ASHA American Speech-Language-Hearing Association

AJSLP

Research Note

Examination of an Automated Procedure for Calculating Morphological Complexity

Carla Wood,[a] Miguel Garcia-Salas,[a] and Christopher Schatschneider[a]

[a] Florida State University, Tallahassee

ARTICLE INFO

Article History:
Received February 2, 2023
Revision received May 8, 2023
Accepted May 25, 2023

Editor-in-Chief: Erinn H. Finke
Editor: Lynne E. Hewitt

https://doi.org/10.1044/2023_AJSLP-23-00044

ABSTRACT

Purpose: The aim of this study was to advance the analysis of written language transcripts by validating an automated scoring procedure using an automated open-access tool for calculating morphological complexity (MC) from written transcripts.
Method: The MC of words in 146 written responses of students in fifth grade was assessed using two procedures: (a) hand-coding of words containing derivational morphemes by trained scorers and (b) an automated analysis of MC using Morpholex, a newly developed web-based tool. Correlational analysis between the different MC calculations was examined to consider the relation between hand-coded derivational morpheme counts and the automated measures. Additionally, all MC measures were compared to a previously gathered rating of writing quality to consider predictive validity between the automated Morpholex score and teachers' ratings of writing quality.
Results: Automated measures of MC had a strong relation (r = .63) with hand-coding of the number of words with derivational morphemes. Additionally, the number of derivational and inflectional and derivational morphemes accounted for a significant amount of the variation in teachers' overall ratings of writing quality.
Conclusion: Automated scoring of MC has potential utility as a valid alternative to hand-coding language samples, which may be valuable for progress monitoring of growth in complexity across repeated samples and measuring components that influence perceived quality of academic writing.

Oral and written language samples are widely accepted as best practice for language assessment (Heilmann et al., 2010; Price et al., 2010) to identify areas to target for treatment and to evaluate progress over time (e.g., Overton & Wren, 2014). Unfortunately, language sample analysis is underutilized in research due to the time demands for conducting, coding, and analyzing (Heilmann et al., 2010; Klatte et al., 2022) and clinically underutilized due excessive time demands for lengthy analyses, lack of access to analysis tools, and lack of reimbursement for such time (Klatte et al., 2022). Some authors have suggested that user-friendly software options may offer practical solutions (Klatte et al., 2022); however, few studies have empirically examined use of automated analysis options, specifically for analysis of morphological forms and use of complex morphological words.

A growing literature base substantiates that students' morphological knowledge and skills play an important role in academic vocabulary skills and reading outcomes in school-age students (e.g., Goodwin et al., 2013; Kieffer et al., 2016). Results of previous studies indicate that *morphological knowledge*, or one's understanding of the morphemic structure of words and the ability to combine morphemes to construct words or to decompose words into smaller meaningful word parts (Kirby & Bowers, 2018; Levesque et al., 2019), explains variation in reading comprehension above and beyond other reading and language measures (Foorman et al., 2012; Kieffer & Lesaux, 2008; Kieffer et al., 2016). Findings of numerous studies suggest that the relation between morphological knowledge and

Correspondence to Carla Wood: carla.wood@cci.fsu.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

reading outcomes holds true for students who are multilingual learners (Goodwin et al., 2013; Kieffer & Lesaux, 2008; Kieffer et al., 2013) and students from varied ability backgrounds, including students with language learning disorders (Fallon & Katz, 2020; Meaux et al., 2020; Wolter & Green, 2013).

Although morphological knowledge is a widely recognized contributing factor to academic language, fewer studies have included or examined measures of morphological complexity (MC) in expressive language tasks. The lack of research on morphological skills based on expressive, generative tasks may be due in part to the time demands of measuring and analyzing MC in students' language output. Even so, it would stand to reason that expressive and receptive knowledge may have different predictive value in assessment.

In response to the time demand barriers to analyzing MC in written transcripts, here, we focus on Morpholex analysis, an open-access web-based tool for assessing MC of written text (Cobb, 2022; Laufer & Cobb, 2020). *MC* is used here to refer to an array of measurements to describe morpheme types in a sample including the number words that include at least one derivational morpheme (e.g., −*ity* in *electricity*), number of words with inflectional morphemes (e.g., −*ed* in *walked*), and the rate at which derivational and inflectional morphemes occur relative to the total number of words. The web-based tool allows for the automated analysis of the number of affixed types of words, percentage by tokens, percentage of base words, percentage of words with inflectional morphemes, and percentage of words with derivational morphemes. Additionally, the automated tool quantifies the number of base words and provides data on the variety of base words and variety of inflectional and derivational morphemes categorized by developmental level and type. These features could lend to the use of MC analysis in clinical and research practices if easier and faster to implement than traditional methods of hand-coding for analyzing words for MC in a lengthy sample (i.e., manually marking inflectional and derivational morphemes in a language sample).

