17. Early prediction of learning outcomes in reading
Arne Lervåg & Monica Melby-Lervåg

1. Introduction
This chapter focuses on the processes involved in the development of children’s decoding skills: the ability to translate printed words into a speech code, typically assessed by the accuracy and speed of reading aloud. If children struggle in developing this ability, they are often characterized as having dyslexia, but since decoding skills are normally distributed in the population, the cutoff for this diagnosis is somewhat arbitrary (Melby-Lervåg et al., 2012). During the first two years of school, the main aim for children is to develop accurate and fluent early reading abilities that will lay the foundation for the main goal of reading – to be able to extract meaning from text. Thus, though insufficient by itself, efficient word reading is, in turn, a necessary condition for the development of reading comprehension. Research on the foundations of learning to read has burgeoned in the last 25 years with important theoretical and practical consequences.

For children in primary school, the most frequently used overarching framework to understand reading development is the simple view of reading. The simple view of reading suggests that reading comprehension is the product of decoding and linguistic comprehension (Gough & Tunmer, 1986). The component we focus on in this chapter is decoding (i.e. the ability to read accurately and fluently). Thus, here the term ‘reading’ is restricted to decoding skills. Here, we will have a closer look at predictors of decoding and also discuss other factors that possibly influence the development of early reading. The second component in the simple view of reading, linguistic comprehension, is the process of deriving semantic information at a word level, deriving meaning from sentences and discourse interpretation (Gough & Tunmer, 1986). As we will see, it has been argued that linguistic comprehension is the starting point for the development of early reading skills.

2. Phonological awareness and letter knowledge as predictors of early reading skills

2.1 Phonological awareness and prediction of early reading: Rhymes, phonemes or both?
Phonological awareness refers to an individual’s ability to reflect upon and manipulate the sound structure of spoken words. This ability has repeatedly been shown to be uniquely related
to the development of early reading skills in longitudinal latent variable studies (De Jong & Van der Leij, 1999; Lervåg et al., 2009; Wagner et al., 1994) (Dutch, Norwegian and US-American samples). The hypothetical reason for this is that phonological awareness tasks tap into how the sound structures of words are represented in the brain (Melby-Lervåg et al., 2012) (meta-analysis with a focus on European and US samples). It is not clear, however, when it comes to the particular phonological representations (e.g. whether it is encoding, storage or retrieval) that are most important for developing accurate and fluent reading. The reason is that such representations can never be directly observed, but only indirectly inferred. What does seem clear, though, is that a lack of phonemically structured phonological representations makes phoneme awareness tasks difficult or impossible to perform. Moreover, since early reading is dependent on this awareness, early reading will be difficult as well (Melby-Lervåg et al., 2012).

A large set of different tasks has been used to assess phonological awareness. These tasks differ in at least two main aspects: First, they vary when it comes to the size of the phonological units that are manipulated: these are either syllables, rhymes or phonemes (Melby-Lervåg et al., 2012). This choice between tasks can be explained by the observation that the development of phonological awareness progresses in a hierarchical fashion throughout childhood. Specifically, a child typically first becomes aware of large sound units such as syllables and then moves on to intermediate units (onsets and rimes), and finally reaches the stage where phoneme awareness is mastered (Carrol et al., 2003) (English sample). Second, the tasks also differ concerning the type of problem to be solved (Melby-Lervåg et al., 2012). Some include more forced-choice tasks (e.g. “Do these two words rhyme?”), while others have more explicit tasks (e.g. “Which word do we create if we remove the /h/ in /hat/?”). This large variety has often led to low correlations between the tasks, and it can be difficult to find out whether different studies actually measure the same underlying construct.

