



This is a repository copy of *Evaluating English-language morphological awareness assessments*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/213273/>

Version: Published Version

Article:

Hudson Kam, C.L. orcid.org/0000-0002-7638-3656, Sadlier-Brown, E., Clark, S. et al. (3 more authors) (2024) Evaluating English-language morphological awareness assessments. *First Language*, 44 (4). pp. 327-344. ISSN 0142-7237

<https://doi.org/10.1177/01427237241245500>

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Evaluating English-language morphological awareness assessments

First Language

1–18

© The Author(s) 2024



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/01427237241245500

journals.sagepub.com/home/fla**Carla L. Hudson Kam** **Emily Sadlier-Brown**

The University of British Columbia, Canada

Shannon Clark

University of Alberta, Canada

Chelsea Jang

The University of British Columbia, Canada

Carrie Demmans Epp

University of Alberta, Canada

Jenny Thomson

University of Sheffield, UK

Abstract

Many studies have shown that morphological knowledge has effects on reading comprehension separate from other aspects of language knowledge. This has implications for reading instruction and assessment: it suggests that children could have reading comprehension difficulties that are due to a lack of morphological knowledge, and thus, that explicit instruction of morphology might be helpful for them, indeed for all children. To find children who might especially benefit from specific instruction in morphology, we would need good tests of morphological knowledge. We evaluated a set of morphological awareness assessments to determine whether they conclusively

Corresponding author:

Carla L. Hudson Kam, Department of Linguistics, The University of British Columbia, 2613 West Mall, Vancouver, BC V6T 1Z4, Canada.

Email: Carla.HudsonKam@ubc.ca

tapped into morphological knowledge, and found that it was not possible to be certain that they were accurately targeting morphological knowledge.

Keywords

Morphological awareness, morphology, methodology, assessment, reading comprehension

Reading comprehension, understanding the message in a text, is a complex process that brings to bear a range of linguistic knowledge (see, e.g. Gough & Tunmer, 1986), including the reader's morphological knowledge (see, e.g. Carlisle, 2000; James et al., 2021). Given that reading comprehension involves extracting the meaning from the text, and morphemes are units of meaning, this makes sense. Imagine a reader encountering the word 'blueish' for the first time: if they know the word 'blue' and the affix '-ish', they will be able to understand it. Although morphology interacts with other aspects of language (e.g. the way the plural -s is pronounced differs depending on the phonological context), some aspects of morphological knowledge are, in principle, separable from knowledge of other aspects of language. In particular, the understanding that words can be decomposed (or broken into) subparts that have their own meanings that contribute to the meaning of the whole and a knowledge of what those meanings are is specifically morphological. This has important implications for reading instruction and assessment. It suggests that explicit instruction of morphological decomposition might be helpful to developing readers (see James et al., 2021), and that there might be children this instruction will be especially helpful for due to particularly weak morphological knowledge. To find such children, indeed to assess the level of morphological knowledge in any reader, we would need good tests of morphological knowledge.

The work described here resulted from a search for just such a test. As part of a larger project, we were searching for an existing assessment that we could use to find readers who were struggling with reading comprehension, at least in part, because of their lack of morphological knowledge of English. In the end, however, when we searched commonly used assessments, we were unable to find one that conclusively tapped into morphological knowledge. All of the tests we examined have been shown to be related to reading comprehension performance and so clearly measure something that is related to reading skill; this 'something' may indeed be morphological knowledge. However, the issue for us is that we cannot be certain that they measure morphological knowledge, and so they are difficult to use for our intended purpose.

What is morphological knowledge?

Morphemes are the smallest units of language with independent meaning. They can be as small as a single sound or comprise multiple syllables. They can occur on their own (*free morphemes*), or occur only as part of a larger form (*bound morphemes*). These two dimensions, length and (in)dependence, are distinct, demonstrated by examples such as

the indefinite article ‘a’ versus the plural marker ‘-s’ (free vs. bound single sounds), or ‘carbon’ versus ‘semi-’ (free vs. bound multisyllabic morphemes). Words that are not decomposable into component morphemes provide no hints as to their meaning. By contrast, words that have component morphemes do. Understanding the meanings that individual bound morphemes carry (or contribute) can help someone understand an unfamiliar word, even in instances where the word has a base they don’t know. Take ‘miffinish’, for example, you probably assume that ‘miffin’ describes a property and that ‘miffinish’ is an adjective. However, as words like ‘fish’ and ‘flourish’ show, not all words that end in ‘-ish’ are adjectives; the same combination of sounds (or letters) can be a morpheme in one word (e.g. in ‘blueish’) but part of the base in another (e.g. in ‘fish’). Syntactic context is a cue to the potential decomposition of a word, as words of different types take or allow different bound morphemes. If you read ‘I have a miffinish’ you would assume that ‘miffinish’ is a noun with ‘-ish’ as part of the base form, whereas if you read ‘I have a miffinish burlicue’ you would assume that ‘miffinish’ is an adjective with the ‘-ish’ affix. You can do this because you understand the effect of the bound morpheme ‘ish’ on the base word.

This -ish example highlights the important relationship between syntax and morphology. For instance, bound morphemes are generally discussed as attaching only to words of specific syntactic categories, for example, ‘-ed’ attaches to verbs. A subset of bound morphemes, derivational morphemes, creates words with new meanings, often in a different syntactic category than the original word, for example, the verb ‘teach’ can be turned into the noun ‘teacher’ with the addition of the agentive morpheme ‘-er’ (cf. ‘do’ and ‘un-do’ where the category doesn’t change). Making matters more complicated, however, English actually has fairly flexible word categorisation that can result in words taking affixes that they apparently shouldn’t, for instance, when we create a new verb out of a noun without adding any apparent derivational morpheme (e.g. proper nouns like *Zwift*, or *Google* can be inflected as in *zwinging* or *googled*).

