Received: 20 December 2022 | Accepted: 27 February 2023

DOI: 10.1111/1467-8578.12464

ORIGINAL ARTICLE

The writing quality scale (WQS): A new tool to identify writing difficulties in students

N. J. Stuart | **A. L. Barnett**

Centre for Psychological Research, Oxford Brookes University, Oxford, UK

Correspondence
N. J. Stuart, Centre for Psychological Research, Oxford Brookes University, Gipsy Lane, Oxford OX3 0BP, UK.
Email: nstuart@brookes.ac.uk

Abstract

Students in higher education (HE) are required to complete a variety of writing tasks for coursework and examinations. However, for some students writing presents a major challenge. In the UK, the availability of tools for specialist assessors to help identify difficulties with the quality of written composition is limited. The aim of this study was to develop a practical new tool, the Writing Quality Scale (WQS) for assessing writing quality in HE, that was both easy to use and did not require specialist or subject knowledge. The reliability and validity of the tool and its ability to identify students in HE who may need support were evaluated by examining scripts from 120 students (60 male) aged 17 to 25 years. The WQS was found to have good inter-rater reliability and was sensitive enough to pick up age differences and differentiate between groups of students with and without dyslexia. The WQS will be a useful tool for specialist assessors in HE to help in the identification of those with poor writing quality and to understand more about the nature of their difficulties.

KEYWORDS
assessment, higher education, SpLDs, writing

Key Points

- In this paper we outline a new tool, the Writing Quality Scale (WQS) which has been developed for assessing the writing quality of students in Higher Education (HE) settings.
- The WQS has been designed as a quick and easy to use tool that can be used alongside the DASH17+ free writing task and does not require specialist or subject knowledge for scoring and interpretation.
- Data on aspects of reliability and validity of the WQS, including its sensitivity to identifying difficulties in students with dyslexia are presented.
- The WQS will help specialist assessors in identifying students who have difficulties with the compositional quality of their writing and to make recommendations for appropriate support that might be required in a HE context.

INTRODUCTION

Writing is an important skill that is taught throughout students' primary and secondary school years. For students in higher education (HE), writing remains the main method of assessment for most disciplines. However, there are some students for whom writing presents a major challenge, with difficulties experienced in transcription (handwriting speed, legibility and spelling) and/or in the composition and quality of their writing.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. *British Journal of Special Education* published by John Wiley & Sons Ltd on behalf of National Association for Special Educational Needs.

This includes students with specific learning difficulties (SpLDs) such as dyslexia. In 2021-22 students with specific learning difficulties (SpLDs) accounted for 6.15% of the HE student population in the UK and for 33% of the student population with a known disability (Higher Education Statistics Authority, 2023).

Identification of difficulties in writing is key to ensuring equal opportunity, and it is important that appropriate adjustments are put in place for students in HE who have these difficulties to ensure they are not disadvantaged compared to their peers. In the UK, specialist teacher assessors work in HE contexts to identify, support and make recommendations for appropriate adjustments for students. These often include using a range of tests in an assessment for SpLDs to establish eligibility for disability provision and/or for providing evidence of the need for additional resources and allowances for students. However, the availability of tools to help assessors in identifying difficulties in writing quality is limited. The aim of this study was to develop and evaluate (for reliability and validity) a practical and easy-to-use criterion-referenced tool to assess writing quality. Bearing in mind the limited time available for assessors and the need to gather information using a range of other tests, this new tool was designed to be applied with another commonly used assessment instrument, the Detailed Assessment of Speed of Handwriting (DASH 17+; Barnett et al., 2010).

Writing is a complex task and four key cognitive processes have been identified (Hayes & Berninger, 2014): a *proposer* (which generates the idea), a *translator* (which selects the appropriate words to represent the idea), a *transcriber* (which converts these words into a written form by using handwriting or typing) and finally an *evaluator* (which involves the working memory and executive function skills that are needed to co-ordinate these processes). Hayes and Berninger's (2014) model also includes the motivation of the writer, the goal of writing, the particular task requirement, the task environment and the resources of the writer, such as their long-term memory, working memory, attention and reading capabilities, in producing a piece of written work. Models such as these help us understand both the component skills involved in writing and the relative contribution of each during development. Research with children has demonstrated the importance of teaching and practising the lower-level transcription skills (that is, handwriting/typing and spelling) to produce letters to form words. Only once these are automated can cognitive resources be freed up to allow the other processes to play a greater role in the planning and production of extended text (Berninger & Amtmann, 2003; Sumner et al., 2014).

Written compositions can be assessed in different ways. First, there are microstructure measures of writing that focus on productivity, complexity, punctuation and spelling at the word and sentence level. Second, there are macrostructure measures that include measures of organisation, structure and cohesion, and can be used at the text level (Wagner et al., 2011). The former can be characterised as the rule-based components of writing; that is, those conventions of standard English that need to be followed for the correct spelling of words and use of punctuation, grammar and word order to construct meaningful sentences. In contrast, the macrostructure measures can be characterised by the 'authorial' component of writing. This relates to the overall coherence and organisation of the written text: how it is structured, how the content and ideas within the text are developed, how vocabulary is selected and used to communicate the intended meaning and contribute to the overall cohesion, and finally how the writer engages their audience with their writing.

