Development of phonological awareness in relation to literacy

An item response theory perspective

Judith Vloedgraven

This study was financed by the Dutch Ministry of Education, Culture and Science.

Cover design: Winterworks Grafisch Ontwerp
Cover photo: Norbert Voskens
Print: Drukkerij Efficiënt, Nijmegen

ISBN 978 90 77529 32 4

Development of phonological awareness in relation to literacy

An item response theory perspective

Een wetenschappelijke proeve op het gebied van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. mr. S. C. J. J. Kortmann,
volgens besluit van het College van Decanen
in het openbaar te verdedigen op woensdag 14 januari 2009
om 13.30 uur precies door

Judith Marie-Therèse Vloedgraven
geboren op 11 april 1981
te Heeten

Promotores:	Dhr. prof. dr. L. T. W. Verhoeven
	Dhr. prof. dr. ir. T. J. H. M. Eggen *Universiteit Twente; Cito*
Manuscriptcommissie:	Dhr. prof. dr. C. A. J. Aarnoutse
	Dhr. dr. P. F. de Jong *Universiteit van Amsterdam*
	Dhr. dr. J. M. McQueen

Contents

Voorwoord

De uitspraak 'het eindpunt kan nooit zo interessant zijn als de weg ernaar toe' past naar mijn mening uitstekend bij mijn promotietraject van de afgelopen jaren. Wat heb ik veel geleerd!

Vier jaar geleden werd ik door Ludo Verhoeven benaderd voor een baan als onderzoeker. Het ging om een contract van een jaar met uitzicht op verlenging, afhankelijk van subsidies van het ministerie. Het doel van het project was een instrument te ontwikkelen waarmee leerlingen met een risico op leesproblemen vroegtijdig gesignaleerd kunnen worden in het basisonderwijs. Ik ben de uitdaging aangegaan en na een jaar liep het contract af. De werkzaamheden tijdens dat eerste jaar smaakten echter naar meer en bovendien, als ik eenmaal ergens aan begin, wil ik het ook graag afmaken. Net voordat ik op zoek wilde gaan naar een andere baan, werd het contract met anderhalf jaar verlengd dankzij extra subsidies van het ministerie van OC&W. Hierdoor werd het mogelijk een grootschalig normeringsonderzoek te organiseren ten behoeve van het signaleringsinstrument. Dit leverde een schat aan data op, genoeg voor het schrijven van meerdere proefschriften. Alleen liep het contract wederom af... Na een aantal weken in spanning te hebben gezeten, kwam er gelukkig opnieuw een verlenging en hiermee uitzicht op een promotie. Het resultaat ligt hier voor u. Wat ik de afgelopen jaren in elk geval heb geleerd is me minder druk te maken om de toekomst, omdat is gebleken dat het altijd wel goed komt. Verder heb ik geleerd uit de vele data die het normeringsonderzoek heeft opgeleverd een keuze te maken bij het schrijven van de artikelen. Liever een aantal aspecten tot in detail onderzocht dan vele onderwerpen maar half (door te promoveren leer je te selecteren). Bijna 70 scholen en ruim 2000 leerlingen hebben meegewerkt aan het 2-jarige normeringsonderzoek. Gedurende deze twee jaren zijn we geholpen door circa 100 toetsleiders. Stelt u zich een uiteindelijk databestand voor bestaande uit ruim een miljoen cellen. Door de enorme omvang van de database heb ik geleerd meer de controle uit handen te geven en meer te vertrouwen op de nauwkeurigheid van anderen.

Naast deze leerervaringen op het persoonlijke vlak, heb ik ook veel kennis opgedaan. Ik herinner me nog mijn sollicitatiegesprek met Ludo, waarin hij zei dat bij dit project een niet alledaagse statistische theorie, de item respons theorie, een grote rol speelde. Hoewel ik daar niet bekend mee was, hoefde ik me daar geen zorgen om te maken. Voor dit project was namelijk samenwerking gezocht met Cito, dat voor de statistische input zou zorgen. Dat is ook zeker gebeurd, maar uiteindelijk heb ik alle statistische analyses toch zelf uitgevoerd. Dit was niet altijd eenvoudig, maar wel enorm leerzaam.

Dan wil ik nu mijn dank uitspreken aan hen die mede bijgedragen hebben aan de totstandkoming van dit proefschrift. Allereerst wil ik mijn beide begeleiders Ludo Verhoeven en Theo Eggen bedanken. In het begin moest ik wennen aan de grote mate van vrijheid die ik kreeg; later leerde ik zelf hulp te vragen op momenten dat het nodig was. Ondanks de drukke werkschema's wisten jullie altijd tijd te vinden om het onderzoek met mij te bespreken en feedback te geven op de teksten, wat heeft bijgedragen aan een vlotte afronding van dit proefschrift. Ludo, ik waardeer enerzijds je deskundige kijk op zaken; jouw sterke inzicht in de structuur van teksten en lijn van argumentatie zijn dit proefschrift zeker ten goede gekomen en anderzijds je positieve kijk op zaken; geen probleem zo groot of jij wist het zo te brengen dat het op het eind niet meer duidelijk was waarom het een probleem was. Theo, hartelijk dank voor je hulp op statistisch gebied, je feedback op alle teksten (en oog voor detail) en de fijne samenwerking. Wie verder van begin tot eind heel nauw bij mijn onderzoek betrokken is geweest, is mijn kamergenoot Jos. Hoewel we totaal verschillend zijn, bleken we prettig complementair doordat we allebei onze eigen sterke kanten hebben. Ik waardeer jouw geduld, geen vraag was jou teveel. Ik heb veel van je geleerd en ik wil je ontzettend bedanken voor je adviezen betreffende de statistiek, inhoud van de artikelen, lay-out en natuurlijk voor de gezelligheid.

Verder wil ik mijn collega's van de vijfde verdieping en in het bijzonder Marieke en Esther bedanken voor de prettige onderbrekingen van de werktijd. Daarnaast dank ik de collega's van het Expertisecentrum Nederlands. Ook al zat ik niet bij jullie in het gebouw, de samenwerking met jullie is me goed bevallen. Mensen uit de dyslexiegroep, het bespreken van de ontwikkelingen in de praktijk was voor mij een waardevolle aanvulling op mijn verder overwegend theoretische bezigheden. Hennie, bedankt voor je adviezen wanneer we de voortgang van het project bespraken. Medewerkers van Cito ben ik veel dank verschuldigd voor de geboden faciliteiten om dit onderzoek goed uit te kunnen voeren. Ik kijk uit naar een voortzetting van onze samenwerking. Daarnaast wil ik alle scholen, leerlingen en toetsleiders bedanken die hebben meegewerkt aan het normeringsonderzoek. Lee Ann Weeks dank ik voor de kritische feedback op mijn schrijfvaardigheid in het Engels.

Maar er is gelukkig veel meer in mijn leven dan een promotieonderzoek. Ik wil een aantal mensen uit mijn naaste omgeving bedanken voor de afleiding en steun in de afgelopen jaren. In dit rijtje wil ik als eerste mijn vriendinnen uit Heeten en Nijmegen noemen voor de geweldige momenten van ontspanning die ik met jullie heb doorgebracht (uit eten, weekendjes weg, samen sporten enz.). Elze en Claudia wil ik op deze plaats in het bijzonder noemen. Bedankt voor de interesse die jullie hebben getoond in mijn onderzoek en voor jullie luisterend oor als ik vertelde over de problemen die zich voor deden. Ik vind het super dat jullie mijn paranimfen zijn! Nu een woord van dank aan de mensen die het allerbelangrijkst voor me zijn. Pap en mam, zonder jullie was ik nooit zover gekomen. En daarmee bedoel ik niet alleen de eigenschappen die ik van jullie heb meegekregen die nodig zijn voor het voltooien van een project als dit, maar vooral de mogelijkheden die jullie me gegeven hebben om me te ontwikkelen en jullie onvoorwaardelijke steun en vertrouwen in mij. Ook mijn broer(tje) Daan en zijn vriendin Marion wil ik bedanken voor de getoonde belangstelling. Jeroen, de laatste dankwoorden zijn voor jou. Als geen ander ken je inmiddels de 'in's en out's' van mijn promotietraject. Ik wil je bedanken voor je interesse, vertrouwen en morele steun tijdens de wat lastigere fasen van dit project. Ook heb ik veel gehad aan je relativeringsvermogen (op momenten dat ik me druk maakte om zaken die achteraf altijd mee bleken te vallen). Daarnaast zorgde het samen doorbrengen van de vrije tijd, bijvoorbeeld de mooie reizen die we maakten, voor de welkome afleiding. Mijn dank aan jou is groot!

Tot slot, de mentale inspanningen van de afgelopen vier jaar vielen niet altijd mee. Maar in combinatie met de fysieke inspanningen gedurende deze periode (vierdaagse, Incatrail Peru, triatlons) hebben ze voor mij de juiste uitdaging gevormd. Samen hebben ze de afgelopen jaren tot een mooie periode van mijn leven gemaakt, die ik zeker niet had willen missen. Terugkomend op de eerste zin van mijn voorwoord kan ik zeggen dat niet alleen het proces waardevol is; ik ben ook ontzettend blij met het eindresultaat. Ik ben nu klaar voor de volgende stap in mijn loopbaan. Ik vertrouw erop bij Cito mijn theoretische inzichten te kunnen vertalen naar de praktijk, waarin ik een nieuwe uitdaging zie.

1

General introduction

A large number of studies across different languages have shown phonological awareness to be the critical factor in reading development (e.g., McBride-Chang & Kail, 2002; Scarborough, 1998; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). There is also evidence that growth in phonological awareness accounts for variance in reading beyond that accounted for by the actual level of phonological awareness (Byrne, Fielding-Barnsley, & Ashley, 2000; Hindson et al., 2005). Given the converging evidence that reading problems can be prevented through early interventions aimed at the improvement of phonological awareness (Ehri et al., 2001; Snow, Burns, & Griffin, 1998; Torgesen et al., 1999), increasing attention has been paid by educators and policy makers to the early identification of reading problems. To be able to identify reading problems early, there is an urgent need for an adequate screening instrument for phonological awareness. Although phonological awareness appeared to be most strongly related to reading in all alphabetic languages studied so far, it is still unclear how phonological awareness can best be measured in practice. A wide variety of tasks have been used to measure phonological awareness and it is evident that it manifests itself in different skills during its course of development. However, it is far from clear whether the different tasks measure the same underlying ability at different points in time. In addition, little is known about the consequences of children's experiences with literacy for the measurement of phonological awareness. During the last decades, the primary interest of most studies concerned the relationships between phonological awareness and reading, while the nature and development of phonological awareness have received only scant attention. Greater insight into the nature of phonological awareness is needed to understand its development and its relations to reading skills. In addition, knowledge concerning the origins of phonological awareness is expected to promote early identification of children at risk for reading problems in order to prevent those children from experiencing reading difficulties and its accompanying social-emotional problems (e.g., Hinshaw, 19092; Shaywitz & Shaywitz, 1993). The present thesis aims to deal with gaps in knowledge about the measurement of phonological awareness across a longer period of time by investigating the nature and development of children's phonological awareness and its relations to literacy throughout the early elementary grades.