There are numerous potential strengths of the automated web-based tool for use in practice by speech-language pathologists and researchers if found to be easy, efficient, and predictive of gold standard measures. Among potential strengths, the time demand of manually marking or designating morphemes in words is expected to be reduced compared to the traditional method of using Systematic Analysis of Language Transcripts (SALT). Relatedly, manually identifying types of morphemes requires training of coders and is subject to human error. As such, an automated web-based tool such as Morpholex, if a valid measure of MC, could reduce the labor-intensive process of identifying types of morphemes and quantifying the rate of use of various types of morphemes for speech-language pathologists and researchers.

Despite potential advantages of Morpholex for clinical and research practices, there are no studies to our knowledge that have systematically compared automated measures of MC generated by Morpholex to more traditional methods for assessing MC. Furthermore, there are no studies to date that have examined the relevance of the output measures for predicting gold standard measures of writing such as teachers' ratings of overall writing quality, which often serve as classroom-based measures of written language for progress monitoring. Writing quality is of interest as ratings of quality based on rubric scoring have been widely used by researchers as a general measure of students' writing performance (e.g., Kim et al., 2018; Olinghouse & Leaird, 2009) and as an outcome variable in intervention studies (e.g., Rosário et al., 2019). Although *quality* represents a multicomponent construct, such ratings are reported to be related to other language and literacy skills (e.g., Kent & Wanzek, 2016). Among skills found to influence quality ratings include vocabulary diversity (Olinghouse & Leaird, 2009), lexical sophistication (Kim et al., 2018; McNamara et al., 2010), and syntactic complexity (Casal & Lee, 2019; Mostafa & Crossley, 2020).

In this study, our primary aim was to conduct a preliminary analysis of the automated measures of MC generated by Morpholex and examine relations between the automated measures and counts of derivational morphemes identified by hand and aggregated using traditional SALT analysis. Additionally, we sought to provide informative data on efficiency and predictive validity of this analysis approach to assist in determining utility for future use in clinical and research practices. Specific research questions included the following.

1. What is the relation between different measures of MC of written language samples including measures calculated through different modalities (e.g., hand-coded, automated analysis)?
2. Are there apparent differences in efficiency between different modalities of measuring MC?
3. Do measures of MC of written words predict teachers' independent ratings of overall writing quality for students from diverse linguistic and ability backgrounds in fifth grade?

Method

The written transcripts used in this study were gathered as baseline data in a larger study that included 1,396 students with ratings of writing quality by teachers in one large school district in southern Florida (Wood & Schatschneider, 2022). Participating schools reported serving a high density of multilingual learners in their school population. This study, approved by the university human subjects committee (HSC #25857), included a random sample of de-identified writing samples from the larger database to allow for a preliminary analysis of the utility of the tool for students who differed in linguistic backgrounds and language ability.

Participants

This study included 146 transcripts randomly selected to sample a subset of written language responses of students from diverse linguistic and ability backgrounds. From the initial random sample of writing samples, duplicates and blank responses were eliminated. The final sample for this study included monolingual English-speaking students (n = 49) and Spanish-English–speaking multilingual learners (n = 97). A subset of students in the sample (38%) were identified by the school district as having a language learning disability (n = 55) and received speech-language services. Students' reported race/ethnicity included 80% Hispanic, 16% Black, 2% White (non-Hispanic), and 2% from mixed racial backgrounds. The data provided by the district indicated that 86% were eligible for free or reduced lunch. All participants were educated in classrooms with monolingual English-speaking students and peers with varied language ability backgrounds including students who received exceptional student support services.

Measures

Automated Morpholex scoring. To obtain the index of MC from Morpholex, the electronic text from the written responses was copied and pasted into the online profiler (Morpholex Affix Profiler v.3.5.1). Morpholex was used to quantify the percentage of affix types, tokens, bases, inflections, and derivations. Furthermore, the output from the profiler provided counts of morpheme occurrences by level of difficulty, which were imported to the database. The measures of morphemes by difficulty level included quantification of base words (Level 1), regular inflectional morphemes (Level 2), early developing derivations (Level 3), and later developing derivations (Levels 4–6). Levels were established according to Bauer and Nation's (1993) MC system. Level 2 includes all inflectional suffixes (e.g., plurals, possessives, superlatives). Level 3 includes the most frequent derivation affixes such as *–able*, −*er*, and *–less*. Level 4 consists of frequent, orthographically regular affixes that often impose pronunciation change, including *–al*, −*ity*, −*ment*, and *–ous* to name a few. Level 5 consists of regular but infrequent affixes (e.g., −*atory* in confirmatory, −*ant* in consultant, and *–dom* in kingdom). Level 6 is made up of affixes that have been categorized in earlier levels; however, in these cases, the addition of the affixes requires irregular truncations or insertions (e.g., truncation of *–ia* in anemic, truncation of *–ate* in attenuable, insertion of *t* in dramatic). In this study, the frequency and regularity of derivational morphemes were not the main focus. As a result, we computed a sum score of the total number of words with derivational morphemes by adding up the occurrences of derivations at Levels 3–6.

