INTERNATIONAL LITERACY ASSOCIATION READING RESEARCH QUARTERLY

Dimensionality of Morphological Knowledge—Evidence from Norwegian Third Graders

Jarl K. Kristensen

Björn Andersson

Centre for Educational Measurement, University of Oslo, Oslo, Norway

Siri S. Bratlie

Department of Education, University of Oslo, Oslo, Norway

Janne V. K. Torkildsen

Department of Special Needs Education, University of Oslo, Oslo, Norway

ABSTRACT

This study aimed to determine the dimensionality of morphological knowledge by examining different sources of variance. According to the Morphological Pathways Framework (*Levesque et al., Journal of Research in Reading, 44, 10-26, 2021*), morphological awareness, morphological analysis and morphological decoding are related, but distinct dimensions of morphological knowledge. However, multidimensionality might also stem from construct-irrelevant variance due to methodological artifacts. We assessed 612 Norwegian third graders on five measures of morphological knowledge and one measure of general vocabulary. Fitting a series of confirmatory factor analysis (CFA) models, we evaluated the dimensionality of morphological knowledge both within and across the five tests. Furthermore, we fitted three structural equation models (SEMs) to explore how different conceptualizations affect the relationship between morphological knowledge and general vocabulary: a five-factor model, a bifactor model, and a higher-order model representing morphological awareness, morphological analysis and morphological decoding. CFAs supported a multidimensional view of morphological knowledge and highlighted the need to account for construct-irrelevant variance. SEM analyses further illustrated that construct-irrelevant variance introduces a confounding element to the relations between morphological knowledge and vocabulary in the test-specific five-factor model, as only the bifactor and higher-order models separate between construct-relevant variance and variance due to methodological artifacts. The bifactor model is useful for separating sources of variance, especially during test development. For research purposes, however, we recommend conceptualizing morphological knowledge in line with *Levesque et al., Journal of Research in Reading, 44, 10-26, 2021*, to increase knowledge of morphological dimensions and their relations to other areas of literacy.

Introduction

In this study, we investigate the dimensionality of morphological knowledge in Norwegian third graders. More specifically, we examine whether tests measuring morphological awareness, morphological analysis and morphological decoding represent a single underlying construct or different dimensions of morphological knowledge. Understanding the dimensionality of the construct is crucial to advance research on morphological knowledge and its relations to other language skills. If the construct is multidimensional, we need to take this into account when comparing results from studies using different measures of morphological knowledge. While it is not a target of the current study, dimensionality may also have implications for the design and evaluation of morphological assessments and interventions.

INTERNATIONAL LITERACY ASSOCIATION

Reading Research Quarterly, 58(3)
pp. 406–424 | doi:10.1002/rrq.497

Morphological knowledge is the ability to recognize, understand, manipulate and produce spoken and written morphemes, the smallest meaning-bearing units of language. It requires knowledge of both the form and meaning of morphemes, as well as the processes through which they can be combined (Nagy et al., 2014). In addition to the term morphological knowledge, the two related terms morphological awareness (e.g., Carlisle, 2010) and morphological processing (e.g., Verhoeven & Perfetti, 2011) are widely used. Morphological awareness refers to explicit morphological knowledge, as it requires conscious reflection on and manipulation of morphemes (Levesque et al., 2021). Morphological processing, on the other hand, refers to the implicit use of morphological knowledge, which may happen at a subconscious level (Bowers et al., 2010; Nagy et al., 2014).

Levesque et al. (2021) introduce the Morphological Pathways Framework. The framework provides a theoretical foundation for morphological knowledge, in which the authors conceptualize it as a multidimensional construct. However, the findings in extant empirical research are mixed. Some studies report evidence of a single dimension of morphological knowledge (e.g., James et al., 2021; Spencer et al., 2015), whereas others propose different dimensions of morphological knowledge such as oral versus written or receptive versus productive (e.g., Jong & Jung, 2015; Tibi & Kirby, 2017). Thus, it is unclear if the theoretical framework suggested by Levesque et al. (2021) is generally applicable across different populations and settings.

Additionally, most studies on the dimensionality of morphological knowledge to date have been conducted in English. Of the 13 studies reviewed in this paper, eight featured English-speaking participants (see Table S1 and the section on empirical studies of dimensionality for more information). However, in their study of English and Korean, Jong and Jung (2015) found evidence of cross-linguistic differences. In English, they found one receptive and one productive dimension, whereas in Korean they found one receptive and two productive dimensions. While this points to possible cross-linguistic differences in dimensionality, it is not clear whether these differences relate to morphological knowledge or a construct-irrelevant source of variance. It is also unclear if such differences exist between languages more closely related than English and Korean.

Dimensionality studies in other areas of language have suggested a developmental trend moving from a single factor that captures language competence in preschool to multidimensional representations in older children (Tomblin & Zhang, 2006). While the existing studies span age ranges from preschool to adulthood, there is a need for further examination of the dimensionality of morphological knowledge in younger children. For morphological tasks which rely on written language, the development from a single factor to multidimensional representations may be affected by the orthographic transparency of the language. Specifically, morphological decoding and analysis may be distinguishable earlier in more orthographically transparent languages where decoding skills place a severe constraint on analysis for only a limited developmental period. Considering the potential impacts of language and age on dimensionality, we aim to add to the current knowledge by examining the construct in Norwegian third graders.

Norwegian Language and Morphology

In many languages, morphology plays an important role in word formation through inflection, derivation and compounding (Gonnerman, 2018), as is also the case in Norwegian. Inflection modifies a word's grammatical features, such as tense (hoppe—hopp*et*, "jump—jump*ed*"), number (blomst—blomst*er*, "flower—flower*s*") or grammatical gender (et fin*t* hus, "a-*neuter* nice-*neuter* house"). Derivation, on the other hand, creates entirely new dictionary words (lexemes), which can change a word's part of speech (spise—spise*lig*, "eat—edi*ble*") and often result in a derived word with a completely different meaning than the base word (tanke—tanke*løs*, "thought—thought*less*"). Compounding also creates new words but does so by joining two bases (sove*rom*, "bedroom"; korrektur*lese*, "proofread") rather than joining a base and a derivational affix.

Norwegian is a Germanic language with a simple verbal morphology (no subject–verb agreement), but a more complex nominal morphology, including three grammatical genders and noun–adjective agreement. Both compounding and the compilation of derivational affixes are widely used as means of word formation in Norwegian. Compounding is a highly productive process in Norwegian, and, thus, closed compounds that can consist of three, four or even more base words are common (e.g., *menneskerettighetsorganisasjon* = human rights organization). Words with three or more derivational affixes are also common (e.g., *u-be-hjelpe-lig* = helpless). A number of the Norwegian derivational affixes are similar to those found in English (e.g., "over-" and "mis-"). Many of the Norwegian derivational affixes and compounds are typical of written language and are, thus, particularly relevant for comprehending and producing academic texts. The Norwegian orthography is morpho-phonetic, and the phoneme–grapheme relationships are more transparent than in English (Seymour et al., 2003). There is a persistent influence of morphology on Norwegian orthography (Lyster, 2002), and morphological features determine the spellings of many words, along with phoneme–grapheme correspondence. Additionally, many high-frequency inflectional and derivational suffixes contain silent letters (e.g., the neuter definiteness marker "-et" /e/ and the common derivation "-lig" /li:/).

The literature on morphological development in Norwegian is scarce and focuses on acquisition of inflections in preschool children (for an overview, see Ribu et al., 2019). One study of past tense acquisition included children up to early primary school age and found that the overwhelming majority of children have reached ceiling performance by age eight (Ragnarsdóttir et al., 1999). Derivational knowledge has only been examined in one study, which showed that for 5-year-olds the mean performance in a derivational task was substantially lower than the performance in similar tasks measuring inflectional knowledge (Grande, 2018). This result supports the common pattern found in studies of many Indo-European languages that inflectional knowledge is typically acquired earlier than derivational knowledge (Kuo & Anderson, 2006).

In sum, there is a need for studies of the acquisition of derivations and compounds in Norwegian. It is especially important to study these morphological skills in children from third grade. Most Norwegian children are skilled decoders by that age (Hagtvet et al., 2006), and consequently, curriculum texts become more complex, including advanced vocabulary with derived and compounded words. The current study, thus, focused on derivational and compound knowledge in Norwegian third graders. Measures of inflectional knowledge were not included, as previous studies indicate near-ceiling performance for nominal inflections before age 3 and for verbal inflections by age 8.

Dimensionality of Morphological Knowledge

Theoretically, the dimensionality of morphological knowledge depends on whether construct-relevant variance relates to one or more morphological skills. However, construct-irrelevant variance might also be a source of multidimensionality, which stems from specific task requirements, formats or content (e.g., Deacon et al., 2008). Hence, tests of morphological knowledge may measure a multidimensional construct of which only one dimension relates to morphology. In the following sections, we review the Morphological Pathways Framework as a theoretical foundation for understanding construct-relevant dimensionality, present potential construct-irrelevant sources of variance, and summarize findings from previous empirical studies on the dimensionality of morphological knowledge.

The Morphological Pathways Framework

A large body of research has shown that morphological knowledge predicts vocabulary, reading fluency and reading comprehension in many languages (James et al., 2021; Manolitsis et al., 2019; McBride-Chang et al., 2005), and that morphological instruction can enhance children's word knowledge and reading development (e.g., Bowers et al., 2010; Carlisle, 2010; Goodwin & Ahn, 2010; Lyster et al., 2016; Reed, 2008; Torkildsen et al., 2022). The Morphological Pathways Framework introduced by Levesque et al. (2021) provides a theoretical model of the mechanisms behind the influence of morphology on other areas of literacy. Furthermore, it provides a theoretical base for viewing morphological knowledge as multidimensional.

In this framework, the authors present three dimensions of morphological knowledge which influence reading comprehension and writing: morphological awareness, morphological decoding and morphological analysis. Morphological awareness is viewed as a metalinguistic skill that involves the conscious reflection on and manipulation of morphemes. Morphological decoding relates to morpho-orthographic segmentation, that is, the recognition of separate morphemes in written words. This is also referred to as form-based skills, as they operate at the level of orthography, or word form (Levesque et al., 2021; Nagy et al., 2014; Torkildsen et al., 2022). Morphological analysis is a morpho-semantic process and involves the recognition of the meaning of separate morphemes within words. This process operates at the level of semantics and is also referred to as meaning-based skills (Levesque et al., 2021; Nagy et al., 2014; Torkildsen et al., 2022). The Morphological Pathways Framework posits reciprocal relations among morphological awareness, morphological decoding and morphological analysis. These three skills represent related, yet distinct, dimensions of the overarching construct of morphological knowledge.