There are also interactions between phonology and morphology. The way a morpheme is pronounced can be affected by the phonological context in which it appears. For example, the plural morpheme ‘-s’ is pronounced as [s] after voiceless consonants (e.g. ‘cats’), as [z] after voiced consonants (e.g. ‘dogs’), and as [ɪz] after some sibilant consonants (e.g. ‘horses’). For many words the pronunciation is dictated by language-wide phonological constraints (as in *cats* and *dogs*), but for some the phonological form of the allomorph could (phonologically) be otherwise (e.g. words that end in [n], as [n] can be followed by a voiceless sibilant, as in ‘*quince*’). As the example of the plural shows, allomorphs (the variants of a morpheme) are sometimes written the same despite the differences in pronunciation (writing the plural forms of *cat*, *dog* and *horse* all involve simply adding an s), but not always (e.g. *fish/fishes*, see also impossible, in-credible and ir-regular, which all involve the same morpheme). This mix of regularities, sometimes the phoneme and other times the morpheme, being represented by a consistent written form, makes English orthography morphophonological (Spencer, 2017).

It might seem that when a child produces a multi-morphemic form it indicates they have some sort of knowledge of the form’s decomposition, but this is not necessarily the case. Indeed, there is reason to believe that children’s early productions of

multimorphemic words are single units in children's minds (see, e.g. the discussion in Bates & Goodman, 1997). A child who says 'kiss' and 'kissed', for instance, may not understand that 'kissed' is actually 'kiss' + 'ed', but instead, just knows two words that happen to sound similar and have similar meanings. Over time, the structure, in this case Verb + ed, is acquired, and the child is then able to create novel multi-morphemic forms. Once a child can do this, we can have some confidence that their knowledge is productive (because it leads to things that are not part of the child's direct experience).¹ Productivity can also be demonstrated using nonce (made-up) words. For instance, Berko (1958) asked 4- to 7-year-old children to produce a novel inflected or derived word and found some evidence for productivity. However, the children's performance was well below ceiling for most items tested, suggesting that even at 7 years of age, children are still learning many English morphemes.

At the same time, children's errors can indicate a fairly sophisticated knowledge of morphology. A fairly well-known example of this can be seen in *overregularisations* (e.g. when a child says 'goed' instead of 'went'). Children can also use morphological productivity to fill lexical gaps, either in the language or in their knowledge of it, for example, a child saying 'I'm horsing' to mean pretending to be a horse, or 'fastening', to mean, going increasingly faster and faster (fast turned into the verb fasten [to increase in speed], plus the verbal affix -ing). (These are examples of child speech collected by the authors.) These facts, that errors can indicate morphological knowledge, while correct answers need not, should be kept in mind as we consider morphological knowledge assessments.

Morphological knowledge and reading comprehension

Many studies now exist showing a relationship between a reader's morphological knowledge and their reading comprehension (e.g. Carlisle, 2000; Deacon et al., 2018). Such studies often measure a construct referred to as morphological awareness (MA), defined by Tighe and Binder (2015, p. 246) as 'a conscious understanding of how words can be broken down into smaller units of meaning', so those were the assessments we set out to evaluate. MA has been shown to be related to reading comprehension above and beyond phonological awareness and direct vocabulary measures (James et al., 2021; Levesque et al., 2017). At the same time, studies suggest a complex relationship between MA and reading comprehension. For instance, MA is correlated with readers' decoding ability for multimorphemic words containing derivational affixes (reflecting the morphophonological nature of English orthography), which is itself predictive of reading comprehension (Singson et al., 2000). MA is also related to children's ability to comprehend multimorphemic words, which then impacts reading comprehension (Levesque et al., 2017). Nevertheless, MA is predictive of reading comprehension over and above any mediated relationships (Levesque et al., 2017).

Assessing tests of MA

In this study, we present an evaluation of a set of MA assessments. Recall that our original goal was to find an assessment that we could use to identify children with poor comprehension that was specifically related to limited knowledge of morphology. To be

useful for that purpose, an MA assessment should clearly measure morphological knowledge. There are several ways to assess the specificity of an assessment. One is to demonstrate statistically that it captures variation other measures do not. Levesque et al. (2017) do this, and show quite conclusively that (their) MA assessments are measuring something distinct to other measures they used. While this is an important property of an assessment, this method only shows that the particular assessments used are distinct, not that they are accurately targeting an underlying construct. That requires a different, complementary, approach, one that is more logical in nature. We take this latter approach, asking whether MA assessments can logically be shown to specifically target morphological knowledge (relevant to comprehension, as this is what we are concerned with).² For it to be the case that MA assessments depend on the child understanding or knowing (at some level) that a word can be broken down into smaller units of meaning, two conditions should be met. First, it should not be possible to be correct without that knowledge, that is, other knowledge (such as vocabulary knowledge), should not be sufficient for correct performance. And second, having morphological knowledge should only lead to correct answers.

In what follows, we describe how we explored the properties of the assessments, what we found and finally what we think our findings mean for MA assessments more broadly, but we start by explaining the logic behind the two kinds of analyses we conducted.

One set of analyses was concerned with the potential for existing word knowledge to influence test performance. As described previously, a person can know a morphologically complex word without knowing that it *is* a morphologically complex word. Furthermore, even if a person does know that a word has component morphemes, they need not rely on that knowledge in the moment: they may be simply retrieving the whole word from memory (Hanna & Pulvermüller, 2014). The ages commonly examined in studies using MA assessments are ages at which children are learning increasing numbers of morphologically complex words, many of which seem to be known as wholes (Anglin, 1993), so it is not inconceivable that a test using known words could be measuring a child's vocabulary rather than their morphological knowledge. Thus, our first set of analyses examine occurrence of test words in language children are exposed to, asking whether each word *could* be known as a whole word or instead had to be understood via a process like morphological problem solving (Anglin, 1993), which entails morphological knowledge. We also examine frequency, as the more frequent the word is, the likelier it is retrieved as a whole (see, e.g. Schreuder & Baayen, 1995). For multiple choice (MC) tests, frequency is even more important: if the target is more frequent than the foils, a correct choice can reflect mere familiarity with the form (Underwood et al., 1972), that is, no understanding is necessary.