Reliability of the Morpholex measures was evaluated in several ways. First, previously typed written samples were electronically copied into Morpholex by two independent research assistants, for which the output and resulting measures were identical. Second, handwritten samples were assigned for direct entry into the web-based Morpholex tool by two independent research assistants. The resulting output matched 86% of the time with relatively minor differences in the numeric output when independent researchers entered the text electronically for duplicate entries, largely due to procedural differences in whether spelling and capitalization errors were maintained that resulted in discrepancies in the output derived. For example, one researcher maintained a capitalization error, whereas the other corrected it according to standard writing conventions, resulting in the word being counted as two unique tokens in the text. Other trivial errors resulted from inconsistencies in the inclusion of portions of the written response that were not in the main body of the student's response (e.g., title, header) and were easily corrected.

Identification of morphologically complex words by hand. To calculate the number of words that contained derivational morphemes by hand, a code (morphologically complex words [MCW]) was entered in the electronic transcript for words containing at least one derivational morpheme (i.e., prefix and suffix), which were included in the word's etymological history and definition as determined by the Merriam-Webster dictionary and the Online Etymological Dictionary. SALT was used to aggregate the number of the MCW word codes to quantify the total number of occurrences of MCW in each student's writing sample. All derivational MCW were counted if the derivational affix was spelled correctly, regardless of the students' accuracy on the stem's spelling. For example, "teacher" and "teecher" both received credit for *–er* in MCW, which is consistent with scoring methods used in previous research (Turnbull et al., 2011). The hand-identified MCW codes were completed as part of the prior study. Coders received instruction on derivational morphemes then practiced identifying MCW on a set of 10 transcripts. After receiving feedback on the practice set, coders received 20 transcripts to code with review and feedback until they reached at least 90% accuracy. The intraclass correlation coefficient across coders was reported to be .98 for a set of 100 transcripts after completion of training.

Ratings of Writing Quality

The ratings of writing quality were previously completed as part of a prior study (Wood & Schatschneider, 2021). Quality rating scores were based on the rubric adopted by the district, which was consistent with the state assessment. The rubric was used to score the written samples on three categories of quality: (a) purpose, focus, and organization; (b) evidence and elaboration; and (c) writing conventions of Standard English. These elements are consistent with components found in established scoring systems such as Wechsler Objective Language Dimensions (Rust, 1996; Wechsler, 2005) and previous studies (Williams et al., 2013).

Each quality component contributed to the total composite rating. For the first category of writing quality, students' writing samples were scored on purpose, focus, and organization. To score the maximum 4 points in this category, the student's written response demonstrated a strong idea with little or no loosely related material, skillful use of transitions, and a logical progression of ideas including an introduction and conclusion. For the second category of writing quality, investigators scored the writing samples on the inclusion of evidence and elaboration. To achieve the maximum number of 4 points, students integrated evidence thoroughly and smoothly using appropriate vocabulary and sentence structure. Finally, for the third quality category rating, investigators rated students' writing on use of conventions of Standard English. To obtain a full 2-point rating in this category, students' responses may have only occasional minor errors in use of Standard English without patterns of errors and generally demonstrate appropriate use of punctuation, capitalization, sentence formation, and spelling. Finally, a composite score was calculated as the sum of the three components. This overall quality of writing rating was aligned with state assessment procedures. The total writing quality rubric score (on a 10-point scale) is purported to reflect original thought, use of text evidence, inferences, implicit understanding, and synthesizing across texts.

Procedures Conducted Prior to This Study

Collection of writing samples. As part of the previous study, classroom teachers administered an interim writing assessment as part of a curriculum-based assessment tool used district wide. Teachers distributed a packet containing two written passages about the benefits of exercise, directions for the writing task, and lined paper to use for a written response. The directions instructed students to read two passages, plan and write a response to a prompt, and revise and edit the response. The first passage, which was seven paragraphs long (one and a half pages double-spaced), focused on unexpected outcomes of fitness. The second passage (two pages in length) was about the benefits of fitness for an individual who was blind. Students were given 120 min to read the two passages and compose a written response explaining the benefits of fitness.

