Developmental Psychology
1998, Vol. 34, No. 6, 1373-1399

0012-1649/98/$3.00

Analyzing the Development of Individual Differences in Terms of Matthew Effects in Reading: Results From a Dutch Longitudinal Study

Janwillem Bast and Pieter Reitsma
Paedologisch Instituut-Vrije Universiteit Amsterdam

The Matthew effect hypothesis provides a theoretical framework to describe the development of individual differences in reading ability. The model predicts an increase of individual differences in reading. Reciprocal relationships between reading and other factors seem to cause these increasing differences. This longitudinal study of 3 years was concerned with uncovering the existence and causes of increasing individual differences in reading in the early elementary grades. Data were analyzed within a structural equation modeling framework. The results clearly indicate increasing individual differences for word recognition skills. For reading comprehension, no such effects could be established for this limited time period. More important, some evidence for interactive relationships between reading and other cognitive skills, behaviors, and motivational factors, hypothesized to cause increasing differences between readers, was found.

It is evident that not all children become proficient readers. How do individual differences in reading performance come into existence? For most children, reading development starts at school with formal education in Grade 1. However, the development of skills necessary to acquire reading skills starts well before the beginning of formal instruction in reading. Already before children enter school, large differences among them exist as a result of innate competencies and the quality and intensity of parental care invested in them (Walberg & Tsai, 1983). Taking initial differences between individuals as a starting point, the question is whether these differences remain stable or whether individuals converge or diverge in level of performance with further instruction in reading.

Stability of individual differences, that is, a relative consistency of individual differences over age, has frequently been found in longitudinal studies of reading (Butler, Marsh, Sheppard, & Sheppard, 1985; Juel, 1994). However, this stability refers only to the rank ordering of participants within a certain population. The absolute performance difference between readers can increase or decline in the course of development as a result of interindividual differences in intraindividual change. Individual differences in reading ability do not seem to disappear with further development or to diminish over the years. Even among college students, large individual differences in their reading skills are still found (Perfetti, 1985). Thus, a decrease in variance with time is not to be expected. Instead of a decline, an increase of interindividual differences with further schooling seems probable. For example, the interindividual differences among the reading achievement levels of 8th to 12th graders are larger than the differences among 1st graders (Daneman, 1991; Williamson, Appelbaum, & Epanchin, 1991). In sum, it seems that the development of individual differences in reading ability can be characterized by the combination of stability of rank orderings and increasing spread.

The Matthew effect model of Stanovich (1986) provides a theoretical framework in which the development of individual differences in reading ability can be described and explained. The *Matthew effect* refers to the phenomenon that, over time, better readers get even better, and poorer readers become relatively poorer. This outcome has reference not only to the different components of reading ability, such as word recognition and reading comprehension skills, but also to the development of cognitive skills related to reading, such as vocabulary and other (meta)linguistic skills. The *Matthew effect model* attempts to account for these fan-spread effects, that is, the increase of individual performance differences over time. The fan spread is, however, simply one component of the Matthew effect phenomenon. The most important feature of the model as proposed by Stanovich (1986) is the underlying developmental pattern that causes this outcome. The phenomenon of increasing achievement differences is hypothesized to be caused by a specific developmental pattern of interrelations between reading skills and other variables.

This hypothesized developmental pattern can be described by two important mechanisms: reciprocal causation and developmental limits. The first mechanism is composed of bidirectional causal relations between reading and other cognitive skills, attitudes, or behaviors. The second mechanism describes the fact that individual differences in a particular cognitive process may be a causal determinant of variation in reading achievement early in development, but at some later point have no further effects on the level of reading efficiency (Stanovich, 1986). Developmental limits can also refer to the fact that individual

Janwillem Bast and Pieter Reitsma, Paedologisch Instituut-Vrije Universiteit, Amsterdam, The Netherlands.

This research was supported by Grant SGW 575-91-002 from the Dutch Organisation of Scientific Research.

We extend our thanks to the children, teachers, and principals of the schools who participated in this project.

Correspondence concerning this article should be addressed to Pieter Reitsma, Paedologisch Instituut-Vrije Universiteit Amsterdam, P.O. Box 303, 1115 ZG Duivendrecht, The Netherlands. Electronic mail may be sent to p.reitsma@psy.vu.nl.

differences for one component of reading do increase during a certain phase in reading development after which a further increase of individual differences can only be detected for another component of reading. For instance, in the beginning of reading development differences in word recognition skills are likely to increase. However, after this initial stage, only increasing differences for reading comprehension skills may be found while differences in word recognition skills are relatively stable.

Because relationships between reading skills and other factors are hypothesized to be developmentally limited, the concept of reciprocal causation has to be framed developmentally (Stanovich, 1986). In other words, one can distinguish reciprocal relationships that are operative throughout reading development and relationships that are effective for only a restricted part of this development.

An important candidate of the latter type is the relationship between phonological processing abilities and word recognition skills. Children with poor phonological processing abilities will have difficulty acquiring word recognition skills and hence read less than children with good phonological processing abilities. Because of reciprocal causal relations, poor readers reading less further impairs development of phonological processing abilities (Perfetti, Beck, Bell, & Hughes, 1987; Stanovich, 1986; Wagner, Torgesen, & Rashotte, 1994). The literature abounds with evidence for the relationship between phonological skills and reading development. Both the influence of phonological skills on reading development and vice versa have been found frequently in empirical research (e.g., Goswami & Bryant, 1990; Sawyer & Fox, 1991; Wagner & Torgesen, 1987). The developmental trend toward word recognition supplemented by direct use of orthography suggests that the relationship between individual differences in phonological processing skills and phonological recoding skills is developmentally limited (Stanovich, 1986). Phonological skills become less important in skilled reading because relatively few unknown words are encountered, but they remain important on those occasions when new words occur. A bidirectional relation between the two in the early phases of reading development, and a diminishing influence at later stages, seems therefore plausible (Perfetti et al., 1987; Wagner, 1988; Wagner et al., 1994).

A similar process may be operative for reading and vocabulary. The size of the listening vocabulary will have a minor influence on reading performance when word recognition skills still have to be developed. The size of the vocabulary will not necessarily predict the reading of single words out of context, but it becomes increasingly important for reading of text. Knowing the meaning of words, especially the relatively infrequent key words in a written passage, is fundamental to reading comprehension. Moreover, children with a wider vocabulary are more able to infer the meanings of new words encountered in a text. There is much research demonstrating the influence of vocabulary on comprehension in reading (e.g., Beck & McKeown, 1991; Daneman, 1991; Whyte, 1993), but it is important to notice that the reverse effect of reading on vocabulary development also has been found. Part of the vocabulary growth seems to take place through learning the meanings of previously unknown words encountered in print (Nicholson & Whyte, 1992; Shu, Anderson, & Zhang, 1995). Through reading, one almost continuously extends and refines one's vocabulary. Therefore, independent reading outside school has been credited with most of the yearly vocabulary growth of students in the middle grades (Nagy, Herman, & Anderson, 1985).

Variables that may be involved in reciprocal relations of the first type, that is, operative throughout reading development, are reading practice and motivational factors. A reciprocally facilitating relationship between reading ability and reading experience comes down to the fact that children who are reading relatively well will read more, and because developing the skill of reading requires much practice, they will read better. The model underlying the influence of the amount of exposure to print on individual differences in reading achievement is one of accelerating skill development by means of practice (Cipielewski & Stanovich, 1992; Stanovich, 1986). The frequency of free reading outside of school has consistently been found to relate to word recognition skills, vocabulary, comprehension ability, and other verbal skills (Anderson, Wilson, & Fielding, 1988; Cipielewski & Stanovich, 1992; A. E. Cunningham & Stanovich, 1990, 1993; Stanovich & West, 1989).

Motivation can be a strong and relatively long-lasting mediator of progress in reading and reading practice. Even when the initial development of cognitive ability is similar across individuals, affective aspects of the skills will often produce great developmental diversity (Fischer, Knight, & Van Parys, 1993). Children who are experiencing difficulties in acquiring reading skills have different histories of success and failure than children who are reading at a higher level. This last group is more likely to enjoy reading and, for example, to choose reading as a leisure time activity, which increases their amount of practice (Stanovich, 1986). Research findings indicate that positive attitudes to reading are associated with the amount of book reading during leisure time and with higher levels of reading achievement (Guthrie & Greaney, 1991; Rowe, 1991; Walberg & Tsai, 1983). This may indicate a developmental bidirectional relation in which readers, being motivated to read, increase their reading skills, which successively adds to their inclination to read.

In sum, there are various candidate components for the Matthew effect model (Stanovich, 1986), but direct empirical evidence for the developmental model that is hypothesized to underlie the increase of individual differences in reading is scarce. In our view, empirical evidence for the Matthew effect model is found when the development of individual differences can be described by increasing variance over time in combination with stable rank orderings, and when this development can be linked to a specific developmental pattern.

Empirical Support for the Matthew Model

The line of reasoning expressed in the Matthew model of Stanovich (1986) is intuitively appealing, and, more important, it seems to provide a theoretical model that fits well with a large body of separate research findings. A crucial issue, however, is whether the Matthew model can be tested empirically, that is, whether hypotheses can be formulated that allow for empirical verification or falsification and that serve to raise or diminish the credibility of the model. Although many studies corroborate different aspects of the hypothesized causes of Matthew effects in reading (see, e.g., Boland, 1993; Burstall, 1977; Butler et al., 1985; Juel, 1988; Klicpera & Schabmann, 1993; Nicholson & Whyte, 1992; Perfetti, Beck, Bell, & Hughes, 1987; Rowe, 1991; Schneider & Näslund, 1993; Wagner, 1988; Wagner et al., 1994; Whyte, 1993), the scope of most of these studies was limited to one or a few of the possible factors involved, and the Matthew effect itself, that is, an increase of individual differences, was not detected. To our knowledge, the only study that explicitly focused on empirical support for the Matthew effect is the study of Shaywitz et al. (1995). However, the more interesting developmental pattern that is hypothesized to underlie this effect was not an issue in this study.

In the longitudinal study of Shaywitz et al. (1995), the reading development of a cohort of 400 students was tracked for 7 years. Two different analysis procedures were used on the composite reading scores: standard score and Rasch score analysis. However, the Matthew effect was operationalized in the same way for both types of metrics, that is, by expecting a positive correlation between initial performance level and linear growth in reading. As argued by Shaywitz et al., with a positive correlation reading scores will fan out with time because initially good readers will have higher growth rates than poor beginning readers. The results of this study revealed no evidence to support fan-spread patterns (Cook & Campbell, 1979), that is, the statistical equivalent of one of the aspects of the Matthew effect model. However, some comments can be made about the design of this study. These comments are obviously also relevant for other studies designed to uncover Matthew effects in reading.

The first issue involves the metric that is used to the examine Matthew effects. Preferably, the measures used are on an interval scale and retain the same psychometric characteristics over time so that the variance of these measures can be modeled. The difficulty is that there are hardly any measures that are relevant across developmental levels because the nature of the reading process changes with development. The standard scores used by Shaywitz et al. (1995) are no solution to this problem. Although standard scores can show how the rank orderings change over time, the raw scores at different ages are transformed to a distribution with the same variance. In this way, the increase of variance is lost in the standardization process, and as a result, this aspect of Matthew effects is undetectable. Moreover, patterns of growth based on standard or Rasch score scales are, in general, highly questionable (Hoover, 1984). In an attempt to deal with the metric issue in our own design, we used multiple measures to be able to detect changes in the measurement properties of the tasks used and to model the latent variance, that is, the variance from which task-specific variance is removed. Although this may not be the final solution to the metric problem, latent-variable structural equation models have considerable advantages if one is not solely interested in the Matthew effect but also in the underlying model.

Second, although the Rasch score analysis uses a more appropriate metric, the operationalization of the Matthew effect by means of a linear growth model may not be appropriate. The main objection to a linear growth model is that although it can lead to a fan-spread pattern, that is, it is an appropriate model if one focuses only on the Matthew effect, it is not an appropriate model if one wants to take the underlying developmental model into account. The use of a linear growth model implies that the relation between initial differences in reading performance and growth of reading skills does not change with development. Only perfect stability of individual differences can be represented. In other words, reading development is characterized by complete stability during the intermediate occasions. However, this characteristic is in conflict with models of reading acquisition and most likely with many, if not all, other developmental processes too.

In other words, the linear growth model is nondevelopmental in its nature and therefore not in accordance with the concept of developmentally limited factors and reciprocal causal relations that are hypothesized to cause the effect. Because different factors are assumed to determine individual differences in reading ability at different developmental levels, a linear growth model is not appropriate to describe growth in reading across developmental levels. A nonlinear growth model in which, apart from the transmission of variance from occasion to occasion, new sources of interindividual variance can be incorporated seems to be more appropriate (Bast & Reitsma, 1997). Such a model may be able to account for the prediction that the causes of individual differences in reading change with development.

Third, in the Shaywitz et al. (1995) study, a composite reading score is used as an outcome measure. This composite score is a weighted combination of word recognition and reading comprehension measures. To formulate a growth model, an important condition of the repeated observations has to be met. The individual skill that is supposed to change during development must retain a comparable meaning over the sequence of observations. Questions of whether the measures used are measuring the same concept in all stages of development with the same unit of measurement and the same reliability can and must be tested empirically (Byrne, Shavelson, & Muthén, 1989; W. R. Cunningham, 1991; Labouvie, 1980; O'Brien & Reilly, 1995). However, no results concerning this test for factorial invariance were reported. It is not likely that the word recognition and comprehension measures retain the same contribution to the composite score throughout reading development. This is especially relevant when developmental limits in the development of word recognition and reading comprehension skills are expected.

Testing the Matthew Model

To find empirical support for Matthew effects in reading, the data collection and analysis techniques should fit this developmental model. However, the Matthew effect model must be regarded as an implicit growth theory, that is, no mathematical growth model is specified concerning the nature of how the observed change occurred. In order to derive testable hypotheses from the Matthew model, further specifications of some unique characteristics are needed. On the basis of these specifications, an appropriate growth model has to be chosen.

The first specification implies that candidates for models of growth should incorporate the expected finding of increasing interindividual variance in combination with stability of the rank ordering of individuals. This would mean that the development of individual differences in reading ability can be described by fan-spread patterns. However, although fan spread provides a description of the expected outcome of the development of reading ability, it gives no mathematical description of the underlying growth processes.

Therefore, the selected statistical growth model should also incorporate the expectation that different factors influence the interindividual variance in reading ability at different times, that is, the influences change with development. The key feature of the Matthew effect model is that systematic individual differences are not solely predicted by time-invariant exogenous predictors of change (such as students' home background, students' innate competencies, environment, and treatment), for which the linear growth model would provide an adequate description, but also by continuous interactions of these variables with reading. Because different factors determine individual differences in reading ability at different developmental levels, and because individual growth lines for each participant are not likely to diverge from the same point, a nonlinear growth model may be more appropriate. Apart from the transmission of variance from occasion to occasion, in such a model new sources of interindividual variance can also be incorporated. Moreover, because the nature of the reading process changes with development, a distinction between growth in word recognition skills and in reading comprehension skills is desirable.