The other set of analyses was concerned with the question of whether incorrect responses could be taken as clear evidence for a *lack* of knowledge. Due to the idiosyncrasies of the English morphological system, there are often multiple ways of forming a word that should be possible, but which don't occur, for example, overregularisations ('runned' instead of 'ran'), adding additional derivational morphemes rather than removing them ('courageousness' instead of 'courage'), or using a derivational affix associated with a subset of words on a word outside that subset ('equalness' instead of 'equality'). These non-existent but logically possible words demonstrate productive knowledge of

the morphological patterns in English – productivity that would allow someone to understand the meaning of a novel (to them) word – but they would be marked as incorrect.

Method

Assessment selection

Our original goal was to find an existing MA assessment that would allow us to identify children with reading comprehension difficulties related to their morphological knowledge, so we looked for a test that had already been shown to correlate with reading performance. (Because of the nature of the studies looking at MA, reading performance generally meant comprehension.) Our search was not conducted using techniques related to meta-analyses or systematic reviews, which are designed to evaluate the findings in a set of papers collectively. Nor were we attempting to find a sample that was statistically representative of assessments. We were not interested in the papers themselves, we were only searching for popular or well-used assessments.

To do this, we performed a series of Google Scholar searches. First, we used the search terms ‘morphological awareness reading’. We then looked at papers listed in the first five pages of results that included typically developing, monolingual English-speaking child populations. Second, we used the search term ‘morphological awareness’ and restricted the results to papers since 2017, this time taking the first three pages of returned results, focusing on the same population (typically developing, monolingual English-speaking children). The motivation behind these two searches was to find the more popular/widely used assessments (papers with more citations show up earlier in Google Scholar searches), while recognising that it can take some time for papers to accumulate citations. These two searches were supplemented by similar searches in the University of British Columbia library general database (Summon). We then performed two additional searches in Google Scholar, one using the terms ‘morphological awareness reading bilingual’, the other using the terms ‘morphological awareness reading dyslexia’. We took the first three papers from each search which studied the effect of MA on reading in child populations. These two additional searches were included so as to not miss a very popular or widely used assessment that is used only in comparative studies.

Together, these searches yielded 89 papers, some of which were later removed because they turned out to not meet the initial search criteria (e.g. they were not on English, or they tested the effect of MA on spelling rather than reading). For the remaining papers, we then determined whether (1) the paper controlled for the effect of vocabulary knowledge on children’s reading ability, and (2) at least one of the studies in the paper exhibited a significant effect (after controlling for vocabulary) of at least one measure of MA on at least one measure of reading in at least one of the populations studied (e.g. there may have been two age groups in the study and MA only had a significant effect for one age group). These selection criteria were included to ensure that the MA assessment was specifically targeting morphological knowledge, not general vocabulary knowledge. Forty-five papers met these final two criteria.

For these 45 papers, we attempted to locate the MA assessments used. In some cases, assessments were included in the paper or appendices; in the remaining cases,

we contacted the corresponding authors to request the original assessment. At the end of this process, we had copies of MA assessments from 25 papers, which contained sufficient methodological detail to understand the test (e.g. sometimes we only had a list of words, which was insufficient for our purposes). These papers are listed in Table 1 of the Supplementary Materials. (Some papers used multiple assessments of different types – examples are only given from the assessments we examined. We did not analyse nonce words or compounds.)

There was a wide variety in the kinds of words and morphemes targeted by the tests: inflectional and derivational morphology, and regular (e.g. ‘weak-weakness’) and irregular (e.g. ‘strong-strength’) forms. They included tasks which involved combining morphemes to form morphologically complex words and decomposing complex words into component parts (usually by providing or identifying the base). Often different types were mixed within the same assessment (e.g. inflectional and derivational items were interspersed in Kruk & Bergman, 2013). (See Supplementary Materials, Table 1, for examples from each assessment.)

Assessments also varied as to whether they used MC questions or instead encouraged other kinds of responses (hereafter, non-MC assessments). A typical non-MC test provided a prompt word such as ‘strong’ and a cue (e.g. a written sentence with a line indicating a missing word), with participants instructed to insert a (morphologically complex) word based on the prompt word. For example, ‘strong: The machine tested Jimmy’s _____’ with ‘strength’ as the target answer. An MC test of the same item would present the child with several options, including ‘strength’ and ask the child to select the correct word. Analogies were also common in the non-MC tests. A child would be presented with a word pair, and asked to generate the correct second member for another word, for example, ‘walk/walked’ and ‘pick’/? Analogy items included sets in which the correct answer had the same phonological form as in the prompt: ‘poison/poisonous, nerve’/? (answer: ‘nervous’,) and items in which the prompt and test words required different morphemes: ‘mad/madness’ and ‘true’/? (answer: ‘truth’). Other tests presented the child with a simple word and a complex word containing the same string of sounds/letters and asked the child if the second (complex) word contained the first (e.g. ‘dirt’ and ‘dirty’). All words that were part of the assessments were extracted for further analysis. For the MC tests, this was the word given to the participant as well as the target and foils (the incorrect choices provided). For the non-MC tests, these were the target word (i.e. the intended response, or the word that the child was asked to assess in some way). Table 1 of the Supplementary Materials notes the type (MC/non-MC) of each assessment.