Transcription of samples. The investigators gathered the students' written responses from their classroom teachers. Research assistants typed the written samples into a Word document to prepare it electronically for graduate research assistants who reviewed the paper copies against the electronic file to check accuracy. A Word document version devoid of formatting was retained for Morpholex analysis. For identification of MCW by hand, the transcripts were formatted using SALT conventions. The investigators ensured all writing samples were transcribed into the electronic database for analysis using the SALT program. Because the SALT program has specific formatting conventions, a check for formatting errors was conducted by another research assistant prior to running SALT analyses.

Writing quality rating. Two raters who were certified teachers working within the partnering school district scored the written samples on each of the three categories of writing quality. The raters had completed extensive training on the writing rubric (provided by the school district), passed an assessment of writing training, and attended monthly training meetings including regular online scoring courses to recalibrate. A randomly selected subsample was blindly double rated by both raters independently. When using the criteria that any point difference is a disagreement, interrater agreement was 70%, 77.5%, and 67.5% for quality subcomponents, respectively. This was above the 60% criteria for ratings of writing quality in published reviews (Graham & Perin, 2007). When considering agreement as a point difference of greater than 1, similar to previous studies (e.g., Koutsoftas, 2016; Koutsoftas & Gray, 2012), an interrater agreement of 100%, 100%, and 97.5% was attained.

Data Analysis

To answer the first research question examining the relation between different measures of MC of written language samples, we first examined descriptive statistics for each approach (i.e., hand-coded MCW and automated analysis of morpheme types and density). We then investigated the relation between measures using a Pearson product–moment correlation coefficient. Preliminary analyses were performed to ensure no violation of the assumptions of normality and linearity.

To explore potential differences in efficiency and accuracy of different modalities, we considered differences in time, consistency between double-entered data by different research assistants, and the time demands in aggregating the data. To quantify the time demands of hand-coding, we randomly selected a subset of 20 samples for a time study. Using a stopwatch, we recorded the total seconds required to hand-code MCW within each transcribed writing sample. To account for the variation in length and complexity of the transcripts, the average number of seconds per word for manual coding was calculated for each sample by dividing the hand-coding time in seconds by the total number of words. We then calculated an overall mean from the average number of seconds per word across the subset to derive an overall average time required to hand-code MCW per word. The overall average time per word was used to estimate the total time required for hand-coding MCW in the complete data set of 146 samples. This was done by multiplying the average time per word by the total number of words in the data set. Additionally, we used the average time per transcript to estimate the time that would be required for large data sets by multiplying the average time per transcript by 300 and 500 samples.

Finally for the third research question, we used multiple regression to examine how a set of the MC measures generated by Morpholex was able to predict teachers' independent ratings of overall writing quality when taken together into the model as a whole. We then examined the relative contribution of each of the variables that make up the model (the number of derived morphemes, the number of inflectional morphemes, and the number of affixed words) to consider which variable in the set was the best predictor of writing quality ratings.

Results

Relations Between Measures of MC

To examine the relation between identifying words with derivational morphemes by hand and automated quantification of MC measures, we first report descriptive statistics for each method (see Table 1). In general, the number of words containing at least one derivational morpheme identified through hand-coding (M = 9.95, SD = 8.07) and Morpholex was similar (M = 9.74, SD = 5.88). A Pearson product correlation coefficient was computed to compare the relationship between hand-coding of MCW and Morpholex's derivational morphemes. There was a large positive correlation between the variables, r = .63, n = 158, and the relationship was significant at $p < .001$.

A Pearson correlation also examined the relation between MC measures and writing quality. The mean writing quality score was 3.39 (SD = 1.67), the mean number of words with at least one inflectional morpheme was 18.37 (SD = 11.31), the mean number of affixed words was 20.72 (SD = 10.68), and the mean ratio of affixed words was 4.46 (SD = 2.94). With one exception, all measures of MC were significantly correlated with teachers' assessments of overall writing quality. Manual coding of MCW had a medium positive correlation with writing quality (r = .40, $p < .001$). Morpholex's measure of words with at least one derivational morpheme (r = .32, $p < .001$), number of words with at least one inflectional morpheme (r = .43, $p < .001$), and number of total affixed words (r = .42, $p < .001$) all had a medium and significant positive correlation with overall writing quality. The percentage of affixed words in relation to the total words used in the writing sample was not correlated with writing quality (r = −.01, $p \geq .999$).