Growth implies some type of comparison across time. These comparisons can refer to different features of psychological measures: the observed means, variances, distributions, relations between measures, patterns, and/or the latent counterparts of all of them. In the Matthew model, reading development is viewed from the perspective of individual differences. Patterns of variability across individuals over time are central. It may be unreasonable to expect that change in group means and change in individual differences have the same causes (Bast & Reitsma, 1997; Kenny & Campbell, 1989; Mandys, Dolan, & Molenaar, 1994). Therefore, no growth model for the means is specified, that is, patterns of change in the means are ignored.

There are different types of growth models possible to describe cognitive growth: polynomial growth curves; linear, exponential, and logistic growth; or growth according to a simplex pattern (see, e.g., Bryk & Raudenbush, 1987; Burchinal & Appelbaum, 1991; Goldstein, 1987; Guire & Kowalski, 1979; McArdle & Aber, 1990; Rogosa & Willett, 1985; Thissen & Bock, 1990; Van Geert, 1991). On the basis of the implicit assumptions about growth expressed in the Matthew effect model, an autoregressive or simplex growth model seems to be best suited to formally represent fan spread in combination with developmentally limited causal factors. Moreover, empirical evidence is available to suggest that an autoregressive model must be preferred above other growth models (Bast & Reitsma, 1997).

The autoregressive or simplex model is particularly well suited to longitudinal series in which there is occasion-to-occasion transmission, that is to say, that the observation at Time 2 depends on the observation at Time 1, and in turn, the observation at Time 3 depends on the observation at Time 2, and so forth. This pattern is well described by a first-order autoregressive or Markov process. When measurement error of the manifest variables is included in the model, the simplex property shifts to the latent level. This model is known as the quasi simplex. Jöreskog (1970, 1979; Jöreskog & Sörbom, 1989; Werts, Linn, & Jöreskog, 1978) formulated the quasi-simplex model in the framework of structural equation models.

The quasi-simplex model implies that it is impossible to obtain a perfect prediction of an ability at time $ti + 1$, from knowledge of the ability at time *ti*. This fact is not due to the lack of reliability of the measures but because influences that systematically alter the ability occur between the two times. However, the correlation between two latent variables at two different times can be high when little systematic influence intervenes between the two times of measurement (Horn & McArdle, 1980). If the events between the two times that produced a simplex pattern could be identified and measured, then the pattern itself could be accounted for in a model in which the intervening influences are added to the initial individual differences to give perfect predictability (Horn & McArdle, 1980). In other words, the benefits of the model are enhanced when one combines autoregression with other variables in a cross-lagged design (McArdle & Aber, 1990).

This article is concerned with the analysis of the development of individual differences in terms of Matthew effects in reading. The present study includes a 3-year longitudinal cross-lagged design in which the development of word recognition skills and reading comprehension skills was measured along with a comprehensive set of relevant factors including characteristics of the home environment and preschool skills, as well as cognitive, behavioral, and motivational characteristics of the participants.

Method

Sample

A sampling scheme with several stages was used in which first a sample of schools was selected and then students were selected from each school. Out of all elementary schools (approximately 1,500) in the wide area of the cities of Amsterdam and Utrecht, a pool of 200 schools was selected that fitted the following two criteria. The first requirement was that schools were representative for all Dutch elementary schools with regard to the proportion of educationally disadvantaged students, students from a non-Dutch background, and class size. Second, only schools that used *Veilig Leren Lezen* (VLL; *Learning to Read Safely*; Mommers, Verhoeven, & Van der Linden, 1990) for reading instruction in the early grades were selected. At the time this study was carried out, this method to teach early reading was used in 75% of all Dutch elementary schools. This selection resulted in 40 schools willing to participate.

Once schools and final kindergarten classes had been selected for participation, each senior kindergarten teacher was asked to nominate a group of 7 children. This procedure resulted in a group of 280 children, consisting of potentially successful beginning readers, average readers, and students who would initially not meet expectations in reading progress. In prospect of the distribution of a standard word recognition test, children with really well-developed beginning reading skills (the upper 25% of the distribution) or with very poor readiness skills (approximately the lowest 10% of the distribution) had to be excluded from the study. Groups of students were excluded from the sampling procedure in order to arrive at a sample of students that could be tracked over the first three grades. Children at risk for nonpromotion or referral to special education were excluded to minimize sample attrition. Children expected to be among the best readers of the group were excluded in an attempt to prevent problems with the ability range of the measurement instruments.

Of the 280 children who began the study in kindergarten, 235 (84%) were evaluated until the end of Grade 3. At the first time of measurement, the mean age of this sample (121 boys and 114 girls) was 74 months (SD = 4 months, minimum = 64, maximum = 88). Sample attrition was due to replacement of students in other schools. The Dutch educational system has separate schools for children with learning and/or behavioral difficulties. A number of the sampled students were referred to these special schools during the study. Other students left with their parents to other cities.

After 3 months of instruction in reading in Grade 1, an assessment battery was administered to the total population of Grade 1 classes of the selected schools. In this manner, the adequacy of the teachers' selections could be determined by comparing the performance of the selected students with the rest of the sample on standardized measures of word recognition skills, receptive vocabulary, and measures of phoneme blending and segmentation. The performance of the target students on the phonological and vocabulary measures was comparable with the performance of the total sample and of reference groups reported in test manuals. The selection of participants was expected to yield a sample of students reading at or just above or under the average level for Grade 1, with less-skilled readers being slightly overrepresented. Comparing the final sample with several reference groups, we were led to the conclusion that the selection procedure had resulted in the intended sample of students. The selected students were tracked through subsequent classroom assignments effected by the schools' normal administrative procedures without influence from the research team.

Procedure

We used multiple instruments within each domain, so that multiple yearly assessments were made of each student's development in reading and of skills, attitudes, and behaviors hypothesized to be related to reading. Of key interest was the development of reading ability in the first three grades of primary education. To track this development, we selected repeated measures of word recognition and reading comprehension. A second group of variables consisted of skills, attitudes, and behaviors hypothesized to be related to reading. Repeated measures of vocabulary, phonological skills, leisure time reading activities, and attitudes towards reading were chosen. A third group of variables consisted of characteristics of the child before the start of formal reading instruction in Grade 1. This set of variables included measures for the general cognitive level of the child, in this case, vocabulary and nonverbal IQ, and emergent reading skills such as letter and word recognition and reading-related phonological skills. In addition to biographical information such as age and gender, some information concerning the child's home background was gathered.

The longitudinal design consisted of seven measurement occasions within a 3-year period. The first wave of data was gathered in June 1991, just before the start of formal education in reading in Grade 1 when all of the children were enrolled in senior kindergarten classes. Because of the amount of change expected to occur in Grade 1, the next three measurements took place in November, March, and June, that is, after about 3, 6, and 9 months of instruction in reading. The data gathering continued with two assessments in Grade 2 (March and June) and was completed in June 1994 with an assessment at the end of Grade 3.

For the seven measurement occasions, a total of 49 tests were used to represent the variables discussed above. The majority of these tests (43) were individually administered; the other 6 tests were group administered to the children. All of the tests were administered by research assistants who received training to ensure standardized assessment procedures. Because we tried to keep each testing session to 30 min or less, tests were administered in two or three sessions on separate days for each measurement occasion. The tasks were administered to all of the children in a fixed order designed to maintain the child's interest, give priority to the most critical measures, and alternate more and less challenging tasks. Before starting each new task, children were provided with practice items. The assignments were not given until the child understood the task requirements. Feedback and modeling were provided for the demonstration and practice trials only. For some of the tests, there were exit rules. When the child failed a certain number of items, the task was terminated. Once the criterion for exit was met, performance on subsequent items was not assessed, assuming failure.

Measurement Occasion 1: Preschool Instruments

Home literacy composite. This scale consisted of nine questions, each scored on a 3-point scale by one of the parents of the selected children. Three questions were about the frequency of parental reading behavior. One question was about the reading of books in general (less than 1 hr, less than 3 hr, or more than 3 hr in an average week), one question was about reading newspapers and/or magazines (less than 1 hr, less than 3 hr, or more than 3 hr in an average week), and one question was about the frequency of reading related to a person's occupation or hobby (less than 1 hr per day, 1 to 2 hr per day, or more than 2 hr per day). Four questions were about the amount of reading material present in the house: the number of books purchased (less than 1 book, 1 book, or more than 1 book a month) and the number of books present in the house (less than 10, less than 50, or more than 50) as well as the number of newspaper and magazine subscriptions (none, one, or more than one) were recorded. The remaining two items rated the frequency of visiting a library with the child (never, occasionally, always) and the frequency of borrowing books from the library (less than 1 book, 1 book, or more than 1 book per week).

Phonological skills. One phoneme blending test (Klanksynthesetoets B; Sas & Wieringa, 1983) and two segmentation tests (Auditieve Woordanalysetoets A en B; Sas & Wieringa, 1983) were administered to determine each child's ability to segment words into phoneme sounds and to blend isolated sounds into words. The first segmentation test consisted of 6 two-phoneme words, 7 three-phoneme words, and 2 four-phoneme words. The second test consisted of 4 four-phoneme words and 6 five-phoneme words. For both tests on each trial, the experimenter pronounced a word and asked the child to segment words into phoneme sounds. There was an exit rule of five consecutive misses. The phoneme blending test consisted of 4 two-phoneme words, 7 three-phoneme words, and 3 four-phoneme words. On each trial, the experimenter gave the constituent phonemes separated by a 1-s pause and asked the child to blend the isolated sounds into words. There was an exit rule of four consecutive misses. Scoring consisted of the total number of words that were correctly analyzed or synthesized. The same tests were used on the second measurement occasion in November of Grade 1.

Verbal working memory. One test was used to measure verbal working memory (Leidse Diagnostische Tests [LDT] Zinnen Nazeggen; Schroots & Van Alphen de Veer, 1976). Children listened to 12 sentences of different length and were asked to repeat them verbatim. Scoring consisted of the number of 124 target words repeated correctly.

Nonverbal IQ. The children were administered 15 items of the LDT Blokpatronen (Schroots & Van Alphen de Veer, 1976). The test assistant provided the child with 15 different two-dimensional patterns of red and white colored squares and asked the child to reproduce those patterns within a certain time limit. Scoring consisted of two points for every item correctly completed within this time limit and one point for a correct item beyond this time limit.

Receptive vocabulary. The children were administered 100 items from the Dutch adaptation of the Peabody Picture Vocabulary Test (Manschot & Bonnema, 1978). The children were told that they would be looking at four picture alternatives while the experimenter said a word out loud. Their task was to choose the picture that best described the meaning of the word the experimenter had spoken out loud.

Letter knowledge and word recognition. Upon entering the study, each student's prereading skills were assessed with two orthographic knowledge tests. The first test assessed the number of 26 lowercase alphabet letters the children could identify by giving the letter-sound or letter-name associations. The second test assessed the number of words they could read out of 13 simple words that are frequently used in beginning reading materials.

Measurement Occasions 2–4: Grade 1 Instruments

Word recognition. All of the tests administered to measure word recognition skills made use of real words as stimuli instead of pseudowords or nonwords. The tasks consisted of sounding out words. Knowledge of the meaning of the word was not assessed. At the first measurement occasion in Grade 1, the first word recognition test used (Caesar Eén-Minuut-Test; Mommers, 1983) consisted of 100 words becoming progressively more difficult to decode. The words of this test were selected from the VLL reading series used at the participating schools. The child was instructed to read aloud the words separately as fast and accurately as possible. Scoring consisted of the number of correct responses within a time limit of 1 min. The second measure (DMT1—the first list of words of the Drie-Minuten-Test; Verhoeven, 1992) consisted of 116 consonant–vowel–consonant (CVC) words. This test was also used on the subsequent measurement moment. The third word recognition test (DMT2—the second list of words of the Drie-Minuten-Test; Verhoeven, 1992) consisted of 116 words with consonant clusters at the beginning or end of the word. This word recognition measure was also used on every subsequent measurement occasion. All of the other word recognition tests that complemented this repeated measure differed only in the set of stimuli used. From the third until the last measurement moment, the Brus Eén-Minuut-Test, Form A or B (Brus & Voeten, 1973) was added as a word recognition measure. This test consists of 116 unrelated regular words that become progressively more difficult. Lists are ordered by difficulty, ranging from simple monosyllabic CVC patterns to polysyllabic items containing blends, digraphs, and vowel variations. For the last two waves (June of Grade 2 and June of Grade 3), the third list of words of the Drie-Minuten-Test (DMT3; Verhoeven, 1992), consisting of 100 more difficult polysyllabic items, was added.

Vocabulary. In November, two tests for receptive vocabulary were group administered to the children. The first test (Taaltoets voor Kinderen Woordenschat Keuze [TVK-WS Keuze]; Van Bon, 1982) consisted of 40 items. The second measure (CITO-Woordenschat, 1991) consisted of 60 items. Both of the tests had the same task format as the Peabody Picture Vocabulary Test. Administration of these tests was repeated in March of Grade 1, March and June of Grade 2, and June of Grade 3. In June of Grade 1, two different tests were used. The first receptive vocabulary test consisted of 26 multiple-choice items (Woordenschattest; Aarnoutse, 1988a). The second productive vocabulary test consisted of 30 items (TVK-WS Produktie; Van Bon, 1982).

Attitudes toward literacy. A 30-item test (Leesattitudeschaal; Aarnoutse, 1988b) was orally administered to the students. Questions concerning attitudes toward reading at home, at school, or during vacations, as well as questions about visiting the library, could be answered by "yes" or "no." Administration of this test was repeated in June of Grade 2 and June of Grade 3.

Reading comprehension. At the end of Grade 1, one reading comprehension test was administered to the children. The test (BELL Form B; Van den Bos, 1992) consists of 39 unrelated sentences that become progressively more difficult. For instance, the wording of one of the first sentences is *John sits on his bicycle,* compared to the last sentence, *When the umpire called the foul, the fanatic fans questioned his decision, by showing their discontent in a dubious way.* After reading each sentence, the children had to choose the one picture out of four that best reflected the meaning of the sentence. Because of confusion about the right alternative for Item 24, this item was removed from the scale score. There was an exit rule of three wrong answers out of the last four items. The BELL (Form B or the parallel Form A) was administered as a reading comprehension test with the same exit rule at every subsequent measurement occasion.