1.1 Phonological awareness and learning to read

1.1.1 Becoming phonologically aware

A deficit in phonological processing abilities is considered to be a plausible cause for the occurrence of reading problems or dyslexia (Bruck & Treiman, 1990; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Phonological processing concerns the use of phonological information, while processing oral and written language (Wagner & Torgesen, 1987). Research has uncovered three types of phonological processing abilities. *Phonological awareness* refers to the access to and an understanding of the sound structure of oral language (e.g., Liberman, Shankweiler, Fischer, & Carter, 1974). *Phonological memory* refers to recoding written information into phonological representations, in order to maintain information efficiently in verbal working memory (Baddeley, 1986). *Phonological naming* can be described as the rapid retrieval of phonological codes from long term memory by recoding written symbols into phonological codes. According to the relevant research literature, of these three kinds of phonological processing abilities, phonological awareness is most strongly related to learning to read (e.g., Wagner et al., 1997).

Recent reviews showed the general sequence of phonological awareness development to be universal across languages (Anthony & Francis, 2005; Ziegler & Goswami, 2005). The development seems to proceed from access to larger phonological units to smaller units of sound. Generally, children become aware of syllables when they are 3 to 4 years old, and they master rhyming skills at the age of 4 to 5. Phoneme awareness skills appeared to be most difficult and children often improve on these skills once they receive literacy instruction. The reason that phoneme awareness skills are most complex is that phonemes are acoustically evanescent (Liberman, Liberman, Mattingly, & Shankweiler, 1980). That is, the ability to manipulate individual phonemes is inherently difficult because the individual phonemes are not audibly distinguishable. For example, when we say the word cat, we do not pronounce the individual phonemes one at a time like "k / ae / t". In that case, the word cat would be pronounced more like "kuh-ae-tuh". Phonemes are abstract representations of language rather than discrete sounds corresponding to the letters.

A large number of definitions of phonological awareness have been proposed, which vary from stringent definitions that include only a few phonological skills to much broader definitions that include a variety of phonological skills (for a review, see Anthony & Lonigan, 2004). According to adherents of the most narrow definition (e.g., Hulme et al., 2002; Yopp, 1988), phonological awareness consists of separate abilities. The various skills differ with respect to linguistic complexity (e.g., words, syllables, onsets and rimes, and phonemes) or the type of cognitive operation required (e.g., detection, blending, segmentation, and manipulation). In contrast, a much broader definition assumes continuity between lower levels of phonological awareness (e.g., rhyming) and higher levels of phonological awareness (e.g., segmentation). Proponents of this approach (e.g., Anthony & Lonigan, 2004; Stanovich, 1992) view phonological awareness as a unidimensional ability. Although there is consensus on the general sequence of phonological awareness development, the issue of how to best conceptualize phonological awareness development across a longer period of time remains to be answered.

1.1.2 Relations between phonological awareness and reading

Over the last decades, accumulating evidence has been provided that phonological awareness is the critical factor in learning to read (e.g., Elbro & Pallesen, 2002; Torgesen et al., 1999). The relationship between phonological awareness and reading appears to be present even after accounting for variance due to factors such as IQ, vocabulary, memory, and social class (Bryant, MacLean, Bradley, & Crossland, 1990). The awareness that words can be divided into single phonemes is necessary to comprehend the alphabetic principle underlying our written language system (Byrne, 1998). The understanding of this process of matching visual symbols to sounds (i.e., phonological recoding) is considered as a prerequisite for the decoding of unknown words, which functions as a self-teaching device (Share, 1995; Ziegler & Goswami, 2005). Nevertheless, there is still no consensus on the size and directions of the relationships between phonological awareness and reading. Establishing the exact relations is complicated by the fact that these relations depend on the specific tasks for phonological awareness and the child's level of development (e.g., Anthony & Lonigan, 2004). As stated by Ziegler and Goswami (2005), floor and ceiling effects, task difficulty, and measurement artefacts have unintended influences on these relations. Furthermore, it might be expected that the relation between phonological awareness and reading changes during the course of development, for example, when children learn to read. How these variables exactly relate to each other throughout the early elementary school years is thus still unclear.

In general, there are three different views about the relations between phonological awareness and reading. The first view is that phonological awareness abilities influence subsequent reading skills. Persuasive evidence for this view comes from longitudinal studies that showed phonological awareness to be a significant predictor of later reading skills (for a review, see Wagner, Torgesen, & Rashotte, 1994) and from intervention studies that showed children to progress in reading abilities due to training programs aimed at the improvement of phonological awareness (e.g., Hatcher et al., 2006; Troia, 1999). Advocates of this view state that the relation between phonological awareness and reading is stable across time. The second view is that phonological awareness develops as a consequence of learning to read, as demonstrated by research that showed illiterate adults and readers of a nonalphabetic script to have no awareness of phonemes (Lukatela, Carello, Shankweiler, & Liberman, 1995; Morais, 1991; Read, Zhang, Nie, & Ding, 1986). The last view is that the relation is bidirectional: the more rudimentary levels of phonological awareness promote the reading development, and, in turn, reading skills may influence the higher levels of phonological awareness (e.g., Perfetti, Beck, Bell, & Hughes, 1987).

1.2 The present study

1.2.1 Aims

It is clear that the measurement of phonological awareness is of great importance for the early identification of reading problems. The proper monitoring of children's phonological awareness abilities over time can result in an even better identification of children with reading problems. Nevertheless, there is still no consensus on how phonological awareness can best be measured in practice and how children's development in this ability can be monitored. The present study was designed to gain greater insight into the possibilities for the construction of a screening instrument for Dutch children's phonological awareness. Before it was possible to develop an adequate phonological awareness test, we needed to investigate some theoretical issues in further detail. An attempt was made to disentangle the issue of the conceptualization of the development of phonological awareness. Moreover, the relations between phonological awareness and reading throughout the early elementary grades were examined more closely.

In order to be able to evaluate phonological awareness, decisions should be made on the content of the phonological awareness measures. This will result in a specification table showing the tasks to be included in the screening instrument for phonological awareness. This table will also provide information about the specific items within each task. In addition, an appropriate measurement model to analyze the data needs to be selected. Given the particular relevance of monitoring children's phonological awareness abilities over time, a model is required that is suitable for exactly this purpose. In the next two sections, the operationalization of phonological awareness and the selection of a measurement model will be successively addressed.

1.2.2 Operationalization of phonological awareness

To be able to measure a broad range of phonological awareness abilities, a large item bank was created. An item bank is a collection of items which are described by certain characteristics (e.g., content domain). For the construction of this item bank, on the one hand, decisions should be taken on the specific tasks (i.e., different indicators) for the assessment of phonological awareness and on the other hand, the specific items within each task.

To select the different tasks, the extent to which a certain indicator represents phonological awareness ability and the predictive value for reading performance were taken into account. Furthermore, we have selected various tasks that according to the literature are known to differ in difficulty to be able to accurately assess a broad range of phonological awareness abilities. According to the literature, a segmentation task best represents phonological awareness, followed by a blending task (van Bon & van Leeuwe, 2003; Yopp, 1988). Furthermore, the predictive validity of both tasks appeared to be good. Adams (1990) and Chard and Dickson (1999) assert that these tasks are relatively difficult and that segmentation will not be acquired before first grade. A task that is considered to adequately suit the ability level of kindergartners is phoneme identification, in which only one phoneme in a word has to be identified (Torgesen, 1998). In addition, the findings of Elbro, Borstrøm, and Petersen (1998) and Høien, Lundberg, Stanovich, and Bjaalid (1995) showed this task to be a powerful predictor of learning to read. Yet, a decision should be taken on the position of the phoneme that has to be identified. Stanovich, Cunningham, and Cramer (1984) showed tasks in which the initial phoneme had to be identified to be the best predictors of reading. Given the fact that it is important to accurately assess the ability of both high- and low-ability individuals, we still looked for a relatively easy and a relatively difficult task. It is generally considered that a rhyming task is one of the easiest tasks for phonological awareness (Adams, 1990; Yopp, 1988). Although rhyming is a less powerful predictor of reading than previous mentioned tasks, results showed rhyming to provide an independent contribution to reading and to be of greater importance for reading than syllables (Høien et al., 1995). Adams found phoneme manipulation to be most difficult as this requires the addition or deletion of phonemes to formulate a new word. Given the present view, measures were developed for the assessment of rhyming, phoneme identification, phoneme blending, phoneme segmentation, and phoneme deletion.

The next step involved the selection of the specific items within each task as regards the Dutch language. In order to control for the familiarity of words in diverse populations, words of high frequency from existing Dutch word frequency lists were selected. The target words were as much as possible drawn from Schrooten and Vermeer's list of word frequencies, which is based on the exposure of language at school instead of language at home (Schrooten & Vermeer, 1994). Other words were preferably drawn from the list of Schaerlaekens, Kohnstamm and Lejaegere (1999), which are considered to be known by six-year-old children. Given that a developmentally appropriate test for phonological awareness should not overreach working memory (Reitsma, 2002), all of the words were both auditorily and visually presented. To be able to present the words as pictures, only concrete words were selected.

Differences in linguistic complexity within tasks appear to influence children's performances on tasks for phonological awareness (Anthony & Francis, 2005). According to Schreuder and van Bon (1989), for example, the consonant-vowel (CV) structure of words is an important determinant. We decided to include monosyllabic words of different CV structures in the item bank to be able to administer items of various difficulty levels. In Table 1.1, the specification table for the content domain of phonological awareness is presented. An overview is provided of the various CV structures that are included in each task.