Time Differences in Assessment of MC

To consider the second research question, we examined the time demands of hand-coding MCW. Recognizing that students' written responses vary in length and complexity, using a randomly selected subset of the transcripts, we calculated the total time required for hand-coding divided by the total number of words to derive an average of 0.69 s per word (with a range of 0.25–1.46 per individual transcript) for determining if each word was morphologically complex or not and inserting the MCW code. By applying the overall average time per word and multiplying it by the average number of total words per transcript in the complete data set of 146 samples, it was estimated that the time required for hand-coding MCW was 154 s per transcript or 375 min for the sample (equivalent to 22,484 s). Additionally, based on the average total time per transcript being 154 s, the time required for manually determining and coding MCW for larger data sets of 300 and 500 samples would be 770 min (46,200 s) and 1,283 min (77,000 s), respectively.

We then considered the unique demands of different methods. Using the open-source web-based tool, the calculation of MC required a trivial amount of time and effort. The text of the Word document was copied and pasted into the online application, and the output on MC was generated in 3 s or less. In contrast, the hand identification of inflectional morphemes in SALT (with insertion of slashes) and manual marking of derivational morphemes required 20 min to 4 hr, with large variation depending

Table 1. Means, standard deviations, and correlations between measures of morphological complexity.

Variable	*M*	*SD*	1	2	3	4	5
1. Hand-coded MCW	9.95	8.07					
2. Morpholex derivational morphemes	9.74	5.88	.63**				
3. Morpholex inflectional morphemes	18.37	11.31	.44**	.48**			
4. Morpholex no. of affixed words	20.72	10.68	.44**	.58**	.89**		
5. Morpholex ratio of affixed words	4.46	2.94	.20*	.58**	.02	.17*	
6. Writing quality	3.39	1.67	.40**	.32**	.43**	.42**	−.01

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. MCW = morphologically complex words.
*Indicates $p < .05$. **Indicates $p < .01$.

on the number of words in the sample, the complexity of the words, and the proficiency of the user. In contrast to hand-coding, with automated analysis, multiple pages of electronic text required no additional time (compared to a short sample) to generate data upon electronically copying and pasting the written response into Morpholex.

Nevertheless, there was a notable difference in time demands between the two methods during the aggregation of the data. The researcher version of SALT allows for triangulation of data analysis across a large number of files exporting to a single data file (i.e., comma-separated values, Excel). In contrast, the Morpholex open-source web-based tool utilizes a single transcript at a time and requires the output be transferred manually to an Excel database.

Predictive Validity

Finally, to address the third research question, which pertained to the predictive relation between a set of measures of MC and teachers' independent ratings of overall writing quality, we conducted a multiple regression analysis using the following Morpholex measures: number of words containing at least one derivational morpheme, number of words with at least one inflectional morpheme, and total number of affixed words (see Table 2). A significant regression equation was found, $F(3, 135) = 11.21$, $p < .001$, and together the predictors accounted for 19.94% of the variance in writing quality. Individually, the number of words containing at least one derivational morpheme ($\beta = .12$, $p = .19$), the number of words containing at least one inflectional morpheme ($\beta = .27$, $p = .10$), and the number of affixed words ($\beta = .10$, $p = .57$) did not contribute unique variance to the prediction of writing quality.

Discussion

Written language samples are reportedly underutilized by clinicians and researchers, in part due to time-intensive nature of traditional coding and analyses (Heilmann et al., 2010; Klatte et al., 2022). For this reason, we were interested in potential time-saving benefits of using the automated analysis tool for morphological analysis of written words. We sought to examine relations between measures and consider the relation to ratings of writing quality to provide some validation for the new methodology for clinicians and researchers. The automated analysis took approximately 3 s or less per sample. Automated calculations of the number of words with derivational morphemes were similar to counts derived using traditional methods. Measures of MC derived from the new methodology significantly predicted teachers' independent ratings of writing quality.

Relation between modalities. This strength of the relation between modalities is partially explained by the

Table 2. Regression results: morphological complexity measures predicting writing quality.