Measurement Occasions 5 and 6: Grade 2 Instruments

The word recognition, vocabulary, attitudes toward reading, and reading comprehension measures were all used in Grade 1 and therefore already have been described in the previous section. Only one new measure for receptive vocabulary was added. The measure for receptive vocabulary (Woordenschattest; Aarnoutse, 1988a) consisted of 26 items. The child had to complete a very short sentence by choosing one word out of four possibilities.

Measurement Occasion 7: Grade 3 Instruments

The word recognition, vocabulary, attitudes toward reading, and reading comprehension measures were all used in Grades 1 and 2 and have been described earlier.

Leisure Time Reading Grades 1–3

The frequency of book reading and comic book reading during leisure time and the frequency of being read to by one of the parents or caretakers at home were assessed from Grades 1–3. The measures were not designed to provide estimates of absolute amounts of time spent reading but as indicators of relative individual differences in exposure to print. During the individual testing sessions, the children were asked questions about leisure time reading activities of the day before. The questions were "Did you read a story yesterday at home?" "Did you read a comic book yesterday at home?" and "Were you read to yesterday at home?" The child could answer "yes" or "no." In Grade 1, these questions were administered three times in March and three times in June. In May, the same questions were group administered four times by the classroom teachers. In Grade 2, seven assessments took place: two times in March, two times in May, and three times in June. In Grade 3, the questions were repeated another three times, supplemented by two questions on leisure time reading behavior in general. The question "How often do you read books at home? (no comics)" as well as "How often do you read comic books or comic magazines at home?" were scored using one of the following response categories: hardly ever, once a month, once a week, two or three times a week, and almost every day.

Statistical Analyses

All analyses were performed within a structural equation modeling framework using the LISREL Version 8.0 computer program (Jöreskog & Sörbom, 1993). The maximum likelihood estimation procedure of the BMDP package (Dixon, 1988) was used for the imputation of missing data (3%), before computing the covariance matrices used to fit the models.

The general structural equation model consists of two parts: the measurement model and the structural equation model. The measurement model specifies how latent variables are related to the observed variables. Because multiple measures of our constructs were administered, we were able to construct latent variables. Latent variables capture the common variance among their indicators and exclude or minimize variance attributable to measurement error or task-specific strategies. In this way we can get more accurate estimates of relations among constructs.

The structural equation model specifies the relationships among the latent variables that are of primary interest. The structural part of the model provides estimates of the unique influence of each latent variable on another latent variable and assigns the total amount of latent variance that can be accounted for by the specified relations.

A structural equation model thus consists of a set of observed, latent, and error variables; parameters that link these types of variables together; parameters specifying the hypothesized structural effects; parameters for variances of and the covariances among the latent variables; and parameters for the variances and covariances of the measurement errors and equation residuals, that is, the variance of the structural part of the model that cannot be explained by the model.

A particular structural equation model is specified by fixing, that is, assigning specified values to the parameters; constraining, that is, specifying parameters to be equal to one or more other parameters; or setting free the parameters of the model. This specification must lead to a model that is identified, that is, a unique solution for the parameters in the model must exist. After specification of a model, the hypothesis is tested that the population covariance matrix of observed variables is equal to the covariance matrix written as a function of the model parameters. The estimation procedure involves finding estimated values for the parameters such that they produce a covariance matrix that differs as little as possible from the observed covariance matrix.

The fit statistic is distributed as chi-square. A nonsignificant value for chi-square indicates a nonsignificant discrepancy between the model and the data. The chi-square statistic is a test of the null hypothesis of exact fit. Browne and Cudeck (1993) proposed to replace this hypothesis by a less implausible interval hypothesis of close fit. The hypothesis that the root mean square error of approximation (RMSEA) is less than or equal to 0.05 is tested, indicating a close model fit. Another measure for the overall fit of the model to the data used in the present study is the normed fit index (NFI). This measure has a value between 0 and 1, with values above .90 indicating acceptable fit (Bentler & Bonett, 1980).

When these fit measures indicate that model fit is unacceptable, a respecification of the model is needed. When the respecified model is nested under the first model, that is, when its set of freely estimated parameters is a subset of those estimated in the first model, the sequential chi-square difference tests procedure can be used. The null hypothesis of no significant difference between the two nested models is tested. The difference between the chi-square statistic values for nested models is itself distributed as chi-square, with degrees of freedom equal to the difference in degrees of freedom for the two models (Anderson & Gerbing, 1988). A large drop in chi-square, compared to the difference in degrees of freedom, indicates that the changes made in the model represent a real improvement. In this way, the best-fitting model can be found.

Every analysis procedure started with maximum-likelihood confirmatory factor analyses on the set of observed measures. These preliminary analyses were performed to explore the data set and determine whether any modifications of the hypothesized measurement structure were needed before hypotheses about the relations between the latent variables were tested.

A series of models was tested for the development of individual differences in word recognition and reading comprehension skills. These models were expected to show two things: (a) that the development of individual differences in reading ability can be characterized by an increase of latent variance over time in combination with stable rank orderings, and (b) that this increase is the result of a specific developmental pattern of which the key features are reciprocal causation and developmental limits.

These two predictions were tested by means of latent-variable simplex growth models. (See Appendix A for more details on the simplex model.) By comparing different specifications of the structural part of these models, it can be determined to what degree the interindividual variance in reading skill at a later time is dependent on the same reading variance at a prior time, that is, stability of individual differences, and to what degree this latent variance increases over time. Moreover, by specifying different lead–lag relationships, that is, the effects of latent variables on a previous time on other latent variables at a later time, it can be determined to what degree the development of the latent reading variance can be explained by these other variables. In the present study, lead–lag relationships were considered suggestive of causal determination.

Results

Descriptive Statistics and Preliminary Analyses

Means and standard deviations for the tasks at each measurement moment are presented in Table 1. Covariances among the variables are presented in Appendixes B and C. The reliabilities of the measures ranged between .64 and .94 with a mean of .88. Exceptions were the test for productive vocabulary with a reliability of .58, and one of the tests for receptive vocabulary (TVK; Cronbach's α = .45). As a result of the way the tests were scored and processed, no internal consistency measures were available for the word recognition measures. However, the split-half and parallel-forms reliabilities reported in the test manuals were above .90 for Grades 1–3 (Mommers, 1983; Van den Bos, Lutje Spelberg, Scheepstra, & De Vries, 1994; Verhoeven, 1992). As could be expected, some variables had a nonnormal distribution of scores caused by ceiling and floor effects. Transformations of these variables improved their distributions but did not change their correlations with other variables. Therefore, all subsequent analyses were performed on raw scores.

As recommended by Anderson and Gerbing (1988) and others, structural equation modeling started with a series of maximum-likelihood confirmatory factor analyses to explore the data set and determine whether any modifications of the hypothesized factor structures were needed before hypotheses about the relations between the latent variables could be tested. For every measurement occasion, an a priori factor structure was specified. In this model, the number of expected factors was determined, and a particular pattern of factor loadings (λs) was specified. For multiple-indicator factors, the factor loadings of one of the measures was fixed at 1 to identify the model; correlations between factors were allowed, but no correlations between the error variances of the measures were allowed. For single-indicator factors, it is typical to fix the measurement error for this one indicator at a plausible value. In the present investigation, for example, for models with a single indicator of reading comprehension, the measurement error was fixed at .1 *SD*, that is, equivalent to a reliability estimate of .9 (Anderson & Gerbing, 1988).

These a priori specified factor structures were tested against other factor models that were theoretically plausible, by means of the chi-square difference procedure. The results of the confirmatory factor analyses were generally consistent with our hypotheses about the nature of the factors underlying performance on this set of measures. In general, the fit of the models was good, as indicated by nonsignificant chi-square statistics and relatively high values for the NFI. However, there were some differences between the a priori specified factor patterns and the results. All of the model changes were necessary to improve model fit or to arrive at the most parsimonious model. In kindergarten and March of Grade 1, one phonological processing skills factor had to be preferred above a two-factor solution with a separate phoneme blending and phoneme segmentation factor.

Table 1

Means and Standard Deviations for All Measures at the Various Occasions of Measurement (N = 235)

Measure	Max.	Kindergarten		Nov. Grade 1		March Grade 1		June Grade 1		March Grade 2		June Grade 2		June Grade 3	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Verbal working memory	124	100.60	14.61												
Receptive vocabulary (Peabody)	100	59.87	8.13												
Nonverbal IQ	30	20.03	3.64												
Letter knowledge	26	3.94	4.43												
Word recognition	13	0.96	1.71												
Home literacy	27	19.35	3.31												
Phoneme blending	14	4.31	4.08	11.51	3.09										
Phoneme segmentation A	15	3.89	4.62	12.42	3.09										
Phoneme segmentation B	10	0.68	1.99	2.69	3.08										
Word recognition (EMT)	100			13.31	4.26										
Word recognition (DMT1)	116			8.26	4.05	27.57	11.90								
Word recognition (DMT2)	116			1.37	2.80	13.66	8.13	25.04	14.27	49.61	18.45	55.99	19.51	70.98	18.21
Word recognition (Brus A/B)	116					15.53	6.70	21.00	9.05	40.54	12.29	42.77	12.90	58.29	13.12
Word recognition (DMT3)	100											40.26	16.18	58.02	17.03
Reading comprehension	38							11.89	5.37	17.09	5.94	17.76	5.45	23.38	5.39
Receptive vocabulary (CITO)	60			44.84	6.69	46.50	5.73			51.55	4.55			55.58	3.28
Receptive vocabulary (TVK-WS)	40			25.09	4.15	27.02	3.39					30.77	3.25	32.24	3.07
Receptive vocabulary (Aarnoutse)	26							21.17	3.35						
Receptive vocabulary (Aarnoutse)	35											25.27	6.20		
Productive vocabulary	30							19.88	3.73						
Attitudes toward reading	30					20.08	5.94					19.32	5.21	20.57	5.99
Leisure time reading	18							5.56	3.44			7.52	3.67	8.26	4.17

Note. The 5 different receptive vocabulary tests are differentiated by adding the reference in parentheses (see the Method section of the text). Nov. = November; Max. = maximum; EMT = Caesar Eén-Minuut-Test; DMT1, DMT2, and DMT3 = the first, second, and third list, respectively, of the Drie-Minuten-Test.

Table 2
Latent Variables at Different Measurement Occasions

	Grade 1			Grade 2		Grade 3
Domain	November	March	June	March	June	June
Phonological skills	X					
Vocabulary	X	X	X	X	X	X
Word recognition	X	X	X	X	X	X
Reading comprehension			X	X	X	X
Attitudes toward reading		X			X	X
Leisure time reading[a]			X		X	X

[a] The leisure time reading variables describe the frequency of free reading during Grades 1, 2, and 3, respectively.

For the leisure time reading variables, a three-factor solution, with one factor for every grade, provided a better description of the data than a six-factor solution, with one factor for the separate months in which the data were gathered.

In sum, the preliminary analyses resulted in 29 factors or latent variables. The six latent variables based on the kindergarten measures were home literacy, nonverbal IQ, reading knowledge, vocabulary, verbal working memory, and phonological processing skills. The other 23 latent variables used in the present study are summarized in Table 2.

Latent-Variable Modeling

Word recognition and phonological processing skills. The first hypothesis to be tested concerned the development of individual differences in word recognition. Could this development be described by stability of individual differences in combination with increasing latent variance? And, more important, to what extent was this development the result of developmentally limited reciprocal causal relation between word recognition and phonological processing skills. Table 3 summarizes the results of the various models fitted to the data.

In the measurement part of the first model, the factor loadings of one of the word recognition measures (DMT2) were set to 1 to identify the model, and the other factor loadings were freely estimated. The common assumption of uncorrelated errors of measurement may be too stringent for analyses of longitudinal data in which the same instruments are completed by the same participants on multiple occasions. If the same measurements are used on multiple occasions, the corresponding error variables will tend to be correlated. Therefore, to get accurate estimates of relations among the latent variables, correlations among the errors of identical word recognition measures at adjacent occasions must be included.

In the structural part of the model, an autoregressive or simplex structure, that is, the effect of a latent variable at a prior time on the same latent variable at a later time (see also Appendix A), was specified for the latent word recognition and phonological processing skills variables.

Moreover, because it is not realistic to expect that the kindergarten variables are totally unrelated correlations between the kindergarten latent variables, reading knowledge, phonological processing skills, vocabulary, nonverbal IQ, verbal working memory, and home literacy were included in the model specification. This first model served as a baseline model. The fit of this model can be found in Table 3 (Model 1). Other model specifications can be compared with this model by means of the sequential chi-square difference tests procedure.

In the measurement part of the second model, the freely estimated factor loadings of the word recognition measures were constrained to be equal. This means that the word recognition factor was required to be measured as the same weighted linear combination of word recognition measures at every time of testing. When this would be the case, the metric problem would be resolved in this study because the same metric would be used at every time of testing. However, the restriction did not hold, as indicated by a significant deterioration of model fit,

Table 3
Phonological Processing Skills (P) and Word Recognition (W): Structural Models

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	RMSEA	*p*	NFI
1. Simplex models	565.90	324				.056	.084	.92
2. Equality constraints	665.14	330	99.24	6	<.001	.066	<.001	.90
3. Exogenous variables→W_1	524.10	317	41.80	7	<.001	.053	.270	.92
4. Deletion of insignificant effects	532.55	322	8.45	5	.133	.053	.270	.92
5a. $W_1 \leftrightarrow P_1$	451.47	319	81.08	3	<.001	.042	.930	.93
5b. $P_1 \rightarrow W_1$	453.50	320	79.05	2	<.001	.042	.930	.93
5c. $W_1 \rightarrow P_1$	451.48	320	81.07	2	<.001	.042	.940	.93

Note. RMSEA = root mean square error of approximation; NFI = normed fit index.

$\Delta\chi^2(6, N = 235) = 99.24$, $p < .001$. This means that the results concerning the increasing variance must be interpreted with some caution. (See Appendix D for more detailed information.)

The next step (Model 3) was adding to the model the lead–lag effect of the kindergarten variables (reading knowledge, vocabulary, phonological processing skills, nonverbal IQ, verbal working memory, and home literacy) on the subsequent latent word recognition variable, and the effect of the kindergarten reading knowledge factor on the Grade 1 phonological processing skills factor. These effects were estimated to see whether a specific relation between phonological skills and reading exists.

The overall fit of the model improved significantly compared to Model 1, $\Delta\chi^2(7, N = 235) = 41.80$, $p < .001$. However, only the phonological processing skills factor and the reading knowledge factor had a significant effect on the adjacent word recognition factor. None of the other variables had a significant effect on the first word recognition factor. Also, the effect of kindergarten reading knowledge on first-grade phonological skills was insignificant. All nonsignificant effects were deleted in the fourth model.