In addition to the word structure, the manner of articulation of consonants in a word appears to influence children's achievements on phonological awareness tasks. In the Netherlands, five manners of articulation are identified (Rietveld & van Heuven, 1997): plosives (b, d, k, p, t), fricatives (f, g, s, v, z), liquids (l, r), nasals (m, n), and glides (h, j, w). This distinction is relevant, because the manner of articulation in words was found to influence the difficulty of the items in phonological awareness tasks (de Graaff, Hasselman, Bosman, & Verhoeven, 2007). In an attempt to make the different tasks comparable concerning the manner of articulation and to administer items of various difficulty levels, we have tried to take care of a proportional distribution of the different articulation manners within each task as much as possible. For each task, we focused on a particular phoneme in the target word. Regarding the rhyming items, the last phoneme was taken into account; regarding the phoneme identification, phoneme blending, and phoneme segmentation items, the initial phoneme was considered; regarding the phoneme deletion items, the phoneme that had to be deleted was considered. It appeared that some manners of articulation occur more frequently than other manners, thus a completely proportional distribution was not possible. In Table 1.1, for each task, the distribution of the articulation manners is presented. The numbers represent the percentages of incidence of the different articulation manners in each task. It should be noted that concerning phoneme blending and phoneme segmentation, we have also tried to represent all manners of articulation in the last phoneme.

Table 1.1
Specification table for the phonological awareness content domain

		Manner of articulation				
Task	CV structure	Plosives	Fricatives	Liquids	Nasals	Glides
Rhyming	CVC	27	24	24	23	2
Phoneme identification	CVC	24	21	18	17	20
Phoneme blending	CVC, CVCC, CCVC, CCVCC, CCCVC, CVCCC	30	31	13	12	14
Phoneme segmentation	VC, CV, CVC, CVCC, CCVC, CCVCC, CCCVC, CVCCC	33	30	13	11	13
Phoneme deletion	CCV, CCCV, CCVC, CVCC, CCVCC, CCCVC	33	20	31	9	7

Another choice concerned the way of administering the phonological awareness items. Given the many advantages of computerized testing, it was decided to administer the items on a computer instead of on paper. An important advantage is the possibility for adaptive testing, in which each child is likely to get different items adjusted to the child's ability level. This means that the scoring system should be automatic and that the items should be administered in a receptive way. In receptive tasks, items have to consist of two or more alternatives. A disadvantage of two alternatives is that more items are needed to reach the same measurement precision than in case of more alternatives. In addition, research has demonstrated that the optimal number of alternatives is three (Haladyna & Downing, 1989), thus it was decided to use three alternatives. With respect to the phoneme blending and phoneme deletion items, it should be mentioned that the distractors contained one or more of the phonemes of the target word in order to measure blending and deletion ability as pure as possible.

1.2.3 Selection of a measurement model

There are two popular psychometrical frameworks for addressing measurement problems in test development: classical test theory (CTT) and item response theory (IRT; e.g., Eggen & Sanders, 1993). CTT focuses attention on the true score of a person on a particular task. The classical test model assumes that the observed test score (number of correctly answered items) consists of the true score and an error score. Classical test models are often denoted as "weak models", because the assumptions of these models are rather easily satisfied by test data. IRT is a theory about the ability of persons and how ability is related to items with certain characteristics. Item response models are denoted as "strong models", given that the underlying assumptions are strict and thus more difficult to be met by test data (Hambleton & Jones, 1993). A first important assumption in IRT is that most of the models require that the different sets of items that compound the test are measuring a single underlying ability. Second, it is assumed that the relation between the underlying ability and the probability of an item response can be described by the item characteristic curve (ICC).

For the measurement of phonological awareness, it is important, on the one hand, to compare task scores of different children and on the other hand, to monitor phonological awareness abilities of individual children over time. However, a number of measurement problems present themselves in doing this. First, given that children develop over time, different tasks must be administered at different points in time to be able to accurately measure children's phonological awareness. The problem is that it is unsure whether the different tasks, administered at different measurement occasions, measure the same underlying ability over time. And when various tasks are administered to different children at the same time, comparison of task scores is also difficult, because the scores do not rely upon a shared metric. A second problem with the measurement of phonological awareness is the unreliability of individual task scores. The precision of measurement depends on the extent to which the phonological awareness measure suits the child's ability level. Imagine a child who achieves a score of zero: this score tells us that the child's ability is low, but contains no information about exactly how low. However, when a child answers some items right and some wrong, the test score provides more information about what a child knows and not knows, and thus gives a more accurate measure of ability. The test scores of the two children contain different amounts of error and are not equally reliable.

Generally, for the measurement of phonological awareness, models from CTT are used. Some advantages of the application of classical test models in comparison to item response models are that smaller sample sizes suffice for the analyses and that the statistical analyses are less complex. However, the aforementioned problems are difficult to resolve within the framework of CTT (e.g., Crocker & Algina, 1986; Lord & Novick, 1968). An important shortcoming is that ability and test characteristics cannot be separated. In CTT, the true score indicates ability, which depends on the particular set of items administered, that is, is test-dependent. Children will obtain lower true scores on difficult tasks and higher true scores on easier tasks. Test-dependent scores are of no value in comparing children who take different tasks, because the scores do not rely upon a shared scale. Another major limitation in CTT is that item statistics (e.g., item difficulty and item discrimination) depend on the group to which the items are administered, that is, are group-dependent. When the sample differs from the population, the usefulness of the item characteristics is reduced. A third problem of CTT concerns the reliability. In CTT, reliability is assumed to be the same for children of various ability levels. But as just mentioned, scores on the same task of children of different ability levels are unequally reliable. A last problem is that CTT focuses attention on the test rather than on the specific items. A model from CTT provides no information about where an item can give the most precise estimation of ability on the ability scale. For the construction of optimal tests, such information is very useful.

It has been demonstrated that in IRT, because of the specific characteristics of this theory, these limitations do not exist any more (e.g., Hambleton, Swaminathan, & Rogers, 1991). This section starts with a brief introduction about the main features of IRT, followed by a description of some important advantages of this theory. As stated previously, items with certain characteristics are related to the ability of persons, which results in one scale for items and ability. In Figure 1.1, three items are represented by their ICC, a curve that describes the relation between ability and children's item performance. The underlying ability, also known as theta, is put on the x-axis. The y-axis is the probability of answering the item right. This function shows that as the ability level increases, the probability of a correct response to an item also increases. Various IRT models diverge with regard to the number of item properties that they take into account. In the one-parameter model, only difficulty parameters are estimated; in the two-parameter model, difficulty and discrimination parameters are estimated; and three-parameter models take into account the effect of item guessing, difficulty, and discrimination. These parameters determine the exact shape of the ICC. The most central is the difficulty parameter. It is the point on the ability scale where the probability of a correct response is 0.5. The more difficult an item is, the more the curve shifts to the right of the ability scale. To illustrate, in Figure 1.1, curve 1 represents the easiest item. Another parameter is the discrimination parameter, which indicates the slope of the curve at its steepest position. This parameter shows how well a certain item discriminates between persons of different ability levels. The steeper the slope, the more discriminating the item is. As can be seen, curve 2 is not only more difficult than curve 1, but is also better able to discriminate between high- and low-ability persons. Another possible parameter is the guessing parameter. This parameter represents a lower asymptote (as shown by curve 3) and indicates the probability of persons with low abilities to answer the item correctly when guessing plays a part.

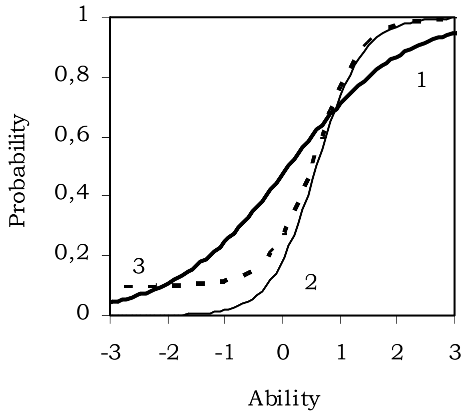

Figure 1.1 *Item characteristic curves for three different items*

Given the specific characteristics of IRT, scores are test-independent and item statistics are group-independent. Due to these useful features, in IRT it is possible to compare scores from different tasks, and to monitor children's phonological awareness abilities over time. Another advantage of IRT is the existence of item information functions. These functions express the contribution that a specific item can make to the precision of the measurement across a range of ability levels. In such a way, a test designer can select those items that are most informative for the ability level of the population of interest. Although CTT has been demonstrated to be valuable in the test development process, it may be clear that IRT offers more possibilities for large-scale investigation including large samples, a great many of items, and several populations and is especially useful for modern testing applications, such as adaptive testing and the monitoring of abilities across time. Therefore, for the present thesis, the data were analyzed within the framework of IRT.

Yet, a decision should be made on the specific IRT model. A main advantage of the one-parameter model is the possibility to apply the conditional maximum likelihood (CML) procedure to estimate item parameters and the sampling independence implied by it (Verhelst & Glas, 1994). In contrast, the use of CML is impossible in the two other models. However, a drawback of the one-parameter model is that it is not very realistic that discrimination indices are the same for all of the items. Therefore, we decided to use the one-parameter logistic model (OPLM), which is a synthesis of the one- and two-parameter model. The most important feature of the OPLM is that difficulty parameters are estimated and discrimination parameters are dealt with as known constants (i.e., discrimination indices can vary, but have discrete values). To this choice, the objection can be raised that no guessing parameter was included in the model, while guessing actually could play a part in the phonological awareness tasks. Yet, there were two main arguments to abandon this guessing parameter. First, the guessing parameter estimation requires much larger samples which is difficult to realize in practice. And even with very large samples, the estimation process may diverge instead of converge, that is, the estimations of the guessing parameter are not consistent. Second, if the phonological awareness items would be administered adaptively, the guessing parameter is not necessary anymore. In adaptive tests, items that provide the highest information value at the current ability estimate are administered and these are certainly not the items in which guessing plays a part. For a more detailed description of the fundamental concepts of IRT and the OPLM, see Appendix A and for example, van der Linden and Hambleton (1997).

1.3 Outline of the present thesis

Several issues on Dutch children's phonological awareness in relation to literacy are addressed in the present thesis. Chapter 2 describes the results of a pilot study, in which the possibilities for the assessment of growth in phonological awareness are explored. Phonological awareness was measured in kindergarten and first grade using four sets of items including rhyming, phoneme identification, phoneme blending, and phoneme segmentation. An initial attempt was made to investigate the nature of phonological awareness and the growth from kindergarten through first grade. To examine these issues, exploratory factor analyses and analyses within the context of IRT were performed.

In Chapter 3, the nature of phonological awareness is investigated in more detail. This paper reports on a study that extended the pilot study with respect to the number of tasks, the number of grades, and the use of larger samples per grade. For the purpose of this study, phoneme deletion items were added to the other sets of items and the phonological awareness items were administered to kindergarten through fourth-grade children. The nature of phonological awareness across different tasks and grades was investigated by means of modified parallel analyses and analyses within the framework of IRT. Furthermore, the acquisition of various phonological awareness skills over time was explored. In addition to these theoretical issues, this study also addresses the suitability of particular tasks for the assessment of children's phonological awareness for different levels of ability.