The different components of writing are typically evaluated or scored using either a holistic or an analytic scale. On a holistic scale, an overall score is given to the whole text; while on an analytic scale, a set of predetermined criteria that are judged to be important for writing are each scored and then an overall score is derived. For assessors, analytic scales provide more detailed information about a student's writing ability, which can help in providing feedback and recommendations for appropriate adjustment, allowances and support. However, analytic scales can vary in the number of criteria they contain, and deciding which criteria (or components of the writing process) to include is important if the tool is to be reliable, valid and appropriate for the age group with whom it is intended to be used.

Furthermore, the way in which the components of writing combine can vary. Crossley and colleagues (Crossley, 2020; Crossley et al., 2014) note that for high school and college students there are different ways to write a higher-quality essay and that these essays may not always share the same linguistic attributes. Lexical sophistication (that is, the extent to which the words chosen are appropriate to the topic and writing genre), syntactic structure (that is, the complexity and variety of the sentences used and clauses and phrases within the sentences) and text cohesion (that is, the connection of ideas at both the sentence level and the paragraph level) are all important linguistic elements of essay writing, but they may combine in different ways to create a successful essay and this can depend on the particular writing genre. Crossley et al.'s (2014) findings on the linguistic features that differentiate high-quality from low-quality essays point to lexical sophistication, syntactic complexity and text cohesion as being important. However, these need to be linked to the other components of writing that may also impact on writing quality, such as the lower-level transcription skills of handwriting/typing (speed and legibility) and spelling, and on following conventions relating to punctuation and capitalisation.

However, for some groups of students, particular components of writing can be problematic. For individuals with dyslexia, for example, spelling is particularly challenging, and in studies of younger children, spelling

14678578, 2023, 2, Downloaded from https://nasenjournals.onlinelibrary.wiley.com/doi/10.1111/1467-8578.12464 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

has been found to impact on the production of text (Sumner et al., 2014). Studies of students with dyslexia in HE have found that these difficulties with spelling persist (Connelly et al., 2006) and that, in addition, difficulties with punctuation, grammar and organising and structuring writing result in poorer-quality essays than those produced by their age-matched peers (Galbraith et al., 2012; Sumner & Connelly, 2020).

In the UK, standards, guidance and training for the assessment of SpLDs is promoted and monitored by the SpLD Assessment Standards Committee (SASC), a standard-setting group concerned with the diagnostic assessment of SpLDs set up to implement the training recommendations of the SpLD Working Group 2005/ DfES guidelines (2005). For the assessment of the writing of individuals over 16 years of age, a sample of 'free writing' is required 'to provide information about qualitative features such as grammar, sentence complexity, coherence, vocabulary choice, spelling accuracy, writing speed and handwriting legibility' (SASC, 2022).

Existing tests such as the DASH 17+ provide a standardised measure of handwriting speed, and the recently developed Handwriting Legibility Scale (HLS; Barnett et al., 2018) can also be used to provide additional information on potential difficulties with legibility. However, other than the Wechsler Individual Achievement Tests (WIAT-III^{UK}; Wechsler, 2017; WIAT-4; Wechsler, 2020), there is, to our knowledge, currently no practical and easy-to-use tool available to assessors in HE to provide an indication of the compositional quality of writing and whether a student may need support.

There are a number of challenges in choosing an appropriate task to assess compositional writing quality of students in HE in the context of an assessment for SpLDs. First, there are different writing genres to consider, and students will be required to produce writing in different genres, both across and within disciplines in HE, as they progress through their course of studies (Nesi & Gardner, 2012). Furthermore, even within a discipline, the type of assignment can vary from essays and literature reviews to case reports and lab reports, all of which have slightly different demands in terms of how the assignment needs to be approached and written. Second, there are time constraints on what can practically be achieved, in terms of the range and number of assessments included within an SpLD assessment session, especially when an assessment of writing is taking place as part of a broader assessment. Finally, a writing task undertaken in an SpLD assessment session is unlikely to directly reflect an actual exam or coursework assignment. The latter, for example, will include time to research, plan, write, edit, revise and proofread.

Wagner et al. (2011, p. 10) note the 'need to assess individual students' strengths and weaknesses at the word, sentence, and discourse level such that intervention can be tailored to meet those needs'. However, the limited availability of tools designed specifically for assessing writing quality in students in HE presents a significant challenge for specialist assessors who need to make specific recommendations for eligibility for support (for example, for Disabled Students Allowance (DSA) in the UK). The Wechsler Individual Achievement Tests (WIAT-III^{UK}, WIAT-4; Wechsler, 2017, 2020) include a sub-test for assessing written expression using an essay composition task, but this takes time to administer and score and, until recently, has been available only for use by psychologists. This still leaves a gap in the availability of suitable tools to assess students in HE. The availability of appropriate tools for identifying difficulties with writing is important, to ensure appropriate support and adjustments are in place for the formal academic qualifications studied in HE.

