2015](#)) highlighting the need for early identification and support. Studies with children who had identified learning disabilities suffered from several limitations. Firstly, the lack of diverse genres assessed; seventy percent of the included studies with children who had an LD used the story narrative genre. Research shows that narrative writing is one of the less structured, less linguistically demanding types of writing ([Beers et al., 2018](#)) and may not provide a reliable indication of writing ability more generally ([Dockrell et al., 2018](#)). Indeed, recent recommendations emphasise the need to assess children with writing difficulties using four or five writing prompts across genres to establish reliable indicators of writing competence ([Wilson et al., 2019](#)). There was no evidence in the current review of studies that used this approach to examine productivity. A significant omission given that a battery of assessments may be a more reliable indicator of at risk writers in Grade 1 than a single measure ([Ritchey & Coker, 2013](#)). Secondly, there was often a lack of typically developing learners as control groups which make comparison across studies more challenging as writing tasks differed in time, genre, and administration. Thirdly, many studies showed methodological weaknesses. During the selection process of this systematic review, many studies that involved students with LD had to be excluded due to small sample sizes and/or insufficient descriptions of the assessment procedures.

4. Limitations

To our knowledge this is the first systematic review to examine metrics of written text production in elementary school and the factors that are associated with children's performance. However, as with any systematic review inclusion and exclusion criteria influence the results. A significant limitation was the lack of data to test our key research questions and interrogate the full range of possible measures (e.g., CIWS see [Romig et al., 2017](#) who examined criterion validity of CBMs) and to produce a meta-analysis

addressing questions related to development of writing productivity and the factors impacting on development. The inclusion research papers published in English, with monolingual children, where most of the included studies were undertaken in the United States limits generalisability to the global community of learners and our understanding of writing productivity across different orthographies (see Dockrell & Connelly, 2021) and different pedagogical approaches. There is a growing literature that highlights the importance of the ways in which writing is taught and it was not possible to capture the influence of pedagogy on productivity. The exclusion of studies with three or fewer participants resulted in the exclusion of some single-subject design studies focusing specifically on LD and there is scope for further reviews which capture these studies. A final limitation was the absence of alternative modalities for written text production in the Boolean search terms, which may have led to fewer mixed-modality studies being identified.

5. Conclusion

Reliable and valid measures of the assessment of written text underpin evidence based teaching and the identification of struggling writers. The review identified sufficient studies to support the use of TNW and CWS as quick, effective, valid and reliable assessments of written expression for children at elementary school. However, there were insufficient studies to provide reliable developmental benchmarks across writing tasks and writing formats and data for other metrics were sparse. Future studies need to establish developmental trends across genres, tasks, writing durations and populations to establish normative benchmarks. To date there has been little attempt to examine the effects of planning or priming of topic knowledge despite its use in classrooms. To use writing productivity as a guide for pedagogy, in assessment protocols and in the identification of struggling writers across the elementary school studies need to capture child- and task-related factors in a reliable and consistent fashion and to consider systematically a wider range of metrics in the evaluation of children's writing products.

Statements and declarations

No funding was received to assist with the preparation of this manuscript. The authors have no conflicts of interest to declare that are relevant to the content of this article.

CRedit authorship contribution statement

Julie Dockrell: Writing – review & editing, Writing – original draft, Supervision, Conceptualization. **Catherine Martin:** Writing – review & editing, Writing – original draft, Conceptualization.

Data Availability

Data will be made available on request.

Acknowledgement

Many thanks to Prof Chloe Marshall for proofreading this article prior to submission.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.asw.2024.100834](https://doi.org/10.1016/j.asw.2024.100834).