In Model 5a the bidirectional relation between word recognition and phonological skills in November Grade 1, and the effect of the first-grade phonological factor on subsequent word recognition, was estimated. Although the overall fit of the model improved considerably, $\Delta\chi^2(3, N = 235) = 81.08$, $p < .001$, this model had to be rejected as an adequate description of the data because the bidirectional effect between phonological skills and word recognition was insignificant. Therefore, a model with a unidirectional effect of the phonological skills factor on the first word recognition factor (Model 5b), and a model with the reverse effect (Model 5c), were estimated. Although Model 5c turned out to be the best-fitting model, the overall fit was only slightly better than Model 5b. However, this model provided a significantly better chi-square than the fourth model. The difference in chi-square was 81.07 in relation to a difference of 2 in degrees of freedom. The RMSEA had a value of .042 (p = .940) for the final model, and the NFI was .93, both indicating a close model fit.

Table 4 shows the parameter estimates of primary interest. Maximum likelihood estimates of all other parameters can be found in Appendix D.

The stability of individual differences is expressed as the correlation between two latent variables adjacent in time. In general, there was a large degree of stability of individual differences. The only exception was the low stability for the first to the second measurement moment. As can be seen in Table 4, the variance of the latent word recognition variables increased with time, except for the last occasion where a decrease was detected.

However, the model was expected to show not only the combination of high stability and increasing latent variance but also that this increase was in part the result of the interrelations between variables. In the first model, only an autoregressive or simplex structure was specified. The final model included the interrelations between word recognition and phonological processing skills. By comparing proportion of variance explained by the first and the final model, it can be concluded that a significant part of the increasing variance was the result of these interrelations between word recognition and phonological processing skills. The percentage of variance explained increased from 27% to 60% for the Grade 1 phonological skills variable. The effect of phonology on word recognition explained 14% and 4% of the variance of the first two word recognition variables.

Table 4
Stability of Individual Differences and Development of Latent Variance for Phonological Processing Skills (P) and Word Recognition (W)

Latent variable	Latent variance	Latent variance explained (%): First model	Final model	Correlation (η_i, η_{i+1})
P	18.57	—	—	—
P_1	4.08	27	60	.34
W_1	3.67	0	14	—
W_2	58.80	18	22	.23
W_3	194.87	77	77	.88
W_4	311.72	67	67	.82
W_5	347.59	93	93	.96
W_6	290.71	87	87	.93

Note. Subscripts to the latent variables refer to the occasion of measurement. P without a subscript was measured in kindergarten. The correlation (η_i, η_{i+1}) represents the stability coefficient of individual differences. Dashes indicate that the variables were not measured at a previous occasion.

Figure 1 summarizes these interrelations in a path diagram. In this figure only the structural part of the model is depicted. Following typical convention, the ovals represent latent variables at different points in time; the paths between the latent variables represent path coefficients. The path coefficients for a given variable (e.g., word recognition, March of Grade 1) represent the predicted change in a latent variable that is associated with a one-unit change in the other variable (e.g., phonological skills, November of Grade 1) when the values of the other variables in the model are constant. In other words, the path coefficients provide estimates of the unique influence of each cause.

Substantive interpretations of the solution depicted in the path diagram suggest several findings of interest. First, after the control for the autoregressive effects (i.e., the effects of the word recognition and phonological processing skills variables measured on a previous time on the same variables at a later time), the effects of vocabulary, nonverbal IQ, verbal working memory, and home literacy on the first latent word recognition variable (W1) were nonsignificant. Only a significant lead–lag relationship (.11), suggestive of causal determination, between the kindergarten phonological processing skills factor and the adjacent word recognition factor was found.

Second, after 3 months of instruction in reading (November of Grade 1) there seemed to be a unidirectional influence of word recognition skills on phonological processing skills (.63).

Third, the Grade 1 phonological skills factor in turn had an effect (1.03) on the March Grade 1 word recognition factor. The influence of the phonological skills factor on subsequent word recognition factors (W3–W6) was not significant.

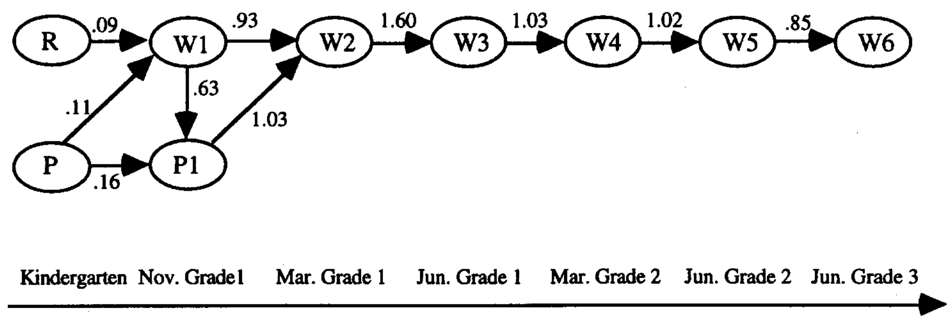

Figure 1. Path diagram of the structural model with phonological processing skills (P), reading knowledge (R), and word recognition skills (W). Nov. = November; Mar. = March; Jun. = June.

In summary, the solutions of the structural equation models indicate that, after other plausible causes have been omitted (autoregressive effects, verbal ability, and allowing for bidirectional relations), strong relationships between phonological skills and word recognition skills exist. The pattern of interrelationships between phonology and word recognition shows that during the onset of reading development, phonological skills have a unidirectional effect on subsequent word recognition, and word recognition skills have a unidirectional effect on phonological skills. This pattern of results can be interpreted as an indication for a developmentally limited reciprocal causal relationship between the two factors. Moreover, the pattern of interrelationships partly explains the development of word recognition skills, which can correctly be described as a fan-spread pattern, that is, increasing individual differences that preserve participants' ordering.

Reading comprehension, word recognition, vocabulary, attitudes toward reading, and leisure time reading. The second set of hypotheses about interrelations between variables that may cause Matthew effects in reading involved the relation between vocabulary, word recognition skills, comprehension in reading, attitudes toward reading, and leisure time reading activities. The same procedures were used as described in the previous section on the first set of hypotheses. Measurement of the development of comprehension in reading started at the fourth measurement occasion at the end of Grade 1. As mentioned before, the leisure time reading factors did not describe how often the students read books and comic books at home for the separate months, but for the entire period in each grade. In the present study, the leisure time reading factors were regarded as a measure of reading practice. Attitudes toward reading were measured once at the end of each grade.

Analyses started with the specification of quasi-simplex models for the vocabulary, decoding, reading comprehension, attitudes toward reading, and leisure time reading variables. The fit of this first model specification can be found in Table 5. The estimation of the first model led to a negative estimate for the variance of the equation residual of the sixth latent vocabulary variable. This improper solution was probably caused by the fact that the value of the parameter in the population was very close to zero. In this situation, a sample estimate may assume an inadmissible value due to sampling fluctuations (Bollen, 1989). Because the estimate for this variance did not depart significantly from zero, this parameter was fixed at zero.

The second model added the effect of the kindergarten home literacy factor on the November of Grade 1 vocabulary factor and on the first reading comprehension factor, and the lead–lag effects of the word recognition and vocabulary factors on the adjacent reading comprehension factors. The fit of the model increased significantly compared to the multivariate simplex models, as can be seen in Table 5, $\Delta\chi^2(9, N = 235) = 217.54$, $p < .001$. All effects were significant except for the effect of the fifth word recognition factor (June of Grade 2) on subsequent comprehension in reading (June of Grade 3) and the expected effect of the home literacy factor on the Grade 1 reading comprehension factor. Moreover, the parameter estimates of the autoregressive effect of the Grade 1 on the Grade 2 reading comprehension factor and of the effect of the Grade 2 on the Grade 3 reading comprehension factor turned into insignificant values. All insignificant effects were deleted in the third model.

In the fourth model, the interrelations between word recognition, leisure time reading, attitudes toward reading, and vocabulary were added. The effects of word recognition and attitudes toward reading on leisure time reading were estimated. Moreover, the effects of word recognition on attitudes toward reading and the effects of leisure time reading activities on vocabulary were added to the model. The fourth model fitted the data significantly better than the previous model, $\Delta\chi^2(12, N = 235) = 132.23$, $p < .001$. However, not every estimated effect was significant. The word recognition factor at the end of second grade did not have an effect on the frequency of leisure time reading activities in third grade. Also, the autoregressive effect of the first on the second attitude factor, as well as the effect of the first word recognition factor on the first attitudes toward reading factor, were not significant. Moreover, the effects of the first and last leisure time reading factors on vocabulary were not significant. As before, all nonsignificant effects were deleted from the model. After these adjustments, model fit did not decrease significantly, $\Delta\chi^2(5, N = 235) = 8.00$, $p = .156$, implying that the more parsimonious model fitted the data equally well. The interval hypothesis of close fit, tested by means of the RMSEA, could not be rejected (RMSEA = .049, $p = .670$), indicating a close model fit.

Table 5
Word Recognition (W), Vocabulary (V), Leisure Time Reading (L), Attitudes Toward Reading (A), Home Literacy (H), and Reading Comprehension (RC): Structural Methods

Model	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	RMSEA	*p*	NFI
1. Simplex models	1,621.63	843				.063	<.001	.81
2. $W_{t-1}\rightarrow RC_t$; $V_{t-1}\rightarrow RC_t$; $H\rightarrow V_1$; $H\rightarrow RC_3$	1,404.09	834	217.54	9	<.001	.054	.088	.84
3. Deletion of insignificant effects	1,413.50	838	9.41	4	.051	.054	.081	.84
4. W_1 and $A_2\rightarrow L_3$; W_3 and $A_5\rightarrow L_5$; W_5 and $A_6\rightarrow L_6$; $W_1\rightarrow A_2$; $W_4\rightarrow A_5$; $W_5\rightarrow A_6$; $L_3\rightarrow V_3$; $L_5\rightarrow V_5$; $L_6\rightarrow V_6$	1,281.27	826	132.23	12	<.001	.049	.680	.85
5. Deletion of insignificant effects	1,289.27	831	8.00	5	.156	.049	.670	.85

Note. Subscripts to the latent variables refer to the occasion of measurement. RMSEA = root mean square error of approximation; NFI = normed fit index.

Table 6 shows the parameter estimates of the final model regarding stability of individual differences, expressed as the correlation between two latent variables adjacent in time, and the development of latent variance. Maximum likelihood estimates of all other parameters can be found in Appendix D. As can be seen in Table 6, there was no systematic increase of latent reading comprehension variance. Moreover, the stability of individual differences was low. In other words, no fan-spread pattern was found for comprehension in reading. A large degree of stability of individual differences was found for vocabulary. It should be noted that for vocabulary the latent variance was not based on a common metric of the manifest variables. Different tests were used for different measurement moments with no possibilities for test equation. Therefore, an interpretation of the development of the latent variance was not warranted. The latent variance of the leisure time reading factor increased every grade. In combination with the moderate stability of individual differences, this means that the gap in the frequency of leisure time reading widened.

Table 6
Stability of Individual Differences and Development of Latent Variance for Reading Comprehension (RC), Vocabulary (V), Attitudes Toward Reading (A), and Leisure Time Reading (L)

Latent variable	Latent variance	Latent variance explained (%)		Correlation
		Model 1	Model 5	(η_i, η_{i+1})
RC_3	25.08	0	36	—
RC_4	31.49	16	25	.26
RC_5	25.60	20	40	.41
RC_6	25.93	18	29	.27
V_1	21.56	34	37	—
V_2	22.85	96	94	.97
V_3	4.24	88	90	.95
V_4	0.98	85	86	.92
V_5	4.10	75	83	.89
V_6	6.65	100	100	1.00
A_2	1.00	0	0	—
A_5	1.00	2	5	—
A_6	2.51	34	40	.58
L_3	0.33	0	10	—
L_5	0.94	39	52	.56
L_6	0.95	38	38	.53

Note. Subscripts to the latent variables refer to the occasion of measurement. The correlation (η_i, η_{i+1}) represents the stability coefficient of individual differences. Dashes indicate that the coefficients could not be calculated because no previous measurements were available.

Figure 2 shows the path diagram of the structural part of the final model. To assist interpretation, the common metric standardized solution is presented. Each latent variable was standardized to a mean of zero and unit variance. This means that, for example, a 1-unit increase in vocabulary in March of Grade 1 led to a corresponding increase of 0.32 in reading comprehension at the end of Grade 1. The amount of the latent variance explained by the final model is shown in Table 6.

First, the path diagram indicates that there were lead–lag relationships between word recognition and vocabulary and comprehension in reading. At the end of Grade 1, the effect of the word recognition factor on reading comprehension was much stronger (.50) than the effect of the vocabulary factor (.32). However, the effect of word recognition on comprehension decreased with time and was absent at the end of Grade 3. In contrast, the effect of vocabulary on reading comprehension became stronger with every grade. As can be seen in Table 6, these variables explain a considerable part of the latent variance of comprehension in reading. In turn, students' vocabulary knowledge was affected by the home literacy factor (.19). The effect of home literacy on subsequent vocabulary explained an additional 3% of the variance.

Second, lead–lag relationships were found between word recognition and leisure time reading activities in first and second grade. However, the effect of attitudes toward reading on the frequency of leisure time reading was stronger and more consistent. The direct effects of attitudes toward reading and word recognition together explained, respectively, 10% and 13% of the variance of the frequency of leisure time reading activities during Grade 1 and Grade 2. Moreover, the Grades 2 and 3 attitudes toward reading factors were affected by word recognition skills. The direct effects of the word recognition factors explained, respectively, 3% and 6% of the variance of the last two attitudes toward reading factors. Furthermore, a positive effect of the second-grade leisure time reading factor on vocabulary knowledge (.19) was found.