In Chapter 4, the focus has shifted from the tasks measuring phonological awareness to the children that were tested. The aim of this study was to characterize phonological awareness development from kindergarten through fourth-grade children. To take care of an adequate model specification, effects of some relevant other variables as letter knowledge, word reading ability, gender, SES, and linguistic diversity were taken into account. To investigate the development of children's phonological awareness, univariate regression analyses were conducted. The study that was described in Chapter 3 yielded an IRT calibrated scale for phonological awareness and this IRT based scale constituted the input for the analyses of this study.

Chapter 5 reports on a longitudinal study on the relations between phonological awareness and word reading in kindergarten through second grade. Autoregressive effects of both phonological awareness and word reading were first examined and the size and directions of the relations between these two variables over time were next determined. In addition, the relative impact of letter knowledge at kindergarten level on the relation between phonological awareness and reading was investigated. It was attempted to minimize the effects of some common sources of model misspecifications. To investigate these issues, several models were tested using structural equation modeling (SEM).

Finally, in Chapter 6 general conclusions are drawn on the basis of the findings of the four studies and integrated into a developmental model of phonological awareness. Moreover, some limitations of the present thesis, recommendations for future research, and some practical applications are presented.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14,* 255-259.

Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology, 96,* 43-55.

Baddeley, A. (1986). *Working memory.* New York: Oxford University Press.

Bruck, M., & Treiman, R. (1990). Phonological awareness and spelling in normal children and dyslexics: The case of initial consonant clusters. *Journal of Experimental Child Psychology, 50,* 156-178.

Bryant, P. E., MacLean, M., Bradley, L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology, 26,* 429-438.

Byrne, B. (1998). *The foundation of literacy: The child's acquisition of the alphabetic principle.* Sussex: Psychology Press.

Byrne, B., Fielding-Barnsley, R., & Ashley, L. (2000). Effects of preschool phoneme identity training after six years: Outcome level distinguished from rate response. *Journal of Educational Psychology, 92,* 659-667.

Chard, D. J., & Dickson, S. V. (1999). Phonological awareness: Instructional and assessment guidelines. *Intervention in School and Clinic, 34,* 261-270.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart and Winston.

de Graaff, S., Hasselman, F., Bosman, A. M. T., & Verhoeven, L. (2007). Cognitive and linguistic constraints on phoneme isolation in Dutch kindergartners. *Learning and Instruction*, doi:10.1016/j.learninstruc.2007.08.001

Eggen, T. J. H. M., & Sanders, P. F. (1993). *Psychometrie in de praktijk* [Psychometrics in practice]. Arnhem: Cito.

Ehri, L. C., Nunes, S., Willows, D., Schuster, B., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250-287.

Elbro, C., Borstrøm, I., & Petersen, D. K. (1998). Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly, 33,* 36-60.

Elbro, C., & Pallesen, B. R. (2002). The quality of phonological representations and phonological awareness: A causal link. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 17-32). Amsterdam: John Benjamins.

Haladyna, T. M., & Downing, S. M. (1989). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education, 2,* 37-50.

Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on comparison of Classical Test Theory and Item Response Theory and their applications to test development. *Educational Measurement: Issues and Practice, 12,* 38-47.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Hatcher, P. J., Hulme, C., Miles, J. N. V., Carroll, J. M., Hatcher, J., Gibbs, S., Smith, G., Bowyer-Crane, C., & Snowling, M. J. (2006). Efficacy of small group reading intervention for beginning readers with reading-delay: A randomized controlled trial. *Journal of Child Psychology & Psychiatry, 47,* 820-827.

Hindson, B. A., Byrne, B., Fielding-Barnsley, R., Newman, C., Hine, D. W., & Shankweiler, D. (2005). Assessment and early instruction of preschool children at risk for reading disability. *Journal of Educational Psychology, 97,* 687-704.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111,* 127-155.

Høien, T., Lundberg, I., Stanovich, K. E., & Bjaalid, I. (1995). Components of phonological awareness. *Reading and Writing: An Interdisciplinary Journal, 7,* 171-188.

Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002). Phoneme awareness is a better predictor or early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology, 82,* 2-28.

Liberman, I. Y., Liberman, A. M., Mattingly, I., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. L. Venezky (Eds.), *Orthography, reading, and dyslexia* (pp. 137-153). Baltimore: University Park Press.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18,* 201-212.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading: Addison-Wesley.

Lukatela, K., Carello, C., Shankweiler, D., & Liberman, I. Y. (1995). Phonological awareness in illiterates: Observations from Serbo-Croatian. *Applied Psycholinguistics, 16,* 463-487.

McBride-Chang, C., & Kail, R. V. (2002). Cross-cultural similarities in the predictors of reading acquisition. *Child Development, 73,* 1392-1407.

Morais, J. (1991). Constraints on the development of phonemic awareness. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy* (pp. 5-27). Hillsdale, NJ: Erlbaum.

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

Read, C., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition, 24,* 31-34.

Reitsma, P. (2002). Precursors of phonemic awareness. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 33-48). Amsterdam: John Benjamins.

Rietveld, A. C. M., & van Heuven, V. J. (1997). *Algemene fonetiek* [General phonetics]. Bussum: Coutinho.

Scarborough, H. S. (1998). Early identification of children at risk for reading disabilities: Phonological awareness and some other promising predictors. In B. K. Shapiro, P. J.

Accardo, & A. J. Capute (Eds.), *Specific reading disability: A view of the spectrum* (pp. 75-119). Timonium, MD: York Press.

Schaerlaekens, A., Kohnstamm, D., & Lejaegere, M. (1999). *Streeflijst woordenschat voor zesjarigen* [Strived vocabulary list for six-year olds]. Lisse: Swets & Zeitlinger.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96,* 265-282.

Schreuder, R., & van Bon, W. H. J. (1989). Phonemic analysis: Effects of word properties. *Journal of Research in Reading, 12,* 59-78.

Schrooten, W., & Vermeer, A. (1994). Woorden in het basisonderwijs. 15.000 woorden aangeboden aan leerlingen [Words in primary education. 15.000 words offered to pupils]. *Studies in meertaligheid, 6,* 1-44.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55,* 151-218.

Shaywitz, S. E., & Shaywitz, B. A. (1993). Learning disabilities and attention deficits in the school setting. In L. J. Meltzer (Ed.), *Strategy assessment and instruction for students with learning disabilities: From theory to practice* (pp. 221-245). Austin, TX: PRO-ED.

Snow, C. E., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington: National Academy Press.

Stanovich, K. E. (1992). Speculations on the causes and consequences of individual differences in early reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 307-342). Hillsdale, NJ: Erlbaum.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38,* 175-190.

Torgesen, J. K. (1998). Catch them before they fall: Identification and assessment to prevent reading failure in young children. *American Educator, 22,* 32-39.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Rose, E., Lindamood, P., Conway, T., et al. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91,* 579-593.

Troia, G. A. (1999). Phonological awareness intervention research: A critical review of the experimental methodology. *Reading Research Quarterly, 34,* 28-52.

van Bon, W. H. J., & van Leeuwe, J. F. J. (2003). Assessing phonemic awareness in kindergarten: The case for the phonemic recognition task. *Applied Psycholinguistics, 24,* 195-219.

van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of modern item response theory.* New York: Springer.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What we have learned in the past four decades? *Journal of Child Psychology and Psychiatry, 45,* 2-40.

Verhelst, N. D., & Glas, C. A. W. (1994). The One Parameter Logistic Model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models, foundations, recent developments and applications* (pp. 215-238). New York: Springer-Verlag.

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.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., et al. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33,* 468-479.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23,* 159-177.

Ziegler, J. C., & Goswami, U. C. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131,* 3-29.

2

Screening of phonological awareness: An IRT approach[1]

Abstract The purpose of the present study was to explore the possibilities for the assessment of growth in phonological awareness of children in kindergarten and first grade. Phonological awareness was measured using four sets of items involving rhyming, phoneme identification, phoneme blending, and phoneme segmentation. The results of an exploratory factor analysis and analyses conducted within the framework of item response theory (IRT) showed one latent ability to underlie the different sets of items. Analyses in terms of the children's ability further showed the phonological awareness measures to be sensitive to growth. The amount of information supplied by the different sets of items depended on the children's level of ability. The conclusion that it is possible to accurately monitor the development of children's phonological awareness in the early elementary grades appears to be justified, and this possibility opens up new perspectives for the early screening for reading problems and dyslexia.

[1] Reference: Vloedgraven, J. M. T., & Verhoeven, L. (2007). Screening of phonological awareness in the early elementary grades: An IRT approach. *Annals of Dyslexia, 57,* 33-50.

2.1 Introduction

During the last few decades, considerable attention has been paid by researchers, educators, and politicians to early screening for reading problems. Several studies have shown that early intervention can prevent later reading problems (Ehri et al., 2001; Slavin, Karweit, & Wasik, 1994). A major obstacle to early screening, however, is that the children have yet to receive formal literacy instruction (Fawcett & Nicholson, 2000). The adoption of a "predictor approach" is therefore called for and the precursors of reading acquisition must be identified.

According to the relevant research literature, one of the strongest predictors of reading skills is phonological awareness. Stanovich (1994) and Elbro (1996) have both suggested that phonological awareness may be even more important than intelligence, vocabulary, and listening comprehension for the prediction of reading development. Phonological awareness refers to access to and an understanding of the sound structure of a spoken language, that is, the awareness that oral language can be broken down into individual words and, in turn, words into individual phonemes (cf. Wagner et al., 1997). Previous research has shown phonological awareness to be strongly related to early reading skills (Bradley & Bryant, 1983; Høien, Lundberg, Stanovich, & Bjaalid, 1995; Liberman, 1973; Perfetti, Beck, Bell, & Hughes, 1987; Wagner & Torgesen, 1987; Wagner, Torgesen, & Rashotte, 1994). There is also evidence that phonological deficits are the critical factor underlying reading problems (Elbro, Nielsen, & Petersen, 1994; Rack, Snowling, & Olson, 1992; Vellutino, Fletcher, Snowling, & Scanlon, 2004). In addition, interventions aimed at the improvement of phonological awareness have been shown to effectively promote learning to read (Ehri et al., 2001; Lundberg, Frost, & Petersen, 1988).