The framework involves different pathways between morphological awareness, morphological analysis and morphological decoding, and other areas of language including text comprehension and generation. Along these paths, we also find connections to word reading, spelling and word knowledge. Morphological awareness is related to knowledge of word meanings through morphological analysis, thus affecting general vocabulary. Specifically, morphological analysis can support inferences about the meanings of morphologically complex words through the meanings of their constituent morphemes (Levesque et al., 2019). Morphological decoding forms the bond between morphological awareness and word reading by enabling letter-sound mapping at the level of morphemes rather than graphemes (Levesque et al., 2021). The relation to spelling is still somewhat unclear, as little research exists in this area. According to Levesque et al. (2021), it is possible that both morphological decoding and morphological analysis are involved.

While the Morphological Pathways Framework provides a theoretical basis for the multidimensionality of morphological knowledge, it is also evident from the literature that researchers measure morphological knowledge with a large number of different tasks with different task requirements (Berthiaume et al., 2018). These requirements are methodological artifacts that introduce construct-irrelevant variance. Hence, they represent confounding factors in research.

Methodological Artifacts

Morphological tasks vary in input and output modality, content, task type, as well as demands on information processing and prior knowledge (Berthiaume et al., 2018; Deacon et al., 2008). Input and output modality concerns whether tasks presentation (input) or responses (output) are oral or written. Berthiaume et al. (2018) describe 10 different task types that are commonly used to measure morphological knowledge: decomposition, definition, lexical decision, derivation, morphological relation judgment, naming, plausibility judgment, spelling, suffix choice, and word analogy. These involve different knowledge demands. Knowledge demands relate to the distinction between awareness and processing. Some tasks, like word analogies, require explicit morphological awareness. Other tasks may rely on implicit morphological processing, for example, word explanations, where morphological analysis may operate at a subconscious level. Finally, tasks may tap into different additional skills such as phonological decoding or general vocabulary, giving rise to construct-irrelevant variance in item responses.

To sum up, multidimensionality in morphological measures can stem from "true" multidimensionality in morphological knowledge. On the other hand, it may also stem from construct-irrelevant variance due to methodological artifacts.

Empirical Studies on the Dimensionality of Morphological Knowledge

Many studies on language and language development utilize measures of morphological knowledge, but few have investigated the dimensionality of the construct explicitly (Goodwin et al., 2017). For the current study, a literature review yielded 13 papers that examined the dimensionality of morphological knowledge (see Table S1 in the Supplementary material for a detailed overview). To evaluate the dimensionality of morphological knowledge, these studies implement a range of statistical models, including single-factor models, correlated traits models, and bifactor models. A single-factor model implies that a single skillset of morphological knowledge underlies test performance. In a correlated traits model, subsets of items or indicators tap into different, correlated factors. Finally, a bifactor model implies that a general factor of morphological knowledge explains the correlation among all items in a test, while there are also specific uncorrelated factors that account for residual correlations among the item scores in separate subtests, beyond what the general factor can explain.

Some previous studies found evidence supporting a unidimensional conceptualization of morphological knowledge (James et al., 2021; Muse, 2005; Spencer et al., 2015; Tibi, 2016; Tibi & Kirby, 2017). Note that Spencer et al. (2015) report analyses of the same data as Muse (2005), and Tibi and Kirby (2017) report on the same data as Tibi (2016). Findings from these studies suggest that morphological knowledge is best represented as a single skillset. Although contrary to the Morphological Pathways Framework at first glance, these findings could relate to the reciprocal nature of morphological awareness, morphological analysis and morphological decoding. Some measures of morphological knowledge may not sufficiently distinguish between the three skills, and different models may provide acceptable fit to the data. For example, the written tasks in Muse (2005) and Spencer et al. (2015) were read aloud by the test administrator and did not require written responses. Thus, they did not test morphological decoding specifically. The written tests of Tibi (2016) did require participants to read, and in some tasks write the answer, thus measuring morphological decoding. Accordingly, Tibi and Kirby (2017), in an extension of the analyses of the data from Tibi (2016), found that a two-factor model also represented the data well. The two factors were related to the oral and written tests, respectively, thus aligning with the theoretical constructs of morphological analysis and morphological decoding.

Other studies have found support for a multidimensional structure of morphological knowledge (González-Sánchez et al., 2018; Jong & Jung, 2015; Levesque et al., 2017; Tighe & Schatschneider, 2015, 2016; Zhang, 2017). Both González-Sánchez et al. (2018) and Jong and Jung (2015) reported separate dimensions of receptive and productive morphological knowledge in their studies. Their studies targeted Spanish children in the last year of preschool (González-Sánchez et al., 2018) and Korean fifth and sixth graders (Jong & Jung, 2015). Tighe and Schatschneider (2015, 2016) studied morphological knowledge in English-speaking Adult Basic Education students and found evidence that a two-factor model separating real words and pseudowords fit the data best. A common finding in all these studies is that response format is a source of multidimensionality. This is not related to morphological knowledge as such, but rather to how we measure it and the additional skills required to respond. This might indicate that potential differences relating to age and language stem from construct-irrelevant sources rather than differences in morphological knowledge.

Levesque et al. (2017) examined morphological knowledge in English-speaking third graders. They measured the theoretically founded skills of morphological awareness, morphological decoding, and morphological analysis. Comparing unidimensional and multidimensional models, they found that a model representing each skill as a separate factor fit the data best. Zhang (2017) found similar results in a study of the morphological knowledge of Singaporean fourth graders speaking both Chinese and English. A two-dimensional model aligning with the theoretical constructs of morphological analysis and morphological decoding fit the data well. These studies, along with Tibi and Kirby (2017), provide support for the theoretical dimensions introduced in the Morphological Pathways Framework.

Goodwin et al. (2017) administered seven morphological tasks to English-speaking seventh and eighth graders. The authors found evidence that a bifactor model performed best, meaning that the tasks measured a general factor of morphological knowledge, as well as seven specific factors related to each of the seven types of tasks. This bifactor model of morphological knowledge was further explored by Goodwin et al. (2021), in which they reported that morphological knowledge was best represented by four skill-related (general) factors as well as task-specific factors. The four general factors align with morphological awareness, morphological analysis and morphological decoding, with the addition of a factor representing morphological-syntactic knowledge. Following an inherent assumption in bifactor models, however, the four factors representing morphological skills are uncorrelated, not taking into account the relations posited in the Morphological Pathways Framework.

When considering a structural model where general vocabulary and reading comprehension were regressed on each factor in the bifactor model for morphology, Goodwin et al. (2017) found that the general factor of morphological knowledge explained most of the variance in both vocabulary and reading comprehension. However, additional variance in reading comprehension was explained by the specific factors of morphological meaning (positive), and morphological spelling and word reading (negative). Additional variance in vocabulary was explained by morphological meaning and word generation (positive), and spelling (negative). These results suggest that general and task-specific morphological skills may have a distinct involvement in different literacy tasks.

While there are differences between studies, there are no consistent patterns relating to language or age. Some differences relate to test format, and in the studies that support a unidimensional view, the tests do not necessarily separate between theoretically founded dimensions. Importantly, the differences underline the importance of separating construct-relevant variance from variance that does not relate to morphological knowledge.

Current Study

The purpose of our study is to investigate the dimensionality of morphological knowledge in Norwegian third graders. We examine whether a unitary construct of morphological knowledge underlies five tests that measure different aspects of the participants' knowledge of morphologically complex words: receptive word knowledge, productive word knowledge, word analogies, spelling and word reading fluency. Furthermore, we examine how different conceptualizations affect the relation between morphological factors and general vocabulary.

This study builds on data from a randomized controlled trial (RCT) of a morphological intervention with Norwegian second graders who were followed until third grade (Torkildsen et al., 2022). Participating students were randomly assigned to an eight-week digital morphology program or an active control group. The program consisted of 40 training sessions targeting derivational morphology (26 common derivational morphemes) and compounding processes in Norwegian. The training targeted both morphological decoding and morphological analysis. In line with this, we developed our five outcome measures to tap both of these constructs, in addition to morphological awareness. Specifically, the tests of receptive and productive word knowledge measure morphological analysis, while the spelling and word reading fluency tests measure morphological decoding. The word analogy test measures morphological awareness. For more information, see the test descriptions in the methods section.

As previous studies have provided evidence both for unidimensionality and multidimensionality, we compare several different models that may represent the construct based on these previous findings. As a part of this investigation, the bifactor analyses of Goodwin et al. (2017) are considered for a new age group and a new language. We also include a higher-order model to examine the dimensional structure suggested by Levesque et al. (2021), including a mediation model similar to those examined by Levesque et al. (2017). Both the bifactor framework and the Morphological Pathways Framework hold promise of producing a deeper understanding of this complex area of language, yet few studies have implemented them to date. Hence, we examine these frameworks in the context of our study, to provide further evidence on their applicability when measuring and analyzing morphological knowledge.

Our study was guided by the following three research questions:

1. Do the five tests of morphological knowledge each measure a unidimensional construct?
2. Is morphological knowledge best represented as a unidimensional or multidimensional construct across the five different tests?
3. How do different models affect the relation between morphological knowledge and general vocabulary?

For research question 1, we hypothesized that each test captures a unidimensional facet of morphological knowledge. Some of the tests, however, include items with features that may influence the measured construct. The test of receptive word knowledge measures morphological knowledge in context as well as in isolation and consists of three different item types (see the methods section for details). Tighe and Schatschneider (2015) examined context versus no context as potential dimensions of morphological knowledge in adults. Although their results did not support this dimensional dichotomy, these might constitute separate dimensions in children. Additionally, the different item types might pose different task demands, thus reflecting different dimensions. The Test of Productive Word Knowledge measures morphological knowledge with real words and pseudowords. Although Jong and Jung (2015) did not find evidence of a real word versus pseudoword division in children, Tighe and Schatschneider (2015, 2016) did find evidence for separate dimensions in adults. Lastly, both the spelling test and the tests of productive and receptive word knowledge measure each specific affix in more than one task. Thus, there is a possibility that the tests of receptive and productive word knowledge may be best represented as multidimensional. Although not representing theoretical dimensions of morphological knowledge, the affix-specific knowledge may cause dependence among items beyond the common variance due to morphological knowledge.