In summary, the solutions of the structural equation models indicate that the level of word recognition skills was associated with positive attitudes toward reading. This implies that good

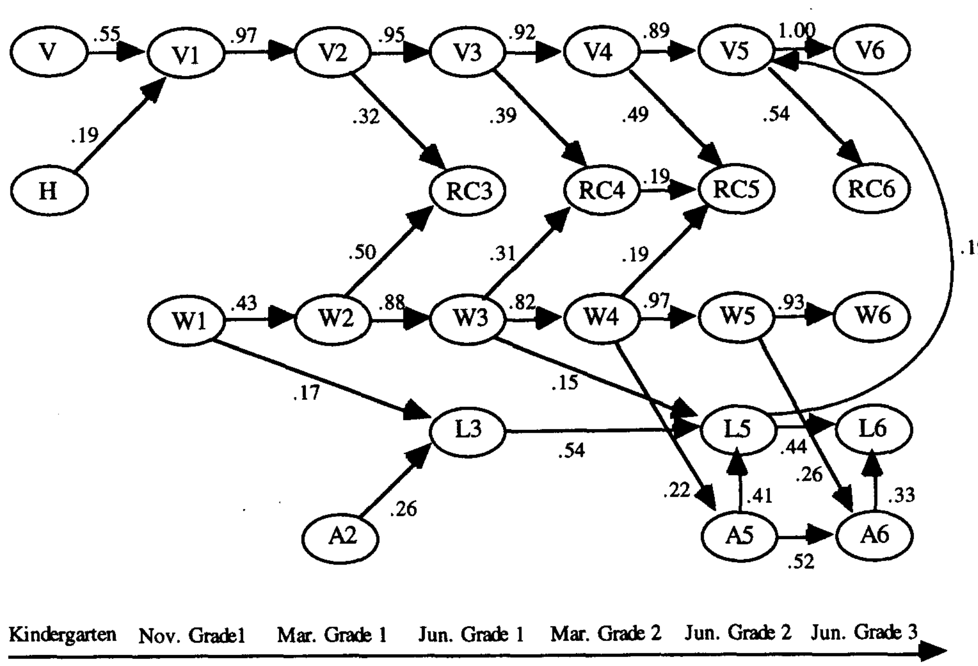

Figure 2. Path diagram of the structural model with home literacy (H), vocabulary (V), reading comprehension (RC), word recognition (W), leisure time reading activities (L), and attitudes toward reading (A). Nov. = November; Mar. = March; Jun. = June.

readers express more positive attitudes toward reading than do poor readers. Furthermore, children with positive attitudes toward reading, and a good word recognition level, tended to read more frequently during leisure time. Moreover, the frequency of independent reading outside school had an impact on vocabulary growth of students at the end of second grade. In view of the consistent effect of vocabulary on subsequent comprehension in reading in the first three grades, it is reasonable to expect that this relation will continue to operate in succeeding grades. Together with the effect of word recognition on leisure time reading and leisure time reading on vocabulary, this pattern of interrelationships can be interpreted as a reciprocal causal relationship between reading development and vocabulary development.

However, this pattern of interrelationships does not lead to the expected increasing individual differences for comprehension in reading. For the latent construct reading comprehension, no fan-spread pattern could be established for this time period. Although there was a small increase of variability from the end of Grade 1 to March of Grade 2, no further enlargement of differences between readers was detected after this time period. In contrast to the development of word recognition skills, the results for reading comprehension show that the hypothesized developmental pattern of the Matthew effect model did not automatically lead to increasing differences.

Discussion

The present study was designed to answer the question of how individual differences in reading ability come into existence. To answer this question, we empirically described the development of differences between readers during the first three grades of primary education. Initially, there were almost no differences with regard to reading ability between children involved in the present study. Most of the selected children could not read a single word before receiving formal reading instruction in Grade 1. However, significant differences in reading ability among the same children were found at the end of Grade 3. This observation led to several questions: How do individual differences in reading ability develop with time? What kind of growth model underlies this development? Which factors are related to the development of individual differences in reading ability, and what is the relative importance of these factors?

The theoretical framework used to describe and explain the development of individual differences in reading ability was the Matthew effect model of Stanovich (1986). In the present study, the development of individual differences in word recognition skills and reading comprehension skills were described separately. For both skills, structural equation models were expected to show an increase of interindividual variability in combination with high stability. They also were expected to show that these increasing differences could be explained by a specific pattern of interrelationships between a comprehensive set of relevant factors.

The Development of Individual Differences in Word Recognition

The results of the present research clearly indicate increasing individual differences in word recognition. In general, the absolute difference in word recognition skills increased with time.

Moreover, the rank ordering of individuals was stable. In other words, a Matthew effect was found for word recognition, that is, initially poor readers remained poor readers during the first three grades and the performance gap relative to good readers became larger over the course of development.

Two remarks concerning this result must be made. First, it should be noted that the latent variance of word recognition skills decreased at the end of Grade 3. However, because the operationalization of word recognition involved not only accuracy but also speed, a limitless increase of performance differences was not realistic. Developmental limits in the development of reading skills were expected. However, when certain levels of word recognition were reached, further enlargement of individual differences could shift to other components of the reading process, such as comprehension skills.

Second, the stability of individual differences seemed to be rather low at the very beginning of reading development. This means that although the absolute performance differences were rapidly increasing, students' rank ordering was changing too. Individual growth lines for each participant seemed not to diverge from the same point. A possible explanation for this low stability is the type of instruction provided in the early stages of reading development, in combination with the time of testing. Because of the large number of schools participating in this study, and the restricted number of test assistants, children of different schools were not tested in the same week and this may have caused some instability of individual differences. During testing, an inventory was made of curricular progress, that is, the relative amount of curricular content of the reading series covered after 3 months of instruction in reading. The data revealed that there were considerable differences among schools with regard to progress in the curriculum and the amount of instruction given. Especially in the early stages of the reading development, a 3–4-week difference in formal instruction in reading can make quite a difference. It could therefore be hypothesized that the autoregressive effect of the first word recognition factor was limited by differences among schools in testing periods and amount of instruction provided. If this assumption is correct, the interaction between school membership and initial growth in word recognition skill would be significant, and this effect would diminish at later occasions of testing. Analysis of variance performed on the repeated measures with school membership as a between-subjects variable did indeed reveal a significant interaction between school membership and growth in word recognition from the first to the second measurement occasion, $F(1, 39) = 1.58$, $p = .023$. For growth in word recognition from the second to the third occasion, no significant interaction effect with school membership was found, $F(1, 39) = 1.34$, $p = .102$.

The development of individual differences in word recognition could adequately be described by an autoregressive or simplex growth model. The typical property of the simplex structure is that the sizes of correlations between measures collected at adjacent occasions are large and decrease systematically as a function of the number of occasions separating two repeated measures. The present data indeed indicate that performances at, for instance, the end of Grade 2 were primarily dependent on performances measured at the prior occasion in Grade 2, and not on performances measured at, for instance, the beginning of Grade 1. Furthermore, the results of the simplex models indicate that there was no complete transfer of variance from occasion to occasion. These results indicate that factors that determine initial skill levels do not necessarily also determine the continued progress of the students in reading. New sources of variance can enter into the development at different times.

The Development of Individual Differences in Reading Comprehension

The expected increase of differences in reading comprehension was not found. Although absolute performance differences in reading comprehension increased from the end of Grade 1 to March of Grade 2, no further enlargement of these differences was detected after this time period. Moreover, the stability of individual differences was low throughout this time period. In other words, no Matthew effects were found for reading comprehension, that is, there was no systematic increase of performance differences and the best performances were made by different students at different times.

A simplex growth model appears to provide a good description of the development of differences in reading comprehension. However, when other variables related to this development were taken into account, some of the autoregressive effects dropped out of this growth model. Individual differences in reading comprehension at the end of Grade 1 were not directly related to those differences in Grade 2. In turn, variability in Grade 2 was not directly related to differences in Grade 3.

It must be acknowledged that the present null finding with regard to Matthew effects for reading comprehension may be due to the fact that no data were obtained after the end of Grade 3. However, reading comprehension continues to develop well beyond this point in time. It would therefore be interesting to see what effects could be obtained when children are followed for an extra couple of years.

The results of the present study—a Matthew effect for word recognition but not for reading comprehension—are in sharp contrast with the study of Shaywitz et al. (1995), where no Matthew effects in reading could be reported. This difference is probably due to the fact that Shaywitz et al. made use of an inappropriate growth model and of a composite reading score. As indicated by the present results, a clear distinction between the development of differences in word recognition and differences in reading comprehension must be made. Thus, combining word recognition and reading comprehension into one single composite reading measure will lead to confounded results with regard to the development of individual differences in reading.

Testing the Matthew Effect Hypothesis

In the present study, two aspects of the Matthew effect hypothesis have been operationalized: increasing differences between readers, and developmentally limited reciprocal causal relationships between reading and other factors. A considerable part of the latent word recognition and reading comprehension variance could not be explained by the underlying growth model. However, the most important assumption of the Matthew effect hypothesis is that the phenomenon of increasing achievement differences is to a great extent caused by a specific developmental pattern of interrelations between reading skills and other variables. In the present study, some of these relationships could be identified.

First, phonological processing skills seemed to enter in a developmentally limited reciprocal relation with word recognition. Differences in phonological skills determine a considerable part of the individual variation in word recognition skill, and vice versa (word recognition skill affects differences in phonological skills). It should be noted that the reciprocal causal relationships between reading and phonological skills could have been demonstrated more convincingly if we had continued our measurement of these later skills after the second measurement moment. However, in the present study, phonological processing skills still could be identified as an influential factor with regard to the initial development of word recognition skills.

The fact that word recognition skills affected the development of phonological skills, even after the autoregressive effect of this variable was taken into account, is in agreement with the notion that phonological processing skills are not the natural consequence of speech production and perception (Morais, Cary, Alegria, & Bertelson, 1979). Because there are no phonemes in the speech stream, the reader has to learn that orthographic symbols are representations of phonemes. Learning to read in an alphabetic language seems to provide the needed instruction. The present results are also in agreement with the study of Wagner, Torgesen, and Rashotte (1994), in which they also found a causal effect of phonological processing skills on word recognition. Moreover, although Wagner et al.'s effect of letter knowledge on phonological processing skills could not be replicated, in the present study the effect of word recognition on phonological skills could be demonstrated, allowing a similar general conclusion that the causal relations between word recognition skills and phonological skills are bidirectional.

Although phonological skills seem to be necessary, they are not sufficient for word recognition. This was demonstrated by the large amount of variance in word recognition skills that remains unexplained. However, it should be noted that the phonological tasks used in the present study covered only relatively explicit awareness of phonemes. It may be that explicit phonemic awareness is not, as such, the precondition for the acquisition of reading skills, but it is the cognitive capacity for becoming aware of phonemic segments during the first stages of the learning process that is crucial for the acquisition of word recognition skills (Morais et al., 1979). This capacity may be indicated by more shallow levels of phonological awareness, such as awareness of alliterations, rhyme, or onset-rhyme components, or may be related to the quality of phonological representations (Elbro, 1996; Fowler, 1991; Wesseling & Reitsma, 1998). No such variables were included in the present study though.

Second, at the end of Grade 1, comprehension in reading was mainly determined by word recognition skills. Poor readers were to a large extent also poor comprehenders. In Grade 2, differences in reading comprehension were due to differences in word recognition and vocabulary. Receptive vocabulary seemed to have no influence on word recognition skills, but at later stages of the reading development it appeared to have a major effect on individual differences in reading comprehension processes. The effect of vocabulary increased in second grade (compared to the effect of word recognition) and was the only significant predictor of reading comprehension at the end of Grade 3. When vocabulary was taken as a measure for general language ability, these results supported the "simple view of reading" (Gough & Tunmer, 1986; Hoover & Gough, 1990). The simple view model holds that both word recognition and listening comprehension are necessary for reading success, but neither is sufficient by itself. However, the proportion of reading comprehension variance that could be explained by the two components in this study was much lower than the results reported by Hoover and Gough (1990), indicating that other factors may have been important too.

Third, home literacy, the frequency of leisure time reading activities, and attitudes toward reading were expected to be important candidates to add to the prediction of differences in reading ability. Home literacy appeared not to have a direct effect on reading. However, home literacy was positively related to vocabulary, indicating that a stimulating home environment with regard to literacy provides children with a richer vocabulary. Thus, with vocabulary as an intervening variable, home literacy had an indirect effect on reading comprehension. In the same manner, no direct effects of leisure time reading and attitudes toward reading were found. In contrast, a reciprocal relationship seemed plausible for reading and vocabulary, mediated by the frequency or volume of reading during leisure time. The results indicate that good readers tended to read more frequently during leisure time than poor readers. These leisure time reading activities were related to differences in the size of the vocabulary at the end of second grade. In turn, vocabulary affected subsequent comprehension in reading.

As an explanation, one could suggest that part of the vocabulary growth takes place through learning the meanings of previously unknown words encountered during leisure time reading activities. As hypothesized by Nicholson and Whyte (1992), only above-average readers seem to gain significantly in vocabulary by incidental learning. Good readers have a deeper knowledge base and a wider vocabulary than average and poor readers and are therefore more able to infer the meanings of new words. To arrive at this conclusion, the assumption has to be made that new words are learned during free reading. However, Carver (1994) showed that the assumption that leisure time reading leads to substantial growth of the vocabulary may not be valid because free reading is likely to involve relatively easy reading materials that contain few, if any, unknown words. Thus, the exact nature of the relationship between reading and vocabulary growth is an interesting topic for future research.

The level of word recognition skill was also associated with positive attitudes toward reading. This means that good readers expressed more positive attitudes toward reading than poor readers. Furthermore, children with positive attitudes toward reading tended to read more frequently during leisure time. These results are in agreement with other studies (e.g., Guthrie & Greaney, 1991; Rowe, 1991; Walberg & Tsai, 1983) with respect to the positive association between attitudes toward reading and the amount of book reading during leisure time, but they fail to replicate the direct effect on levels of reading achievement.

Although confirmation was found for some of the predicted interrelationships between reading and reading-related variables, other relationships that can be derived from the Matthew effect model could not be confirmed. First, is the relationship between phonology and word recognition, as predicted by the Matthew effect model, developmentally limited? In cognitive developmental models of growth in reading ability, it is assumed that beginning readers go through different phases in which they adopt different word recognition strategies, from logographic reading to a final stage in which alphabetic reading is supplemented by orthographic word recognition (Ehri, 1991; Rayner & Pollatsek, 1989; Wimmer & Goswami, 1994). The developmental trend suggests that the relationship between individual differences in phonological awareness and phonological recoding skills is developmentally limited. The results of the present research indicate that the influence of phonological skills on word recognition skills diminishes with time. It should be noted that the issue of developmental limits could not be fully addressed in the present study because the phonological tasks were only administered at the first two moments of measurement. Moreover, the claim expressed in the Matthew model—that the relationship between reading ability and attitudes toward reading and leisure time reading (as a measure for reading practice) is operative throughout reading development—could not be affirmed. However, these relationships may be restricted to reading development in the middle grades, where comprehension processes become more dominant than word recognition processes and a higher frequency of reading for fun is more likely.

Second, an implication of the Matthew effect model is that not only differences in reading ability but also differences in reading-related skills such as, for instance, vocabulary will increase. In other words, as a consequence of the notion of reciprocal causal relationships, the increase of interindividual variability not only refers to the different components of reading ability but also to the development of individual differences in skills related to reading. Although relationships that are suggestive of reciprocal causation were found, in general no increase of differences between students with regard to reading-related skills was found. Of the selected reading-related abilities, an increase of interindividual variance was found only for leisure time reading activities. In contrast to the claims of the Matthew effect model, no increasing differences regarding attitudes toward reading and vocabulary were found.

The considerable stability of individual differences in reading performance from the beginning of the first to the end of the third grade suggests that the progress made by children in the reading lessons in first grade is decisive for long-term success. A possible reason is that the learning prerequisites to a great extent determine not only early success but also the continued progress of the students in reading. It is not known to what extent differences in learning prerequisites can be attributed to constitutional factors.