2.1.1 Problems with the measurement of phonological awareness

Researchers have encountered several problems with the measurement of phonological awareness. The first problem concerns the content validity of phonological awareness as a theoretical construct. A wide variety of tasks have been used to measure phonological awareness: rhyming tasks, phoneme counting tasks, sound comparison tasks, blending tasks, segmentation tasks, and deletion tasks. There is ample evidence that these tasks differ in difficulty: rhyming tasks appear to be the easiest while tasks that require the manipulation of phonemes appear to be the most difficult (Adams, 1990; Chard & Dickson, 1999; Stanovich, Cunningham, & Cramer, 1984; Yopp, 1988). Just how these tasks relate to each other is far from clear. According to some researchers, the various aspects of phonological awareness measured by the tasks may actually reflect a single latent ability (Anthony & Francis, 2005; Anthony & Lonigan, 2004; Anthony et al., 2002; Stahl & Murray, 1994; Stanovich et al., 1984). In contrast, Yopp (1988) has argued that the construct of phonemic awareness consists of two highly related factors that nevertheless differ in the number of cognitive operations that they require: a simple phonemic awareness factor, which requires one operation, and a compound phonemic awareness factor, which requires an extra operation while holding the results of the first operation in memory. Muter, Hulme, Snowling, and Taylor (1997) have provided evidence for two other distinct factors and shown rhyming ability to be separate from segmentation ability. Høien et al. (1995) found three basic components to characterize phonological awareness: a phoneme factor, a syllable factor, and a rhyming factor. Clearly, there is still no consensus on the structure of phonological awareness.

Another problem with the measurement of phonological awareness is that the measures are often inaccurate. Inaccuracy problems may be caused by the fact that the suitability of a specific task depends on the child's level of development (Anthony & Lonigan, 2004; Chard & Dickson, 1999; Schatschneider, Francis, Foorman, Fletcher & Mehta, 1999). According to Hambleton, Swaminathan, and Rogers (1991), standard errors are only small when the difficulty of a test fits the ability of the examinee. Phonological awareness skills appear to lie along a continuum of increasing difficulty. By the end of kindergarten, for example, children have generally developed the ability to rhyme (Chard & Dickson, 1999). If a rhyming task is then administered in first grade, most of the children will obtain a maximum score and, in this case, the exact level of each child's phonological awareness is still unknown. Conversely, a phonological awareness task may be too difficult at times. In a study by de Jong and van der Leij (1999), for example, no evidence was found for a relation between phonological awareness in kindergarten and reading performance in first grade probably due to the fact that two of the three tasks appeared to be too difficult for the kindergartners. These examples illustrate that if a task is not administered at the proper moment in a child's development, inaccurate measurement will be the result.

A related problem with the measurement of phonological awareness is that growth in this ability is hard to establish. If the children's abilities are not accurately measured, growth also cannot be accurately assessed. One possible solution to this problem is to administer different tasks of phonological awareness at different points in time (i.e., different developmental levels). However, comparison of task scores is made difficult, if not impossible, by the use of different scales and no demonstration of functional relations between the different scales. It is thus difficult to measure growth in phonological awareness, and this is a major problem for the identification of children who are at risk for reading problems. Several studies have shown that the measurement of phonological awareness and growth in this capacity are critical for the early identification of reading problems (Byrne, Fielding-Barnsley, & Ashley, 2000; Hindson et al., 2005; Spector, 1992). Growth in phonological awareness appeared to account for variance in reading in addition to that accounted for by the actual level of phonological awareness ability. Not only children's reading abilities but also their phonological abilities should thus be monitored during the development of beginning literacy as only the proper monitoring of children's (pre)literacy skills can enable the early identification of reading problems and dyslexia (Vellutino et al., 2004; Vellutino, Scanlon, & Lyon, 2000).

In sum, there are some major problems with the measurement of phonological awareness. First, it is unclear how the tasks used to measure the different aspects of phonological awareness relate to each other. Second, inaccurate measurement is a problem. And third, it is hard to measure growth. Most of the relevant studies use models from classical test theory (CTT) to assess the level of phonological awareness and predict the acquisition of beginning reading ability. However, the problems just described are difficult to resolve within the framework of CTT. A first problem with CTT is that scores have been found to depend on the particular set of items administered (i.e., be test-dependent). Another problem is that item parameters are group-dependent (i.e., characteristics such as item difficulty and discriminatory capacity appear to depend on the group to which the items are administered). Once again, these problems make it difficult to compare scores from different tasks (Hambleton & Jones, 1993). Even if the same task is completed by the child on different occasions (i.e., points in development), score comparison is still difficult because the accuracy of the measurement can vary across time.

An alternative approach is item response theory (IRT) or what is also known as latent trait theory (Hambleton et al., 1991). The distinctive feature of IRT models is that they relate item responses to ability: the difficulty of the items and the ability of persons are scaled on the same metric. Two assumptions hold for the specification of IRT models. First, it is assumed that the ability to be measured is unidimensional. Second, it is assumed that the relation between the latent trait and the probability of a correct response on a particular item can be described by the item characteristic curve (ICC). This curve is defined by one or more parameters, which determine the exact shape of the ICC. IRT has several advantages over CTT. A first advantage is that the estimated ability is test-independent, provided the different tasks are constructed from an IRT calibrated item bank. A second advantage is that the item parameter estimates are independent of the sample from which they are obtained. These two advantages make it possible to compare scores from different tasks. Another major advantage of IRT is the possibility to show the contribution of particular items and tasks to the assessment of ability (Lord, 1977). For the construction of early screening tasks, then, the test designer can select those items that provide the most information with regard to a particular ability and thereby develop the most accurate measures.

2.1.2 The present study

In the previous sections, the importance of measuring growth in phonological awareness was highlighted. Two methodological points appeared to be of particular importance. First, the construct of phonological awareness, as measured by various sets of items, has to be unidimensional. Second, the measures used to monitor the development of phonological awareness need to be sufficiently sensitive to growth (Kaminski & Good, 1996). The present study attempts to answer two questions related to these two methodological points. The first question is whether the different sets of items intended to measure phonological awareness appear to reflect a single underlying ability or several related abilities. For this purpose, the underlying structure of phonological awareness will be addressed from an IRT perspective. The second question is whether the items intended to measure phonological awareness can be used to measure growth from kindergarten through first grade. As already mentioned, the use of inaccurate measures is a major problem for the assessment of phonological awareness.

With regard to the first question, an initial attempt to identify the underlying structure of children's phonological awareness by the use of IRT was already undertaken by Schatschneider et al. (1999). The results of a factor analysis and the fit of an IRT model suggested that phonological awareness can be conceived as a unitary construct. A limitation on the study by Schatschneider et al., however, is that a rhyming task was not included. This means that the authors could neither confirm nor reject the findings of Muter et al. (1997) who found evidence suggesting that rhyming ability and segmentation ability may be separate. In the present study, we therefore administered four different types of items, which included rhyming items, to examine the underlying structure of phonological awareness. Schatschneider et al. also tested children speaking English while in the present study children speaking Dutch participated. It is the question whether the nature of phonological awareness is expected to be different in these two languages. Given the great overlap in phonological principles, it can be hypothesized that the sequence of phonological awareness development (i.e., from large units of sound to small units of sound) is the same for languages like English and Dutch (Ziegler & Goswami, 2005). Because of the fact that the relation between phonological awareness and reading is bidirectional (cf. Perfetti et al., 1987), it is important to look at the differences between the orthographies of the two languages as well. Seymour, Aro, and Erskine (2003) concluded on the basis of a cross-linguistic comparison of different orthographies that Dutch and English orthography share a complex syllabic structure, but differ in orthographic depth. Because the orthographic depth in Dutch is evaluated to be smaller than in English, we expect Dutch children to be faster in developing phonological awareness without a change in the underlying structure of phonological awareness. Educational environment neither seems to alter the structure of phonological awareness, because both in the Netherlands and in England a phonics teaching method is primarily used. Given the fact that the orthographic depth only seems to influence the rate of development in phonological awareness and not the underlying structure, the present study can by and large be seen as a replication of the study by Schatschneider et al. with, as central hypothesis, a unidimensional structure of phonological awareness.

Related to the issue of the unidimensional or multidimensional structure of phonological awareness is the issue of the relative difficulty of the different sets of items. As already noted, the various sets of items used to measure phonological awareness have been found to differ in difficulty. However, the exact differences between the various item sets are still open to investigation. In addition to these differences between tasks, differences in linguistic complexity within tasks appear to influence phonological awareness (Anthony & Francis, 2005). For example, according to Schreuder and van Bon (1989) the consonant-vowel (CV) structure is an important determinant. Therefore, as a next step, we have investigated the differences in difficulty of various CV structures.

The second question to be investigated is whether the items, measuring phonological awareness, are able to measure growth from kindergarten to grade 1. As mentioned earlier, the lack of accurate measures for the assessment of growth in phonological awareness is a major problem. If the measures used in the present study appear to be sensitive to growth in phonological awareness, then the accuracy for the different sets of items and the complete set of items will be examined for a range of ability scores. Results will show which set(s) of items are of importance in assessing the ability of kindergartners and first graders.

2.2 Method

2.2.1 Participants

A total of 172 children in their second year of kindergarten (KG) and 173 grade one children (G1) were randomly selected from 12 elementary schools in the east part of the Netherlands. The total sample included 177 boys and 168 girls. All of the children spoke Dutch and were from a variety of socioeconomic backgrounds. In the Dutch educational system, children visit school from the moment they are 4 years old, after which they spend 2 years in kindergarten. After these 2 years, children enter first grade. In kindergarten, literacy education is generally limited to some language games to stimulate phonological awareness and beginning literacy. Formal instruction in reading and spelling starts in first grade and from that moment, explicit instruction in phonics is offered.

The children were tested in April or May of 2005. At the time of testing, the mean age of the kindergartners was 6 years and 1 month (*SD* = 4.4 months); the mean age of the first graders was 7 years and 1 month (*SD* = 3.9 months).

2.2.2 Materials

To select tasks for phonological awareness, we have looked at the extent to which a task represents phonological awareness ability. In addition, the predictive value for reading performance was taken into account. Furthermore, we have selected various tasks that according to the literature are known to differ in difficulty to be able to accurately assess the ability of both high- and low-ability individuals. Taking these criteria into consideration, the following four tasks were selected: rhyming, phoneme identification, phoneme blending, and phoneme segmentation (Adams, 1990; Chard & Dickson, 1999; Høien et al., 1995; Vellutino & Scanlon, 1987; Yopp, 1988). The tasks consisted of high frequency monosyllabic words containing two, three, four, or five phonemes. The CV structure of the target words varied. The target words were selected from current Dutch word frequency lists (Schaerlaekens, Kohnstamm, & Lejaegere, 1999; Schrooten & Vermeer, 1994). Given that a developmentally appropriate test for phonological awareness should not overload working memory (Reitsma, 2002), all of the words were presented both auditorily and visually. In all cases, the presented pictures were previously named to be certain that the correct names were associated with the pictures.