References

- Abbott, R. D., & Berninger, V. W. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary-and intermediate-grade writers. *Journal of Educational Psychology, 85*(3), 478.
- (*) Allen, A. A., Jung, P. G., Poch, A. L., Brandes, D., Shin, J., Lembke, E. S., & McMaster, K. M. (2019). Technical adequacy of curriculum-based measures in writing in Grades 1–3. *Reading & Writing Quarterly, 36*(6), 563–587. <https://doi.org/10.1080/10573569.2019.1689211>.
- (*) Alves, R. A., Castro, S. L., & Olive, T. (2008). Execution and pauses in writing narratives: Processing time, cognitive effort and typing skill. *International Journal of Psychology, 43*(6), 969–979. <https://doi.org/10.1080/00207590701398951>.
- American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.).
- (*) Beers, S. F., Berninger, V., Mickail, T., & Abbott, R. (2018). Online writing processes in translating cognition into language and transcribing written language by stylus and keyboard in upper elementary and middle school students with persisting dysgraphia or dyslexia. *Learning Disabilities: A Multidisciplinary Journal, 23*(2), 70–86. <https://doi.org/10.18666/ldmj-2018-v23-i2-9008>.
- (*) 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: An Interdisciplinary Journal, 24*(2), 183–202.
- (*) Berninger, V. W., Abbott, R. D., Augsburger, A., & Garci, N. (2009). Comparison of pen and keyboard transcription modes in children with and without learning disabilities. *Learning Disability Quarterly, 32*(3), 123–141.
- Berninger, V. W., Nielsen, K. H., Abbott, R. D., Wijsman, E., & Raskind, W. (2008). Gender differences in severity of writing and reading disabilities. *Journal of School Psychology, 46*(2), 151–172. <https://doi.org/10.1016/j.jsp.2007.02.007>
- Berninger, V., & Swanson, H. (1994). Modifying Hayes and Flowers' model of skilled writing to explain developing writing. *Advances in cognition and educational practice. Children's writing: Toward a process theory of the development of skilled writing, 2*, 1–30.