Predictor	*b*	*SE b*	β	*R*	Fit
(Intercept)	2.01**	0.29			
Morpholex derivational morphemes	0.03	0.03	.12	.32**	
Morpholex inflectional morphemes	0.04	0.02	.27	.43**	
Morpholex no. of affixed words	0.02	0.03	.10	.42**	
					$R^2 = .199$**
					95% CI [.08, .30]

Note. A significant *b*-weight indicates that the beta-weight and semipartial correlation are also significant. *b* represents unstandardized regression weights. β indicates the standardized regression weights. *r* represents the zero-order correlation. CI = confidence interval.
**Indicates $p < .01$.

fact that the measures represent different aspects of MC. The available measures in Morpholex reflect complexity in slightly different ways, attributing to differences in number of words with derivational morphemes identified manually. For example, some of the Morpholex measures reflect the number of unique occurrences, where the manual identified did not differentiate between unique types and reoccurrences of the morpheme. Furthermore, other Morpholex measures reflect complexity based on percentage of words containing inflectional or derivational morphemes, and the hand identification of MCW was implemented to quantify the occurrence of derivational morphemes specifically. Moreover, it is possible and likely that hand-coding of occurrence of derivational morphemes could be more subject to human error, contributing to discrepancies between modalities. Overall, the large positive significant relation between MC measures supports the notion that the different indices of MC are reflecting related aspects of a larger constellation of skills that are presumed to represent underlying morphological knowledge and skills.

Efficiency. The time-saving aspects, compared to identifying and coding by hand, make an automated tool such as Morpholex well suited for clinical use due to its efficiency (particularly for lengthy text), open access, and lack of subscription requirements. For research use, the potential magnitude of advantages is less conclusive regarding the efficiency of one tool over the other for assessing MC. While the automated aspect of Morpholex makes it faster than hand-coding within SALT, the output is more laborious when managing and aggregating the resulting data. At the time of this study, the aggregation of data across transcripts had to be done by hand (e.g., in Excel), which took additional time, but the cost–benefit analysis of time would be dependent on the length of the transcripts and the number of transcripts analyzed. With Morpholex, the time demands may occur in aggregation of data or pooling together the data obtained from individual transcripts. In SALT, the time demands are on the front-end in formatting and identifying morphemes manually, while aggregation of data is faster than Morpholex if the researcher version of SALT is available to the user and utilized. Both methods are impacted by the number of transcripts to be analyzed, but in different ways. For a small number of samples, time demands in aggregating data across transcripts may not be a deterrent to use; however, for large data sets, the efficiency advantages of Morpholex could be more fully realized with an alternative aggregation tool to extract the data from the website.

Validation of complexity with gold standard. While MC of writing may have a variety of broad uses in clinical and research practices, we were interested in the predicative validity in relation to performance on standardized assessments and a social perception of writing quality. The finding that MC measures significantly predicted independent ratings of overall writing quality provides support for the construct validity or meaningfulness of the measures generated by the automated analysis tool for assessing MC of students' written responses. Considering that researchers, clinicians, and educators use writing quality ratings as a common indicator of students' writing performance (e.g., Kim et al., 2018; Olinghouse & Leaird, 2009) and the fact that three Morpholex measures explained nearly 20% of the variance in ratings of quality are important findings. This provides useful evidence to support decision making regarding use.

Although *quality* is a multidimensional or multicomponent construct, it is not surprising that MC would significantly predict quality ratings since such ratings have previously been reported to vocabulary diversity (Olinghouse & Leaird, 2009) and lexical sophistication (Kim et al., 2018; McNamara et al., 2010). It seems reasonable that the use of morphologically complex words would explain some of the variance in teachers' perceptions of lexical sophistication and the breadth of vocabulary across the written language sample. Furthermore, the fact that the use of inflectional morphemes accounted for additional unique variance in ratings of quality is not surprising given that inflectional morphemes are related to grammaticality, a core component of rubrics used for rating quality.

Limitations

Although the current results were considered pioneering work in modalities for assessment of written MC, the findings should be interpreted cautiously, recognizing limitations. To ensure authenticity and social validity, the study utilized an established rubric for quality ratings that was commonly used throughout the district. However, the rubric itself had certain inherent limitations, such as its compulsory use of General American English (GAE) for rating quality. This particular aspect of the rubric was considered a weakness, as it may not accurately reflect the writing skills of individuals who do not use GAE. Although non-GAE speakers are not expected to use a written form of their dialect in fifth-grade classrooms, teachers' preference for GAE reflects a negative bias when assessing students who speak non-GAE in oral contexts.

It is important to note that the time required for hand-coding was based on an experienced coder identifying morphologically complex words. Therefore, the average time per word may underestimate the time required for new coders, and it does not account for training or practice time. Additionally, there is no widely accepted "gold standard" for assessing MC in students' written responses. While manual coding in SALT is used in research, it is unclear how widely it is used. Furthermore, it is unclear if licensed software analysis options are widely accessible or used by clinicians in practice, so the efficiency comparison may not accurately reflect real-world situations.