The aim of this study was therefore to address the limited availability of suitable tests for assessing writing quality in students in HE as part of an SpLD assessment. The first aim was to develop a new analytic tool specifically for use with 17- to 25-year-olds by specialist teacher assessors working in a HE context. Specifically, the new tool was developed:

- to identify students in HE who may have difficulties with writing for academic assignments;
- to be easy to use and not require specialist subject knowledge to apply the criteria, score and interpret results;
- to make use of an existing sample of free writing (DASH 17+) routinely collected in an assessment session.

The second aim was to evaluate aspects of the reliability and validity of the tool and its ability to identify students in HE who may need support.

METHODS

Development of the writing quality scale (WQS)

The Writing Quality Scale (WQS) was developed in response to feedback from specialist teacher assessors about the lack of appropriate tools to measure writing quality. The design of the WQS drew on various sources of information. This included previous literature on the predictors of writing quality in college students' essays (that is, lexical sophistication, syntactic structure and text cohesion; Crossley, 2020; Crossley et al., 2014), and the difficulties encountered in writing by students with SpLDs such as dyslexia (Galbraith et al., 2012; Sumner et al., 2014). Together with a review of existing tests (that is, WIAT-III^{UK}, WIAT-4; Wechsler, 2017, 2020) this informed the selection of criteria that would be appropriate for use with an older age group in HE. To meet the aim of developing a practical tool that did not require additional administration time within an SpLD assessment

14678578, 2023, 2, Downloaded from https://nasenjournals.onlinelibrary.wiley.com/doi/10.1111/1467-8578.12464 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

session, the WQS is designed to be used alongside the DASH 17+. The DASH 17+ is recommended for identifying handwriting difficulties (Castiglione, 2021) and is widely used by specialist teacher assessors in the UK to obtain information about handwriting speed. It includes a range of tasks and is standardised for use with 17- to 25-year-olds. Scripts from the DASH 17+ ten-minute free writing task were used in the development of the WQS.

The WQS is an analytic scale which involves the assessment of six criteria (see Table 1). Each is given a score between 1 and 4 and these are summed to yield a total score (ranging from 6 to 24), with higher scores indicating poor writing quality. In addition, once an overall WQS score has been calculated, a cut-off score has been established to help the assessor in identifying poor writing quality performance. This followed the procedure of using the sample group mean plus one standard deviation. This is considered an appropriate method for establishing a cut-off for a screening tool (Cascio et al., 1988). This then made it possible to identify whether a script fell into the high (indicating poor writing quality), medium or low category.

Grammar was originally considered as a separate criterion. However, after scoring a sample of 15 DASH 17+ free writing scripts in the development phase, it was felt that this criterion did not differentiate free writing scripts for this age group, and it was instead included within the *Sentence structure* criterion. To assist with scoring, a set of instructions, with examples, for the application of the criteria to the DASH 17+ free writing scripts was produced. This was developed through discussion among the research team, following the scoring of sample scripts used in the development of the criteria.

An expert panel was then formed to review and evaluate the content, structure and scoring of the updated version. This included six experts in writing development and assessment from different professions (test authors, writing researchers and specialist teacher assessors). They were provided with a set of instructions for using the WQS and an example script that had been scored and included notes on how the criteria had been applied. They were provided with a further two scripts and were requested to apply the WQS criteria to these scripts and then complete a feedback form. This asked, for each of the six criteria, whether they were clear, how easy it was to apply each criterion to the DASH 17+ free writing script and whether they had any additional comments. The expert panel were also asked whether they thought that the WQS would enable assessors (in a HE setting) to make a quick assessment of the quality of the written composition using the DASH 17+ free writing task. Their feedback was very positive about the tool and its value for assisting in the assessment of writing quality. A finalised set of criteria, instructions and examples was then prepared.

TABLE 1 Writing quality scale criteria.

Criteria	Brief description
1 Content and development	Development of content to form a coherent text that is clear, and which also engages the reader
2 Structure and organisation	Arrangement and ordering of ideas to form a cohesive text
3 Vocabulary	Appropriateness to convey the intended meaning and to retain the reader's interest
4 Sentence structure	Production of meaningful and well-formed sentences that conform to the grammatical 'rules' (or conventions) of standard English
5 Punctuation (including capitalisation and apostrophes)	Correct use to aid reading of the text
6 Spelling	Accuracy of words used

Participants

One hundred and twenty scripts were drawn from the UK stratified sample for the DASH 17+. These came from 60 male students and 60 female students aged 17 to 25 years (mean age = 20). The sample was selected from a range of schools, colleges and universities across the UK, including England, Scotland, Wales and Northern Ireland. The numbers in each of the three age groups used in the DASH 17+ norms are shown in Table 2. Further demographic information, provided by each participant, is shown in Tables 3 and 4. The number of participants from each geographical region in the UK is shown in Table 3. Each participant categorised themselves as belonging to one of the races/ethnic groups shown in Table 4.