Corpus analyses

We cannot know whether any particular word is known to any particular child, but we can examine whether the target words are likely to have occurred in children’s input and so are potentially known. To assess this, we searched in four corpora for evidence of the target (and foil) words contained in the MA assessments. Counts of word occurrences were used to represent the potential for exposure to target words.

Table 1. Summary of the selected sections from the corpora.

Corpus	Number of sources	Publication dates	Number of words
COCA – Juvenile Fiction	2,042	1990–2017	2,984,664
COCA – Children’s Magazines	2,043	1990–2019	2,242,297
COCA – Family Movies	49	1990–2019	309,667
Cornell Movie–Dialogs Corpus – Family Movies	16	1939–1998	66,112
Children’s Picture Books	445	1885–2022	199,877
CHILDES (publication date = recording)	21	1962–2005	3,826,525

Data from a subset of the Corpus of Contemporary American English (COCA; Davies, 2008), the family movies in the Cornell Movie–Dialogs Corpus (Danescu-Niculescu-Mizil & Lee, 2011), a selection of children’s picture books known to be read to children under 3 years (Hudson Kam & Black, n.d.; Hudson Kam & Matthewson, 2017), and a selection of CHILDES corpora (MacWhinney, 2000) were analysed. These corpora were selected because they contain language that children are exposed to, and so, provide a more realistic picture of children’s potential knowledge than an adult-language based corpus would (e.g. the Enron email dataset or Common Crawl which is composed of resources from the internet and contains pornography).³

COCA (Davies, 2008-) contains approximately 1 billion words from texts and transcribed speech. We selected subgenres from this corpus based on the types of materials that children would likely have exposure to (juvenile fiction, children’s magazines and family movies). We used data from the family movies section of the Cornell Movie–Dialogs Corpus (Danescu-Niculescu-Mizil & Lee, 2011), which contains transcripts of movie conversations. The children’s picture book collection that we analysed contains 445 books all intended for young children (Hudson Kam & Black, n.d.). Table 1 provides the number of sources, publication dates and number of words for the sections of the various corpora that were used.

To count the words, the corpora were first divided into tokens. A token could contain any combination of the following symbols: letters, digits, hyphens and apostrophes. For counts of the words from the MA tests, an occurrence would be counted if the token exactly matched the word disregarding its case. This means that contractions and hyphenations of the word were excluded from these counts. The counts for each word were divided by the overall count for each corpus and multiplied by 1,000 to obtain the word frequency per 1,000 words.

Each corpus has different formatting. This was taken into account in the scripts to ensure consistent counting methods across corpora.⁴ For example, contractions in COCA (Davies, 2008-) are separated into two tokens (e.g. ‘can’t’ is ‘can n’t’), but they are not in the other corpora. In both cases, contractions were counted as one word. In addition, some of the tokens in the COCA texts were replaced with @ symbols by the distributors as a measure to discourage unauthorised redistribution of the corpus. These tokens were excluded from the counts because the underlying text was unknown.

For the CHILDES (MacWhinney, 2000) analyses, we used a subset of the English–NA corpora. Included corpora all had a target child aged 1–8 years who was typically developing, and naturalistic and spontaneous speech by people other than the target child. Files with some naturalistic and some non-naturalistic speech were included if we could easily restrict our analysis to the naturalistic situations (e.g. there were multiple files divided by task). In addition, the target child(ren) had to be verbal at some point in the data collection, this ruled out corpora with exclusively preverbal children. As we were interested in input, we excluded the children’s own productions. Speech produced by other children in the context was included, as that would be potential input. The list of corpora included in these analyses is presented in Table 2 of the Supplementary Materials.

Other possible responses

Incorrect responses should indicate that the child does not (yet) have an understanding of the morphological system of English. However, this will only be the case if the correct response is the *only possible* response. To get a sense of whether this was indeed the case, we looked at a subset of the non-MC assessments and the four MC assessments and asked whether there were any other morphologically possible responses that were not captured by the test’s target responses. One of the authors, a trained linguist, read the test questions and considered whether there were any other possible answers that could be considered correct by any pattern/rule of English morphology. To illustrate, in one of the examples above, the target answer is ‘strength’; however, a productive pattern (or rule) in English is the formation of a noun by adding ‘-ness’ to an adjective (e.g. ‘weak’ + ‘-ness’=the noun ‘weakness’). A child using this rule with ‘strong’ would say ‘strongness’, an answer that would reveal knowledge of the patterns of English morphology. This example is based on an exception to a regular pattern, but for other items the alternatives emerge from subsets of the language, each with their own pattern. As an illustration, in English, roots of Germanic origin and roots of Latin origin form nouns in different ways, only one of which is correct for that particular stem, for example, ‘equal’ and ‘equality’ not ‘equalness’ (Latin), but ‘natural’ and ‘naturalness’ not ‘naturality’ (Germanic). This analysis looked for these kinds of possible (but technically incorrect) responses. The other possibilities also included words or phrases that exist, but which weren’t designated as the target answer – for example, ‘Please. This weather is very _____’ target answer= ‘pleasant’, but ‘pleasing’ would also be semantically and grammatically possible – and answers that were possible on an alternative understanding of part of the item, e.g. ‘Perform. Tonight is the last _____’ target= ‘performance’, however ‘performer’ could also make sense. Note that the alternatives generated in this step are likely to be an underestimate of actual possibilities. Something similar was done for the MC assessments: items were examined to see if any of the foils could also potentially be correct according to the patterns in English.