An alternative assumption is that the children are blocked in their further development by their initial modest success in learning to read (Klicpera & Schabmann, 1993). This explanation is supported by the results of the present study. An important implication of the finding that a simplex growth model underlies the development of reading skills is that the factors that determine initial skill levels do not necessarily also determine the continued progress of the students in reading. New sources of variance enter into the development of individual differences. A few of these subprocesses have been identified in the present study. These sources of variance can probably be influenced by teachers and parents who train children in these subprocesses. But, as stated by Rutter (1983) and Snow (1989), the only way to substantially reduce individual differences in scholastic attainment is to severely restrict and impair the schooling of the most advantaged students. If this is true, educational attempts to decrease differences between readers must be the wrong objective. A better goal would be, of course, to lead less able students to levels of (functional) reading skill that are minimally required in present-day society.

References

Aarnoutse, C. A. J. (1988a). *Begrijpend leestest en leesattitudeschaal bestemd voor groep 4 van het basisonderwijs* [Test for Reading Comprehension and Reading Attitude Scale for Grade 2 of elementary school]. Nijmegen, The Netherlands: Berkhout.

Aarnoutse, C. A. J. (1988b). *Begrippentest (woorden) en woordenschattest: Tests voor begrijpend lezen bestemd voor groep 3 van het basisonderwijs* [Test for Concepts, Vocabulary, and Reading Comprehension for Grade 1 of elementary school]. Nijmegen, The Netherlands: Berkhout.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103,* 411–423.

Anderson, R. C., Wilson, P. T., & Fielding, L. G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly, 23,* 285–303.

Bast, J. W., & Reitsma, P. (1997). Matthew effects in reading: A comparison of latent growth curve models and simplex models with structured means. *Multivariate Behavioral Research, 32,* 135–167.

Beck, I., & McKeown, M. (1991). Conditions of vocabulary acquisition. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 789–814). New York: Longman.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Boland, T. (1993). The importance of being literate: Reading development in primary school and its consequences for the school career in secondary education. *European Journal of Psychology of Education, 8,* 289–305.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. Scott Long (Eds.), *Testing structural equation models* (pp. 136–162). London: Sage.

Brus, B. T., & Voeten, M. J. M. (1973). *Een Minuut Test: Verantwoording en handleiding* [One Minute Decoding Test, Form A and B: Rationale and manual]. Nijmegen, The Netherlands: Berkhout.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin, 101,* 147–158.

Burchinal, M., & Appelbaum, M. I. (1991). Estimating individual developmental functions: Methods and their assumptions. *Child Development, 62,* 23–43.

Burstall, C. (1977). The Matthew effect in the classroom. *Educational Research, 21,* 19–25.

Butler, S. R., Marsh, H. W., Sheppard, M. J., & Sheppard, J. L. (1985). Seven-year longitudinal study of the early prediction of reading achievement. *Journal of Educational Psychology, 77,* 349–361.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105,* 456–466.

Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behavior, 26,* 413–437.

Cipielewski, J., & Stanovich, K. E. (1992). Predicting growth in reading ability from children's exposure to print. *Journal of Experimental Child Psychology, 54,* 74–89.

CITO-Woordenschat. (1991). *Toets voor woordenschat aan het begin van de basisschool* [Vocabulary test at the beginning of elementary school]. Arnhem, The Netherlands: CITO.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Chicago: Rand McNally.

Cunningham, A. E., & Stanovich, K. E. (1990). Assessing print exposure and orthographic processing skill in children: A quick measure of reading experience. *Journal of Educational Psychology, 82,* 733–740.

Cunningham, A. E., & Stanovich, K. E. (1993). Children's literacy environment and early word recognition subskills. *Reading and Writing, 5,* 193–204.

Cunningham, W. R. (1991). Issues in factorial invariance. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 106–113). Washington, DC: American Psychological Association.

Daneman, M. (1991). Individual differences in reading skills. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 512–538). New York: Longman.

Dixon, W. J. (1988). *BMDP statistical software manual* (Vol. 2). Berkeley: University of California Press.

Ehri, L. C. (1991). Development of the ability to read words. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 383–417). New York: Longman.

Elbro, C. (1996). Early linguistic abilities and reading development: A review and a hypothesis. *Reading and Writing: An Interdisciplinary Journal, 8,* 453–485.

Fischer, K. W., Knight, C. C., & Van Parys, M. (1993). Analyzing diversity in developmental pathways: Methods and concepts. In R. Case & W. Edelstein (Eds.), *The new structuralism in cognitive development: Theory and research on individual pathways* (pp. 33–56). Basel, Switzerland: Karger.

Fowler, A. E. (1991). How early phonological development might set the stage for phoneme awareness. In S. Brady & D. Shankweiler (Eds.), *Phonological processes in literacy. A tribute to Isabelle Y. Liberman* (pp. 97–118). Hillsdale, NJ: Erlbaum.

Goldstein, H. (1987). *Multilevel models in educational and social research.* London: Charles Griffin.

Goswami, U., & Bryant P. (1990). *Phonological skills and learning to read.* Hove, England: Erlbaum.

Gough, P., & Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7,* 6–10.

Guire, K. E., & Kowalski, C. J. (1979). Mathematical description and representation of developmental change functions on the intra- and inter-individual levels. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 89–100). New York: Academic Press.

Guthrie, J. T., & Greaney, V. (1991). Literacy acts. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 68–96). New York: Longman.

Hoover, H. D. (1984). The most appropriate scores for measuring educational development in the elementary schools: GE's. *Educational Measurement: Issues and Practices, 3,* 8–14.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing, 2,* 127–160.

Horn, J. L., & McArdle, J. J. (1980). Perspectives on mathematical/statistical model building (MASMOB) in research on aging. In L. W. Poon (Ed.), *Aging in the 1980s* (pp. 503–541). Washington, DC: American Psychological Association.

Jöreskog, K. G. (1970). Estimation and testing of simplex models. *British Journal of Mathematical and Statistical Psychology, 23,* 121–145.

Jöreskog, K. G. (1979). Statistical estimation of structural models in longitudinal developmental investigations. In J. R. Nesselroade & P. B. Baltes (Eds.), *Longitudinal research in the study of behavior and development* (pp. 303–351). New York: Academic Press.

Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7—User's reference guide.* Chicago: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8—User's reference guide.* Chicago: Scientific Software International.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from 1st through 4th grade. *Journal of Educational Psychology, 80,* 437–447.

Juel, C. (1994). *Learning to read and write in one elementary school.* New York: Springer-Verlag.

Kenny, D. A., & Campbell, D. T. (1989). On the measurement of stability in over-time data. *Journal of Personality, 57,* 445–481.

Klicpera, C., & Schabmann, A. (1993). Do German-speaking children have a chance to overcome reading and spelling difficulties? A longitudinal survey from the second until the eighth grade. *European Journal of Psychology of Education, 8,* 307–323.

Labouvie, E. W. (1980). Identity versus equivalence of psychological measures and constructs. In L. W. Poon (Ed.), *Aging in the 1980s* (pp. 493–502). Washington, DC: American Psychological Association.

Mandys, F., Dolan, C. V., & Molenaar, P. C. M. (1994). Two aspects of the simplex model: Goodness of fit to linear growth curve structures and the analysis of mean trends. *Journal of Educational and Behavioral Statistics, 19,* 201–215.

Manschot, W., & Bonnema, J. T. (1978). *Handleiding Nederlandse normering van de Peabody Picture Vocabulary Test* [Manual and Dutch norms of the Peabody Picture Vocabulary Test]. Lisse, The Netherlands: Swets & Zeitlinger.

McArdle, J. J., & Aber, M. S. (1990). Patterns of change within latent variable structural equation models. In A. Von Eye (Ed.), *Statistical methods in longitudinal research, Volume I: Principles and structuring change* (pp. 151–224). San Diego, CA: Academic Press.

Mommers, M. J. C. (1983). *Caesar-één-minuut-test: Handleiding en verantwoording* [Caesar-One-Minute-Decoding Test: Manual and rationales]. Tilburg, The Netherlands: Zwijsen.

Mommers, M. J. C., Verhoeven, L., & Van der Linden, S. (1990). *Veilig Leren Lezen* [Learning to read safely]. Tilburg, The Netherlands: Zwijsen.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a consequence of phones arise spontaneously? *Cognition, 7,* 323–331.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20,* 233–253.

Nicholson, T., & Whyte, B. (1992). Matthew effects in learning new words while listening to stories. In D. J. Leu & C. K. Kinzer (Eds.), *Literacy research, theory, and practice: Views from many perspectives* (pp. 499–503). Chicago: National Reading Conference.

O'Brien, R. M., & Reilly, T. (1995). Equality in constraints and metric-setting measurement models. *Structural Equation Modeling, 2,* 1–12.

Perfetti, C. A. (1985). *Reading ability.* New York: Oxford University Press.

Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33,* 283–319.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading.* Englewood Cliffs, NJ: Prentice-Hall.

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50,* 203–228.

Rowe, K. J. (1991). The influence of reading activity at home on students' attitudes towards reading, classroom attentiveness and reading achievement: An application of structural equation modelling. *British Journal of Educational Psychology, 61,* 19–35.

Rutter, M. (1983). School effects on pupil progress: Research findings and policy implications. *Child Development, 54,* 1–29.

Sas, J. F., & Wieringa, C. J. (1983). *Het deelvaardigheidsonderzoek ten behoeve van het voorbereidend en aanvankelijk lees—en spellingonderwijs* [Tests of subskills for use in preparational and beginning instruction in reading and spelling]. Groningen, The Netherlands: Wolters-Noordhoff.

Sawyer, D. J., & Fox, B. J. (Eds.). (1991). *Phonological awareness in reading: The evolution of current perspectives.* New York: Springer-Verlag.

Schneider, W., & Näslund, J. C. (1993). The impact of early metalinguistic competencies and memory capacity on reading and spelling in elementary school: Results of the Munich Longitudinal Study on the Genesis of Individual Competencies (Logic). *European Journal of Psychology of Education, 8,* 273–287.

Schroots, J. J. F., & Van Alphen de Veer, R. J. (1976). *Leidse diagnostische tests: Handleiding* [Leiden Diagnostic Tests: Manual]. Lisse, The Netherlands: Swets & Zeitlinger.

Shaywitz, B. A., Holford, T. R., Holahan, J. M., Fletcher, J. M., Stuebing, K. K., Francis, D. J., & Shaywitz, S. E. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study. *Reading Research Quarterly, 30,* 894–906.

Shu, H., Anderson, R. C., & Zhang, H. (1995). Incidental learning of word meanings while reading: A Chinese and American cross-cultural study. *Reading Research Quarterly, 30,* 76–95.

Snow, R. E. (1989). Aptitude, instruction and individual development. *International Journal of Educational Research, 13,* 869–881.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21,* 360–407.

Stanovich, K. E., & West, R. F. (1989). Exposure to print and orthographic processing. *Reading Research Quarterly, 24,* 402–433.

Thissen, D., & Bock, R. D. (1990). Linear and nonlinear curve fitting. In A. von Eye (Ed.), *Statistical methods in longitudinal research, Volume II: Time series and categorical longitudinal data* (pp. 289–318). San Diego, CA: Academic Press.

Van Bon, W. H. J. (1982). *Taaltests voor kinderen: Handleiding* [Language tests for children: Manual]. Lisse, The Netherlands: Swets & Zeitlinger.

Van den Bos, K. P. (1992). *De BELL-Test. Begrijpend lezen en—luisteren. Experimentele versie* [The BELL-Test for reading and listening comprehension, Experimental version]. Groningen, The Netherlands: Rijksuniversiteit.

Van den Bos, K. P., Lutje Spelberg, H. C., Scheepstra, A. J. M., & De Vries, J. R. (1994). *De Klepel: Een test voor de leesvaardigheid van pseudowoorden—Verantwoording, handleiding, diagnostiek en behandeling* [The Klepel: Test for the decoding of pseudowords—Rationale, manual, diagnosis, and treatment]. Nijmegen, The Netherlands: Berkhout.

Van Geert, P. (1991). A dynamic systems model of cognitive and language growth. *Psychological Review, 98,* 3–53.

Verhoeven, L. (1992). *Drie-Minuten-Toets* [Three Minutes Decoding Test]. Arnhem, The Netherlands: CITO.

Wagner, R. K. (1988). Causal relations between the development of phonological processing abilities and the acquisition of reading skills: A meta-analysis. *Merrill-Palmer Quarterly, 34,* 261–279.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin, 101,* 192–212.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology, 30,* 73–87.

Walberg, H. J., & Tsai, S. L. (1983). Matthew effects in education. *American Educational Research Journal, 20,* 359–373.

Werts, C. E., Linn, R. L., & Jöreskog, K. G. (1978). A simplex model for analyzing academic growth. *Educational and Psychological Measurement, 37,* 745–756.

Wesseling, R. F. J. Q., & Reitsma, P. (1998). *Quality of phonological representations in prereaders and reading acquisition.* Manuscript submitted for publication.

Whyte, J. (1993). Longitudinal correlates and outcomes of initial reading progress for a sample of Belfast boys. *European Journal of Psychology of Education, 8,* 325–340.

Williamson, G. L., Appelbaum, M., & Epanchin, A. (1991). Longitudinal analyses of academic achievement. *Journal of Educational Measurement, 28,* 61–76.

Wimmer, H., & Goswami, U. (1994). The influence of orthographic consistency on reading development: Word recognition in English and German children. *Cognition, 51,* 91–103.

Appendix A

Quasi-Simplex Models for the Present Longitudinal Data

In the present study the development of the latent reading variables is described by the quasi-simplex model. In the quasi-simplex model the observed variables at time i (Y_i) are assumed to be related to their corresponding latent variables (η_i) and measurement error (ϵ_i) by the equation

$$Y_i = \eta_i + \epsilon_i. \quad \text{(A1)}$$

The simplex structure among the latent variables can be stated as

$$\eta_{i+1} = B_i\eta_i + \zeta_{i+1}, \quad \text{(A2)}$$

where B_i is the true regression weight and ζ_i are the residuals of the latent variables. The covariance matrix implied by the quasi-simplex model is defined by the following equations:

$$\text{var}(y_i) = \text{var}(\eta_i) + \text{var}(\epsilon_i), \quad \text{(A3)}$$

$$\text{var}(\eta_{i+1}) = \beta^2_{i+1,i}\,\text{var}(\eta_i) + \text{var}(\zeta_{i+1}), \quad \text{(A4)}$$

$$\text{var}(\eta_1) = \text{var}(y_1), \quad \text{(A5)}$$

and

$$\text{cov}(\eta_i, \eta_{i+1}) = \beta_{i+1,i}\,\text{var}(\eta_i). \quad \text{(A6)}$$

The first assumption of the Matthew effect hypothesis is that the development of individual differences in reading ability can be described by fan-spread patterns. As can be derived from Equations A4 and A6, increasing latent variance in combination with stable rank orderings is only obtained when the variance of the equation residuals ζ (known as Ψ) is relatively small and β is greater than one. For example, when the latent variance at the second measurement occasions is 20, $\beta_{2,3}$ connecting the second with the third latent variable is 1.5, and Ψ_3 is 5, following Equation A4, the latent variance at the third occasion will be $1.5^2 \cdot 20 + 5 = 50$. In this case, the stability of individual differences, expressed as the correlation between the two latent variables is .95 following Equation A6: $\text{cov}(\eta_2, \eta_3)/\sqrt{(\text{var}(\eta_2) \cdot \text{var}(\eta_3))}$.