Eighty-nine percent of the students had attended a state school and 11% had attended a private school. Students were taking a variety of courses: 17% A levels, 18% vocational qualifications, 50% BA/BSc, 6% PGDip, 4% MA/MSc and 5% PhD. Students taking vocational courses were working towards obtaining qualifications in a range of subjects including childcare, animal management, business studies, computing, media and sport studies. Those taking academic courses were working across a broad range of subjects including the arts and humanities, sciences and healthcare, mathematics and technology.

TABLE 2 Number of male and female participants in the sample.

Age (years)	Male	Female	Total
17–18	21	20	41
19–21	20	20	40
22–25	19	20	39
Total	60	60	120

14678578, 2023, 2, Downloaded from https://nasenjournals.onlinelibrary.wiley.com/doi/10.1111/1467-8578.12464 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Percentage of participants by geographical region.

Region	%
North of England	23.2
Midlands and east of England	20.0
London and south-east of England	38.4
South-west of England	6.2
Wales	2.0
Scotland	6.2
Northern Ireland	4.0
Total	100

TABLE 4 Percentage of participants according to race/ethnic group.

Race/ethnic group	%
White	82.0
Black (Black African, Black Caribbean, Black other)	6.0
Asian (Indian, Pakistani, Bangladeshi, Asian other)	7.0
Other (Chinese, mixed race, other)	5.0
Total	100

Scripts from an additional 28 participants (11 male, 17 female) with a mean age of 19 years from the DASH 17+ standardised sample were also scored (see Table 5). These students were reported to have dyslexia by the students themselves and by their dyslexic support tutors. They were matched to an age and gender comparison group selected from the larger sample described in Table 2.

Measures

The WQS was applied to the free writing scripts collected as part of the UK stratified sample for the DASH 17+ between January 2008 and March 2010.

Detailed assessment of speed of handwriting (DASH 17+)

The DASH 17+ includes four main tasks to measure speed of handwriting production, but only the scripts for the 10-minute free writing task were used in this study. In this task participants are required to write on the topic of 'My Life'. They are given some time before writing to generate ideas and writing prompts are available during the writing period. Students are instructed to write using their 'everyday' handwriting and to write continuously for 10 minutes. Students were tested in groups of six to 25. A further sample of 12 students were included who were re-tested after an interval of one to two weeks. For this study, the DASH 17+ free writing scripts were first transcribed and typed up (preserving spelling errors and crossed-out words, and noting instances of illegible words) to reduce bias in scoring from poor handwriting (Graham & Weintraub, 1996; Greifeneder et al., 2010).

TABLE 5 Number of male and female participants with a self-reported diagnosis of dyslexia.

Age (years)	Male	Female	Total
17–18	8	7	15
19–21	3	9	12
22–25	0	1	1
Total	11	17	28

Writing quality scale (WQS)

The DASH 17+ scripts were then given a score for each of the six WQS criteria and a total WQS score was calculated (range 6–24), with higher scores indicating poor quality writing. The scores were assigned by a trained rater who was a postgraduate student in psychology.

Data analysis

The data are presented in two sections. The first examines aspects of reliability (internal consistency, inter-rater, test–retest) and the second examines aspects of validity (construct, differential) of the WQS.

Analysis was conducted using IBM SPSS for Windows Version 25. Effect sizes for independent t-tests are reported where small ($d=0.2$), medium ($d=0.5$) and large ($d=0.8$) following Cohen (1992), and for Pearson correlations as small ($d=0.1$), medium ($d=0.3$) and large ($d=0.5$). The significance level for statistical tests was taken as $p<0.05$. Where ANOVAs were significant, post hoc tests were computed and effect size was measured using partial eta square where 0.02 is considered to be a small effect size, 0.13 a medium effect size and 0.26 a large effect size.

RESULTS

Descriptive data (means and standard deviations) for the overall WQS scores obtained from the three age groups are shown in Table 6. A total mean score of 14.59 (SD = 2.73) was obtained for the 17- to 25-year-old sample. However, the mean scores decreased as age increased, indicating that writing quality increased with age. The differences between the three age groups were statistically significant, F (2,117) = 5.606, $p=0.005$, $\eta^2=0.087$. Tukey post hoc analysis revealed that the mean difference between the 17- to 18-year-old age group and the two older groups was statistically significant: 19- to 21-year-old group (1.46, 95% CI [0.07, 2.84], $p=0.037$); 22–25 year old group (1.86, 95% CI [0.47, 3.26], $p=0.005$). There was no

14678578, 2023, 2, Downloaded from https://nasenjournals.onlinelibrary.wiley.com/doi/10.1111/1467-8578.12464 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 6 Writing Quality Scale mean and standard deviations for age groups and total sample.

Age (years)	Number (n)	Mean	Standard deviation
17–18 years	41	15.68	2.86
19–21 years	40	14.23	2.58
22–25 years	39	13.82	2.42
Total	120	14.59	2.73

significant difference between the 19- to 21- and 22- to 25-year-olds.