Results and interpretation

Frequency analyses

As described earlier, even when the target answer is a multimorphemic form the child can know it as a whole. Thus, it is important to ask whether a target form is potentially

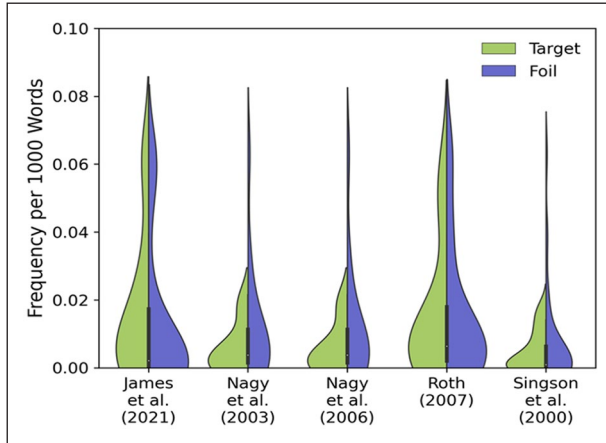


Figure 1. Frequency of target and foil words per 1,000 words by paper.

known. We ask this by examining whether the target forms are likely to have occurred in children's input. We present the results separately for assessments using MC and non-MC type tests, as the potential impact of frequency is different (and more straightforward) for MC tests than it is for tests using other measures.

Looking first at the MC tests, of the 72 total test items (used in the assessments in the five papers), only 5 (6.9%) of the target words did not occur in the corpora we searched. Thus, most of the target words are potentially known as whole lexical items. The foils, by contrast, were less likely to be known: 29 of the 154 foils (18.8%) did not occur in the corpora we searched. This imbalance exists even if we restrict our analysis to just speech used around young children (the CHILDES corpora, MacWhinney, 2000): 29.1% of targets versus 46.8% of foils don't occur.

That was only about occurring versus not, and collapsed the assessments into one analysis. But what about actual frequency? And what if we break it down by paper? Figure 1 shows the average occurrence (per 1,000 words, averaged across the four corpora) for each target and foil, separately for each of the five papers using MC tests that we examined. Target frequency, on average, differs from foil frequency in children's input ($U=4637.5$, $Z=1.98$, $p=.047$, $r=.01$). To understand which test sets may be impacted by this difference, we compared the target with foil frequency for each source (Table 2). There was a significant difference in frequency for the assessment in the James et al. (2021) paper, where the targets ($M=0.023$, $SD=0.0414$) had a higher frequency than the foils ($M=0.019$, $SD=0.0438$). This means that a child could succeed if all they were relying on was recognition memory for the words in the test, as the targets would have stronger memory traces than the foils.

To make matters more complicated, for at least some items, pre-existing knowledge of the words seems to be *necessary* for successful performance. Example 1 is from a test used by Singson et al. (2000). This test involves no morphological composition (or decomposition): the child is given a sentence with a missing word and asked to select between four possible related words, only one of which 'makes a good sentence' (p. 246):

Table 2. Two-tailed Mann–Whitney test results comparing target and foil frequencies for each paper.

Paper	<i>U</i>	<i>Z</i>	<i>p</i>	<i>r</i>
James et al. (2021)	330.5	2.13	.033	.03
Nagy et al. (2003)	26.0	1.00	.315	.05
Nagy et al. (2006)	26.0	1.00	.315	.05
Roth (2007)	192.0	0.22	.829	.01
Singson et al. (2000)	554.5	0.51	.612	.01

Ex. 1 Farmers _____ their fields. a) fertilize, b) fertilization, c) fertility, d) fertilizer

Presumably, the logic behind this as a test of *morphological* knowledge is the relationship between word forms and their syntactic category. To someone who knows the words already, ‘fertilize’ is the obvious correct choice and the other three options are clearly incorrect because the blank requires a verb and ‘fertilization’, ‘fertility’ and ‘fertilizer’ are nouns. This is supposed to be apparent due to their endings; however, there are English verbs ending in ‘-er’, for example, ‘administer’, ‘alter’, ‘anger’; ‘-y’, for example, ‘amplify’, ‘deploy’, ‘destroy’; ‘-ty’, for example, ‘dirty’ and even ‘-ity’, for example, ‘pity’. Given that there are only so many combinations of sounds available, and those that form morphemes often occur as non-morphemic sequences too, the nominal endings in the foils are only recognisable as such if the word is *already* known. The result of this is that it becomes a vocabulary test rather than a test of morphological knowledge.^{5,6}

The non-MC tests present the child with various tasks, but most give the child an input form and ask the child to produce a new word that is based on it. Sometimes these involve composition of morphologically complex forms, other times they involve decomposition. The triggers or guides for the target word vary. Often the child is given a word (the prompt) followed by a sentence missing a word and they are supposed to do something with the prompt word to complete the sentence, as in Example 2 (from Kruk & Bergman, 2013). Analogical tasks are also common: for example, the child is given a pair of words that are related in some way, and another word that they are supposed to present the appropriate pair for, as in Example 3 (this kind of task is used by Deacon et al., 2014, among others):

Ex. 2 ‘say walked’. They went for a long _____. (target: walk)

Ex. 3 dog/dogs person/ _____. (target: people)

Of the 661 non-MC test items we examined, only 7 (1.1%) of the target words did not occur in the corpora. If we restrict our analyses to input available to the youngest children (i.e. the CHILDES corpora, MacWhinney, 2000), the number of target words that don’t occur in the input rises to 123 (18.6%). Even the most restrictive analysis then shows that a large proportion of the target words are potentially known to school-aged children, since many may already be known to preschoolers.