- (* Coker, D. L., Jr., Ritchey, K. D., Uribe-Zarain, X., & Jennings, A. S. (2018). An analysis of first grade writing profiles and their relationship to compositional quality. *Journal of Learning Disabilities, 51*(4), 336–350. <https://doi.org/10.1177/0022219417708171>.
- (* Corkett, J. K., & Benevides, T. (2016). iPad versus handwriting: Pilot study exploring the writing abilities of students with learning disabilities. *Journal of International Special Needs Education, 19*(1), 15–24.
- Critical Appraisal Skills Programme. (2018). Critical Appraisal Skills Checklists. Retrieved from (<https://casp-uk.net/casp-tools-checklists/>).
- Crowhurst, M. (1990). Teaching and learning the writing of persuasive/argumentative discourse. *Canadian Journal of Education, 15*(4), 348–359.
- Dockrell, J. E., & Connelly, V. (2015). The role of oral language in underpinning the text generation difficulties in children with Specific Language Impairment. *Journal of Research in Reading, 38*(1), 18–34. <https://doi.org/10.1111/j.1467-9817.2012.01550.x>
- Dockrell, J. E., Connelly, V., & Arfe, B. (2019). Struggling writers in elementary school: Capturing drivers of performance. *Learning and Instruction, 60*(1), 75–84. <https://doi.org/10.1016/j.learninstruc.2018.11.009>
- Dockrell, J. E., & Connelly, V. (2021). Capturing the Challenges in Assessing Writing. In T. Limpo, & T. Olive (Eds.), *Executive Functions and Writing*. UK: Oxford University Press.
- Dockrell, J. E., Connelly, V., Walter, K., & Critten, S. (2018). The role of curriculum based measures in assessing writing products. In B. Miller, P. McCardle, & V. Connelly (Eds.), *Writing development in struggling learners: Understanding the needs of writers across the lifecourse* (pp. 182–197). Leiden, Boston: Brill.
- Dockrell, J. E., Lindsay, G., Connelly, V., & Mackie, C. (2007). Constraints in the production of written text in children with specific language impairments. *Exceptional Children, 73*(2), 147–164. <https://doi.org/10.1177/001440290707300202>
- 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. <https://doi.org/10.1016/j.learninstruc.2014.01.008>
- (* Edman, E. W. (2012). *Exploring writing accuracy and writing complexity as predictors of highstakes state assessments*. Indiana University of Pennsylvania.
- Espin, C. A., 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), Article 208217.
- Espin, C., Wallace, T., Campbell, H., Lembke, E. S., Long, J. D., & Ticha, R. (2008). Curriculum-based measurement in writing: predicting the success of high-school students on state standards tests. *Exceptional Children, 74*(2), 174–193.
- Fearrington, J. Y., Parker, P. D., Kidder-Ashley, P., Gagnon, S. G., McCane-Bowling, S., & Sorrell, C. A. (2014). Gender differences in written expression curriculum-based measurement in third through eighth grade students. *Psychology in the Schools, 51*(1), 85–96. <https://doi.org/10.1002/pits.21733>
- Gansle, K. A., Noell, G. H., Van Der Heyden, 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., Noell, G. H., Van Der Heyden, A. M., Slider, N. J., Hoffpauir, L. D., & Whitmarsh, E. L. (2004). An examination of the criterion validity and sensitivity to brief intervention of alternate curriculum-based measures of writing skill. *Psychology in the Schools, 41*(3), 291–300.
- (* Gansle, K. A., Van Der Heyden, 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.
- Gillespie, A., & Graham, S. (2014). A Meta-Analysis of Writing Interventions for Students With Learning Disabilities. *Exceptional Children, 80*(4), 454–473. <https://doi.org/10.1177/0014402914527238>
- Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews* (2nd ed.). Los Angeles, USA: SAGE.
- 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.
- Hall-Mills, S. (2010). *Linguistic feature development in elementary writing: Analysis of microstructure and macrostructure features in a narrative and an expository genre* [Unpublished thesis]. Florida State University.
- Hall-Mills, S., & Apel, K. (2015). Linguistic feature development across grades and genre in elementary writing. *Language, Speech, and Hearing Services in Schools, 46*(3), Article 242255. https://doi.org/10.1044/2015_LSHSS-14-0043
- Hammill, D.D., & Page, S.C. (2009). Test of Written Language (4 ed.). Austin, TX: Pro-Ed.
- Higgins, J., & Thomas, J. (2021). Cochrane Handbook for Systematic Reviews of Interventions Version 6.2. *The Cochrane Collaboration*. (www.training.cochrane.org/handbook/current).
- (* Jewell, J., & Malecki, C. K. (2005). The utility of CBM written language indices: An investigation of production-dependent, production-independent, and accurate production scores. *School Psychology Review, 34*(1), 27–44.
- Joanna Briggs Institute. (2020). Critical Appraisal Tools. Retrieved from (<https://jbi.global/critical-appraisal-tools>).
- (* Kim, Y.-S., Al Otaiba, S., Folsom, J. S., Greulich, L., & Puranik, C. (2014). Evaluating the dimensionality of first-grade written composition. *Journal of Speech, Language and Hearing Research, 57*(1), 199–211. [https://doi.org/10.1044/1092-4388\(2013\)12-0152](https://doi.org/10.1044/1092-4388(2013)12-0152).
- Kim, Y.-S., Al Otaiba, S., Puranik, C., Folsom, J. S., Greulich, L., & Wagner, R. K. (2011). Componential skills of beginning writing: An exploratory study. *Learning and Individual Differences, 21*(5), 517–525.
- Kim, Y.-S., Al Otaiba, S., Wanzek, J., & Gatlin, B. (2015). Toward an understanding of dimensions, predictors, and the gender gap in written composition. *Journal of Educational Psychology, 107*(1), 79–95. <https://doi.org/10.1037/a0037210>
- Kim, Y.-S., Al Otaiba, S., & Wanzek, J. (2015). Kindergarten predictors of third grade writing. *Learning and Individual Differences, 37*, 27–37.
- (* Kim, Y.-S., Puranik, C., & Al Otaiba, S. (2015). Developmental trajectories in writing skills in first grade: Examining the effects of SES and language and/or speech impairments. *The Elementary School Journal, 115*(4), 593–613.
- Koutsoftas, A. D. (2018). Writing-process products of fourth- and sixth-grade children: A descriptive study. *Elementary School Journal, 118*(4), 632–653. <https://doi.org/10.1086/697510>
- 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. [https://doi.org/10.1044/0161-1461\(2012\)11-0018](https://doi.org/10.1044/0161-1461(2012)11-0018)
- Koutsoftas, A. D., & Gray, S. (2013). A structural equation model of the writing process in typically-developing sixth grade children. *Reading and Writing, 26*(6), 941–966. <https://doi.org/10.1007/s11145-012-9399-y>
- Llaurado, A., & Dockrell, J. E. (2019). Children's plans for writing: Characteristics and impact on writing performance. *Journal of Literacy Research, 51*(3), 336–356. <https://doi.org/10.1177/1086296x19859516>
- (* Mackie, C., & Dockrell, J. E. (2004). The nature of written language deficits in children with SLI. *Journal of Speech, Language and Hearing Research, 41*(3), 1469–1483.
- Malecki, C. K., & Jewell, J. (2003). Developmental, gender, and practical considerations in scoring curriculum-based measurement writing probes. *Psychology in the Schools, 40*(4), 379–390. <https://doi.org/10.1002/pits.10096>
- 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.
- NCS Pearson (2012). Aimsweb Technical Manual. Bloomington, MN: NCS Pearson.
- 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*, 37–50. <https://doi.org/10.1037/a0013462>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews, 10*(89).
- Piercy, W., & Dockrell, J. E. (2023). Assessing children's writing products using curriculum-based measures of writing (CBM-W). *Educational and Child Psychology, 40* (1), 44–60. <https://doi.org/10.53841/bpsecp.2023.40.1.44>
- (* 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. [https://doi.org/10.1044/1058-0360\(2008\)012](https://doi.org/10.1044/1058-0360(2008)012).