The number of scripts scoring in the low (scores of 6–12), medium (scores of 13–16) and high (scores of 17–24) WQS categories are shown in Table 7. The high category is equivalent to more than one standard deviation above the mean of the sample of 120 participants (that is, 14.59+2.73=17.32 and rounded down to a score of 17) and indicates poor writing quality.

Reliability

Internal consistency

Cronbach's (1951) alpha coefficient was used to establish whether the six criteria in the WQS were all measuring writing quality. The Cronbach's alpha coefficient was 0.70 (n=120) for the WQS and is at the recommended level for the reliability of the overall internal consistency of the scale.

Inter-rater reliability

The trained rater scored all 120 scripts and the first author independently scored 24 (20%) of them. Inter-rater reliability for the total WQS score, following categorisation of the scores as low, medium and high, was 83.33% and inter-rater reliability using Cohen's kappa was good: k=0.669 (95% CI, 0.375 to 0.963) p=<0.001.

Test–retest reliability

The stability of scores on the WQS on different occasions was assessed by analysing a sample of 12 students who were tested twice at an interval of one to two weeks. A positive Pearson's r correlation of r (10)=0.62, p=0.033 (Cohen's d=0.38) was found. However, the mean score for the retest scripts (mean=13.58, SD=2.61) was lower (indicating better quality) than the original mean test scores (mean=15.50, SD=2.24) and this difference was significant (t (11)=3.086, p=0.010 (Cohen's d=0.89)). Scores improved on all scripts, except for two that received the same score and one script where writing quality decreased.

TABLE 7 Number and percentage of scripts with low, medium and high Writing Quality Scale category scores.

Writing quality scale category	Number (n)	%
Low (6–12)	25	21
Medium (13–16)	75	62
High (17–24)	20	17
Total	120	100

Validity

Construct validity

A principal component analysis (PCA) of the six component scores was undertaken to establish whether using the WQS assessed one or more components of writing quality. A PCA was run on the six criteria in the WQS for 120 participants. The PCA revealed two components that had eigenvalues greater than one and which explained 42.13%, and 19.61% of the total variance, respectively. The two-component solution explained 61.73% of the total variance. A varimax orthogonal rotation was employed to aid interpretability (see Table 8) with *Spelling*, *Punctuation* and *Sentence structure* in the first component and *Content and development*, *Structure and organisation* and *Vocabulary* in the second component.

Differential validity

Further evidence for validity was sought by comparing a different group of students with a reported diagnosis of dyslexia, who would be expected to differ on writing quality from age- and gender-matched peers without dyslexia taken from the main sample.

Half of the 28 scripts (11 male) for the group with dyslexia fell into the high category (indicating poor writing quality), with 36% in the medium category and 14% in the low category. This compares with 18% of the age- and gender-matched scripts falling in the high category and 54% and 28% in the medium and low categories, respectively. A chi-square test of independence revealed a significant difference in the distribution across the categories for the group with dyslexia compared to the age- and gender-matched group (χ^2 (2, N=56)=6.60, p=0.04).

The total WQS score (mean=16.89, SD=3.48) for the group with dyslexia was significantly higher than for the age- and gender-matched control group (mean=14.43, SD=2.52) indicating poorer writing quality (t (54)=3.04, p=0.01) with a large effect size (Cohen's d=0.81) indicating that the magnitude of difference between the total WQS score for the two groups was meaningful. Performance on the individual WQS criteria for the individuals with dyslexia and their age- and gender-matched groups are shown in Table 9. The group with dyslexia had higher scores (lower performance) than the age- and

14678578, 2023, 2, Downloaded from https://nasenjournals.onlinelibrary.wiley.com/doi/10.1111/1467-8578.12464 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 8 Principal component analysis of the six Writing Quality Scale criteria.

Writing quality scale criteria	Rotated component coefficients	
	Component 1	Component 2
Spelling	0.854	−0.081
Punctuation	0.815	0.256
Sentence structure	0.676	0.481
Content and development	−0.072	0.826
Vocabulary	0.259	0.654
Structure and organisation	0.161	0.585

TABLE 9 Mean scores (and standard deviations) for the Writing Quality Scale criteria for the group with dyslexia and the age and gender match group.

	Dyslexia $n=28$	Age and gender match $n=28$	p*	Cohen's d
Content and development	2.57 (0.92)	2.14 (0.80)	0.07	0.30
Structure and organisation	3.11 (0.63)	2.54 (0.69)	0.01	0.41
Vocabulary	2.64 (0.68)	2.29 (0.53)	0.03	0.30
Sentence structure	3.11 (0.69)	2.71 (0.71)	0.04	0.30
Punctuation	3.00 (0.77)	2.71 (0.66)	0.16	0.20
Spelling	2.46 (1.11)	2.04 (0.88)	0.13	0.23
Total	16.89 (3.48)	14.43 (2.52)	0.01	0.81

*Bonferroni significance level for multiple comparisons = $p<0.01$.

gender-matched group on all six criteria. However, only *Structure and organisation* was close to being statistically significant when a Bonferroni significance level of $p<0.01$ for multiple comparisons was used.