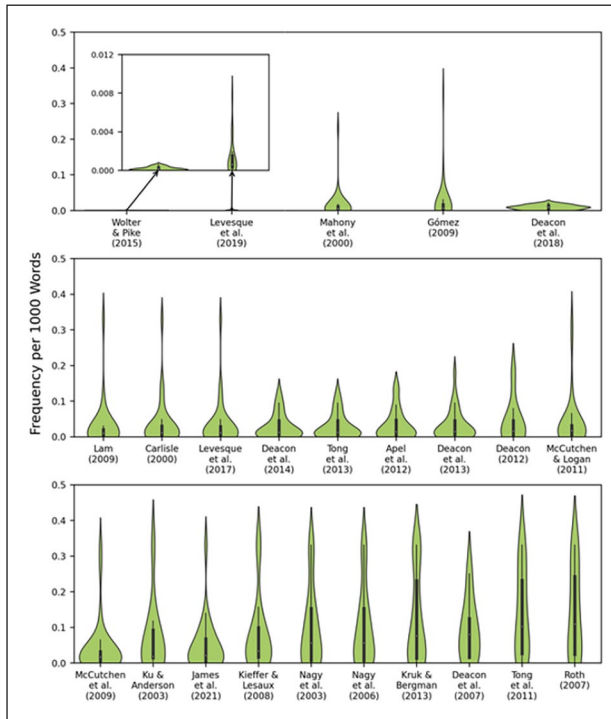


Figure 2. Frequency of target words per 1,000 words by paper.

As with the MC tests, here again it turns out that successful performance could, at least some of the time, rest on existing knowledge of the tested words. We can see this in the (very clever) analogy tests. They frequently involve irregulars as input or target, or changes in the actual morpheme used (e.g. pian-ist and paint-er), with the result that the child has to understand the meaning of the words in the test – they can't just blindly create a new word following the pattern in the prompt. However, they still do not depend on any knowledge of the *subcomponents* of the words involved. A child who knows the meanings of the whole words will know that they mean related things and can use this knowledge to 'get' what word they should be supplying – without any understanding of the underlying morphological system. Indeed, for items including irregulars, they *cannot* be answered correctly by a child who does not already know the irregular form.

Exactly how to think about word frequency for the target words in the non-MC tests is not straightforward, as there are no forms in 'competition' for selection, and so comparisons between input and target items make little sense. But it is still worth looking at average frequencies (per 1,000, averaged across the four corpora) for the target words, shown in Figure 2. Frequencies are shown for each target word, separately for each paper we drew (an) assessment(s) from. The figure is arranged according to the mean frequency of the words in each assessment, from left to right and top

to bottom. This is done to make the y-axes consistent within a panel, while still allowing readers to see the range within each assessment.

We can see from the figure that there is variation in the frequencies. Some assessments are more likely than others to be testing children on words they are likely to know already. The five papers shown at the top of the figure use the lowest frequency words. But even in these assessments most of the words do occur in the corpora we examined.

Other possible responses

This analysis looked for other potential responses (i.e. other things the child could have replied) that would have made sense given the general patterns of English morphology, but which for some reason or other turn out not to be correct. In total, 12 non-MC and 5 MC assessments were examined: all were found to have other descriptively correct answers for at least some items. For the MC tests, this usually affected only one or two items, but for the non-MC assessments, substantial portions of each assessment were found to have alternative possible answers. Example 4 is from Apel et al. (2012). In this item, the target or expected answer is ‘length’, but ‘longness’ would follow a more general regularity in English. In Example 5 (from Carlisle, 2000), the child is expected to do decomposition; however, the same meaning can be created by adding an additional derivational morpheme instead. Example 6, from McCutchen and Logan (2011), is especially interesting, as there is actually an existing word that fits the sentence. This latter type was not very common, but it did occur in more than one assessment. Example 7 is an MC item from Nagy et al. (2003). The expected response is ‘directions’, but ‘directing’ could also make sense (as a gerund, e.g. in a theatre), and ‘directed’ could even work (in reference to a person or group of people already established as being directed):

Ex. 4 long: He used the ruler to measure the table’s _____. target answer: length other possibility: longness

Ex 5 courageous: The man showed great _____. answer: courage other possibility: courageousness

Ex 6 examine: Lee went to the doctor for an _____. answer: examination other possibility: exam

Ex 7 He listened carefully to the _____. a) directing, b) directed, c) directions, d) directs.

Discussion

At this point we can ask whether MA tests measure what we would want them to measure? The answer is: maybe. That they measure something related to reading comprehension performance is clear. What exactly that is is not, however. It could be morphological knowledge but it need not be. On one hand, the tests all primarily use target words that are potentially known to children. The target words do vary in frequency between tests, and some tests use target words that are quite uncommon in the corpora we examined, making it less likely that the target words are actually known by many children prior to participation in the study. Other assessments use much more frequent words, indeed, one MC test included target words that were significantly more frequent than the foils, such

that a child could perform well on the test just by selecting the option that seemed more familiar to them, regardless of whether they understood the meanings. Thus, we cannot be sure that correct answers result from morphological knowledge, MA assessments could be testing more sophisticated vocabulary, or testing it in a different way, than the dedicated vocabulary assessments. It is difficult to say either way.

On the other hand, incorrect answers cannot always be assumed to indicate a lack of morphological knowledge. The analysis of other possible responses showed that all of the tests we looked at included items with other potential answers. Thus, children who knew something about the morphology of English could perform poorly despite their knowledge. Indeed, for the items where the correct answers are exceptions, rather than forms that follow the general morphological patterns, a child can only answer correctly if they already know the word in question, that is, correctness is a measure of existing vocabulary knowledge rather than knowledge of the systematic morphology. But it is knowledge of the *system* (or at least, some of its parts) that enables a child (or any reader) to understand a novel morphologically complex word.