Implications and Future Directions

The findings suggest that automated scoring of MC has potential utility as an alternative to identifying morpheme types manually and coding for morphologically complex words by hand, particularly for derivational morphemes. The results demonstrate that Morpholex is an efficient and valid method for assessing MC of transcripts. As such, the results of this study have implications for clinical use of the tool. The automated process may encourage implementation of analysis of MC into research and clinical practice. Considering the relation between the MC metric and the performance on the standardized measure of vocabulary and teachers' ratings of writing quality, speech-language pathologists and researchers may want to consider adding MC indices into informal assessment protocols. It is considered a strength that the study sample included students from diverse ability and linguistic backgrounds, which may support generalizability of findings on relations of writing quality to a variety of diverse and inclusive classrooms, but additional studies are needed to further validate the tools and methodology.

Recognizing that morphological knowledge contributes to reading comprehension (Carlisle, 2000; Levesque et al., 2019) and use of morphologically complex words is associated with superior ratings of writing quality, additional studies are warranted to further explore the use of informal assessment indices such as Morpholex for measuring generative morphology skills. Future studies are needed to further validate such indices through multiple sources of evidence including examination of content and construct validity, consistency of the index over time, and evaluation of the ability to detect growth or skill development over time. Given the efficiency of the automated analysis, the tool could have utility in clinical and research practice by reducing or alleviating the time demand burdens. The efficiency and open-access may enhance the feasibility and likelihood of clinical use for measuring change over time and progress monitoring of MC across the school year or multiple grades using repeated written language classroom-based assessment measures.

Data Availability Statement

Data used in this study are available on request from the authors.

Acknowledgments

The collection of the written language samples used in this study was supported by the Institute of Education Sciences, U.S. Department of Education, Grant R305L180019 to Florida State University. Miguel Garcia-Salas was supported by Grant R305B200020 from the Florida Center for Reading Research at Florida State University, funded by the Institute of Education Sciences, U.S. Department of Education. The opinions expressed are those of the authors and do not represent views of the Department of Education.

References

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography, 6,* 253–279.

Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing, 12,* 169–190. https://doi.org/10.1023/A:1008131926604

Casal, J. E., & Lee, J. J. (2019). Syntactic complexity and writing quality in assessed first-year L2 writing. *Journal of Second Language Writing, 44,* 51–62. https://doi.org/10.1016/j.jslw.2019.03.005

Cobb, T. (2022). Counting affixes with Morpholex: A response to McLean and Stoeckel. *Reading in a Foreign Language, 34*(1), 165–171.

Fallon, K. A., & Katz, L. A. (2020). Structured literacy intervention for students with dyslexia: Focus on growing morphological skills. *Language, Speech, and Hearing Services in Schools, 51*(2), 336–344. https://doi.org/10.1044/2019_LSHSS-19-00019

Foorman, B. R., Petscher, Y., & Bishop, M. D. (2012). The incremental variance of morphological knowledge to reading comprehension in grades 3–10 beyond prior reading comprehension, spelling, and text reading efficiency. *Learning and Individual Differences, 22*(6), 792–798. https://doi.org/10.1016/j.lindif.2012.07.009

Goodwin, A. P., Huggins, A. C., Carlo, M. S., August, D., & Calderon, M. (2013). Minding morphology: How morphological awareness relates to reading for English language learners. *Reading and Writing, 26*(9), 1387–1415. https://doi.org/10.1007/s11145-012-9412-5

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*(3), 445–476. https://doi.org/10.1037/0022-0663.99.3.445

Heilmann, J., Miller, J. F., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology, 19*(2), 154–166. https://doi.org/10.1044/1058-0360(2009/08-0024)

Kent, S. C., & Wanzek, J. (2016). The relationship between component skills and writing quality and production across developmental levels. *Review of Educational Research, 86*(2), 570–601. https://doi.org/10.3102/0034654315619491

Kieffer, M. J., Biancarosa, G., & Mancilla-Martinez, J. (2013). Roles of morphological awareness in the reading comprehension of Spanish-speaking language minority learners: Exploring partial mediation by vocabulary and reading fluency. *Applied Psycholinguistics, 34*(4), 697–725. https://doi.org/10.1017/S0142716411000920

Kieffer, M. J., & Lesaux, N. K. (2008). The role of derivational morphology in the reading comprehension of Spanish-speaking English language learners. *Reading and Writing, 21*(8), 783–804. https://doi.org/10.1007/s11145-007-9092-8

Kieffer, M. J., Petscher, Y., Proctor, C. P., & Silverman, R. D. (2016). Is the whole greater than the sum of its parts? Modeling the contributions of language comprehension skills to reading comprehension in the upper elementary grades. *Scientific Studies of Reading, 20*(6), 436–454. https://doi.org/10.1080/10888438.2016.1214591