The selected statistical growth models should also incorporate the expectation that different factors influence the interindividual variance in reading ability at different times, that is, the influences change with development. In the case of the multivariate simplex model, different lead–lag effects are expected for a restricted time period. In the present research lead–lag relationships, that is, the effects of latent variables at a previous time on other latent variables at a latter time, are considered suggestive of causal determination. The value of the lead–lag relationships between exogenous (ξs) and adjacent endogenous latent variables (ηs) are represented by coefficient γ. The lead–lag relationships between two succeeding endogenous latent variables (ηs) are represented by coefficient β. The lead–lag effects are expected to explain part of the variance of the equation residuals (Ψ) of the simplex models for the development of individual differences in reading ability.

(*Appendixes continue*)

Appendix B

Observed Covariance Matrix: Kindergarten Measures, Phonological

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13
1	9.57												
2	4.25	9.49											
3	5.26	4.30	9.55										
4	5.24	6.44	6.11	16.38									
5	4.70	6.38	5.52	14.34	18.15								
6	2.03	4.16	2.29	7.52	7.08	7.85							
7	6.26	5.80	7.23	11.21	8.70	5.32	44.88						
8	9.94	8.56	10.83	18.59	14.53	7.89	71.45	141.71					
9	6.43	5.89	7.52	11.92	8.77	5.61	49.25	86.30	66.20				
10	7.42	6.78	6.57	15.76	12.28	7.76	49.26	86.55	57.43	81.91			
11	10.04	10.93	9.43	25.33	21.30	12.62	78.76	138.06	93.65	120.34	203.80		
12	7.86	6.89	6.24	15.76	15.69	8.01	55.50	104.43	63.54	82.48	135.63	151.27	
13	11.40	9.84	9.07	22.94	22.65	12.29	78.95	152.02	93.31	121.58	201.39	207.29	340.56
14	8.19	6.57	5.54	16.89	14.89	8.66	53.25	101.82	62.02	85.75	139.33	141.31	209.90
15	11.67	8.97	8.92	22.00	20.52	10.82	69.22	128.35	79.58	108.30	177.89	176.29	265.08
16	13.07	9.47	8.99	24.18	22.94	12.50	74.63	146.11	88.29	116.82	195.24	206.37	319.66
17	6.84	5.15	5.54	15.94	14.93	7.21	45.98	91.67	52.63	72.95	119.39	137.14	200.61
18	8.89	7.52	6.64	21.57	19.72	11.34	59.10	116.96	67.52	98.90	156.28	177.56	265.65
19	7.40	7.26	4.58	19.20	18.38	8.73	59.04	115.96	68.62	94.39	156.54	176.37	276.03
20	3.60	5.30	5.03	5.56	5.07	3.17	7.79	9.82	6.88	7.41	13.88	8.14	11.23
21	0.94	2.27	1.33	1.65	1.64	1.03	2.16	2.77	1.84	2.03	4.24	1.98	2.80
22	3.03	5.71	4.79	3.91	3.38	3.12	6.25	7.93	5.28	4.30	7.39	6.32	7.87
23	2.45	3.75	2.56	4.91	4.18	2.94	6.00	7.83	5.00	6.43	11.08	8.55	11.16
24	0.47	0.57	0.76	0.83	0.34	0.47	1.33	1.73	1.59	1.82	3.62	1.83	2.44
25	0.43	0.76	−3.04	−0.81	−0.33	−0.89	2.61	0.96	1.92	1.12	−1.38	4.69	11.15
26	2.22	5.77	4.61	7.33	8.02	3.09	21.96	33.05	20.25	20.03	31.82	32.82	50.29
27	1.19	0.34	0.56	2.23	2.38	0.60	5.91	10.83	6.57	9.28	12.25	11.70	16.41
28	0.46	0.74	0.62	0.67	0.66	0.70	1.09	1.96	0.89	3.51	3.74	5.66	8.26

Note. November of Grade 1: 1 = phoneme segmentation A, 2 = phoneme segmentation B, 3 = phoneme blending, 4 = DMT1, 5 = EMT, 6 = 13 = DMT2; June of Grade 2: 14 = Brus B, 15 = DMT3, 16 = DMT2; June of Grade 3: 17 = Brus A, 18 = DMT3, 19 = DMT2; Kindergarten: vocabulary, 26 = verbal working memory, 27 = nonverbal IQ, 28 = home literacy. DMT1, DMT2, and DMT3 = the first, second, and third list, respectively.

Processing Skills, and Word Recognition (N = 235)

14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
166.50														
190.90	262.17													
228.69	287.79	381.17												
143.77	177.15	222.46	174.63											
190.85	241.19	296.04	205.28	290.43										
192.50	239.67	313.55	214.26	284.74	331.95									
8.48	10.15	11.41	4.31	6.65	4.00	21.34								
2.24	2.01	3.02	1.29	1.88	0.39	6.50	3.95							
6.72	9.76	8.95	4.00	5.54	2.48	13.16	4.20	16.64						
10.29	11.74	10.77	7.67	10.68	7.35	7.38	2.99	5.94	19.60					
2.29	2.51	2.37	0.65	1.68	0.57	2.20	0.66	1.65	4.02	2.92				
6.24	9.67	13.53	5.21	9.90	7.98	4.26	1.56	3.25	3.79	−0.24	66.23			
31.05	53.96	63.13	35.92	47.40	46.72	12.08	3.18	11.58	4.01	1.17	34.07	213.80		
12.43	15.10	15.31	9.33	14.36	11.96	1.99	0.51	0.97	3.23	0.73	3.03	−0.22	13.26	
6.88	9.90	10.74	5.89	9.91	8.04	1.02	0.48	1.53	1.17	−0.28	3.24	1.97	0.98	10.95

DMT2; March of Grade 1: 7 = Brus A, 8 = DMT1, 9 = DMT2; June of Grade 1: 10 = Brus B, 11 = DMT2; March of Grade 2: 12 = Brus A,
20 = phoneme segmentation A, 21 = phoneme segmentation B, 22 = phoneme blending, 23 = letter knowledge, 24 = word recognition, 25 =
respectively, of the Drie-Minuten-Test; EMT = Caesar Eén-Minuut-Test; Brus A and Brus B = Form A and B of the Brus Eén-Minuut-Test,

(*Appendixes continue*)

Appendix C

Observed Covariance Matrix: Word Recognition, Vocabulary, Leisure Time Reading, Reading Comprehension, and Attitudes Toward Reading

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	7.84																					
2	7.52	16.36																				
3	7.08	14.34	18.14																			
4	5.61	11.92	8.77	66.13																		
5	7.89	18.59	14.53	86.30	141.57																	
6	5.32	11.21	8.70	49.25	71.45	44.83																
7	12.62	25.33	21.30	93.65	138.06	78.76	203.60															
8	7.76	15.76	12.28	57.43	86.55	49.26	120.34	81.83														
9	12.29	22.94	22.65	93.31	152.02	78.95	201.39	121.58	340.22													
10	8.01	15.76	15.69	63.54	104.43	55.50	135.63	82.48	207.29	151.12												
11	12.50	24.18	22.94	88.29	146.11	74.63	195.24	116.82	319.66	206.37	380.79											
12	8.66	16.89	14.89	62.02	101.82	53.25	139.33	85.75	209.90	141.31	228.69	166.33										
13	10.82	22.00	20.52	79.58	128.35	69.22	177.89	108.30	265.08	176.29	287.79	190.90	261.91									
14	8.73	19.20	18.38	68.62	115.96	59.04	156.54	94.39	276.03	176.37	313.55	192.50	239.67	331.62								
15	7.21	15.94	14.93	52.63	91.67	45.98	119.39	72.95	200.61	137.14	222.46	143.77	177.15	214.26	174.46							
16	11.34	21.57	19.72	67.52	116.96	59.10	156.28	98.90	265.65	177.56	296.04	190.85	241.19	284.74	205.28	290.14						
17	1.24	1.83	2.59	1.05	0.12	1.38	1.01	1.21	2.12	2.27	5.25	4.85	4.69	3.29	3.29	2.22	17.19					
18	0.31	0.44	1.68	2.34	1.62	3.30	1.32	3.65	10.10	4.79	9.36	5.84	7.06	9.12	2.89	8.29	9.44	44.80				
19	0.87	0.34	0.43	0.01	−1.07	0.34	1.53	2.67	5.37	2.70	8.12	6.55	6.30	8.06	4.64	8.25	5.55	9.78	11.51			
20	1.44	0.71	1.46	2.20	1.63	2.56	3.70	4.91	12.43	8.45	16.96	10.86	10.53	14.23	9.11	14.37	9.40	21.36	11.26	32.85		
21	0.70	1.66	1.47	5.47	8.80	4.90	8.96	5.97	14.28	8.09	15.54	8.87	13.17	11.59	7.47	11.82	2.53	7.93	3.09	8.20	11.25	
22	0.93	−0.17	−0.79	2.71	5.90	4.26	7.32	6.31	15.83	8.87	16.46	9.56	13.60	11.09	6.62	11.75	2.16	8.92	3.82	9.35	5.04	13.95
23	0.71	0.19	1.58	3.25	3.01	2.78	4.28	2.81	9.66	6.41	13.08	7.53	9.83	8.85	5.51	8.53	7.83	16.95	7.38	17.95	7.22	7.28
24	0.37	0.32	0.72	0.36	0.58	1.84	4.20	2.29	10.75	7.40	10.93	6.66	12.26	11.61	6.13	11.98	3.62	8.05	3.89	8.02	3.37	4.76
25	3.91	7.93	7.76	22.19	34.28	21.38	40.37	26.57	57.31	36.40	58.93	38.13	52.27	48.97	33.52	47.18	4.02	13.83	4.37	13.96	10.24	8.85
26	0.91	0.81	0.72	1.63	1.45	1.70	2.18	1.78	4.21	4.47	4.35	3.51	6.32	2.89	1.28	5.39	4.38	7.77	3.55	7.09	3.11	3.80
27	0.45	1.11	1.33	3.27	3.43	3.44	4.25	3.02	8.50	5.48	9.04	4.83	6.48	7.52	5.51	8.34	5.00	9.29	3.73	10.80	4.82	4.88
28	−0.09	0.12	0.12	0.33	0.31	0.21	0.56	0.15	0.17	−0.60	−0.23	−0.51	−0.16	−0.71	−0.59	−0.64	0.11	−0.21	−0.32	−0.17	0.12	0.14
29	0.60	1.00	1.07	0.92	0.57	0.66	2.55	0.97	−0.13	−0.03	0.40	−0.04	0.88	−0.59	0.30	−0.99	0.50	−0.25	−0.27	0.59	0.21	0.28
30	0.53	0.92	0.69	1.33	1.36	0.71	1.96	0.86	1.69	0.98	2.77	1.33	2.24	0.92	0.48	0.79	0.21	0.20	−0.67	0.16	0.11	−0.15
31	0.45	0.84	0.72	2.16	2.61	1.44	4.69	2.58	5.93	3.44	4.40	3.91	5.23	3.68	3.11	3.24	0.62	0.85	−0.49	0.17	0.59	0.51
32	0.02	0.48	0.88	−0.07	0.55	−0.07	1.49	0.92	1.91	0.38	2.54	1.52	2.56	1.07	1.05	0.64	−0.13	0.71	−0.68	0.51	0.76	0.39
33	−0.24	0.76	0.64	1.97	2.35	1.33	4.33	2.79	2.87	1.56	2.88	2.64	4.34	1.92	1.49	2.01	0.61	0.79	−0.05	0.63	0.57	0.11
34	1.27	3.02	2.28	4.84	6.09	4.50	10.89	6.26	11.98	8.72	9.54	8.16	11.22	10.69	10.25	11.40	2.60	0.80	−0.05	1.85	2.68	−0.30
35	3.30	6.28	4.66	20.26	28.66	16.76	36.78	22.94	44.82	28.26	47.92	32.78	41.85	39.91	29.34	38.53	4.08	6.10	4.27	7.33	5.99	5.51
36	2.77	4.96	5.97	9.20	17.61	10.03	27.44	16.89	37.19	23.46	42.88	25.60	30.79	27.14	19.46	29.53	3.88	9.52	4.92	9.82	5.60	7.29
37	1.80	3.46	2.84	8.28	13.83	9.34	22.83	14.70	28.00	19.48	31.94	22.18	30.36	25.57	18.71	27.04	6.29	13.65	6.32	13.27	8.42	6.95
38	−0.50	1.33	0.16	2.82	5.44	2.92	8.99	6.80	14.23	10.90	18.80	13.21	14.55	14.30	10.73	14.73	6.09	13.81	4.59	8.39	6.63	6.65
39	2.29	1.53	1.55	2.49	5.36	3.84	9.82	5.80	13.96	7.45	13.00	7.66	10.37	10.18	8.04	9.05	−1.11	4.60	1.52	1.68	0.90	2.23
40	0.51	3.22	1.72	8.86	12.31	6.78	15.28	10.57	20.90	11.55	15.68	12.16	16.98	11.22	11.44	13.77	−0.85	5.92	−0.21	−0.01	2.16	1.51
41	0.54	4.11	3.63	11.14	14.54	10.52	24.94	15.35	34.06	23.34	36.17	27.11	34.15	32.60	29.58	34.24	2.15	1.84	1.34	3.33	2.82	−0.03
42	−0.89	−0.81	−0.33	1.92	0.96	2.61	−1.38	1.12	11.15	4.69	13.53	6.24	9.67	7.98	5.21	9.90	8.80	20.79	6.34	20.62	11.25	10.15
43	0.70	0.67	0.66	0.89	1.96	1.09	3.74	3.51	8.26	5.66	10.74	6.88	9.90	8.04	5.89	9.91	1.43	4.74	1.34	2.45	1.26	1.90

Appendix C (*continued*)