Whether or not these results are important depends on the goal of the researcher. We are interested in assessing productive morphological knowledge of the sort that enables a child to interpret a word they don't already know. This requires a demonstration that a child can understand a complex word that they don't yet know or understand, something that may not be accurately captured by tests we looked at. The MC tests appear to depend too much on existing word knowledge. The same is true for most of the non-MC tests. Indeed, even if we could be sure that the non-MC assessments were tapping into morphological knowledge, because they require either production or some kind of explicit judgement, they depend on abilities, or at least a depth of knowledge, that go beyond what we ourselves are interested in assessing.⁷

However, some of the tasks could be perfectly appropriate for other researchers' goals. For a researcher who is interested in whether a child knows that two words are related, the non-MC tests that require the child to generate a response based on a prompt word can provide that evidence. And the analogy tests demonstrate that the child knows that both words share some aspects of meaning (in addition to phonological material), since they depend on understanding the meaning relationship between the input pair. Moreover, knowing that two different words can be related to each other *because* they share parts is not unimportant in emerging knowledge of morphology. For someone who is interested in a child's ability to explicitly analyse components, a quite different task would be required. But note that the abilities underlying these tasks would be related to reading performance in different ways, that is they are not simply measuring a child's ability to figure out the meaning of a novel morphologically complex word. Certainly, a child who can analyse and explain the parts of a complex word has the requisite morphological knowledge to understand novel words, but that level of sophistication isn't necessary.

We do not want to leave readers with the impression that the MA assessments we looked at are useless, they are not, they just cannot be said to conclusively target *productive* morphological knowledge. This leaves the question of what kind of task would have the properties we seek. To some degree that might depend on the age of the children being tested. Anglin (1993) shows that preschool-aged children do not know many

derived words, nor are they actively learning many during that time period. Thus, for younger ages real words might suffice. Once children are school-aged, however, the learning of derived words takes off, making real words less reliable as indicators of morphological knowledge (Anglin, 1993). At the same time, younger school-aged children are less likely to be able to figure out the meaning of a derived word by decomposing it explicitly. Taken together, these points suggest that a task that uses nonce words (e.g. a wug-type task) and MC responses (because they control the possibilities) might be the most appropriate for our purposes. Importantly, the nonce words would have to be introduced in a context that includes (e.g. using pictures) or at least suggests (e.g. syntactic context) a meaning and the test forms should change or extend the meaning, otherwise it would not clearly demonstrate that children could both recognise and understand the morphemes, two components of morphological knowledge required to interpret a novel word.

Final thoughts

We set out to find an assessment that had already been used to include in a project. In our search, we found some patterns that we felt were worth sharing with the broader research community, providing food for thought about what MA assessments really measure, and their relationship to what any particular researcher wants to understand in using them.

Author contributions

Carla L. Hudson Kam: Conceptualisation; Data curation; Funding acquisition; Methodology; Project administration; Supervision; Writing – original draft; Writing – review & editing.

Emily Sadlier-Brown: Data curation; Investigation; Writing – original draft; Writing – review & editing.

Shannon Clark: Data curation; Formal analysis; Project administration; Writing – review & editing.

Chelsea Jang: Data curation; Investigation; Writing – review & editing.

Carrie Demmans Epp: Formal analysis; Methodology; Project administration; Resources; Supervision; Writing – original draft; Writing – review & editing.

Jenny Thomson: Conceptualisation; Writing – review & editing.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This paper draws on research supported by the Social Sciences and Humanities Research Council (SSHRC PG ‘Ensuring Full Literacy in a Multi-cultural and Digital World’).

ORCID iD

Carla L. Hudson Kam  <https://orcid.org/0000-0002-7638-3656>

Supplemental material

Supplemental material for this article is available online.

Notes

1. Exactly what leads a learner to discover sublexical structure is not entirely clear, but it seems to be related to vocabulary structure and size (see Marchman & Bates, 1994).
2. Morphological awareness (MA) as defined includes awareness or conscious understanding (see, e.g. Carlisle, 2000; Tighe & Binder, 2015), however, as awareness is not necessary to understand a novel or unfamiliar word, we were not concerned with this aspect of MA.
3. An important caveat of these analyses is the fact that we were searching for strings. As a result, we may have overcounted some items due to homographs, for example, 'still' corresponds to an adverb and an adjective which are quite different in meaning. Also note that some of the papers used the same tests, or almost the same tests. We did not exclude duplicates because we are presenting data by paper, so duplicates do not bias the results.
4. Scripts can be accessed via GitHub: <https://github.com/EdTeKLA/MorphologicalAwarenessWordFrequencies>.
5. All four forms, the target and foils, occur in the corpora, although the target is the least frequent of the four.
6. There were also (infrequent) questions that could be answered using only world knowledge, for example, 'Fourth. Two plus two equals _____' target: four (from Kieffer & Lesaux, 2008).
7. It might seem that production requires deeper knowledge than comprehension. However, correct production can occur in the absence of comprehension (Papafragou et al., 2007).

References

- Anglin, J. M. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, 58(10), v–165.
- Apel, K., Wilson-Fowler, E. B., Brimo, D., & Perrin, N. A. (2012). Metalinguistic contributions to reading and spelling in second and third grade students. *Reading and Writing: An Interdisciplinary Journal*, 25(6), 1283–1305.
- Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language and Cognitive Processes*, 12, 507–584.
- Berko, J. (1958). The child's learning of English morphology. *Word*, 14, 150–177.
- Carlisle, J. F. (2000). Awareness of the structure and meaning of morphologically complex words: Impact on reading. *Reading and Writing*, 12(3), 169–190.
- Danescu-Niculescu-Mizil, C., & Lee, L. (2011). *Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs*. Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, 76–87, Portland, Oregon.
- Davies, M. (2008-). *The corpus of contemporary American English (COCA)*. <https://www.english-corpora.org/faq.asp>
- Deacon, S. H. (2012). Sounds, letters and meanings: The independent influences of phonological, morphological and orthographic skills on early word reading accuracy. *Journal of Research in Reading*, 35(4), 456–475.
- Deacon, S. H., Benere, J., & Pasquarella, A. (2013). Reciprocal relationship: Children's morphological awareness and their reading accuracy across grades 2 to 3. *Developmental Psychology*, 49(6), 1113–1126.