The two most common tasks concerning the size of phonological units have been rhyme awareness tasks and phonemic awareness tasks. These two tasks are rather different in nature and difficulty. Unfortunately, many previous studies merge these two tasks and labeled them ‘phonological awareness tasks.’ However, a considerable controversy in the field has been centered on the question whether one of them is more important than the other in predicting early reading skills (Bryant, 2002; Hulme et al., 1998; Hulme et al., 2002). One prominent theory in this context is the grain size theory (Goswami & Bryant, 2016; Ziegler & Goswami, 2005). This theory postulates that children learn to read by becoming aware of the letters that correspond with onsets and rhymes in language. According to this theory, awareness of phonemes comes later, possibly as a consequence and in reciprocity with learning to read. Further, it should also be noted that there is an alternative view suggesting that phonological awareness is best considered as a unitary construct and that the skills are
so strongly interrelated and correlated that it is not meaningful to separate them into different constructs (Anthony et al., 2002; Anthony & Lonigan, 2004). Thus, according to this unitary theory, the division between phonemes and rhymes is artificial, since both are a part of the same construct.

The grain size theory and the theory of phonological awareness as a unitary construct are, however, challenged by studies that have shown that phoneme awareness is a stronger predictor of reading than rhyme awareness, both cross-sectionally and longitudinally, at least from the age range of 6 to 7 years, (Muter et al., 1998; Muter et al., 2004) (English samples). A review also showed that the mean bivariate correlation between phoneme awareness and early reading was larger than that of rhyme awareness and early reading ($R = 0.57$ and $R = 0.43$) and that children with dyslexia had significantly more problems with phoneme awareness than with rhyme awareness ($d = -1.73$ and $d = -0.93$, respectively) (Melby-Lervåg et al., 2012). It seems that there is now a consensus that phoneme awareness is critical for learning to read, and rhyme awareness is an important precursor ability, but has little or no influence on early reading beyond phoneme awareness.

2.2 Is letter knowledge separable from phoneme awareness?

Letter knowledge is perhaps a more obvious predictor of early reading ability than phoneme awareness because children need to know letters to be able to decode at all. Indeed, it has been repeatedly shown that letter knowledge is a strong and unique predictor of early reading ability (De Jong & Van der Leij, 1999; Lervåg et al., 2009; Wagner et al., 1994) (Dutch, Norwegian and US-American samples). While studies vary regarding their separation of letter-name and letter-sound knowledge most of them find that these two dimensions are best conceptualized as a unitary construct (Lervåg et al., 2009; Muter et al., 2004). Out of the few studies that have reported separate measures of letter–name and letter–sound knowledge obtained from the same children, most show that both name and sound knowledge are correlates of reading ability (Caravolas et al., 2001; McBride-Chang, 1999) (English samples).

Still, even if it may seem obvious that letter knowledge is a predictor of early reading, the explanation for this association might not be straightforward. Most obvious, of course, is that knowledge of letter–sound–links is pivotal to the child’s discovery of the alphabetic principle (i.e. how individual speech sounds (phonemes) in oral words are represented by letters in printed words (graphemes). However, it is also plausible that letter knowledge taps into the visual-phonological learning that is critical for early reading (Lervåg et al., 2009) (Norwegian sample). This hypothesis has been tested in studies focusing on paired-associate learning (i.e. associating figures with a nonsense word). The results of these studies vary, but some have shown that visual-phonological learning is a concurrent predictor of early reading (Hulme et al., 2007). From a longitudinal perspective however, there seems to be little support
that visual-phonological learning is a longitudinal unique predictor of early reading (Lervåg et al., 2009).

Another issue that has been debated in the field is whether phoneme awareness contributes independently to letter knowledge. Castles and Coltheart (2004) claimed that studies did not unequivocally show this. One of the reasons for the failure to demonstrate a causal link was that studies were not able to reveal separate contributions of phoneme awareness and letter knowledge to reading ability. However, several studies now demonstrate an independent contribution of phoneme awareness over and above letter knowledge (and vice versa). For example, (Lervåg et al., 2009) report standardized path coefficients of 0.34 for the relation between early reading and letter knowledge, and of 0.32 for the relation between early reading and phoneme awareness. Similarly, in the study by Muter et al., (2004) standardized path coefficients to early reading 0.25 for letter knowledge and 0.15 for phoneme awareness. Still, it is important to note that phoneme awareness does not come into play until right before the child is able to decode. Moreover, phoneme awareness is also enhanced in reciprocity with early reading skills such that learning letters refines phoneme awareness (Wagner et al., 1994). Further, there is also a study that found no independent contribution of phonological awareness beyond letter knowledge (De Jong & Van der Leij, 1999). It should be mentioned, however, that in this study phonological awareness comprised measures of both phoneme and rhyme awareness.