Regarding research question 2, we hypothesized that a common construct of morphological knowledge underlies item responses across all five tests. This could align with studies that support morphological knowledge as a unidimensional construct (James et al., 2021; Muse, 2005; Spencer et al., 2015; Tibi, 2016; Tibi & Kirby, 2017). However, the tests also differ in task demands (e.g., comprehension, production, analogies, reading fluency, and writing). Thus, we hypothesized that the tests may measure other test-specific skills as well as the common factor of morphological knowledge. Additionally, two of our morphological tests measure morphological decoding, two tests measure morphological analysis skills, and one test measures morphological awareness. Hence, we examine whether morphological awareness, morphological decoding and morphological analysis are separate dimensions of morphological knowledge, in line with Levesque et al. (2021).

Finally, with regard to research question 3, the literature points towards strong relationships between morphological knowledge and vocabulary (e.g., McBride-Chang et al., 2005; Nagy et al., 2006). Hence, we hypothesized that potential dimensions of morphological knowledge should have significant positive relations to general vocabulary. However, the results of Goodwin et al. (2017) suggested that these relations are different if morphological knowledge is accounted for in a general factor (i.e., in a bifactor model). Thus, for the bifactor model, we hypothesized that general morphological knowledge, as well as specific receptive and productive word knowledge (morphological analysis) have a significant and positive relation to general vocabulary (which is measured by a meaning-based definition task), whereas the specific skills related to word analogies, reading fluency and spelling (morphological awareness and decoding) have non-significant or negative relationships to general vocabulary, in accordance with Goodwin et al. (2017). In line with Levesque et al. (2021), we expected a significant and positive relation between morphological analysis and vocabulary, as well as an indirect effect of morphological awareness through analysis, in the mediation model.

Methods

Design

The participants in the intervention study were tested before starting the program (pre-test), directly after the program (post-test), and at follow-up, which was approximately 9 months after the pre-test. The current study analyzes data from the follow-up, which was administered during the participants' first semester in the third grade. The decision to use the data from third grade was made to include a word analogy test, which was only administered at this grade level, as a measure of morphological awareness. This enabled us to examine as many potential theoretical and empirical dimensions of morphological knowledge as possible.

Participants

The participants were 612 third graders (n = 325 girls, n = 286 boys, and n = 1 with missing information) from 12 schools in the eastern part of Norway. The approximate mean age was 8.34 years (SD = 0.3). All students in each classroom were invited to participate, with a positive response rate of 93%. Schools were recruited by municipality officials, who were instructed to select schools with different characteristics (average SES and proportion of language minority students) to help make the sample representative of schools in the area. The morphology training program required that schools had access to iPads for all children participating; hence the schools were not randomly selected. Across schools, the proportion of mothers with a university education ranged from 28.3% to 95.7% (mean for the whole sample = 72.9%), and the proportion of students with a language minority background ranged from 2.8% to 93.6% (mean for the whole sample = 28.8%). Ethical approval to conduct the study was granted by the Norwegian Centre for Research Data.

Measures

All measures were administered as part of a larger test battery, either individually or in groups (full classes). As there is a lack of standardized tests of morphological knowledge in Norwegian, these five tests of morphological knowledge were developed within the project. All tests were piloted in several rounds with approximately 200 children who did not participate in the current study. As mentioned, these tests were selected to measure learning outcomes in the intervention study, not primarily to assess dimensionality. However, we include information on how the measures relate to theoretical and empirical perspectives on dimensionality in the description of each test. Table 1 provides an overview of the measures, including Cronbach's α (ranging from .80 to .96).

The derivations targeted in the intervention program were selected based on frequency information from language corpora, utility and familiarity from a pilot study of 100 fourth graders. The fourth graders' knowledge of 96 derivations was rated on a scale from 0–2 where 0 indicated no knowledge, 1 indicated some knowledge (often highly specific), and 2 indicated more advanced general knowledge. To ensure that the derivations were not only already mastered by most second graders but also not too advanced for them, we selected 26 derivations in which 40–70% of fourth graders demonstrated at least some knowledge. For more information on morpheme selection, see Torkildsen et al. (2022). All the words used in the morphological measures were multimorphemic (e.g., consisting of an affix and a base word). Half of the test items in measures 2–5 contained *exposed words* (i.e., words that were included in the app training sessions) and half of the test items contained *unexposed words* (i.e., words that were not included in the training, but which contained trained affixes). The word analogy test did not contain any exposed words.

Morphological Awareness

Morphological awareness was measured with the word analogy test, in line with Levesque et al. (2017). The test focuses on extracting the bases of derived words, that is, words which are made up of a derivational affix and a base (for an example, see measure 1 in Table 1). This requires knowledge of morpheme boundaries and segmentation. As both presentation and response are given orally, the test does not rely on morphological decoding. The test, adapted from Brinchmann et al. (2016) and Bryant et al. (1997), consists of 15 items. The test administrators first presented a derived word containing a given affix and extracted the base from the derived word. Then another derived word containing the same affix was presented, and the children were prompted to extract the base. The test was administered individually, and item scores were binary (0, 1).

Morphological Analysis

Morphological analysis was measured with two tests that focus on the meaning of words, similarly to Levesque et al. (2017) and Goodwin et al. (2021). The test of receptive word knowledge (see measure 2 in Table 1) measures comprehension of morphologically complex words and consists of 48 multiple choice items covering 20 affixes (each appearing in 2 tasks with different base words), 6 compound words and 2 words with multiple affixes. The

TABLE 1
Overview of the Tests of Morphological Knowledge and General Vocabulary

Measure	Task example(s)	Items (final)	Cronbach's α
1) Word analogy test	"I say the word *typical*, then change it to *type*. We can also change the word *magical* to …"	15 (14)	.80
2) Test of receptive word knowledge	*Word combination*: "Which part can you put after *de-* to make a real word? [*sit, pict, shake, song*] *Cloze tasks*: "Janne wanted to stay in the student council. She hoped for a (…)*election*. [*re, new, well, after*]" *Picture tasks*: "Press the picture that shows *overexertion*."	48 (26)	.85
3) Test of productive word knowledge	*Real words*: "What does *machinist* mean?" *Pseudowords*: "What could *busist* have meant, if it were/had been a real word?"	18 (13)	.80
4) Spelling test	"It was a happy *reunion*. Write *reunion*."	24 (24)	.92
5) Word reading efficiency test	Timed reading of randomized lists of morphologically complex words (i.e., not sorted by difficulty). 30 second time limit.	4 (3)*	.96
6) WISC-IV vocabulary subscale	"What is a thief?"	3 (3)**	.84

Note. All examples are translated from Norwegian. Cronbach's α reported for final item sets. *Four lists of 48 words each (sum scores), all four used in the within-test model, three of the lists retained in the across-tests models. **Three parcels of 12 items each (sum scores).

test was administered digitally. Tasks were presented orally and in writing, in a multiple-choice format with one correct option and three distractor options. The tasks were divided into three different types: morpheme choice tasks, cloze tasks and picture tasks. In the morpheme choice tasks participants were asked to match an affix with a base to form a real word. In the picture selection tasks, participants identified the most appropriate picture in response to a morphologically complex word. The cloze tasks required participants to select an affix to fill a blank in a sentence. Examples of the three task types are given in Table 1. While cloze and picture tasks provided context through the sentences and pictures, the morpheme choice tasks did not. The test was administered in group sessions. Item scores were binary (0, 1).

The second measure of morphological analysis was the Test of Productive Word Knowledge (see measure 3 in Table 1). The test measures the ability to define morphologically complex words and pseudowords. The test covers six affixes, with three items for each affix and, thus, a total of 18 items. Each affix was presented as part of a real word in two of the tasks and as part of a pseudoword in the third task. Pseudowords were created by adding an affix to a regular Norwegian base, creating a nonexistent but plausible word (e.g., *bussist* = busist, which could mean "a person who drives/rides/likes buses"). The test was administered individually, with oral presentation and oral responses. Partial scoring in three categories was used (0, 1, 2). Two points were awarded for synonyms or precise explanations of the meaning of a word and one point was awarded for definitions that reflected only vague knowledge of the word's meaning. Pseudoword explanations were scored for knowledge of what an affix does to the meaning of a word.

Morphological Decoding

Morphological decoding was measured with two tests focusing on the written form of words, in line with Levesque et al. (2017) and Goodwin et al. (2021). The spelling test (see measure 4 in Table 1) measures the ability to spell morphologically complex words with nontransparent spelling patterns. The test consists of 24 morphologically complex words covering 11 derivational affixes, each included in two items, and two items with compound words. The words were first presented in the context of a sentence and then repeated in isolation. The children were then asked to write the target word of each item on a sheet of paper. The test was administered in groups and partial scoring was used. 0–3 points were given for words with derivations (1 point, respectively, for the correct spelling of the affix, the correct spelling of the base word, and writing the morphemes together with no space between, following Norwegian orthographic rules) and 0–2 points for compound words (1 point respectively for correct spelling of the base words and writing of the compounds together with no space between).

The second measure of morphological decoding was the word reading efficiency test (see measure 5 in Table 1), which measures word reading fluency and accuracy. It consists of four lists, each containing 48 morphologically complex words, covering both derivations and compound words. The children were asked to read as many words aloud from each list as they could in 30 seconds. Children were instructed to read the words in the order they were presented, but if unable to read an attempted word, children could skip to the next word on the list. The test was administered individually and sum scores from each list were used for analyses. Note that while there are 192 items across the four lists, we only have four sum score indicators. No items were excluded in the process.

General Vocabulary (Word Definitions)

Using word definitions as a proxy for general vocabulary, we measured this construct with the Vocabulary subtest from the Norwegian 2009 version Wechsler Intelligence Scale for Children® Fourth Edition (WISC-IV; Wechsler, 2009). See measure 6 in Table 1 for an example. This test measures the ability to explain the meaning of words. It consists of 36 items. The test was administered individually, according to the manual. As specified in the manual, the test was discontinued after five consecutive errors. Items were parceled into three sum scores which were used as indicators of general vocabulary in the analyses.