DISCUSSION

The WQS was developed to provide a practical tool to assess writing quality alongside an existing measure of handwriting speed as part of an SpLD assessment. The WQS was initially designed for use by specialist teacher assessors working in HE settings to identify students aged 17 to 25 years with, or at risk of having, difficulties in the quality of their written composition. Furthermore, it was designed not to require specialist or subject knowledge for its administration and interpretation.

The starting point was to develop a set of criteria for an analytic scale which were then used to assess free writing samples obtained from administration of the free writing task in the DASH 17+, a standardised task that is widely used in the UK to assess handwriting speed. It is important to note that the DASH 17+ free writing task was not designed as a task to measure writing quality; however, the application of the WQS criteria to the 'My Life' free writing scripts was found to be sensitive to age differences in writing quality. Students were also found to write on a variety of topics; for example, their family, their holidays, hobbies or life at university.

One of the challenges in developing a tool that is quick and easy to use is ensuring that it is also reliable. In terms of its reliability and the extent to which all the items on a scale measure one construct, the internal consistency of the WQS was good, with an acceptable Cronbach's alpha of 0.70. Inter-rater reliability as measured by Cohen's kappa was in the good range (Altman, 1990) for WQS scores falling into the high, medium and low categories. One of the difficulties of scoring scripts for writing quality is that, unlike an assessment of spelling, for example, the assessor is required to make a subjective judgement. Brown et al. (2004) note, in a review of studies of writing assessment tools, that exact agreement between raters is typically between 40% and 60% with kappas of between 0.70 and 0.80 typically reported. These are consistent with those reported for the WQS, with a kappa of 0.70, although exact agreement on the WQS was higher at 83%.

In terms of validity, first the content and structure were supported by the feedback from the expert panel. In terms of construct validity, a two-component solution was generated with each component accounting for 42.13% and 19.61% of the variance. Interpretation of the data is consistent with there being two separate components or independent dimensions to writing quality: the rule-based aspects, or conventions of writing (that is, *Spelling*, *Punctuation* and *Sentence structure*) and the 'authorial' component of writing (that is, *Content and development*, *Structure and organisation* and *Vocabulary*). The latter criteria are more concerned with the development and structure of ideas and the overall meaning of the script. *Content and development* and *Structure and organisation* are traditionally included in macrostructure measures of overall text coherence and cohesion, but the inclusion of *Vocabulary* in this study could be seen as recognising the important role that the choice of vocabulary can play in the coherence and cohesion of a written text, and that may not always be captured in more traditional microstructure measures of vocabulary use and diversity. The retention of the six criteria in the WQS allows for further exploration of where particular difficulties with the rule-based conventions of writing or the 'authorial' components may lie. The inclusion of *Spelling* in the criteria, while typically viewed as a transcription skill, was important because it can impact on the meaning of what is written and can interact with sentence structure and punctuation. It can also be compared with performance on standardised spelling tests which would form part of an SpLD assessment.

A key aim in developing the WQS was for it to identify those students who might find the demands of academic writing challenging in HE. The differential validity of the WQS tool was examined by checking the ability of the WQS tool to differentiate between groups of students with

14678578, 2023, 2, Downloaded from https://nasenjournals.onlinelibrary.wiley.com/doi/10.1111/1467-8578.12464 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and without dyslexia. The group with dyslexia had lower performance on the total WQS score and on each of the six criteria compared to their age- and gender-matched peers. Previous studies investigating the writing performance of students with dyslexia in HE have found significant differences in spelling performance compared to peers (Connelly et al., 2006; Galbraith et al., 2012; Sumner & Connelly, 2020). Performance on spelling in the WQS was lower in the group with dyslexia than their peers, but this difference was not statistically significant and may be explained by the group with dyslexia having smaller vocabularies and/or adopting a strategy of using high-frequency and common/familiar words that they were confident they knew how to spell. Such a strategy could also explain the lower performance score on the *Vocabulary* measure found in the WQS for the group with dyslexia, although the difference was not statistically significant, as the vocabulary used is limited to a smaller range of words that the student is confident that they know how to spell.

Limitations

The scores from the WQS are criterion-referenced and, as for any individual test, will need to be interpreted alongside the results from other tests administered and observations made during the SpLD assessment, together with information from and about the student. This should include current or past difficulties and in particular difficulties that might impact on writing (for example, in reading and spelling single words, understanding written language, spoken language production and understanding, and motor difficulties). The information from the WQS should enable assessors to have a more comprehensive picture of any difficulties in writing quality, which can then be evaluated in the wider context of the SpLD assessment.

To help in identifying poor writing quality, the overall WQS scores were categorised into high (indicating poor writing quality), medium and low categories. The cut-off score that was established to categorise scores that fell in the high (poor quality) range was the mean score from the main sample plus the standard deviation. Given the quality of the scripts falling within this category, the cut-off score of 17 would appear to be appropriate for identifying students with poor writing quality and in need of additional support. However, as with any other tool that includes cut-off scores, care should be taken with students who are at or just below the cut-off score, and further validation of this cut-off is required.