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *The Modern Language Journal, 102*(1), 120–141. https://doi.org/10.1111/modl.12447

Kirby, J. R., & Bowers, P. N. (2018). The effects of morphological instruction on vocabulary learning, reading, and spelling. In R. Berthiaume, D. Daigle, & A. Desrochers (Eds.), *Morphological processing and literacy development: Current issues and research* (pp. 217–243). Routledge. https://doi.org/10.4324/9781315229140-10

Klatte, I. S., van Heugten, V., Zwitserlood, R., & Gerrits, E. (2022). Language sample analysis in clinical practice: Speech-language pathologists' barriers, facilitators, and needs. *Language, Speech, and Hearing Services in Schools, 53*(1), 1–16. https://doi.org/10.1044/2021_LSHSS-21-00026

Koutsoftas, A. D. (2016). Writing process products in intermediate-grade children with and without language-based learning disabilities. *Journal of Speech, Language, and Hearing Research, 59*(6), 1471–1483. https://doi.org/10.1044/2016_JSLHR-L-15-0133

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)

Laufer, B. & Cobb, T. (2020). How much knowledge of derived words is needed for reading? *Applied Linguistics, 41*(6), 971–998. https://doi.org/10.1093/applin/amz051

Levesque, K. C., Kieffer, M. J., & Deacon, S. H. (2019). Inferring meaning from meaningful parts: The contributions of morphological skills to the development of children's reading comprehension. *Reading Research Quarterly, 54*(1), 63–80. https://doi.org/10.1002/rrq.219

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication, 27*(1), 57–86. https://doi.org/10.1177/0741088309351547

Meaux, A. B., Wolter, J. A., & Collins, G. G. (2020). Forum: Morphological awareness as a key factor in language-literacy success for academic achievement. *Language, Speech, and Hearing Services in Schools, 51*(3), 509–513. https://doi.org/10.1044/2020_LSHSS-20-00064

Mostafa, T., & Crossley, S. A. (2020). Verb argument construction complexity indices and L2 writing quality: Effects of writing tasks and prompts. *Journal of Second Language Writing, 49,* Article 100730. https://doi.org/10.1016/j.jslw.2020.100730

Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth grade students. *Reading and Writing, 22*(5), 545–565. https://doi.org/10.1007/s11145-008-9124-z

Overton, S., & Wren, Y. (2014). Outcome measurement using naturalistic language samples: A feasibility pilot study using language transcription software and speech and language therapy assistants. *Child Language Teaching and Therapy, 30*(2), 221–229. https://doi.org/10.1177/0265659013519251

Price, L. H., Hendricks, S., & Cook, C. (2010). Incorporating computer-aided language sample analysis into clinical practice. *Language, Speech, and Hearing Services in Schools, 41*(2), 206–222. https://doi.org/10.1044/0161-1461(2009/08-0054)

Rosário, P., Högemann, J., Núñez, J. C., Vallejo, G., Cunha, J., Rodríguez, C., & Fuentes, S. (2019). The impact of three types of writing intervention on students' writing quality. *PLOS ONE, 14*(7), Article e0218099. https://doi.org/10.1371/journal.pone.0218099

Rust, J. (1996). *The manual of the Wechsler objective language dimensions (WOLD): UK edition*. The Psychological Corporation.

Turnbull, K., Deacon, S. H., & Kay-Raining Bird, E. (2011). Mastering inflectional suffixes: A longitudinal study of beginning writers' spellings. *Journal of Child Language, 38*(3), 533–553. https://doi.org/10.1017/S030500091000022X

Wechsler, D. (2005). *Wechsler Individual Achievement Test (WIATT-II)*. Pearson; Harcourt Assessments.

Williams, G. J., Larkin, R. F., & Blaggan, S. (2013). Written language skills in children with specific language impairment. *International Journal of Language & Communication Disorders, 48*(2), 160–171. https://doi.org/10.1111/1460-6984.12010

Wolter, J. A., & Green, L. (2013). Morphological awareness intervention in school-age children with language and literacy deficits. *Topics in Language Disorders, 33*(1), 27–41. https://doi.org/10.1097/TLD.0b013e318280f5aa

Wood, C., & Schatschneider, C. (2021). Differential growth in writing quality of students in fifth grade from diverse backgrounds. *Reading and Writing Quarterly, 38*(2), 168–183. https://doi.org/10.1080/10573569.2021.1926027

Wood, C., & Schatschneider, C. (2022). Growth in written academic word use in response to morphology-focused supplemental instruction. *Reading and Writing, 35*(2), 399–426. https://doi.org/10.1007/s11145-021-10187-w