Analyses

Our analyses consisted of three distinct parts, which reflected research questions 1, 2, and 3, respectively. All analyses were conducted in R (R Core Team, 2020), using the packages psych (Revelle, 2021) for descriptive statistics and lavaan (Rosseel, 2012) for the factor analyses and structural equation modeling. The proportion of missing data ranged from 5% to 9% for the models within tests. For the models across tests, the proportion was 4%. Models were estimated based on the observed pairwise information between pairs of variables to minimize the loss of information due to missing responses. We compared this procedure to listwise deletion, and there was no impact on any conclusions of the study.

Descriptive Statistics

Table 2 provides descriptive statistics for total scores on each test. As the data come from an RCT study, we show the statistics for the experimental and control groups separately. The table reports on both the pre-test in second grade and the follow-up in third grade, which was the measurement point in focus in the current study. For item-level statistics, see Tables S2–S7 in the supplementary materials. The patterns of means, standard deviances,

TABLE 2
Descriptive Statistics (Sum Scores) for All Measures at the Pre-Test in Second Grade and Follow-Up in Third Grade

Test	Group	*n* P/F	*M* P/F	*SD* P/F	Range P/F	Skewness P/F	Kurtosis P/F
Word analogy	E	NA/290	NA/7.97	NA/3.52	NA/14	NA/−0.68	NA/−0.27
	C	NA/292	NA/7.89	NA/3.31	NA/15	NA/−0.48	NA/−0.27
Receptive knowledge	E	308/276	17.38/22.98	5.30/8.31	33/41	0.69/0.07	0.62/−0.66
	C	282/278	16.80/21.43	5.54/7.54	33/46	0.78/−0.23	0.73/−0.29
Productive knowledge	E	298/291	11.83/17.74	6.10/6.51	28/30	0.16/−0.25	−0.80/−0.54
	C	287/291	11.53/15.91	5.97/6.39	29/28	0.24/−0.01	−0.28/−071
Spelling	E	306/293	47.95/54.37	11.14/9.21	68/68	−1.34/−1.30	3.27/3.33
	C	297/288	47.88/52.59	11.17/10.20	68/70	−1.60/−1.51	4.22/4.20
Word reading	E	298/291	21.43/39.28	15.61/25.07	97/152	1.35/0.94	2.82/1.03
	C	286/290	22.34/37.65	18.28/25.63	156/183	2.39/1.30	11.29/3.20
Vocabulary	E	298/291	16.83/19.25	5.14/5.71	31/33	0.18/0.37	0.12/−0.08
	C	287/290	17.24/18.87	5.28/5.74	36/36	0.38/0.33	0.98/0.22

Note. C = control group, E = experimental group, F = follow-up, P = pre-test. MCWA was administered at follow-up only.

skewness and kurtosis are similar across the groups, with the exception of the kurtosis of the word reading efficiency test in the control group at pre-test. The experimental group had larger increases in means overall than the control group from the pre-test to the follow-up.

We tested measurement invariance between the experimental and control groups on all exposed items of the tests. These are the test items that contain morpheme combinations that the experimental group has experienced through the tasks in the intervention. As mentioned, the word analogy test does not contain any exposed items. We used chi-square difference tests to test the null hypotheses of measurement invariance against lack of measurement invariance with a Bonferroni-corrected significance level of 0.0125. The results indicated that all exposed items functioned equally for participants in the experimental group and the control group (see Table 3).

Research Question 1

Research question 1 concerned the overall item quality and dimensionality in each of the tests separately. This study provides the first in-depth psychometric evaluation of the tests. Hence, we went through several steps before arriving at the final models. In the first step, correlations between item scores and total scores for each test were calculated, and items with $r < .3$ were excluded from further analyses (Nunnally & Bernstein, 1994). This resulted in the exclusion of five of the 48 items in the test of receptive word knowledge and one of the 15 items in the word analogy test. These items represented noise, likely due to too high difficulty and unintended item features. For example, one item in the test of receptive word knowledge had the target "løsbart". This word can mean either "solvable" (correct response) or "false mustache" in Norwegian, and the only difference lies in the pronunciation. The response options were pictures, of which one could be mistaken to depict a false mustache. Hence, a large number of children chose the confounding distractor. Removing these items did not change the substantive interpretation of the underlying constructs, nor did they change the possible dimensional structures of either test that were evaluated in the subsequent analyses.

TABLE 3
Measurement Invariance Tests

Model	x^2	*Df*	Δx^2	Δdf	*p*
Receptive (p)	505.663	618			
Receptive (f)	615.547	648	47.431	30	.023
Productive (p)	131.710	134			
Productive (f)	169.073	149	26.843	15	.030
Spelling (p)	719.084	526			
Spelling (f)	826.630	572	60.488	46	.074
Word reading (p)	17.739	8			
Word reading (f)	17.750	10	0.011	2	.994

Note. (p) = partial invariance, exposed items free to vary. (f) = full invariance, all items restricted.

In the second step, we considered unidimensional confirmatory factor analysis (CFA) models for each test and evaluated the model fit. Note that for the spelling test and the test of productive word knowledge, the models were specified with correlated residuals between items containing the same affix. We used polychoric correlations with the diagonally weighted least squares (DWLS) estimator for the ordinal data and the ML estimator for continuous data. All the unidimensional models fit the data well, indicating that a single construct underlies responses to each test. As there was a very large amount of indicators across the tests, we decided to exclude items with standardized factor loadings that were below .4 from further analyses. This choice was made to reduce complexity and facilitate the analyses across tests. Note that this item exclusion was carried out after establishing unidimensionality for each test. As cutoff values for considering a factor loading salient vary in the literature (e.g., Brown, 2015), a strict cutoff was deliberately chosen to reduce the vast amount of indicators in the final models containing all tests. This second analysis step resulted in the exclusion of another 17 items from the test of receptive word knowledge (in addition to the five items excluded in step 1), retaining 26 items. From the Test of Productive Word Knowledge, we removed five items, keeping 13 items. We did not exclude any further items from the word analogy test. The spelling test and word reading efficiency test were also kept intact, as there was no factor loading < .4 in the models for these tests. Note, however, that we excluded one indicator from the word reading efficiency test at a later stage, outlined in the next section. The number of items, original and final, are reported in Table 1. The exclusion of items, though substantial, did not affect the substantive or statistical interpretations of the constructs. It did, however, increase the fit indices and coefficient alphas to some extent. For the sake of brevity and continuity, we present the results for the final models based on the reduced item sets in the next chapter, as these are the item sets we use in the subsequent analyses across tests.

Research Question 2

To address research question 2, models containing the retained items from all tests were evaluated to examine dimensionality across the measures. We fit a series of nested CFA models: one-factor (morphological knowledge); three-factor (morphological awareness, morphological analysis, and morphological decoding); five-factor (test-specific); and higher-order (morphological awareness, morphological analysis and morphological decoding). Note that in the higher-order model, morphological awareness is represented as a first-order factor, since we tested this construct with a single test. Figures S1–S4 show conceptual illustrations of these models. The observed indicators from the word reading efficiency test were very highly correlated (ranging from .86 to .90), which caused empirical underidentification in the initial analyses across tests. Thus, to estimate the models, we removed one variable. Because of the high correlations, this did not change the substantive interpretation of the Word Reading factor, nor its contribution to the models across tests. We used chi-square difference tests to select among the models, with a significance level of 0.05. In the last step of the measurement models, we fit a bifactor model to unravel the common and specific variance of the measures (conceptual illustration in Figure S5). We used polyserial correlations and the DWLS estimator in estimation since we had a combination or ordinal and continuous item scores (Olsson et al., 1982). When assessing model fit, we focus primarily on the SRMR. Most data in our analyses are ordinal, and recent studies have suggested that the SRMR is more appropriate to use than fit statistics such as the RMSEA when analyzing ordinal observed variables (e.g., Shi et al., 2020). A value of the SRMR lower than 0.08 indicated a good model fit (Hu & Bentler, 1999). For the bifactor model, we assessed the dominance of the general factor by the explained common variance (Rodriguez et al., 2016).

Research Question 3

To address research question 3, we fitted three structural equation models (SEMs). The first was based on the five-factor model, with general vocabulary regressed on each of the factors. The second model was a mediation model based on the higher-order model. In line with the Morphological Pathways Framework, we specified direct paths from morphological awareness to morphological analysis, morphological decoding and vocabulary, as well as indirect paths from morphological awareness to vocabulary through analysis and decoding. The third model was based on the bifactor model where vocabulary was regressed on the general factor and each of the specific factors.

Results

Individual Test Models (Research Question 1)

Item-level descriptive statistics are listed in Tables S2–S7 in the Supplementary materials. Table 4 displays the fit statistics of the individual unidimensional models for each test. The chi-square tests of model fit were significant for all models except Productive Word Knowledge, but this was not unexpected given the large sample. The SRMR values were all < .08, indicating a good model fit and supporting the hypotheses of unidimensionality within the tests. Recall, however, that we specified the Productive Word Knowledge and spelling models with residual correlations

TABLE 4
Model Fit for Individual Models of Morphological Tests

	n	x^2 (robust)	*Df*	*p*	SRMR	RMSEA	CFI	TLI
Word analogy	580	89.669	64	.019	.055	.026	.990	.986
Receptive knowledge	554	363.697	299	.006	.058	.020	.984	.982
Productive knowledge	582	62.919	56	.245	.036	.015	.997	.996
Spelling	586	621.186	240	< .001	.055	.052	.956	.949
Word reading	583	17.451	2	< .001	.006	.115	.995	.986

among items containing the same affix. This indicates that there is some multidimensionality in the form of shared affix-specific variance within these models.

Models Across Tests (Research Question 2)

Fit statistics for the models that included all morphological tests are reported in Table 5. All models showed significant chi-square values. Again, this was not unexpected due to the large sample size. The unidimensional model provided the least good fit to the data, with SRMR = .083 exceeding the recommended cut-off value of .08. The less restricted models all provided a good fit (see Table 5). To compare the model fit further, we conducted chi-square difference tests for the one-factor, three-factor, five-factor, and higher-order models. The one-factor model is the only model that can be compared directly with the bifactor model using a chi-square difference test, as the other models are not nested in the bifactor model (e.g., Mansolf & Reise, 2017). Hence, we conducted a separate chi-square difference test for the one-factor and bifactor models. The hypothesis testing procedure showed that the five-factor model had a superior fit compared with the other models and that the bifactor model was preferred to the one-factor model. The results of all chi-square difference tests are reported in Table 5 (see the last three columns).