Rhyming. Three pictures were shown to the children. The target word was then presented auditorily (via the computer) and the children were asked to select the word that rhymed with the target word. All of the target words were CVC words. Each child was given 30 items.

Phoneme identification. Three pictures were presented to the children. The target phoneme was then pronounced along with a word that started with the same phoneme. The child's task was to select the picture that started with the same sound that the target word started with. Only consonants were used as target phonemes and articulated as sounds. All of the target words were CVC words. Each child was given 30 items.

Phoneme blending. Three pictures were presented to the children. The isolated phonemes from the target word were then pronounced in their correct order. The child's task was to select the picture that represented the target word. To be able to measure blending ability as purely as possible, distractors contained one or more of the phonemes of the target word. The target words consisted of three, four, or five phonemes with different CV structures. Each child was given 40 items.

Phoneme segmentation. The target word was presented visually as well as auditorily. The child was asked to say the phonemes of the target word separately in the correct order. Word length was two, three, four, or five phonemes. Each child was given 40 items.

2.2.3 Procedure

All of the tasks were administered individually and presented on a computer. For kindergartners, the tasks were administered in two sessions of about 20 minutes each, because of their relatively short attention spans. First graders were tested in one session, which took approximately 20 minutes to complete. The rhyming task was presented to the kindergartners only because it is well known that this task is the easiest phonological awareness task and most suitable for kindergartners (Adams, 1990; Chard & Dickson, 1999). In addition, Schatschneider et al. (1999) have shown identification of the first sound in a word to provide a poor estimate of phonological ability for first-grade children because the task is too easy for this age. Therefore, this task was also administered in kindergarten only.

Three practice items preceded each task to familiarize the children with the testing procedure. After each practice item, the experimenter provided feedback on the correctness of the child's response. If the child gave an incorrect answer, the correct answer was provided.

As mentioned earlier, phonological awareness was measured by four different tasks. An item bank has been constructed that contained four sets of items representing these four different tasks (i.e., 45 rhyming items, 45 phoneme identification items, 60 phoneme blending items, and 60 phoneme segmentation items). Because it was not feasible to present all of the items to all of the children, we used a structural incomplete design, called the anchor-test design (Petersen, Kolen, & Hoover, 1989). Therefore, all of the items of a task were divided in three modules and each child was given two of the three modules of a task (i.e., booklet). To be able to administer all of the items from the item bank, different groups of children were given different booklets (i.e., different combinations of modules). Characteristic of this design is the link between booklets: the different booklets have certain anchor items in common. And by the use of these anchor items, it is possible to present all of the items at one scale of measurement. The design of the study is presented in Figure 2.1.

	Rhyming			Identification			Blending			Segmentation		
Grade	A	B	C	D	E	F	G	H	I	J	K	L
KG	X	X		X	X		X	X		X	X	
KG		X	X		X	X		X	X		X	X
KG	X		X	X		X	X		X	X		X
G1							X	X		X	X	
G1								X	X		X	X
G1							X		X	X		X

Figure 2.1 *Anchor-test design for rhyming, phoneme identification, phoneme blending, and phoneme segmentation*

2.2.4 Statistical analyses

The four sets of items intended to measure phonological awareness were submitted to several analyses in order to establish their underlying structure. As a first step, we performed an exploratory factor analysis on the matrix with tetrachoric correlations of the items. The factor analysis was conducted using the minimum residuals (MINRES) method (Harman & Jones, 1966). The MINRES method minimizes the sum of squared residuals, resulting in a matrix of factor loadings.

The next step in the analyses involved the use of IRT models. The phonological awareness items were calibrated using the one-parameter logistic model (OPLM). The OPLM combines the attractiveness of the mathematical properties of the one-parameter model with the flexibility of the two-parameter model. In the OPLM, the difficulty parameters are estimated and discrimination parameters are dealt with as known constants (i.e., discrimination indices can vary, but have discrete values). The discrimination indices were based on a geometric mean of 3. For a detailed description of the fundamental concepts and the favourable practical applications of the OPLM, see Verhelst and Glas (1994).

To estimate the item parameters, we used the conditional maximum likelihood (CML) procedure. A one-way ANOVA was conducted, followed by a Tukey's test, to establish the differences between the difficulty parameters for the four sets of items. Person parameters (ability) were estimated by means of the weighted maximum likelihood (WML) procedure. To assess the ability distributions of the two populations, the marginal maximum likelihood (MML) method was used. This method was applied by fixing the item parameters and estimating a mean and a standard deviation in a normal distribution. The MML method is preferred over the WML method for the estimation of means and standard deviations in the population because the individual WML estimators tend to overestimate means and standard deviations due to measurement error. Thereafter, the Cohen's *d* was used to measure the strength of growth from kindergarten through first grade.

2.3 Results

2.3.1 Underlying structure of phonological awareness

First, a matrix with tetrachoric correlations was computed for all of the items. Next, a two-factor analysis on this matrix of correlations using MINRES was conducted. On the basis of the factor loadings for the items on both factors, the eigenvalues were then computed. This resulted in a powerful first factor, which extracted 82% of the total variance. The second factor accounted for 18% of the total variance. It should be noted that tetrachoric correlations have relatively large standard errors (Brown, 1977). In the case of a small sample size, this complicates identification of the correct number of factors. The results of the analyses with MINRES should therefore be interpreted with caution. The large percentage of the variance explained by the first factor and the significant difference in the contribution of the second factor can nevertheless be seen as evidence for unidimensionality.

Assuming unidimensionality, we examined whether the OPLM fits the data. To assess model fit, both the item-oriented statistics and an overall statistic were computed. First, the OPLM can be used to determine if the individual items fit the same latent trait. For each item, an indication of the fit into the model is provided by the *p* value. A formal means to judge the distribution of the *p* values for all of the items is not available (Verhelst, Glas, & Verstralen, 1995). However, it is certain that a majority of low *p* values indicate model violations, and it is desirable that the frequencies of the *p* values are rectangularly distributed at the interval [0, 1]. It appeared that none of the items had a *p* value lower than .05. Moreover, the distribution of the frequencies was fairly balanced across the interval, showing model fit. Second, additional information about the model fit can be provided by the overall R1c-test. The R1c value was 584.01 (df = 537, p = .08), which suggests that the different sets of items can be included in the same scale.

When the OPLM fits the data, the invariance of the ability and item parameters can be established. Ability invariance means that the estimated ability of each person does not depend on the specific set of items administered. Invariance of item statistics means that the item parameters derived from the model are independent of the specific sample (Hambleton et al., 1991). To assess whether the OPLM fits the data, we investigated these two properties. First, we dealt with the question of ability invariance. The items were divided into two subsets: even and odd items. For each child, the ability parameters were then estimated for the two subsets of items. Thereafter, a scatter plot of the pairs of ability estimates was made (see Figure 2.2a). If the ability estimates are invariant, the plot should demonstrate a straight line. As can be seen, a strong linear association was indeed found to hold between the ability estimates for the even and odd items.

The invariance of the item statistics was next examined by determining the associations between the difficulty parameters estimated from two different samples. For this purpose, the total sample was split into two subpopulations: boys and girls. The estimated difficulty parameters for the boys were then plotted against the estimated difficulty parameters for the girls. The corresponding scatter plot is presented in Figure 2.2b. As can be seen, the relation between the difficulty parameters for the two different samples appears to be linear, which clearly indicates the invariance of the item parameters. These results provide strong evidence for the assumption of a unidimensional underlying structure for phonological awareness.

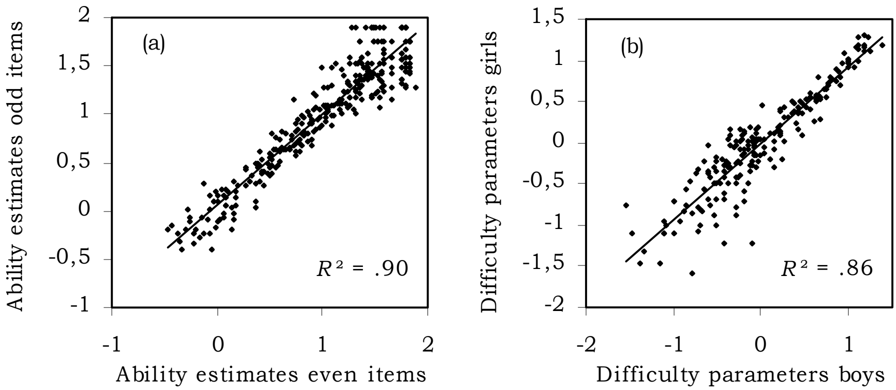

Figure 2.2 *Invariance of ability and item parameters: (a) ability estimates even and odd items; (b) difficulty parameters boys and girls*

2.3.2 Item parameters

The preceding results showed phonological awareness to be well represented by a single underlying scale. The next question is just how the various sets of items measuring phonological awareness relate in terms of difficulty. In Table 2.1, the average difficulty and discrimination parameters for the four sets of items measuring phonological awareness are presented from least to most difficult. The rhyming items turned out to be the easiest, and the phoneme segmentation items turned out to be the most difficult. The items measuring phoneme blending and phoneme identification occurred in between. A one-way ANOVA showed the differences in difficulty to also be significant, F (3, 192) = 188.78, $p < .01$. Multiple comparisons were next conducted using the Tukey procedure to determine which pairs of item sets differed significantly from each other. The analyses showed all of the pairs of item sets to differ significantly, with the exception of the difference between the phoneme identification items and the phoneme blending items. As can be seen in Table 2.1, the various sets of items also differ in their capacity to discriminate between high- and low-ability individuals. Items measuring phoneme segmentation turned out to be most discriminating while items measuring rhyming ability turned out to be least discriminating.