- Deacon, S. H., Holliman, A. J., Dobson, G. J., & Harrison, E. C. J. (2018). Assessing direct contributions of morphological awareness and prosodic sensitivity to children's word reading and reading comprehension. *Scientific Studies of Reading, 22*(6), 527–534.
- Deacon, S. H., Kieffer, M. J., & Laroche, A. (2014). The relation between morphological awareness and reading comprehension: Evidence from mediation and longitudinal models. *Scientific Studies of Reading, 18*, 432–451.
- Deacon, S. H., Wade-Woolley, L., & Kirby, J. (2007). Crossover: The role of morphological awareness in French immersion children's reading. *Developmental Psychology, 43*(3), 732–746. <https://doi.org/10.1037/0012-1649.43.3.732>
- Gómez, G. E. R. (2009). *The role of morphological awareness in bilingual children's first and second language vocabulary and reading* [PhD dissertation]. University of Toronto.
- Gough, P., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6–10.
- Hanna, J., & Pulvermüller, F. (2014). Neurophysiological evidence for whole form retrieval of complex derived words: A mismatch negativity study. *Frontiers in Human Neuroscience, 8*, Article 886. <https://www.frontiersin.org/articles/10.3389/fnhum.2014.00886>
- Hudson Kam, C. L., & Black, A. (n.d.). *Examining possible differences in book reading input to girls versus boys*.
- Hudson Kam, C. L., & Matthewson, L. (2017). Introducing the infant bookreading database (IBDb). *Journal of Child Language, 44*, 1289–1308.
- James, E., Currie, N. K., Tong, S. X., & Cain, K. (2021). The relations between morphological awareness and reading comprehension in beginner readers to young adolescents. *Journal of Research in Reading, 44*(1), 110–130.
- Kieffer, M., & Lesaux, N. (2008). The role of morphology in the reading comprehension of Spanish-speaking English Language Learners. *Reading and Writing, 21*, 783–804. <https://doi.org/10.1007/s11145-007-9092-8>
- Kruk, R. S., & Bergman, K. (2013). The reciprocal relations between morphological processes and reading. *Journal of Experimental Child Psychology, 114*(1), 10–34.
- Ku, Y. M., & Anderson, R. C. (2003). Development of morphological awareness in Chinese and English. *Reading and Writing, 16*(5), 399–422.
- Lam, K. Y. Y. (2009). *The effects of morphological awareness on reading in Chinese and English among young Chinese children: A longitudinal study* [MA thesis]. University of Toronto.
- Levesque, K. C., Kieffer, M. J., & Deacon, S. H. (2017). Morphological awareness and reading comprehension: Examining mediating factors. *Journal of Experimental Child Psychology, 160*, 1–20.
- 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.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.
- Mahony, D., Singson, M., & Mann, V. (2000). Reading ability and sensitivity to morphological relations. *Reading and Writing, 12*, 191–218.
- Marchman, V., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21*, 339–366.
- McCutchen, D., & Logan, B. (2011). Inside incidental word learning: Children's strategic use of morphological information to infer word meanings. *Reading Research Quarterly, 46*, 334–349.

- McCutchen, D., Logan, B., & Biangardi-Orpe, U. (2009). Making meaning: Children's sensitivity to morphological information during word reading. *Reading Research Quarterly, 44*, 360–376.
- Nagy, W., Berninger, V. W., & Abbott, R. D. (2006). Contributions of morphology beyond phonology to literacy outcomes of upper elementary and middle-school students. *Journal of Educational Psychology, 98*(1), 134–147.
- Nagy, W., Berninger, V. W., Abbott, R. D., Vaughan, K., & Vermeulen, K. (2003). Relationship of morphology and other language skills to literacy skills in at-risk second-grade readers and at-risk fourth-grade writers. *Journal of Educational Psychology, 95*(4), 730–742.
- Papafragou, A., Li, P., Choi, Y., & Han, C-H. (2007). Evidentiality in language and cognition. *Cognition, 103*, 253–299.
- Roth, L. M. (2007). *Orthographic and morphological processing as predictors of reading achievement in grade 3 children* [dissertation]. Library and Archives Canada.
- Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131–154). Lawrence Erlbaum Associates.
- Singson, M., Mahony, D., & Mann, V. (2000). The relation between reading ability and morphological skills: Evidence from derivational suffixes. *Reading and Writing, 12*, 219–252.
- Spencer, A. (2017). Morphophonological operations. In A. Spencer & A. M. Zwicky (Eds.), *The handbook of morphology* (pp. 123–143). Portico. <https://doi.org/10.1002/9781405166348.ch6>
- Tighe, E. L., & Binder, K. S. (2015). An investigation of morphological awareness and processing in adults with low literacy. *Applied Psycholinguistics, 36*(2), 245–273. <https://doi.org/10.1017/S0142716413000222>
- Tong, X., Deacon, S. H., & Cain, K. (2013). Morphological and syntactic awareness in poor comprehenders: Another piece of the puzzle. *Journal of Learning Disabilities, 47*(1), 22–33.
- Tong, X., Deacon, S. H., Kirby, J. R., Cain, K., & Parrila, R. (2011). Morphological awareness: A key to understanding poor reading comprehension in English. *Journal of Educational Psychology, 103*(3), 523–534.
- Underwood, B. J., Patterson, M., & Freund, J. S. (1972). Recognition and number of incorrect alternatives presented during learning. *Journal of Educational Psychology, 63*, 1–7.
- Wolter, J. A., & Pike, K. (2015). Dynamic assessment of morphological awareness and third-grade literacy success. *Language, Speech, and Hearing Services in Schools, 46*(2), 112–126.