The five-factor model fit the data well (SRMR = .061, see Table 5 for further fit indices). The standardized factor loadings in the five-factor model ranged from .373 to .987. For a list of all standardized factor loadings for the five-factor model, see Table S8. Although the five-factor model pointed to a multidimensional construct, the factor correlations shown in Table 6 were quite high overall. This may indicate that the tests capture common variance across all the tests as well as the specific variance related to each separate test.

While the five-factor model fit significantly better in the model comparison, the higher-order model fit the data equally well in terms of SRMR (SRMR = .061, additional fit statistics in Table 5). The standardized factor loadings for the first-order factors ranged from .372 to .987 (see Table S9 for a complete list). For the second-order factors, the standardized loadings on analysis were .918 (productive) and .919 (receptive), and the loadings on decoding were .717 (word reading) and .968 (spelling). The

TABLE 5
Fit Statistics and Model Comparisons for Models across Tests

CFA	x^2 (robust)	Df	$p(x^2)$	SRMR	RMSEA	CFI	TLI	Δx^2	Δdf	$p(\Delta x^2)$
5-factor	3645.271	3037	< .001	.061	.018	.971	.970			
HO	3659.751	3040	< .001	.061	.019	.970	.969	11.468	3	< .01
3-factor	3836.307	3044	< .001	.063	.021	.962	.960	84.082	4	< .001
1-factor	5298.151	3047	< .001	.083	.035	.892	.888	295.470	3	< .001
Bifactor	3828.046	2967	< .001	.062	.022	.959	.956			
1-factor								1222.400	80	< .001
SEM										
5-factor	3916.832	3272	< .001	.060	.018	.970	.969			
HO	4017.891	3278	< .001	.061	.020	.966	.964			
Bifactor	4082.400	3201	< .001	.061	.022	.959	.957			

Note. Δ x2 is based on standard x2 values, not robust.

TABLE 6
Factor Correlations

Five-factor model				
	Receptive	Productive	Analogy	Spelling
Receptive	1			
Productive	.844	1		
Analogy	.646	.687	1	
Spelling	.629	.622	.658	1
Word reading	.457	.404	.542	.694
Higher-order model				
	Awareness	Analysis		
Awareness	1			
Analysis	.721	1		
Decoding	.693	.696		

Note. All correlations are significant at $p < .001$.

correlations among awareness, analysis, and decoding were medium to high, in line with the reciprocal relations suggested in the Morphological Pathways Framework (see Table 6).

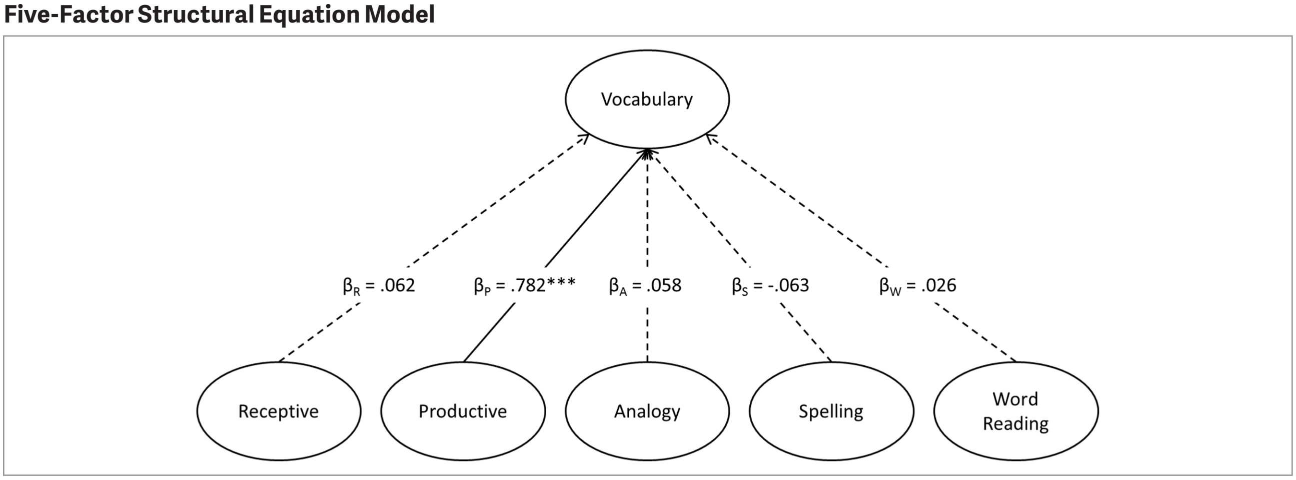
Within the framework of these models, however, it is not possible to examine the common and specific variances of the tests simultaneously. Hence, in accordance with Goodwin et al. (2017), we proceeded with a bifactor model to illuminate the construct of morphological knowledge further. The bifactor model fit the data well (SRMR = .062, see Table 5 for all fit indices). The chi-square difference test showed that the bifactor model fit significantly better than the one-factor model. All factor loadings on the general factor were significant, ranging between .254 and .689. For the specific factors, however, two indicators showed non-significant factor loadings. For a complete list of factor loadings for the bifactor model, see Table S10 in the supplementary material. Seventy-nine percent of the factor loadings on the general factor were ≥ .4, indicating a high degree of overlap between items from the different tests. This overlap was also spread out among the tests, so no test showed less overlap than others. The estimated explained common variance was .63. This value indicates that both the general and specific factors contribute to explaining the variance in the indicators (Rodriguez et al., 2016).

In sum, the results of our analyses clearly favored a multidimensional view of morphological knowledge. However, there was substantial ambiguity regarding how this multidimensionality should be represented. While the five-factor model provided the best fit among the nested models, the factors may be contaminated by substantial amounts of construct-irrelevant variance. The higher-order and bifactor models also provide an excellent fit and can help us separate the construct-irrelevant variance from variance related to morphological knowledge. To further disentangle dimensionality of morphological knowledge, we chose to proceed with all three models in the final part of our analyses.

Structural Equation Models (Research Question 3)

In the final part of our analyses, we expanded each model to a SEM. In these models, general vocabulary, as measured by the WISC-IV Vocabulary subtest, was regressed on each of the morphological factors. The goal of these analyses was not to investigate the relationship between morphological knowledge and vocabulary per se, but rather to demonstrate what kind of information the measurement models can provide. The five-factor SEM (Figure 1) fit the data well (SRMR = .060, see Table 5 for other

FIGURE 1
Five-Factor Structural Equation Model



Note. Indicators are left out for readability. ***$p < .001$

fit statistics). Inspecting the standardized regression coefficients, Productive Word Knowledge was the only factor with a significant relation to general vocabulary ($\beta_P = .782$, $p < .001$). Note that the predictors in this model were substantially correlated, which inflates the standard errors of the estimated regression coefficients. Thus, the model provides little information concerning the relations of morphological factors to vocabulary.

The higher-order SEM (Figure 2) fit the data well (SRMR = .061, see Table 5). In line with the Morphological Pathways Framework, morphological awareness directly affected both morphological analysis ($\beta_{a1} = .845$, $p < .001$) and morphological decoding ($\beta_{b1} = .799$, $p < .001$). Morphological analysis was also related to vocabulary ($\beta_{a2} = 1.002$, $p < .001$). There were no direct effects of morphological decoding or morphological awareness on vocabulary, nor any indirect effect of awareness through decoding. There was, however, a significant indirect effect of awareness through analysis ($\beta_{a1*a2} = .846$, $p < .001$). Since there was no direct effect of morphological awareness on vocabulary, the relation between them was fully mediated through morphological analysis. This provides additional support for the theoretical relations of the Morphological Pathways Framework.

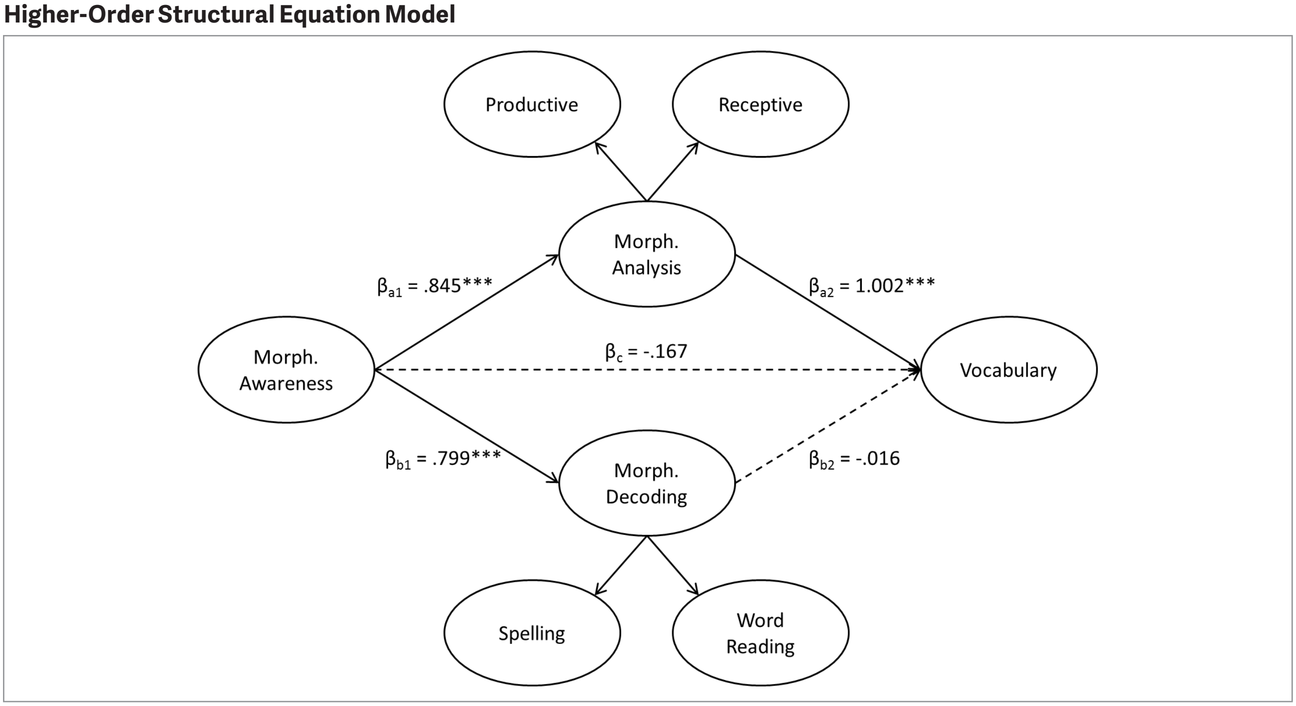
Finally, the bifactor SEM (Figure 3) also fit the data well (SRMR = .061, see Table 5). In this model, the general factor of morphological knowledge had the strongest relation to general vocabulary ($\beta_G = .664$, $p < .001$), followed by the specific productive factor ($\beta_{SP} = .527$, $p = .001$; see Figure 3). The specific receptive factor ($\beta_{SR} = .339$, $p = .003$) was also positively associated with general vocabulary, whereas the relation to the specific word reading factor ($\beta_{SW} = -.136$, $p = .010$) was negative. The specific factors of spelling ($\beta_{SS} = -.049$, $p = .280$) and analogy ($\beta_{SA} = .121$, $p = .077$) were not significantly related to general vocabulary. The results were similar to those found by Goodwin et al. (2017) with the exception of reading and spelling. In their study, reading was not significantly related to vocabulary, whereas spelling had a negative relation.