Table 2.1
Average difficulty and discrimination parameters for the four sets of items measuring phonological awareness

	Difficulty parameter		Discrimination parameter	
Set of items	*M*	*SD*	*M*	*SD*
Rhyming	-0.69	0.36	2.33	0.68
Phoneme blending	-0.16	0.32	2.85	0.78
Phoneme identification	-0.08	0.25	2.57	0.63
Phoneme segmentation	0.77	0.30	4.98	1.04

In addition to differences between the four sets of items, effects of CV structure within tasks have been investigated. With respect to phoneme blending, items were divided into three sets of items: (1) CVC; (2) CVCC and CCVC; and (3) CCVCC, CCCVC, and CVCCC. Items of the first item set appeared to be the easiest and items of the third item set appeared to be the most difficult. However, a one-way ANOVA showed the differences in difficulty not to be significant, F (2, 52) = 2.17, $p = .124$. With respect to phoneme segmentation, items were divided into five item sets: (1) CV and VC; (2) CVC; (3) CCV and VCC; (4) CCVC and CVCC; and (5) CCVCC, CCCVC and CVCCC. Table 2.2 shows the average difficulty parameters for the various CV structures and ranks the different item sets from least to most difficult. Results of an ANOVA analysis revealed that differences in difficulty between the various CV structures were significant, F (4, 49) = 29.16, $p < .01$. This leads to the next question, that is, between which pairs of item sets are the differences significant? Results of Tukey's test showed most of the pairs to differ significantly, except the distinction between CV, VC and CVC; between CCV, VCC and CCVC, CVCC; and between CCVC, CVCC and CCVCC, CCCVC, CVCCC. Considering these findings, we may conclude that merely lengthening a word does not have an effect on the difficulty of segmentation. Differences are only significant when a pair of items differs in the distribution of consonant clusters. Another finding that confirms the effect of consonant clusters is that the difference between CVC and CCV, VCC words turns out to be significant, with the latter being the most difficult, despite the fact that they are similar in word length. The existence of one or more consonant clusters in a word thus appears to complicate the performance in a segmentation task.

Table 2.2
Average difficulty parameters for the different CV structures within phoneme segmentation

	Difficulty parameter	
Set of items	*M*	*SD*
CV, VC	0.37	0.13
CVC	0.56	0.15
CCV, VCC	0.81	0.23
CCVC, CVCC	0.97	0.20
CCVCC,CCCVC, CVCCC	1.10	0.13

2.3.3 Growth in phonological awareness

To establish whether the phonological awareness measures are sensitive to growth, the progress of the children from kindergarten to first grade was investigated. The results concerning growth should be interpreted cautiously because the subjects are separate groups of kindergartners and first graders. Despite this, results are expected to give valuable indications about the development in phonological awareness. The ability distributions for the kindergartners (*M* = 0.582, *SD* = 0.485) and first graders (*M* = 1.677, *SD* = 0.341) are presented in Figure 2.3 below the *x*-axis. As can be seen, the first graders improved importantly. The Cohen's *d*, which is an objective measure of the strength of growth, was 2.60 (Cohen, 1988). Given that an effect of 0.80 is interpreted as a large effect, an effect of 2.60 can be judged to be a substantial effect.

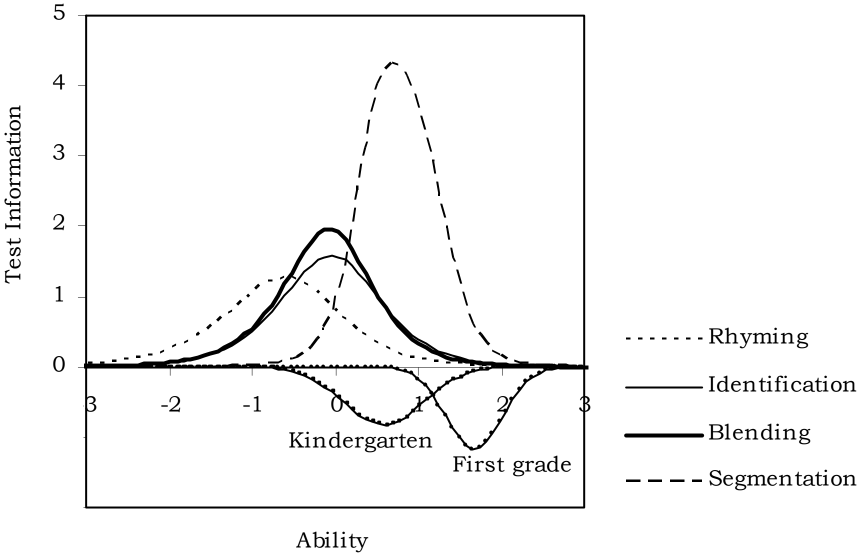

Figure 2.3 *Test information functions for the four sets of items measuring phonological awareness*

2.3.4 Information functions and accuracy of ability estimates

The difficulty and discrimination parameters for the different sets of items used to measure phonological awareness were just compared. A restriction on these comparisons is that they lack information on which task or tasks may be most useful for the population of interest. As already mentioned, one of the advantages of IRT is the possibility to show the contributions of particular items and sets of items. This can be realized by calculating test information functions (for a detailed description of the concept of information, see Appendix A). These functions link information from both the difficulty and discrimination parameters. In such a manner, it is possible to specify just how well a task estimates ability across the total distribution of ability. In other words, information functions indicate the accuracy of measurement of the different tasks for different ability levels. Information functions frequently diverge across the range of ability scores; a task is possibly more informative for high-scoring individuals than for low-scoring individuals or the other way round. The analogous measure for information functions in CTT is reliability. Nevertheless, reliability in IRT can not be compared with reliability in CTT without any problems. Because in IRT reliability is different for each point of the latent ability scale, it is also called local reliability. However, it is possible to transform the local measurement precision in IRT to the classical measure of reliability. Before we continue with accuracy of measurement in IRT, first we will mention the classical indices for reliability because these indices are easier to interpret. The reliability of each task from a CTT framework was determined using the MAcc coefficient (Verhelst et al., 1995). For rhyming the MAcc appeared to be .83, for phoneme identification .91, for phoneme blending .96, and for phoneme segmentation .99. All of the tasks appeared to be sufficiently accurate in estimating phonological awareness skills.

As a next step, the information functions for each of the four sets of items were thus computed. Given that the amount of information provided by a set of items is influenced by the number of items, we have computed the mean information per item in each set of items. The four information functions are presented in Figure 2.3. In the same figure, the ability distributions for both subpopulations (i.e., kindergartners and first graders) are plotted below the *x*-axis. This gives the opportunity to see at a glance which specific set of items is most informative at a particular level of ability; the higher the information function, the more accurate the estimates of ability are at a given point of the ability scale. The maximum of the information function for phoneme segmentation appears to be highest and occurs at approximately 0.75 on the ability scale, which shows the segmentation items to be most informative at that point of the ability scale. In addition, relative to the information functions for rhyming, phoneme identification, and phoneme blending, the information function for phoneme segmentation has moved to the right along the ability axis. This means that this task more accurately estimates the ability of kindergartners with a higher ability and of first graders. However, items measuring phoneme segmentation provide weaker estimates for those children at the highest end of the ability range. Although the items from the three other sets are generally less informative, they nevertheless provide more information about the relevant capacities of the kindergartners with lower levels of ability than the phoneme segmentation items.

In addition to the information functions for the different sets of items, the total test information function for all of the items was also calculated (see Figure 2.4). As can be seen, the four sets of items together provide the most precise estimates of children with an ability score of about 0.5, which corresponds to the average ability score for kindergartners. The total test does not provide an accurate estimate of the ability of first graders with an average or above average ability score. However, the estimates for kindergartners at the lower end of the ability distribution are still satisfactory.

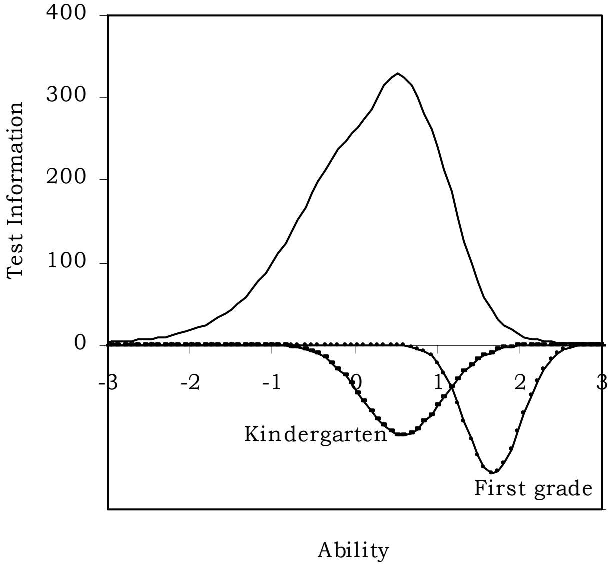

Figure 2.4 *Total test information function*

2.4 Conclusions and discussion

Several conclusions can be drawn on the basis of the present results. With respect to the underlying structure of phonological awareness, an exploratory factor analysis showed one latent ability to underlie the different sets of items used to measure phonological awareness within the context of the present study. The results showed the first factor to account for a large percentage of the variance and an enormous difference in the contribution of the first and second factor, which are findings highly indicative of a single underlying factor (i.e., the unidimensionality of phonological awareness). The assumption of unidimensionality was further investigated using a model based on IRT. Both the item-oriented and overall statistical tests showed the OPLM to fit the data. Ability invariance and item parameter invariance were also demonstrated, which supports the conclusion that the various sets of items used to measure phonological awareness indeed reflect one and the same latent ability. This result is in accordance with the outcomes of the study by Schatschneider et al. (1999) who tested English-speaking children. As we expected, differences in orthographic depth between English and Dutch did not influence the underlying structure of phonological awareness. In contrast to Schatschneider et al., we included a rhyming task. The results of the present study further support Treiman's (1985) claim that although rhyming deals with larger linguistic units than phonemes, the cognitive operations needed to rhyme also require awareness of abstract speech representations.

Given the indications that one latent ability underlies the different sets of items, the next issue to be addressed was the relative difficulty of the different sets of items. The results of the ANOVA and Tukey analyses indeed showed the sets of items to differ in difficulty. The rhyming items appeared to be the easiest and the phoneme segmentation items appeared to be the most difficult with the phoneme blending and phoneme identification items occurring in between. The differences between all of the pairs of item sets were significant, with the exception of the difference between the sets of items of phoneme blending and phoneme identification. These results show the cognitive task requirements for the sets of items on the unidimensional phonological awareness scale to clearly differ. The finding that the rhyming items were the easiest is in agreement with the findings of many other studies (Adams, 1995; Chard & Dickson, 1999; Stanovich et al., 1984; Yopp, 1988). The present findings also confirm the findings of previous research showing the extreme difficulty of phoneme segmentation and the intermediate difficulty of phoneme blending and phoneme identification (Høien et al., 1995).

Furthermore, our examination of the relative difficulty of various CV structures within tasks showed no effect for phoneme blending, which may be due to the relative ease of this task. However, differences in CV structure in the segmentation task appeared to be significant: the longer the word, the more difficult it was to segment that word in separate phonemes. Closer inspection of the significant differences between all pairs of items revealed that longer words were only harder to segment when one or more consonant clusters were added. This finding is in agreement with previous research (Arnqvist, 1992; Schreuder & van Bon, 1989; Treiman & Weatherston, 1992).