Altogether, research on phonological awareness, letter knowledge and early reading is a mature research field, and there are a number of studies and reviews to support the main findings in this area. Of course, observational prediction studies cannot draw a conclusion about causality. It is therefore important to note that there are also a number of training studies that support the main issues discussed here (for reviews, see (National Early Literacy Panel (US), 2008; National Reading Panel (US), 2000)). Notably, it also seems clear that there is a strong coherence between the findings in studies of children with dyslexia and prediction studies of typically developing children (Melby-Lervåg et al., 2012; Snowling & Melby-Lervåg, 2016): The strongest predictors of individual differences in reading abilities of typically developing children belong to those domains in which children with dyslexia struggle most, namely phoneme awareness and letter knowledge.

3. Rapid automatized naming (RAN) and beginning reading skills

Beyond phoneme awareness and letter knowledge, rapid automatized naming (RAN) has been a consistent predictor of beginning reading skills (Lervåg et al., 2009; Lervåg & Hulme, 2009) (Norwegian and English samples). Typical RAN tasks include a sheet with either colours, objects, digits or letters that are (semi)randomly repeated across several lines. The task is to say aloud all items as quickly as possible, beginning at the upper left corner and ending at the
bottom right corner of the sheet. The time it takes to finish naming all items predicts beginning reading skills both concurrently and longitudinally, also across different orthographies (Araújo et al., 2015; Caravolas et al., 2012) (meta-analyses with a focus on European and US samples). Furthermore, RAN is able to predictively differentiate between people who are diagnosed with dyslexia and people who are not (Araújo & Faisca, 2019) (meta-analysis with a focus on European and US samples). RAN is correlated with phoneme awareness, letter knowledge and verbal short-term memory, but it is able to predict the development of beginning reading skills beyond these other predictors (De Jong & Van der Leij, 1999; Lervåg et al., 2009).

Several theories have been proposed to explain the relationship between RAN and beginning reading skills. What seems to be a consensus is that RAN taps into a basic mechanism that is causally relevant for learning to read. It is less clear, however, what this mechanism actually is. One popular suggestion is that RAN reflects the speed at which phonological representations can be retrieved from long-term memory (Wagner & Torgesen, 1987; Wimmer et al., 2000). However, this is not to say that RAN taps into phonological representations in the same way as phonological awareness does since the correlations between typical tasks related to these domains are only low to moderate. In addition, RAN has been consistently shown to explain variations in reading development beyond phoneme awareness and verbal short-term memory tasks. For instance, Lervåg and Hulme (2009) found that non-alphanumeric RAN (colours and objects) measured before the start of formal reading instruction explained reading development one year later after formal reading instruction had started (English sample). Additionally, alphanumeric RAN (letters and digits) explained further growth in text-reading efficiency between the second and fourth grade beyond earlier reading skills, phoneme awareness and IQ. Findings like these have led researchers to the conclusion that RAN taps into the integrity of the neural circuits involved in object identification and naming and that these neural circuits are recruited as critical components of the children’s developing visual word recognition system (Lervåg & Hulme, 2009).

Several studies have shown that serial RAN seems to be a better predictor than discrete naming, particularly amongst more experienced readers. While both serial RAN and discrete naming seem to predict beginning reading beyond each other, only serial RAN seems to predict reading among older children and adults beyond discrete naming and not vice versa (Altani et al., 2018; Georgiou & Parrila, 2020; van den Boer & de Jong, 2015) (Greek, English and Dutch samples). According to Altani et al. (2020) this difference reflects a shift from dealing with words at a micro level (word-by-word reading), as beginning readers do, to dealing with words at a macro level (multi-word reading), as more experienced readers do. Reading at a macro level is thought to be partially determined by the ability to simultaneously deal with several items in a cascaded fashion (Altani et al., 2020). The ability to deal with several items
simultaneously in cascades is also assumed to be an ability taxed by serial RAN but not by discrete naming (Altani et al., 2020; Protopapas et al., 2018). Thus, both serial and discrete RAN predict beginning reading presumably because they reflect fast access and retrieval of phonological representations. Serial RAN, in addition, reflects word processing in readers who read serially at a macro level.