To sum up, the five-factor model, while empirically sound, provided little information about morphological knowledge and its relation to vocabulary. The higher-order model provided more information, particularly about the relations between the morphological constructs. It does not, however, allow us to investigate the specific variance of the first-order factors related to morphological decoding and analysis. Finally, the bifactor model provided information about construct-relevant and construct-irrelevant variance of morphological knowledge and allows us to investigate the specific variance within tests, as well as the common variance related to morphological knowledge.

Discussion

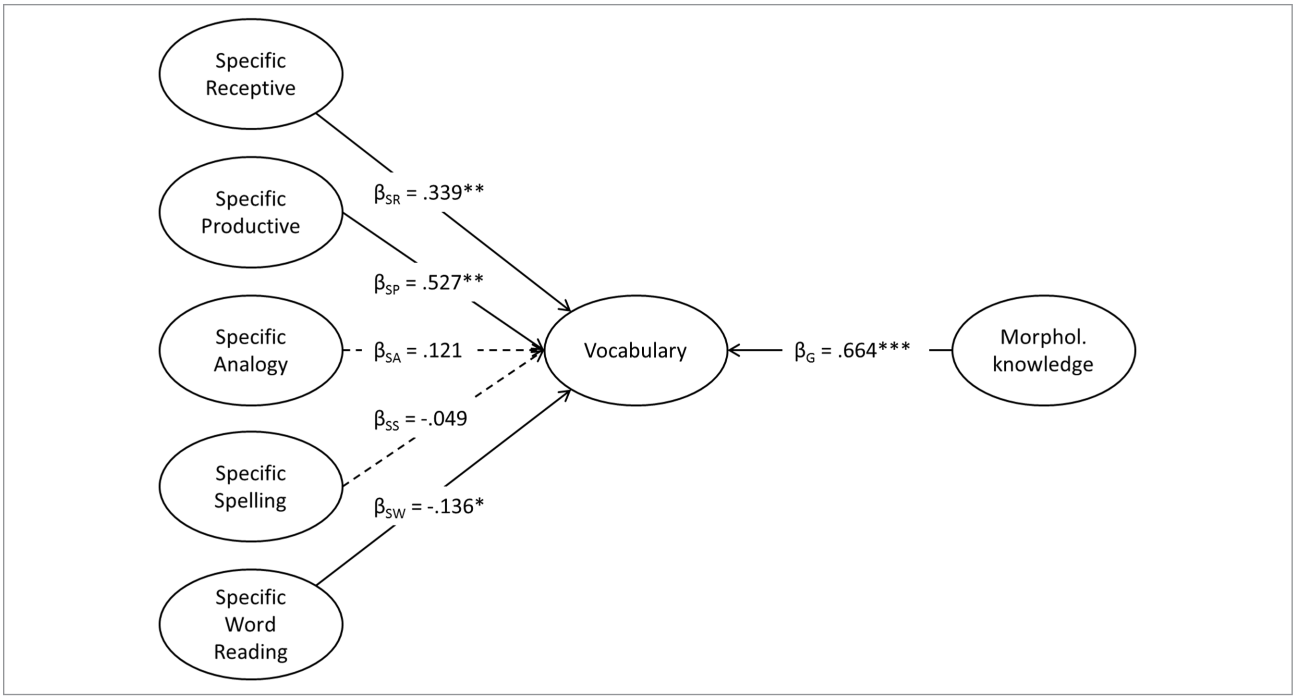
The present study evaluated the dimensionality of morphological knowledge in Norwegian third graders, both within and between tests that require different skills in addition to morphological knowledge. Moreover, the study investigated how different conceptualizations affect the

FIGURE 2
Higher-Order Structural Equation Model



Note. Indicators are left out for readability. ***$p < .001$

FIGURE 3
Bifactor Structural Equation Model



Note. Indicators are left out for readability. *$p < .05$, **$p < .01$, ***$p < .001$

relations between morphological dimensions and general vocabulary. We examined test-related dimensions, theoretical dimensions (morphological awareness, morphological analysis and morphological decoding), and general and specific dimensions relating to construct-relevant and construct-irrelevant variance. Our results show that each of the five tests measures a unidimensional construct. When analyzed together, the tests are best represented as multidimensional. The findings from the measurement models alone, however, are ambiguous as to whether a five-factor, higher-order, or bifactor model is most appropriate. When general vocabulary is regressed on each factor in the models, the different models imply different relationships between the morphological factors and general vocabulary. Below we discuss the findings related to each of our three research questions in turn.

RQ1: Are the Constructs Measured by the Separate Tests Unidimensional?

The individual test models support unidimensionality within tests. This means that the potential dimensions related to the lexical status, contextual cues and item types are not supported by our analyses within tests. Although previous studies investigated these dimensions across different tests, there was the possibility that subsets of items in our test could function as different subtests. The test of receptive word knowledge contains both tasks with context and tasks without context. In line with Tighe and Schatschneider (2015), we found no evidence for separate contextual dimensions in our sample, indicating similarities across age groups and languages in this regard. Neither did we find any evidence for separate dimensions relating to the three item types in the receptive test. In the Test of Productive Word Knowledge, we found no evidence of dimensionality relating to real words versus pseudowords. This is contrary to the findings of Tighe and Schatschneider (2015, 2016), but in line with Jong and Jung (2015), perhaps indicating that children are more inclined to accept pseudowords in line with real words than adults. The test of productive word knowledge and the spelling test did however require correlated residuals of items measuring the same affix, indicating some dimensionality related to specific affix knowledge. In the productive test, the children are asked to provide definitions of words. For example, if a child knows that *alveaktig* (elflike) means “similar to an elf”, they would also likely be able to infer that the pseudoword *høneaktig* (henlike) could mean “similar to a hen”. Hence, it is not surprising that the residuals are correlated for items containing the same affix. Similarly, in the spelling test, if a child knows that *endelig* (*final*) is spelled with a silent g at the end, they are probably more likely to remember the silent g in *fredelig* (peaceful).

It is perhaps not surprising that each test measures a unidimensional construct, with the exception of the correlated residuals, in the productive test and the spelling test. Although the items in our analyses vary in lexical status (words vs. nonwords) in the productive test, and contextual cues (the presence vs. absence of a linguistic/image context) and item types in the receptive test, the same specific task demands are posed within each test. This might point to the task demands having a greater impact on dimensionality than item characteristics within a test. Nevertheless, this step was important to establish unidimensionality within tests and avoid potential confounding in further analyses.

RQ2: Is Morphological Knowledge Best Represented as a Unidimensional or Multidimensional Construct Across the Different Tests?

Considering the models incorporating all tests, the results of this study do not support a strictly unidimensional construct of performance on different tests of morphological knowledge. This indicates that using a single measure of morphological knowledge, whether in assessment or research, could impart an incomplete picture of children's morphological skills, at least in Norwegian. Moreover, morphological knowledge may be confounded with other skills such as decoding or general vocabulary, making claims of the effect of morphological knowledge uncertain. Hence, morphological knowledge should be measured across different tests that allow us to separate the common variance attributable to morphological knowledge from the specific variance due to other skill requirements inherent in the tests. This can provide a deeper understanding of the morphological knowledge and enhance comparisons across studies.

Our results indicate that a five-factor model fits the data very well, and significantly better than the three-factor and higher-order models, similar to the findings of Goodwin et al. (2017). This provides evidence of similarities in English and Norwegian, and across primary and middle school. Our finding that receptive and productive knowledge make up two of these factors is also in line with González-Sánchez et al. (2018) and Jong and Jung (2015), indicating similarities with Spanish children in preschool as well as with Korean fifth and sixth graders. This separation of receptive and productive knowledge may, however, represent construct-irrelevant variance due to differences between general comprehension and language production, rather than separate dimensions of morphological knowledge. A critical drawback of the five-factor model is that it does not separate construct-relevant and irrelevant variance. Thus, multidimensionality could be a consequence of tests measuring other skills in addition to morphological knowledge. While more research is needed to strengthen any conclusions, one potential source of multidimensionality is the methodological artifacts inherent in the set of tests. This could explain some of the differences found across studies thus far, since different tasks may pose different demands of both morphological knowledge and other linguistic skills. Furthermore, these demands may vary across age groups and languages, potentially explaining why some studies find evidence of unidimensionality and others of multidimensionality. Another drawback is that a test-specific conceptualization of morphological knowledge implies that every test measures a separate morphological dimension, thus disabling comparisons of results from studies using different measures of morphological knowledge. These drawbacks make a correlated traits model like the five-factor model an ill-advised choice for research.

Although the five-factor model provides a closer fit than the higher-order model in terms of chi-square difference, the latter also fits the data very well with an equal value of SRMR. This model is theoretically founded in the Morphological Pathways Framework (Levesque et al., 2021), and provides similar results to those found by Levesque et al. (2017), Zhang (2017), and in parts by Tibi and Kirby (2017). This provides further evidence of similarities rather than discrepancies across languages and age groups. A key benefit of using a higher-order model rather than a three-factor (correlated factors) model to represent the morphological dimensions is that it allows us to separate out the construct-relevant variance in the second-order factors. In this respect, the addition of a second measure of morphological awareness would have strengthened our model. One drawback of the model is that we cannot investigate the construct-irrelevant variance directly to assess additional sources of variance within tests. It does, however, enable us to examine the relations between morphological dimensions according to the Morphological Pathways Framework.