- 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), 891-119. <https://doi.org/10.1080/10573569.2013.741957>
- Ritchey, K. D., & Coker, D. L., Jr. (2014). Identifying writing difficulties in first grade: An investigation of writing and reading measures. *Learning Disabilities Research & Practice*, 29(2), 54-65. <https://doi.org/10.1111/ldrp.12030>
- Romig, J. E., Therrien, W. J., & Lloyd, J. W. (2017). Meta-Analysis of Criterion Validity for Curriculum-Based Measurement in Written Language. *The Journal of Special Education*, 51(2), 72-82. <https://doi.org/10.1177/0022466916670637>
- Salas, N., Birello, M., & Ribas, T. (2021). Effectiveness of an SRSD writing intervention for low- and high-SES children. *Read Writ*, 34, 1653-1680. <https://doi.org/10.1007/s11145-020-10103-8>
- (*) 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), Article 324339.
- Sénéchal, M., Hill, S., & Malette, M. (2018). Individual differences in grade 4 children's written compositions: The role of online planning and revising, oral storytelling, and reading for pleasure. *Cognitive Development*, 45, 92-104. <https://doi.org/10.1016/j.cogdev.2017.12.004>
- SmitsEngelsman, B. C. M., & VanGalen, G. P. (1997). Dysgraphia in children: Lasting psychomotor deficiency or transient developmental delay? *Journal of Experimental Child Psychology*, 67(2), 164-184. <https://doi.org/10.1006/jecp.1997.2400>
- Sumner, E., Connelly, V., & Barnett, A. L. (2013). Children with dyslexia are slow writers because they pause more often and not because they are slow at handwriting execution. *Reading and Writing*, 26(6), 991-1008. <https://doi.org/10.1007/s11145-012-9403-6>
- Truckenmiller, A. J., Eckert, T. L., Coddling, R. S., & Petscher, Y. (2014). Evaluating the impact of feedback on elementary aged students' fluency growth in written expression: A randomized controlled trial. *Journal of School Psychology*, 52(6), 531-548. <https://doi.org/10.1016/j.jsp.2014.09.001>
- Wagner, R. K., Puranik, C. S., Fooman, B., Foster, E., Wilson, L. G., Tschinkel, E., & Kantor, P. T. (2011). Modeling the development of written language. *Reading and Writing: An Interdisciplinary Journal*, 24(2), 203-220. <https://doi.org/10.1007/s11145-010-9266-7>
- Wechsler, D. (2009) (WIAT-III).. *Wechsler Individual Attainment Test (Third edition.)*. San Antonio, TX: Pearson,
- Weissenburger, J., & Espin, C. A. (2005). Curriculum-based measures of writing across grade levels. *Journal of School Psychology*, 43, 153-169. <https://doi.org/10.1016/j.jsp.2005.03.002>
- Weiss, S. L., Brinkley, J., & Bock, J. (2019). Patterns of growth: Cluster analysis of written expression curriculum-based measurement. *Education and Treatment of Children*, 42(2), 161-184. <https://doi.org/10.1353/etc.2019.0008>
- Wilson, J., Chen, D., Sandbank, M. P., & Hebert, M. (2019). Generalizability of automated scores of writing quality in grades 3-5. *Journal of Educational Psychology*, 111(4), Article 619640. <https://doi.org/10.1037/edu0000311>
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, Illinois: Riverside.

Catherine Martin is a South African Speech and Language Therapist with a special interest in paediatric language and literacy.

Julie Dockrell (FRCSLT, FAcSS) is Professor of Psychology and Special Needs at the UCL, Institute of Education, qualified as both a clinical and educational psychologist. Her research focuses on the application of evidence-based research to support children's learning. She has published widely on language development and difficulties.