Several studies have also searched for a reciprocal relationship by addressing the question whether reading also predicts the development of RAN. There could be several reasons for why reading might be able to predict RAN. One is that the automaticity of letter and digit recognition could promote the development of good alphanumeric skills. Another is that left-to-right visual scanning skills and visual attention, both of which are presumably refined by reading experience, could make the RAN tasks more familiar and thereby easier to execute. While there are studies supporting the first idea, showing that letter knowledge can predict the development of alphanumeric RAN skills (Lervåg & Hulme, 2009; Peterson et al., 2018), the second idea is not empirically supported. For instance, Protopapas et al. (2013) showed that reversing the RAN task from left-to-right to right-to-left did not change RAN’s predictive value for reading skills (Greek sample). There is also a study indicating that early reading skills predict the development of non-alphanumeric RAN (Powell & Atkinson, 2020) (English sample), but this link currently seems rather uncertain since this study was based on observed variables. It is well known that the measurement error embedded in observed variables can affect the results of these cross-lagged panel models (Little, 2013). Thus, the potential issue of a reciprocal relationship seems unresolved. However, one should not yet fully discard the idea that reading experience can modulate the ability to simultaneously deal with several items in a cascaded fashion and that this in turn may promote the development of RAN skills. Still, additional empirical evidence is needed to support this hypothesis.

One problem of all the studies examining reciprocal relationships between RAN and reading is that none of them has attempted to separate within-subject from between-subject variance in their statistical models. It is expected that much of the correlation between RAN and reading originates from correlations at the between-subject level (i.e. stable trait-like between-person variance). Therefore, these models do not tell us if reading actually affects the development of RAN or if the reading-RAN association is only a product of the stable parts of reading and RAN. Some more recent statistical models like the random intercept-cross lagged model (RI-CLM) allow us to look at the predictors of change within an individual (the individual serves as their own control) (Usami et al., 2019). These models should be preferred when addressing questions about reciprocal relationships.

4. Are there other cognitive predictors of early reading than ‘the big three’?
As we have seen, there are three predictors that have been repeatedly demonstrated to be most influential for early reading skills: phonological awareness, letter knowledge and rapid automatized naming. How much variation in early reading skills they explain together, however, varies between studies. In latent variable analyses, the explained variance in early reading based on measures taken in pre-readers ranges from 40% (Näslund & Schneider, 1991) (German sample), 45% (De Jong & Van der Leij, 1999) (Dutch sample) and 52% (Lervåg et al., 2009) (Norwegian sample) to 56% (Wagner et al., 1994) (US sample). Accordingly, there is still variance left to be explained by other predictors, and we will now examine some of the potential candidates.

4.1 Verbal short-term memory
It has long been suggested that verbal short-term memory plays an important role for developing early reading skills. In particular, this notion has been linked to Baddeley and Hitch’s (Baddeley & Hitch, 1974) working memory model. According to Gathercole and Baddeley (2014), verbal short-term memory is important for early reading because early reading involves holding sounds in memory while deciphering letters. A poor verbal short-term memory could therefore restrain the development of a phonological early reading strategy (Gathercole & Baddeley, 2014). Therefore, earlier, verbal short-term memory was typically included in a broader phonological processing construct, for instance, in Wagner and Torgesens’ (1987) seminal review (Wagner & Torgesen, 1987).