Measure	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
23	20.69																				
24	7.50	10.58																			
25	12.26	7.27	38.41																		
26	6.55	5.67	6.13	9.41																	
27	10.22	4.72	10.04	4.63	10.74																
28	−0.42	−0.37	−0.28	0.04	0.36	1.47															
29	0.35	0.07	−0.45	0.09	0.48	0.66	2.56														
30	−0.14	−0.43	−0.36	−0.33	0.13	0.67	1.39	2.40													
31	0.22	0.00	0.87	0.44	0.68	0.55	0.66	0.94	2.84												
32	0.06	0.04	1.14	0.63	0.33	0.57	0.65	0.71	0.94	3.41											
33	0.30	0.55	1.07	0.83	0.65	0.29	0.62	0.59	0.77	0.97	1.82										
34	0.78	1.45	3.40	1.61	2.81	0.85	1.69	1.64	2.49	2.55	2.13	17.42									
35	6.51	4.82	16.73	4.45	4.60	0.26	0.47	0.47	0.21	1.08	1.53	1.40	28.79								
36	8.81	3.84	14.17	3.69	5.64	−0.20	0.26	0.90	−0.07	0.15	0.44	−0.07	11.15	35.34							
37	11.62	7.92	17.09	6.17	8.06	0.31	0.16	−0.32	0.14	0.06	0.74	1.28	9.94	12.78	29.69						
38	9.33	5.58	13.92	6.26	6.51	0.16	−0.84	−0.65	0.60	0.94	0.97	1.74	6.28	9.28	10.88	29.09					
39	1.02	−0.16	0.99	0.97	1.12	1.32	1.78	1.75	1.96	1.46	0.51	1.65	2.98	4.87	1.18	0.07	35.34				
40	1.03	1.05	7.37	1.32	2.06	1.26	0.60	1.08	2.72	2.38	2.44	4.92	4.30	2.62	2.57	4.37	4.18	27.17			
41	1.69	1.70	9.24	0.89	2.92	1.07	1.27	0.87	2.86	3.03	2.26	11.66	5.25	1.84	3.82	3.73	4.14	16.27	35.88		
42	16.65	8.43	13.88	8.91	10.41	−0.09	0.25	0.66	2.10	2.14	1.28	2.61	6.06	7.64	12.37	12.38	0.43	6.88	8.25	66.17	
43	3.54	2.51	3.94	1.72	2.69	0.05	0.57	0.28	0.27	0.30	0.11	0.44	3.13	1.43	2.47	3.55	3.87	1.08	0.90	3.24	10.94

Note. November of Grade 1: 1 = DMT1, 2 = DMT2, 3 = EMT; March of Grade 1: 4 = DMT2, 5 = DMT1, 6 = Brus A; June of Grade 1: 7 = DMT2, 8 = Brus B; March of Grade 2: 9 = DMT2, 10 = Brus A; June of Grade 2: 11 = DMT2, 12 = Brus B, 13 = DMT3; June of Grade 3: 14 = DMT2, 15 = Brus A, 16 = DMT3; November of Grade 1: 17 = CITO-Woordenschat, 18 = TVK-WS Keuze; March of Grade 1: 19 = CITO-Woordenschat, 20 = TVK-WS Keuze; June of Grade 1: 21 = Woordenschattest, 22 = TVK-WS Produktie; March of Grade 2: 23 = CITO-Woordenschat; June of Grade 2: 24 = TVK-WS Keuze, 25 = Woordenschattest; June of Grade 3: 26 = CITO-Woordenschat, 27 = TVK-WS Keuze; March of Grade 1: 28 = leisure time reading; May of Grade 1: 29 = leisure time reading; June of Grade 1: 30 = leisure time reading; March of Grade 2: 31 = leisure time reading; May of Grade 2: 32 = leisure time reading; June of Grade 2: 33 = leisure time reading; June of Grade 3: 34 = leisure time reading; June of Grade 1: 35 = reading comprehension; March of Grade 2: 36 = reading comprehension; June of Grade 2: 37 = reading comprehension; June of Grade 3: 38 = reading comprehension; March of Grade 1: 39 = attitudes toward reading; June of Grade 2: 40 = attitudes toward reading; June of Grade 3: 41 = attitudes toward reading; Kindergarten: 42 = Peabody vocabulary, 43 = home literacy. Measures 1–16 are all word recognition tests (see definitions for Appendix B), and Measures 17–27 are vocabulary tests referred to in the Method section of the text.

(*Appendixes continue*)

Appendix D

Annotated LISREL Output of the Two Final Models

In Table D1, the LISREL estimates (maximum likelihood) for all parameters of the final structural equation model involving phonological skills and decoding (Model 5c, Table 3) are presented. The standard errors of these estimates are in parentheses. When no standard errors are provided, the parameter has not been estimated but was fixed to a certain value. Lambda Y is the factor loadings matrix of the observed measures and the endogenous η variables (P_1–W_6). P refers to the phonological skills factor and W to the latent word recognition factor, and the subscript$_i$ indicates the measurement occasion as listed in Table 2 (see also Figure 1). See Appendix B for the definitions of the observed variables (1–28). The capital letters P, V, H, and R (without subscripts) refer to the phonological skills factor, vocabulary, home literacy, and the reading knowledge factor in kindergarten, respectively. Lambda X is the factor loadings matrix of the kindergarten measures and the exogenous ξ variables (P to H). The beta matrix contains the structural parameters between the endogenous η variables, and the gamma matrix the structural parameters between the exogenous ξ variables and the endogenous η variables. The phi matrix is the covariance matrix of the latent exogenous exogenous ξ variables. The psi matrix is the covariance matrix of the equation residuals. Theta epsilon is the error covariance matrix of observed measures related to the η variables, and theta delta is the error covariance matrix of observed measures related to the ξ variables.

Similarly, in Table D2 the LISREL estimates (maximum likelihood) for all parameters of the final structural equation model with decoding, vocabulary, leisure reading, and attitudes toward reading (Model 5, Table 5) are presented. See Appendix C for the definitions of the observed variables (1–43) in Table D2. Latent factor W (with subscripts to indicate measurement occasion) refers to word identification skills, V to vocabulary, L to leisure time reading, RC to reading comprehension, and A to attitudes toward reading. V without a subscript is the latent vocabulary factor in kindergarten, and H refers to the home literacy factor in kindergarten.

Table D1
Parameter Estimates (Model 5c, Table 3)

Observed variable	Lambda Y						
	P_1	W_1	W_2	W_3	W_4	W_5	W_6
1	1.00						
2	1.07 (0.13)						
3	1.10 (0.13)						
4		2.06 (0.16)					
5		1.90 (0.16)					
6		1.00					
7			0.83 (0.03)				
8			1.46 (0.05)				
9			1.00				
10				0.62 (0.02)			
11				1.00			
12					0.67 (0.02)		
13					1.00		
14						0.66 (0.02)	
15						0.83 (0.02)	
16						1.00	
17						0.73 (0.02)	
18							0.97 (0.03)
19							1.00
	Lambda X						
	P	R	V	Memory	IQ	H	
20	1.00						
21	0.34 (0.03)						
22	0.71 (0.05)						
23		1.00					
24		0.24 (0.05)					
25			7.72				
26				13.88			
27					3.46		
28						3.15	

Table D1 (*continued*)

Beta

	P_1	W_1	W_2	W_3	W_4	W_5
P_1		0.63 (0.09)				
W_2	1.03 (0.43)	0.93 (0.41)				
W_3			1.60 (0.07)			
W_4				1.03 (0.06)		
W_5					1.02 (0.03)	
W_6						0.85 (0.03)

Gamma

	P	R
P_1	0.16 (0.03)	
W_1	0.11 (0.04)	0.09 (0.04)

Phi

	P	R	V	Memory	IQ	H
P	18.57 (2.15)					
R	7.80 (1.41)	16.65 (3.44)				
Vocab	0.54 (0.31)	0.43 (0.30)	1.00 (0.10)			
Memory	0.90 (0.32)	0.30 (0.30)	0.32 (0.08)	1.00 (0.10)		
IQ	0.54 (0.31)	0.94 (0.31)	0.11 (0.07)	0.00 (0.07)	1.00 (0.10)	
Home	0.40 (0.31)	0.31 (0.30)	0.13 (0.07)	0.05 (0.07)	0.09 (0.07)	0.99 (0.10)

Psi

P_1	W_1	W_2	W_3	W_4	W_5	W_6
1.61	3.17	46.08	43.94	103.46	25.28	36.94
(0.39)	(0.55)	(4.96)	(5.26)	(11.59)	(4.56)	(4.81)

Theta epsilon

Variances

1	2	3	4	5	6	7	8	9	10
5.48	4.79	4.58	0.76	4.95	4.19	3.75	15.98	7.43	7.28
(0.62)	(0.58)	(0.58)	(0.52)	(0.64)	(0.41)	(0.57)	(2.06)	(0.97)	(1.15)
11	12	13	14	15	16	17	18	19	
9.38	12.45	29.11	13.88	22.55	34.80	19.62	16.51	36.98	
(2.57)	(1.86)	(4.22)	(1.76)	(2.84)	(4.27)	(2.36)	(2.90)	(4.44)	

Covariances

6, 9	9, 11	7, 12	11, 13	10, 14	12, 17	13, 16	15, 18	16, 19
0.11	1.39	1.17	0.92	1.40	4.63	4.36	3.87	13.45
(0.42)	(1.00)	(0.67)	(1.96)	(0.90)	(1.45)	(2.82)	(2.05)	(3.17)

Theta delta

Variances

20	21	22	23	24	25	26	27	28
2.75	1.74	7.26	2.93	1.94	6.61	21.34	1.32	1.10
(0.93)	(0.20)	(0.83)	(2.95)	(0.25)				

(*Appendixes continue*)

Table D2
Parameter Estimates (Model 5, Table 5)

Lambda Y

Observed variable	W_1	W_2	W_3	W_4	W_5	W_6	V_1	V_2	V_3	V_4	V_5
1	1.00										
2	2.10 (0.17)										
3	1.90 (0.16)										
4		1.00									
5		1.46 (0.05)									
6		0.83 (0.03)									
7			1.00								
8			0.62 (0.02)								
9				1.00							
10				0.67 (0.02)							
11					1.00						
12					0.66 (0.02)						
13					0.83 (0.02)						
14						1.00					
15						0.73 (0.02)					
16						0.97 (0.03)					
17							0.45 (0.06)				
18							1.00				
19								0.45 (0.04)			
20								1.00			
21									0.94 (0.14)		
22									1.00		
23										4.32	
24											1.00
25											1.80 (0.24)

Lambda Y

	V_6	L_3	L_5	L_6	RC_3	RC_4	RC_5	RC_6	A_2	A_5	A_6
26	0.74 (0.08)										
27	1.00										
28		1.00									
29		2.02 (0.34)									
30		2.06 (0.34)									
31			1.00								
32			1.00 (0.17)								
33			0.90 (0.14)								
34				3.96							
35					1.00						
36						1.00					
37							1.00				
38								1.00			
39									5.64		
40										4.95	
41											3.59

Lambda X

	V	H
42	7.72	
43		3.15

Beta

	W_1	W_2	W_3	W_4	W_5
W_2	1.72 (0.28)				
W_3		1.61 (0.07)			
W_4			1.03 (0.06)		
W_5				1.02 (0.04)	
W_6					0.85 (0.03)
L_3	0.05 (0.02)				
L_5			0.01 (0.00)		
RC_3		0.33 (0.04)			
RC_4			0.13 (0.02)		
RC_5				0.05 (0.02)	
A_5			0.01 (0.00)		
A_6				0.02 (0.00)	

	V_1	V_2	V_3	V_4	V_5
V_2	1.00 (0.12)				
V_3		0.41 (0.05)			
V_4			0.44 (0.05)		
V_5				1.82 (0.18)	
V_6					1.27 (0.13)
RC_3		0.33 (0.07)			
RC_4			1.07 (0.21)		
RC_5				2.50 (0.33)	
RC_6					1.35 (0.20)

Table D2 (*continued*)

Beta

	L_3	L_5	RC_4	A_2	A_5	A_6
V_5		0.41 (0.13)				
L_3				0.15 (0.05)		
L_5	0.91 (0.21)				0.40 (0.08)	
L_6		0.44 (0.09)				0.20 (0.04)
RC_5			0.17 (0.06)			
A_6				0.83 (0.10)		

Gamma

	V	H
V_1	2.57 (0.37)	0.87 (0.31)

Phi

	V	H
Vocab	1.00 (0.10)	
Home literacy	0.13 (0.07)	0.99 (0.10)

Psi

W_1	W_2	W_3	W_4	W_5	W_6	V_1	V_2	V_3
3.61	48.15	43.81	103.76	25.34	36.70	13.60	1.31	0.42
(0.63)	(5.08)	(5.23)	(11.58)	(4.48)	(4.76)	(2.57)	(1.88)	(0.26)
V_4	V_5	V_6	L_3	L_5	L_6	RC_3	RC_4	RC_5
0.14	0.71	—	0.30	0.45	0.59	16.15	23.60	15.47
(0.05)	(0.24)	—	(0.09)	(0.13)	(0.07)	(1.82)	(2.53)	(1.76)
RC_6	A_2	A_5	A_6					
18.42	1.00	0.95	1.50					
(2.09)	(0.10)	(0.10)	(0.17)					

Theta epsilon

Variances

1	2	3	4	5	6	7	8	9
4.25	0.51	5.16	7.43	15.93	3.89	8.88	7.46	27.91
(0.42)	(0.65)	(0.71)	(0.96)	(2.06)	(0.58)	(2.52)	(1.15)	(4.05)
10	11	12	13	14	15	16	17	18
13.28	34.70	14.12	22.64	37.06	19.55	16.52	12.90	23.23
(1.86)	(4.21)	(1.75)	(2.81)	(4.45)	(2.35)	(2.90)	(1.28)	(2.85)
19	20	21	22	23	24	25	26	27
6.94	10.00	7.52	9.71	2.07	6.38	24.88	5.72	3.94
(0.76)	(1.39)	(0.75)	(0.95)		(0.65)	(2.50)	(0.58)	(0.50)
28	29	30	31	32	33	34	35	36
1.14	1.20	0.99	1.84	2.42	1.01	1.74	3.00	3.00
(0.12)	(0.19)	(0.18)	(0.21)	(0.26)	(0.13)			
37	38	39	40	41				
3.00	3.00	3.53	2.71	3.59				

Covariances

1, 4	2, 5	4, 7	6, 10	7, 9	8, 12	11, 14	10, 15	13, 16
0.09	0.01	1.42	1.33	0.47	1.44	14.28	4.49	3.98
(0.43)	(0.61)	(0.99)	(0.68)	(1.95)	(0.91)	(3.24)	(1.45)	(2.04)

Theta delta

Variances

42	43
6.61	1.10

Received March 10, 1997
Revision received April 14, 1998
Accepted April 14, 1998 ■