The WQS was used to assess a narrative descriptive task: the 'My Life' free writing task from the DASH 17+. It is acknowledged that this task places relatively minor demands on recalling information and converting it into appropriate written language, unlike formal written assignments in HE. However, in the context of a full SpLD assessment session, the topic should not result in too much anxiety over thinking about what to write. The 'My Life' task also has the advantage of typically generating topics that most assessors will be familiar with, and for which they will be able to make judgements on the sophistication of the vocabulary used, sentence construction and organisation and coherence of the ideas presented and developed. However, in interpreting WQS scores, assessors will need to be mindful of the particular assessment demands of the discipline that the student is or will be studying and whether a regional accent and/or dialect may impact on their writing. In addition, it might be helpful to look at a sample of academic writing (if available) from the student to help inform decisions about the type of support that might be recommended.

Implications

The WQS provides specialist assessors with a tool for obtaining information to help identify potential difficulties in the quality of students' written composition and to assist in recommending further support and adjustments for them. Further assessment might be required in the case of those with very high scores on the WQS (indicating poor quality writing). The WQS tool was sensitive enough to identify students with dyslexia. The development of the WQS is ongoing, and future work will help to determine whether it is useful for more diverse groups of students (such as those with English as a second language) in addition to those with other SpLDs associated with writing difficulties (for example, developmental language disorder, developmental co-ordination disorder), how it relates to different types of academic assessments, and whether it could also be used in secondary school with 14- to 16-year-olds. An administration and scoring guide to support the application and interpretation of the criteria has also been developed.

CONCLUSIONS

This study has described the development and initial evaluation of the WQS for use in an SpLD assessment of students aged 17 to 25 years in HE to address a gap in the availability of assessment tools for this age group. The WQS was found to have good inter-rater reliability and statistical evidence of differential validity was found. The WQS provides a tool that can help assessors identify the category (low, medium or high) into which a student's writing quality falls, and can indicate whether the overall writing quality is likely to be of concern. It can also be used to identify areas of relative strengths and difficulties, whether in the content and development of what is written, its structure and organisation, the appropriateness of the vocabulary used, the complexity and variety of sentence structures used, or the correct use of punctuation and spelling to convey the intended meaning. Used in conjunction with

14678578, 2023, 2, Downloaded from https://nasenjournals.onlinelibrary.wiley.com/doi/10.1111/1467-8578.12464 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the results from other tests administered in an SpLD assessment, observations, and the student's personal history, the WQS can provide a valuable additional assessment tool. HE institutions may vary as to where the responsibility falls for supporting the student in developing their writing skills (for example, central provision or within the academic discipline) but it is hoped that by improving the ability to identify students with writing difficulties, those students who need support will be able to access it.

ACKNOWLEDGEMENTS

The authors would like to thank Karen Gurney for her contribution and members of the expert panel for reviewing the WQS and for their helpful comments.

FUNDING INFORMATION

This research was funded by an Oxford Brookes University Research Excellence award to the first author.

CONFLICT OF INTEREST STATEMENT

The authors declare no potential conflicts of interest with respect to the research, authorship and/or publication of this article. Anna Barnett is the principal author of the DASH17+ published by Pearson.

DATA AVAILABILITY STATEMENT

A copy of the WQS scoring sheet and manual is available on application to the corresponding author. The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

Institutional research ethics approval had been obtained for the collection of scripts as part of the DASH 17+ standardisation project.

ORCID

N. J. Stuart https://orcid.org/0000-0002-2438-9669
A. L. Barnett https://orcid.org/0000-0003-3800-9468

REFERENCES

Altman, D.G. (1990) *Practical statistics for medical research*, 1st edition. Chapman and Hall/CRC. Available from: https://doi.org/10.1201/9780429258589

Barnett, A.L., Henderson, S., Scheib, B. & Schulz, J. (2010) *Detailed assessment of speed of handwriting 17+*. London, UK: Pearson Assessment.

Barnett, A.L., Prunty, M. & Rosenblum, S. (2018) Development of the handwriting legibility scale (HLS): a preliminary examination of reliability and validity. *Research in Developmental Disabilities*, 72, 240–247. Available from: https://doi.org/10.1016/j.ridd.2017.11.013

Berninger, V.W. & Amtmann, D. (2003) Preventing written expression disabilities through early and continuing assessment and intervention for handwriting and/or spelling problems: research into practice. In: Swanson, H.L., Harris, K.R. & Graham, S. (Eds.) *Handbook of learning disabilities*. New York: Guilford Press, pp. 345–363.

Brown, G.T.L., Glasswell, K. & Harland, D. (2004) Accuracy in the scoring of writing: studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing*, 9, 105–121. Available from: https://doi.org/10.1016/j.asw.2004.07.001

Cascio, W.F., Alexander, R.A. & Barrett, G.V. (1988) Setting cutoff scores: legal, psychometric, and professional issues and guidelines. *Personnel Psychology*, 41, 1–24. Available from: https://doi.org/10.1111/j.1744-6570.1988.tb00629.x

Castiglione, L. (2021) *Assessing the need for access arrangements in examinations: sixth edition*. Evesham: PATOSS (Professional Association of Teachers of Students with Specific Learning Difficulties).