Prior studies, however, provide little support for a unique relationship between verbal short-term memory and early reading. Specifically, the review by Melby-Lervåg et al. (2012) rather supports the idea that verbal short-term memory measures can also be seen as an index for the quality of phonological representations. Accordingly, the tight link to phonological representations might underlie the importance of verbal short-term memory for early reading rather than the memory part in itself. Finally, there are also several longitudinal studies indicating that verbal short-term memory (as measured by various span measures, e.g. words, digits and nonwords) does not uniquely explain early reading development on top of the ‘big three’ predictors (Lervåg et al., 2009; Näslund & Schneider, 1991; Wagner et al., 1994). Cunningham et al. (2021), however, found that verbal short-term memory predicted later early reading beyond phoneme awareness with the limitation that this study did not include rapid naming (English sample).

4.2 Vocabulary and language comprehension
It has also been suggested that vocabulary explains variation in reading ability in addition to phoneme awareness, letter knowledge and rapid naming. Theoretically, this can be explained with the lexical quality hypothesis (Perfetti, 2007), which claims that high-quality word representations (including not only phonology, but also morphology and semantics) are
mutually supportive and critical for early reading development. Thus, the better and broader the representation of a word, the easier it is to decode it. This theory is in line with the observation that nonwords are much more difficult to decode than real words.

The empirical evidence supporting this view is mixed. In studies using latent variables, vocabulary hardly explains any variance in early reading development beyond the ‘big three’. In a study by Lervåg et al. (2009), which traced reading development in first and second grade, vocabulary did not show a direct influence on top of the other predictors. A similar finding was also reported in a cross-linguistic longitudinal study by Furnes and Samuelsson (Furnes & Samuelsson, 2011) and in a study by Muter et al. (2004). However, a longitudinal latent variable study by Hecht et al. (2000) showed significant unique contributions of kindergarten vocabulary skills to the prediction of early reading in 1st, 2nd and 3rd grade (between 4% and 13% of explained variance), but this work was based on a sample with a low socio-economic status (in England).

Still, even if vocabulary and language comprehension are not directly important for the development of early reading skills, they might play an indirect role as a foundation for the precursor skills of early reading. Language comprehension and the precursors for early reading are highly related but studies using confirmatory factor analyses have shown that they are different constructs (Hjetland et al., 2020; Hulme et al., 2015; Storch & Whitehurst, 2002; Torppa et al., 2016). According to the lexical restructuring hypothesis (Walley et al., 2003), vocabulary growth enhances the development of clearer representations for phonological units in childhood. This enables the child to develop awareness for increasingly smaller linguistic units (syllables, rhymes) with phoneme awareness forming the endpoint. An alternative view is proposed by the segmentation theory (Boada & Pennington, 2006). This theory postulates that auditory perception and speech processing, rather than vocabulary growth, form the starting point of phonological skills and word reading ability. According to this hypothesis, vocabulary growth can be a consequence of the development of more fine-grained phonological representations. Thus, in line with the segmentation theory, phonological representations become more detailed because segmentation of speech sounds drives the emergence of phonological awareness, and this also enables vocabulary growth.

Based on the currently available literature (Hjetland et al., 2020; Hulme et al., 2015; Snowling et al., 2019; Storch & Whitehurst, 2002; Torppa et al., 2016), it is difficult to draw conclusions about the direction of the relationship between the precursors of early reading and linguistic comprehension. An important part of this problem is that it is difficult to find autoregressors (i.e. previous skills related to the same construct) for phonological skills in young children. What is clear, however, is that language comprehension and the precursors for early reading are highly intertwined and closely related. We currently consider it as more likely that there is a reciprocal rather than a unidirectional relationship between them.
4.3 Executive functions
Executive functioning has also been suggested to be important for early reading and several authors have argued that early reading problems cannot be fully understood in terms of phonology. According to this view, executive functioning influences early reading because early reading involves parallel activation of lexical and phonological routes, which places cognitive demands on executive functions (Ober et al., 2020) (meta-analysis with a focus on European and US samples).