In the bifactor model, 79% of the factor loadings on the general factor were of a magnitude indicating overlap in the variance of items across the tests. Furthermore, most indicators have significant positive loadings on their respective specific factors (see Table S10). Along with an estimated explained common variance of .63, this means that the tests measure unique skillsets in addition to the common factor, and the bifactor model allows us to separate the common and specific variance of the tests. Similarly to the higher-order model, the extraction of construct-relevant variance is a crucial point if we wish to compare findings across studies, as the interpretation of relationships between morphological knowledge and other skills in language and literacy tasks depends on what causes the variance in morphological knowledge. Although we tested a model with a single general factor in our study, a bifactor model could also incorporate multiple general factors to further account for multidimensionality (e.g., Goodwin et al., 2021). The assumption that factors are uncorrelated is a drawback of the bifactor model, however. For example, the Morphological Pathways Framework cannot be tested in a bifactor model, since it does not allow for relations between morphological dimensions.

RQ3: How do Different Models Affect the Relation Between Morphological Knowledge and General Vocabulary?

The results from the three structural models imply very different relations between morphological knowledge and general vocabulary. The five-factor SEM in our study suggests that productive word knowledge is the only dimension of morphological knowledge that is related to general vocabulary, as measured by a word definition test. The lack of any relationship among general vocabulary and the other morphological factors might be due to the factor correlations disguising the unique contributions of each factor. This makes the five-factor model less informative, and further strengthens our claim that a correlated factor model is not suited for research on morphological knowledge and the relations of its facets to other areas of language and literacy.

The higher-order SEM is more informative, as we are able to examine relations among the morphological factors, as well as their relations to vocabulary. Supporting the theoretical pathways posited by Levesque et al. (2021), we found that morphological analysis was strongly related to vocabulary, whereas morphological decoding had no direct relation. The relation between morphological awareness and vocabulary was fully mediated through analysis, which is also in line with the theory. Thus, we found evidence of specific mechanisms within dimensions of morphological knowledge that influence the relations to other linguistic skills, exemplified with vocabulary in our study.

The results from our bifactor SEM analysis closely resemble those found by Goodwin et al. (2017). While we cannot draw any firm conclusions, the similarity in relations between the morphological factors and vocabulary provides support for the interpretation that these are generalizable patterns that apply to different languages and age groups. A model with three general factors, in line with Levesque et al. (2021), or four, as in Goodwin et al. (2021), might have been even more informative, but this was beyond the scope of our study. Even if such a model had been possible, the assumption of uncorrelated factors would prohibit an investigation of potential relations between the general factors. Still, the bifactor model provides the opportunity to examine the relations of specific factors to other linguistic skills. This is of importance when developing assessments, as it provides information on which skills we are measuring in addition to morphological knowledge.

In sum, our results show that morphological knowledge in Norwegian third graders is a multidimensional construct and that we need to account for construct-irrelevant variance due to methodological artifacts to get a clear representation of the construct. The five-factor model cannot separate construct-relevant and construct-irrelevant variance. Hence, it does not provide a clear view of whether the separate factors are due to different dimensions of morphological knowledge, or due to methodological artifacts such as tests measuring other language skills in addition to morphological knowledge. The bifactor model is well suited to separate construct-relevant and construct-irrelevant variance and accounts for multidimensionality as a methodological artifact. Thus, it provides an excellent framework for examining the overarching construct of morphological knowledge. A substantial drawback is that all factors in a bifactor model are uncorrelated, so in a theoretical model with three general factors representing morphological awareness, morphological analysis and morphological decoding, we would have to assume that these dimensions are unrelated. This assumption does not align with theory. Our results support the theoretical structure proposed by Levesque et al. (2021). To represent this structure, a higher-order model provides the best alternative, allowing us to remove construct-irrelevant variance while still enabling relations among the different factors.

Implications for Assessment and Research

It is clear from the findings of the present study that the associations between morphological knowledge and other language and literacy skills depend on how morphological knowledge is conceptualized. There is no doubt about the major differences in interpretation when comparing the five-factor, higher-order and bifactor SEMs in the current study. This implies that the use of different measures and different models may lead to confusion or misinterpretation if we are not careful in how we interpret results. Furthermore, the bifactor model might remove some of the confounding factors by separating the construct-relevant variance from that which is irrelevant. If we aim to investigate general morphological knowledge, it would be favorable to remove the variance related to other constructs, whether these represent specific morphological skills or other linguistic or task-related abilities. On the other hand, if our aim is to examine the relations among different morphological skills, we should turn to a higher-order model to enable relations between these factors. The bifactor model might be especially informative in test development. To further our understanding of morphological knowledge, however, we recommend representing the three dimensions of morphological awareness, morphological analysis and morphological decoding, in line with Levesque et al. (2021).

Our results indicate that in Norwegian, at least, morphological knowledge can be differentiated into morphological awareness, morphological analysis and morphological decoding from a relatively early age. Morphological decoding does require that the children have mastered basic word reading and spelling skills. Given that this assumption is met, we recommend measuring all three constructs to get a complete picture of children's morphological knowledge. To separate potential confounding information, we should use a model that separates construct-relevant variance from variance attributable to sources other than morphological knowledge, for example, a higher-order model.

Regarding the construction of interventions, we should take into account that morphological awareness, morphological analysis and morphological decoding might require different supporting skills, such as general vocabulary (base word knowledge) and decoding or spelling skills. According to the Morphological Pathways Framework, growth in morphological awareness will impact both morphological analysis and morphological decoding. Furthermore, improving morphological analysis will increase word knowledge and comprehension, whereas morphological decoding can enhance word reading and spelling. Thus, morphological interventions can aim to enhance language development broadly, or be tailored to affect specific skills, for example, reading or spelling.

Limitations and Future Research

To help us understand the source of variance in different dimensions of morphological knowledge and shed further light on the interpretation of factors, future research should aim to investigate the relationship between morphological factors and a wide variety of linguistic skills, such as reading comprehension, reading fluency, spelling, and listening comprehension. One particular limitation of the current study is the lack of a reading comprehension measure. Including a measure of reading comprehension would have provided additional context for factor interpretation, especially in the case of the specific word reading and spelling factors of the bifactor model, as well as the morphological decoding factor of the higher-order model. Additionally, general vocabulary was measured only with the vocabulary subtest of WISC-IV, a word definition test. A broader construct of vocabulary, for example, including a test of receptive vocabulary, would have been preferable.

Another limitation of the current study relates to the extensive exclusion of items, particularly from the test of receptive word knowledge. While the item exclusion did not change the substantive or statistical interpretations of the constructs measured, the analyses should be replicated in an independent sample to examine the generalizability of our models and results. This would also help refine the measures we developed in the project for use in future studies. Reducing the number of items will decrease the effort required from the children, as well as the time needed for testing, provided that validity holds for the intended use of the test scores.

The study supports the conclusion of Goodwin et al. (2017) that the bifactor model can help separate between construct-relevant and construct-irrelevant variance. Since the bifactor model also represents task-specific variance explicitly, it can contribute information about what we are measuring in addition to morphological knowledge. Future research should investigate how such specific factors related to general measures of skills such as word reading, spelling, and reading comprehension, as this could be informative for test development. Our results also support the Morphological Pathways Framework of Levesque et al. (2021). This provides preliminary evidence that the skills underlying the three theoretical constructs of this framework emerge relatively early in Norwegian, and perhaps in other alphabetic languages such as English. To strengthen the generalizability of the findings, future research should investigate whether the framework can be extended to similar languages as well as languages with different writing systems or distributions of morphemes (derivations, compounds and inflections), such as Chinese or Hebrew. Future research should also include children in preschool and early primary school to shed further light on the age of onset for the different morphological skills.

Acknowledgments

The authors thank the students, teachers, school leaders, and municipalities who participated. They thank Engagelab, especially Richard Nesnass, Ole Smørdal, Siri Jønnum and Jesús Gonzáles Torres, for their contributions to app development. They also thank Bente E. Hagtvet, Sol Lyster, Anita Lopez-Pedersen, Kari-Anne B. Næss, Riikka-Maija Mononen, Ona Bø Wie, Charles Hulme, and Catherine Snow for their valuable contributions to the design and implementation of the RCT. They are grateful to Terje Ulv Throndsen, Stein Malmo, Germán García Grande, and Nina Melsom Kristensen for research assistance, and master's students at the Department of Special Needs Education, University of Oslo, for their help with data collection and scoring. This work was supported by the Research Council of Norway, Grant no. 24033.

Conflicts of Interest

The authors have no known conflict of interest to disclose.

REFERENCES

Berthiaume, R., Bourcier, A., & Daigle, D. (2018). Morphological processing tasks and measurement issues. In R. Berthiaume, D. Daigle, & A. Desrochers (Eds.), *Morphological processing and literacy development* (pp. 48–87). Routledge.

Bowers, P. N., Kirby, J. R., & Deacon, S. H. (2010). The effects of morphological instruction on literacy skills: A systematic review of the literature. *Review of Educational Research*, *80*(2), 144–179. https://doi.org/10.3102/0034654309359353

Brinchmann, E. I., Hjetland, H. N., & Lyster, S. A. H. (2016). Lexical quality matters: Effects of word knowledge instruction on the language and literacy skills of third-and fourth-grade poor readers. *Reading Research Quarterly*, *51*(2), 165–180. https://doi.org/10.1002/rrq.128

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications.

Bryant, P., Nunes, T., & Bindman, M. (1997). Backward readers' awareness of language: Strengths and weaknesses. *European Journal of Psychology of Education*, *12*(4), 357–372. https://doi.org/10.1007/BF03172798

Carlisle, J. F. (2010). Effects of instruction in morphological awareness on literacy achievement: An integrative review. *Reading Research Quarterly*, *45*(4), 464–487. https://doi.org/10.1598/RRQ.45.4.5

Deacon, S., Parrila, R., & Kirby, J. (2008). A review of the evidence on morphological processing in dyslexics and poor readers: A strength or weakness? In G. Reid, A. J. Fawcett, & F. Manis (Eds.), *The SAGE handbook of dyslexia* (pp. 212–238). SAGE Publications Ltd.