Cohen, J. (1992) A power primer. *Psychological Bulletin*, 112(1), 155–159. Available from: https://doi.org/10.1037//0033-2909.112.1.155

Connelly, V., Campbell, S., MacLean, M. & Barnes, J. (2006) Contribution of lower order skills to the written composition of college students with and without dyslexia. *Developmental Neuropsychology*, 29(1), 175–196. Available from: https://doi.org/10.1207/s15326942dn2901_9

Cronbach, L.J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334. Available from: https://doi.org/10.1007/BF02310555

Crossley, S. (2020) Linguistic features in writing quality and development: an overview. *Journal of Writing Research*, 11, 415–443. Available from: https://doi.org/10.17239/jowr-2020.11.03.01

Crossley, S.A., Roscoe, R. & McNamara, D.S. (2014) What is successful writing? An investigation into the multiple ways writers can write successful essays. *Written Communication*, 31(2), 184–214. Available from: https://doi.org/10.1177/0741088314526354

Galbraith, D., Baaijen, V., Smith-Spark, J. & Torrance, M. (2012) The effects of dyslexia on the writing processes of students in Higher Education. In: Torrance, M., Alamargot, D., Castelló, M., Ganier, F., Kruse, O., Mangen, A. et al. (Eds.) *Learning to write effectively: current trends in European research*. Bingley: Emerald Group, pp. 167–171.

Graham, S. & Weintraub, N. (1996) A review of handwriting research: progress and prospects from 1980 to 1994. *Educational Psychology Review*, 8(1), 7–87. Available from: https://doi.org/10.1007/BF01761831

Greifeneder, R., Alt, A., Bottenberg, K., Seele, T., Zelt, S. & Wagener, D. (2010) On writing legibly: processing fluency systematically biases evaluations of handwritten material. *Social Psychological and Personality Science*, 1(3), 230–237. Available from: https://doi.org/10.1177/1948550610368434

Hayes, J.R. & Berninger, V.W. (2014) Cognitive processes in writing: a framework. In: Arfé, B., Dockrell, J. & Berninger, V. (Eds.) *Writing development in children with hearing loss, dyslexia, or oral language problems: implications for assessment and instruction*. Oxford: Oxford University Press, pp. 3–15. Available from: https://doi.org/10.1093/acprof:oso/9780199827282.003.0001

Higher Education Statistics Authority. (2023) 'HE student data: who's studying in HE', 9 February. Available from: https://www.hesa.ac.uk/data-and-analysis/students/table-15

Nesi, H. & Gardner, S. (2012) *Genres across the disciplines: student writing in higher education (Cambridge applied linguistics series)*. Cambridge: Cambridge University Press.

SASC (SpLD Assessment Standards Committee). (2022) *SASC – format for a post-16 diagnostic assessment report for specific learning difficulties*, May. Available from: https://sasc.org.uk/Downloads.aspx

SpLD Working Group 2005/DfES guidelines (2005) Available from: https://www.sasc.org.uk/(S(2j2fuq45af3c3qafusymmw45))/SASCDocuments/SpLD_Working_Group_2005-DfES_Guidelines.pdf

Sumner, E. & Connelly, V. (2020) Writing and revision strategies of students with and without dyslexia. *Journal of Learning Disabilities*, 53(3), 189–198. Available from: https://doi.org/10.1177/0022219419899090

Sumner, E., Connelly, V. & Barnett, A.L. (2014) The influence of spelling ability on handwriting production: children with and

14678578, 2023, 2, Downloaded from https://nasenjournals.onlinelibrary.wiley.com/doi/10.1111/1467-8578.12464 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

without dyslexia. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 40(5), 1441–1447. Available from: https://doi.org/10.1037/a0035785

Wagner, R.K., Puranik, C.S., Foorman, B., Foster, E., Wilson, L.G., Tschinkel, E. et al. (2011) Modelling the development of written language. *Reading and Writing*, 24(2), 203–220. Available from: https://doi.org/10.1007/s11145-010-9266-7

Wechsler, D. (2017) *The Wechsler individual achievement test. Third UK edition. (WIAT-IIIUK)*. London, UK: Pearson Assessment.

Wechsler, D. (2020) *The Wechsler individual achievement test. Fourth edition. (WIAT-4)*. San Antonio, TX, USA: Pearson Assessment.

How to cite this article: Stuart, N.J. & Barnett, A.L. (2023) The writing quality scale (WQS): A new tool to identify writing difficulties in students. *British Journal of Special Education*, 50, 258–267. Available from: https://doi.org/10.1111/1467-8578.12464

14678578, 2023, 2, Downloaded from https://nasenjournals.onlinelibrary.wiley.com/doi/10.1111/1467-8578.12464 by Test, Wiley Online Library on [24/07/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License