Evidence in support of this view was provided in a meta-analysis, which found a moderate correlation between early reading and executive functions, and also showed that this link was rather stable across different moderators (Ober et al., 2020). However, a concurrent correlation alone does not necessarily imply that executive functioning is important for the development of early reading. When it comes to longitudinal associations, results are scarce. A study with latent variables has shown that a general working memory factor predicts early reading in first grade, but this association is not specific since this study does not include measures of rapid naming or phonological awareness (Preßler et al., 2014) (German sample). Another study comprising observed variables showed that executive functions (measured by attentional control) in kindergarten predict first-grade reading ability, but this work did not include rapid naming (Segers et al., 2016) (Dutch sample). Similarly, other work supporting a predictive role of executive functioning for early reading also lacks all or some of the ‘big three’ (Colé et al., 2014; Foy & Mann, 2013) (French and US-American samples).

4.4 Nonword repetition
A number of studies indicate that children with dyslexia experience problems with nonword repetition. Nonword repetition is widely used as a cognitive marker of dyslexia and as a predictor of early word reading skills (Goulandris et al., 2000) (English sample). The main explanation for why nonword repetition should be related to early reading is that it reflects how efficiently a person can process phonological information (Melby-Lervåg & Lervåg, 2012). Nonword repetition is considered to be highly dependent on output phonology (Snowling & Hulme, 1989), and may therefore be a marker of dyslexia. However, if nonword repetition is an important marker of dyslexia, it should also be a strong predictor of early reading, since individuals with dyslexia typically have problems in the domains that explain individual differences in early reading (Melby-Lervåg et al., 2012).

A meta-analysis of studies that have examined nonword repetition skills confirms that individuals with dyslexia struggle particularly with nonword repetition (mean differences in $d = -0.82$) which is also moderately correlated with early reading ($R = 0.33$) (Melby-Lervåg & Lervåg, 2012). However, the severity of the nonword repetition problem varies significantly between studies, and the most important predictor of this variability is linguistic
comprehension. While nonword repetition has been used mainly as a marker of dyslexia there are few longitudinal prediction studies with typically developing children. One of the few available studies found that nonword repetition goes together with measures of verbal short-term memory in a latent construct but does not predict variation in early reading beyond the ‘big three’ (De Jong & Van der Leij, 1999). In a recent study, however, a latent variable capturing nonword repetition at 5 years was shown to predict early word reading at 6 years beyond phoneme awareness and an autoregressor variable for early reading (Cunningham et al., 2021). Still, in the latent variable study of Hulme et al. (2015) (English sample) it was shown that it was the variance that nonword repetition had in common with broader language skills like vocabulary, sentence structure etc. and not the unique variance that it shared with word repetition and articulation at 3.5 years beyond broader language skills that predicted later decoding skills indirectly through the big three (phoneme awareness, letter knowledge and RAN) and reading comprehension directly.

4.5 Nonverbal IQ
Nonverbal IQ of one standard deviation below the population average was traditionally used as an exclusion criterion for dyslexia following the assumption that early decoding was related to nonverbal IQ (Stuebing et al., 2002). However, a meta-analysis of 46 studies addressing the validity of this diagnostic classification showed considerable decoding performance overlap between the IQ-discrepant and IQ-consistent poor readers, indicating that nonverbal IQ is not strongly related to early reading (Stuebing et al., 2002). Furthermore, longitudinal studies have also shown that nonverbal IQ cannot explain development in early reading beyond the ‘big three’ (De Jong & Van der Leij, 1999; Lervåg et al., 2009).

5. Prediction of early reading skills in different orthographies
As alphabetic orthographies differ in how consistently letters (graphemes) represent sounds (phonemes), it is important to ask which impact this has on early reading development. One major difference between orthographies is how quickly children master fluent and accurate reading skills. Comparing 14 European orthographies, Seymour et al. (2003) found that in relatively consistent orthographies (like Finnish), children had fluent and accurate reading skills (> 80%) at the end of Grade 1, while in English – which is seen as a less consistent orthography – children had much lower reading accuracy levels (34%). A study by Caravolas et al. (2013) confirmed this result in a longitudinal sample comparing early reading development in three different orthographies: two relatively consistent (Spanish and Czech) and one inconsistent orthography (English). While the children in relatively consistent orthographies revealed a similar growth pattern with a clear acceleration after the onset of formal reading instruction, the children learning an inconsistent orthography revealed a more even and slower growth in their reading skills.
As the growth pattern of emerging reading skills seems to differ as a function of how consistent an orthography is, it is also important to ask if predictors of reading development differ as a function of orthographic consistency. In other words, the question is: Do the same predictors predict reading development in consistent and in inconsistent orthographies? This has been a controversially discussed issue (Caravolas et al., 2012; Share, 2008). Some authors have, for example, suggested that phoneme awareness is less important and rapid automatized naming (RAN) is more important in consistent compared to inconsistent orthographies (Wimmer et al., 2000).