Gonnerman, L. M. (2018). A linguistic analysis of word morphology. In R. Berthiaume, D. Daigle, & A. Desrochers (Eds.), *Morphological processing and literacy development* (pp. 3–15). Routledge.

González-Sánchez, L., García, T., Areces, D., Fernández, E., Arias-Gundin, O., & Rodríguez, C. (2018). Validación del Instrumento de Evaluación de la Conciencia Morfológica Oral (IECMO). *European Journal of Education and Psychology*, *11*(2), 107–122. https://doi.org/10.30552/ejep.v11i2.225

Goodwin, A. P., & Ahn, S. (2010). A meta-analysis of morphological interventions: Effects on literacy achievement of children with literacy difficulties. *Annals of Dyslexia*, *60*, 183–208. https://doi.org/10.1007/s11881-010-0041-x

Goodwin, A. P., Petscher, Y., Carlisle, J. F., & Mitchell, A. M. (2017). Exploring the dimensionality of morphological knowledge for adolescent readers. *Journal of Research in Reading*, *40*(1), 91–117. https://doi.org/10.1111/1467-9817.12064

Goodwin, A. P., Petscher, Y., & Tock, J. (2021). Multidimensional morphological assessment for middle school students. *Journal of Research in Reading*, *44*(1), 70–89. https://doi.org/10.1111/1467-9817.12335

Grande, G. G. (2018). *Morphological awareness in Norwegian preschoolers. [Master's thesis, University of Oslo]*. http://urn.nb.no/URN:NBN:no-66745

Hagtvet, B. E., Helland, T., & Lyster, S. A. H. (2006). Literacy acquisition in Norwegian. In R. M. Joshi & P. G. Aaron (Eds.), *Handbook of orthography and literacy* (pp. 15–30). Routledge.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

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. https://doi.org/10.1111/1467-9817.12316

Jong, Y. O., & Jung, C. K. (2015). Pedagogical significance of morphological awareness in Korean and English. *English Language Teaching*, *8*(8), 79–93. https://doi.org/10.5539/elt.v8n8p79

Kuo, L. J., & Anderson, R. C. (2006). Morphological awareness and learning to read: A cross-language perspective. *Educational Psychologist*, *41*(3), 161–180. https://doi.org/10.1207/s15326985ep4103_3

Levesque, K. C., Breadmore, H. L., & Deacon, S. H. (2021). How morphology impacts reading and spelling: Advancing the role of morphology in models of literacy development. *Journal of Research in Reading*, *44*(1), 10–26. https://doi.org/10.1111/1467-9817.12313

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. https://doi.org/10.1016/j.jecp.2017.02.015

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

Lyster, S. A. H. (2002). The effects of morphological versus phonological awareness training in kindergarten on Reading development. *Reading & Writing*, *15*(3–4), 261–294. https://doi.org/10.1023/A:1015272516220

Lyster, S. A. H., Lervåg, A. O., & Hulme, C. (2016). Preschool morphological training produces long-term improvements in reading comprehension. *Reading and Writing*, *29*(6), 1269–1288. https://doi.org/10.1007/s11145-016-9636-x

Manolitsis, G., Georgiou, G. K., Inoue, T., & Parrila, R. (2019). Are morphological awareness and literacy skills reciprocally related? Evidence from a cross-linguistic study. *Journal of Educational Psychology*, *111*(8), 1362–1381. https://doi.org/10.1037/edu0000354

Mansolf, M., & Reise, S. P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence*, *61*, 120–129. https://doi.org/10.1016/j.intell.2017.01.012

McBride-Chang, C., Cho, J. R., Liu, H., Wagner, R. K., Shu, H., Zhou, A., Cheuk, C. S., & Muse, A. (2005). Changing models across cultures: Associations of phonological awareness and morphological structure awareness with vocabulary and word recognition in second graders from Beijing, Hong Kong, Korea, and the United States. *Journal of Experimental Child Psychology*, *92*(2), 140–160. https://doi.org/10.1016/j.jecp.2005.03.009

Muse, A. E. (2005). *The nature of morphological knowledge. [Doctoral dissertation, Florida State University]*. DigiNole: FSU's Digital Repository https://diginole.lib.fsu.edu/islandora/object/fsu%3A180396

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. https://doi.org/10.1037/0022-0663.98.1.134

Nagy, W. E., Carlisle, J. F., & Goodwin, A. P. (2014). Morphological knowledge and literacy acquisition. *Journal of Learning Disabilities*, *47*(1), 3–12. 10.1177%2F0022219413509967

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill.

Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. *Psychometrika*, *47*, 337–347. https://doi.org/10.1007/BF02294164

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/

Ragnarsdóttir, H., Simonsen, H. G., & Plunkett, K. (1999). The acquisition of past tense morphology in Icelandic and Norwegian children: An experimental study. *Journal of Child Language*, *26*(3), 577–618. https://doi.org/10.1017/S0305000999003918

Reed, D. K. (2008). A synthesis of morphology interventions and effects on Reading outcomes for students in grades K–12. *Learning Disabilities Research & Practice*, *23*(1), 36–49. https://doi.org/10.1111/j.1540-5826.2007.00261.x

Revelle, W. (2021). *Psych: Procedures for personality and psychological research*. Northwestern University, Evanston, Illinois, USA https://cran.r-project.org/package=psych

Ribu, I. S., Simonsen, H. G., Løver, M. A., Strand, B.-M. S., & Kristoffersen, K. E. (2019). N-LARSP: A developmental language profile for Norwegian. In M. J. Ball, P. Fletcher, & D. Crystal (Eds.), *Grammatical profiles: Further languages of LARSP* (pp. 1–48). Multilingual Matters.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*, 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology*, *94*, 143–174. https://doi.org/10.1348/000712603321661859

Shi, D., Maydeu-Olivares, A., & Rosseel, Y. (2020). Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(1), 1–15. https://doi.org/10.1080/10705511.2019.1611434

Spencer, M., Muse, A., Wagner, R. K., Foorman, B., Petscher, Y., Schatschneider, C., Tighe, E. L., & Bishop, M. D. (2015). Examining the underlying dimensions of morphological awareness and vocabulary knowledge. *Reading and Writing.*, *28*, 959–988. https://doi.org/10.1007/s11145-015-9557-0

Tibi, S. (2016). *Cognitive and linguistic factors of reading Arabic: The role of morphological awareness in reading. [Doctoral dissertation, Queens University Kingston]*. QSpace: Queen's Scholarship & Digital Collections https://qspace.library.queensu.ca/bitstream/handle/1974/14674/Tibi_Sana_T_201607_PhD.pdf.pdf?sequence=1

Tibi, S., & Kirby, J. R. (2017). Morphological awareness: Construct and predictive validity in Arabic. *Applied PsychoLinguistics*, *38*, 1019–1043. https://doi.org/10.1017/S0142716417000029

Tighe, E. L., & Schatschneider, C. (2015). Exploring the dimensionality of morphological awareness and its relations to vocabulary knowledge in adult basic education students. *Reading Research Quarterly*, *50*(3), 293–311. https://doi.org/10.1002/rrq.102

Tighe, E. L., & Schatschneider, C. (2016). Modeling the relations among morphological awareness dimensions, vocabulary knowledge, and Reading comprehension in adult basic education students. *Frontiers in Psychology*, *86*(7), 395–409. https://doi.org/10.3389/fpsyg.2016.00086

Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research*, *49*, 1193–1208. https://doi.org/10.1044/1092-4388(2006/086)

Torkildsen, J. V. K., Bratlie, S. S., Kristensen, J. K., Gustafsson, J.-E., Lyster, S.-A. H., Snow, C., Hulme, C., Mononen, R.-M., Næss, K.-A. B., López-Pedersen, A., Wie, O. B., & Hagtvet, B. (2022). App-based morphological training produces lasting effects on word knowledge in primary school children: A randomized controlled trial. *Journal of Educational Psychology*, *114*(4), 833–854. https://doi.org/10.1037/edu0000688

Verhoeven, L., & Perfetti, C. A. (2011). Morphological processing in reading acquisition: A cross-linguistic perspective. *Applied Psycho-Linguistics*, *32*(3), 457–466. https://doi.org/10.1017/S0142716411000154

Wechsler, D. (2009). *WISC-IV norsk versjon. Manual del 1*. NCS Pearson, Inc.

Zhang, D. (2017). Multidimensionality of morphological awareness and text comprehension among young Chinese readers in a multilingual context. *Learning and Individual Differences*, *56*, 13–23. https://doi.org/10.1016/j.lindif.2017.04.009

Submitted January 19, 2022
Final revision received March 1, 2023
Accepted March 5, 2023

JARL K. KRISTENSEN (corresponding author) is a Doctoral Research Fellow at the Centre for Educational Measurement, University of Oslo, Oslo, Norway; email: jarlkkristensen@gmail.com

BJÖRN ANDERSSON is an Associate Professor at the Centre for Educational Measurement, University of Oslo, Oslo, Norway; email: bjorn.andersson@cemo.uio.no

SIRI S. BRATLIE is a Postdoctoral Research Fellow at the Department of Education, University of Oslo, Oslo, Norway; email: s.s.bratlie@iped.uio.no

JANNE V. K. TORKILDSEN is a Professor at the Department of Special Needs Education, University of Oslo, Oslo, Norway; email: janneto@isp.uio.no

Supporting Information

Additional supporting information may be found in the online version of this article on the publisher's website:10.1002/rrq.497/suppinfo

Table S1. Literature Review.
Table S2. Descriptive Statistics—Test of Receptive Word Knowledge.
Table S3. Descriptive Statistics—Test of Productive Word Knowledge.
Table S4. Descriptive Statistics—Word Analogy Test.
Table S5. Descriptive Statistics—Spelling Test.
Table S6. Descriptive Statistics—Word Reading Efficiency Test.
Table S7. Descriptive Statistics—WISC-IV Vocabulary.
Table S8. Factor Loadings, Five-Factor Model.
Table S9. Factor Loadings, Higher Order Model.
Table S10. Factor Loadings, Bifactor Model.
Figure S1. One-Factor Model.
Figure S2. Three-Factor Model.
Figure S3. Five-Factor Model.
Figure S4. Higher Order Model.
Figure S5. Bifactor Model.