In the last decade, a handful of studies have addressed this issue. First, the concurrent studies of Vaessen et al. (2010) and Ziegler et al. (2010) concluded that to a large extent the same predictors are important in both consistent and inconsistent orthographies (Vaessen et al., 2010; Ziegler et al., 2010). In their longitudinal study, Caravolas and colleagues (2012, 2013) went even further, claiming that there are universal cognitive prerequisites for learning to read in all alphabetic orthographies, even though the rate of learning differs between consistent and inconsistent orthographies. This claim was based on the observation that letter knowledge, phoneme awareness and RAN predicted growth in all three orthographies. Interestingly, Caravolas et al. (2013) found that these three predictors were associated with growth in reading at different stages in the developmental process. While letter knowledge and phoneme awareness were associated with reading at the very beginning of the developmental process, RAN was associated with the later acceleration of reading growth.

Still, not all longitudinal results are so clear cut. For example, Furnes and Samuelson (2009) found that phoneme awareness predicted reading only at the end of Grade 1 in the relatively consistent Norwegian and Swedish orthographies but continued to predict reading in the more inconsistent English orthography in Grade 2. In contrast, RAN was a consistent predictor in all three orthographies. The finding that phoneme awareness only predicted reading very early in the consistent orthographies and longer in the inconsistent English orthography fits well with the work of Caravolas et al. (2013). They found that learning to read takes longer in inconsistent orthographies and that phoneme awareness was associated only with early reading development. As it takes longer to develop fluent and accurate reading skills in inconsistent orthographies, it is to be expected that phoneme awareness will predict reading development for a longer time as well.

Compared to the studies above, the data of Landerl and colleagues (Landerl et al., 2019) draw a less consistent picture. While RAN turned out to be a reliable predictor of reading development in both consistent and inconsistent orthographies, phoneme awareness predicted later reading skills in French and German (inconsistent and consistent, respectively) but not in English, Dutch and Greek (inconsistent, consistent and consistent, respectively). On the basis of these results, they concluded that while RAN seems to be a consistent and
universal predictor of reading development, the predictive power of phoneme awareness might have been overstated in the literature. It should be noted, however, that the study of Landerl et al. (2019) started after the onset of formal reading instruction, whereas the studies of Caravolas et al. (2012, 2013) started before formal reading onset. As phoneme awareness is most strongly associated with very early reading development, this discrepancy might explain why Landerl et al. (2019) did not find an association with reading development in all of these orthographies. Nevertheless, it is not in line with a large body of literature that phoneme awareness did not predict reading development in English.

6. Prediction of early reading skills based on socio-economic background

It is well known that there is a moderate to high positive correlation between socioeconomic status (SES) and reading (Sirin, 2005) (meta-analysis with a focus on European and US samples) (see chapter 14). This is the case not only for reading comprehension, but also for basic early reading skills (Buckingham et al., 2014). Hence, children living in poverty have been found to achieve a considerably lower level of literacy than their peers with higher SES (Buckingham et al., 2014). Indicators of SES typically include income, educational level, employment and living conditions (Buckingham et al., 2014).

An important issue is whether there is a causal relationship between SES and early reading or whether there is only a correlation. To disentangle this, three issues need to be clarified. The first is whether SES can explain development or growth in early reading beyond the autoregressor (i.e. early reading ability itself). The second is whether SES also explains development in early reading beyond the ‘big three’ predictors. Third, it remains to be specified what it would mean that SES explains growth beyond the autoregressor and the ‘big three’ predictors. Possible interpretations are complicated by the fact that SES has often been called a fuzzy construct because it comprises a large range of factors, including genes and environments. Presumably around 50% of the relationship between SES and academic achievement can be explained by genetic factors (Krapohl & Plomin, 2015) (English sample) (see chapter 6). What the rest can be attributed to is not clear. Potential candidates could be stress of poverty, parental home literacy practices, early diet and nutrition that are important for brain development and instructional quality in schools (see chapter 14).

SES is typically not able to explain development in early reading when the autoregressor and the ‘big three’ predictors are controlled for (Buckingham et al., 2014). Thus, in general, a causal relationship between SES and early reading development does not seem likely. However, there are exceptions in the literature. One study of children in Romania (Dolean et al., 2019) provided evidence that SES explained actual growth in early reading beyond rapid naming, letter knowledge, and phoneme awareness. Variation in SES was large in this sample, ranging from children living in severe poverty to children from affluent families.
Thus, poverty alone does not explain why SES was related to growth in early reading skills. Notably, SES was modeled as a latent construct consisting of family income, mother's education, parents' employment, and living conditions. This is a more differentiated model of SES compared to most other studies which might have given SES greater predictive strength. Interestingly, the relationship between SES and early reading was mediated through school absence. Similar findings have also been demonstrated in other studies in the context of high levels of poverty (Herbers et al., 2012) (US-American sample).

7. Conclusion
In this chapter, we have looked at various potential predictors of beginning reading development by reviewing longitudinal observational studies. Although the title of this overview refers to prediction, we have gone beyond prediction in our descriptions and discussions. With prediction, we typically refer to merely foreseeing a future outcome. Thus, when we are concerned with pure prediction, we are not interested in why certain factors predict something. Instead, we are merely interested in finding a set of variables that predicts as much variation in an outcome variable as possible (Pedhazur, 1997; Shmueli, 2010). When predicting reading development, however, we are not only interested in identifying a set of predictors associated with later reading skills, but we are also interested in explaining why they predict reading skills. To this end, we need to apply theories and see to what degree they fit the empirical evidence. These theories can explain how the predictors reflect causes for the ease or the learning experience of a particular child (e.g. less distinct phonological representations will cause difficulties in learning to read).

If a variable, measured before a child can read, is able to predict variation in later reading development, we often say it fits with a causal theory of reading development. The same is the case if a variable predicts later reading skills beyond earlier reading skills or actual growth in reading skills. In this way, longitudinal observational studies can help us to test theories about reading development and thereby inform us about why certain skills are important for learning to read. It should be noted, however, that observational studies cannot prove empirically that the predictors are causally related to reading development. Nevertheless, they can help us to generate theories that can be used as a framework for designing randomised control trials (see chapter 19). Such intervention studies in turn can help us to test causal theories and understand how we can support children on their way to becoming literate (Lervåg, 2020). It should also be noted that there are more theories about predictors of reading development than those we describe here (see chapter 1). We describe and discuss only those theories that we consider as most debated and/or best supported by the recent literature.
Our overview indicates that phoneme awareness, letter knowledge and rapid automatized naming (RAN), measured just before the children start learning to read, are the most reliable predictors of reading across several studies and across consistent and inconsistent orthographies. While letter knowledge and phoneme awareness seem most important at the very beginning of learning to read, RAN continues to be associated with growth in reading skills, probably as a function of the serial information processing aspects inherent to both RAN and reading. In addition, some studies indicate that SES might have a direct impact on reading development, at least in samples including children living in poverty (Dolean et al., 2019). That being said, there are several other variables that are associated with beginning reading skills, but they do not seem to predict reading beyond the 'big three.' Nevertheless, these variables can have indirect effects on reading, and this is demonstrated, for instance, for measures related to vocabulary and linguistic comprehension (Hjetland et al., 2020; Hulme et al., 2015).

Suggestions for further reading