The second issue to be addressed in the present study was whether the phonological awareness measures were also sensitive to growth. The strength of growth from kindergarten to first grade was indicated by Cohen's *d*, which appeared to be 2.60 and can thus be interpreted as a substantial effect. It is thus possible to measure growth in phonological awareness during the development of beginning literacy. The accuracy of the various sets of items across the spectrum of kindergartners and first graders with different degrees of ability was determined by investigating the information functions of the four sets of items and the ability distributions of both subgroups. On the basis of these results, the assumption that the appropriateness of a particular task depends on the level of child development (Anthony & Lonigan, 2004; Chard & Dickson, 1999; Schatschneider et al., 1999) received support. Our results indeed showed the usefulness of the various sets of items to depend upon the difficulty of the items and the abilities of the child. The IRT model showed the phoneme segmentation items to provide the most information about ability. Closer inspection of the information function showed the phoneme segmentation set of items to estimate the ability of higher scoring kindergartners and lower scoring first graders most accurately; for lower scoring kindergartners and higher scoring first graders, however, the estimates were less accurate. Although the information provided by rhyming performance, phoneme blending, and phoneme identification is relatively low, inclusion of these sets of items in addition to phoneme segmentation items in an instrument for early screening may be critical as exactly these aspects of phonological awareness appear to be most informative for those children at the lower end of the ability continuum.

Information functions are determined to a great extent by the discriminating power of the items. As mentioned earlier, segmentation items are the most discriminating. From the information function of segmentation, we can derive that this is especially valid for kindergartners and lower scoring first graders. This can be explained by the fact that segmentation items best suited the ability level of these children. A striking result was that the discriminating power of the phoneme segmentation items decreased substantially as children improve their phonological awareness ability during first grade. Due to the start of literacy education with explicit instruction in phonics in first grade, children generally master the ability to segment words into phonemes in the course of first grade. At the end of this year, segmentation items are too easy for most of the children and are thus not able any more to differentiate between high- and low-ability children.

When the total test information for all of the items is examined, we can conclude that the most accurate estimates are obtained for the average kindergartner. The four sets of items together adequately measure the ability of lower scoring kindergartners. However, as the children's abilities increase during first grade, the ability estimates become less and less accurate. These results suggest that inclusion of the four sets of items in a screening instrument can be recommended but that another set of more difficult items should also be included to improve the accuracy of measurement for first graders in particular. Adams (1990), for example, has described the different levels of difficulty for phonemic awareness and found phoneme manipulation to be most difficult as this requires the addition or omission of phonemes to formulate a new word.

In sum, the results of the present study have shown that it is possible to measure growth in phonological awareness. The various sets of items used to measure phonological awareness could be placed along a single ability scale and were found to measure changes in phonological awareness (i.e., growth). However, a refinement of the ability scale is necessary to attain more accurate ability estimates for the higher end of the ability range.

The findings of the present study have some important implications for the early screening of reading problems and dyslexia. The results show that the development of phonological awareness can be accurately monitored. As already stated, several studies showed growth in phonological awareness to add unique information to the prediction of reading, which highlights the importance of monitoring the development of phonological awareness. When McBride-Chang, Wagner, and Chang (1997) investigated the development of phonological awareness, they found evidence for Matthew effects (Stanovich, 1986) of prereading skills. That is, children who started with a higher level of phonological awareness tended to improve their level of skill more quickly than children who started with a lower level of phonological awareness. Future research within the framework of IRT will show whether these results can be confirmed. If such Matthew effects are indeed found for phonological awareness, then a successful early start can be seen to be paramount and the value of early screening for reading problems thus reinforced. However, we have seen that the discriminating power of phonological awareness tasks decreases in the course of first grade. It is important to note that, despite adding a more difficult task to try to improve the accuracy of the measures in first grade, the value of screening of phonological awareness for the prediction of reading will steadily decrease as children's abilities increase. This finding is in concordance with the conclusion of de Jong and van der Leij (1999) that the predictive value of phonological awareness tasks is limited to the early phases of learning to read in the Netherlands. As a consequence, for the early identification of reading problems, it is thus of major importance to assess children's phonological awareness in kindergarten and in the first half of first grade. And most important is the measurement of growth in phonological awareness. Like intervention in reading is used to help in the distinction between reading difficulties caused by cognitive deficits and those caused by instructional deficits (Vellutino et al., 1996), intervention in phonological awareness can aid in the same way. It is clear that phonological abilities should be monitored to be able to identify children who show a phonological deficit given the fact that, even after receiving intervention, they hardly improve.

As already mentioned, the prediction of future reading skills clearly depends on the accuracy with which prereading skills are measured. The results of the present study have confirmed earlier findings showing the precision of the measurement by a set of items to depend on the child's actual level of ability. Valuable information has been provided on the utility of particular sets of items for kindergartners and first graders with different levels of ability. These findings further suggest that the influence of various aspects of phonological awareness on later reading skill may be constrained. This issue certainly merits further study and longitudinal study in particular to show which tasks best predict reading skill at different moments in a child's development. Greater research on this topic is also of major importance for improved early screening of children who are possibly at risk for reading failure.

The present study can be seen as a first attempt to investigate the underlying structure of phonological awareness as regards Dutch language from an IRT perspective. In a follow-up study, we will enlarge the sample size and also collect data about reading to be able to relate phonological awareness measures with reading measures.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14,* 255-259.

Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology, 96,* 43-55.

Anthony, J. L., Lonigan, C. J., Burgess, S. R., Driscoll Bacon, K., Phillips, B. M., & Cantor, B. G. (2002). Structure of preschool phonological sensitivity: Overlapping sensitivity to rhyme, words, syllables, and phonemes. *Journal of Experimental Child Psychology, 82,* 65-92.

Arnqvist, A. (1992). The impact of consonant clusters on preschool children's phonemic awareness: A comparison between readers and nonreaders. *Scandinavian Journal of Psychology, 33,* 29-35.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 301,* 419-421.

Byrne, B., Fielding-Barnsley, R., & Ashley, L. (2000). Effects of preschool phoneme identity training after six years: Outcome level distinguished from rate response. *Journal of Educational Psychology, 92,* 659-667.

Brown, M.B. (1977). Algorithm AS 116: The tetrachoric correlation and its standard error. *Applied Statistics, 26,* 343-351.

Chard, D. J., & Dickson, S. V. (1999). Phonological awareness: Instructional and assessment guidelines. *Intervention in School and Clinic, 34,* 261-70.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology, 91,* 450-476.

Ehri, L. C., Nunes, S., Willows, D., Schuster, B., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250-287.

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

Elbro, C., Nielsen, I., & Petersen, D. K. (1994). Dyslexia in adults: Evidence for deficits in non-word reading and in the phonological representations of lexical items. *Annals of Dyslexia, 44,* 205-226.

Fawcett, A. J., & Nicolson, R. I. (2000). Systematic screening and intervention for reading difficulty. In N. A. Badian (Ed.), *Prediction and prevention of reading failure* (pp. 57-85). Baltimore: York Press.

Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on comparison of Classical Test Theory and Item Response Theory and their applications to test development. *Educational Measurement: Issues and Practice, 12,* 38-47.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Harman, H. H., & Jones, W. H. (1966). Factor analysis by minimizing residuals (MINRES). *Psychometrika, 31,* 351-369.

Hindson, B. A., Byrne, B., Fielding-Barnsley, R., Newman, C., Hine, D. W., & Shankweiler, D. (2005). Assessment and early instruction of preschool children at risk for reading disability. *Journal of Educational Psychology, 97,* 687-704.

Høien, T., Lundberg, I., Stanovich, K. E., & Bjaalid, I. (1995). Components of phonological awareness. *Reading and Writing: An Interdisciplinary Journal, 7,* 171-188.

Kaminski, R. A., & Good, R. H. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review, 25,* 215- 227.

Liberman, I. (1973). Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society,* 23, 65-77.

Lord, F. M. (1977). Practical applications of item characteristic curve theory. *Journal of Educational Measurement, 14,* 117-138.

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly, 23,* 263-284.

McBride-Chang, C., Wagner, R., & Chang, L. (1997). Growth modeling of phonological awareness. *Journal of Educational Psychology, 89,* 621-630.

Muter, V., Hulme, C., Snowling, M., & Taylor, S. (1997). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of Experimental Child Psychology, 65,* 370-396.

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

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 221-262). New York: American Council on Education.

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly, 27,* 28-53.

Reitsma, P. (2002). Precursors of phonemic awareness. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 33-48). Amsterdam: John Benjamins.

Schaerlaekens, A., Kohnstamm, D., & Lejaegere, M. (1999). *Streeflijst woordenschat voor zesjarigen* [Strived vocabulary list for six-year olds]. Lisse: Swets & Zeitlinger.

Schatschneider, C., Francis, D. J., Foorman, B. F., Fletcher, J. M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology, 91,* 467-478.

Schreuder, R., & van Bon, W. H. J. (1989). Phonemic analysis: Effects of word properties. *Journal of Research in Reading, 12,* 59-78.

Schrooten, W., & Vermeer, A. (1994). Woorden in het basisonderwijs. 15.000 woorden aangeboden aan leerlingen [Words in primary education. 15.000 words offered to pupils]. *Studies in meertaligheid, 6,* 1-44.

Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology,* 94, 143-174.

Slavin, R. E., Karweit, N. L., & Wasik, B. A. (1994). *Preventing early school failure: Research, policy, and practice.* Boston: Allyn & Bacon.

Spector, J. E. (1992). Predicting progress in beginning reading: Dynamic assessment of phonemic awareness. *Journal of Educational Psychology, 84,* 353-363.

Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology, 86,* 221-234.

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. (1994). Romance and reality. *Reading Teacher, 47,* 280-291.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38,* 175-190.

Treiman, R. (1985). Onset and rimes as units of spoken syllables: Evidence from children. *Journal of Experimental Child Psychology, 39,* 161-181.

Treiman, R., & Weatherston, S. (1992). Effects of linguistic structure on children's ability to isolate initial consonants. *Journal of Educational Psychology, 84,* 174-181.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What we have learned in the past four decades? *Journal of Child Psychology and Psychiatry, 45,* 2-40.

Vellutino, F. R., & Scanlon, D. M. (1987). Phonological coding, phonological awareness and reading ability: Evidence from a longitudinal and experimental study. *Merrill-Palmer Quarterly, 33,* 321-363.

Vellutino, F. R., Scanlon, D. M., & Lyon, G. R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition of reading disability. *Journal of Learning Disabilities, 33,* 223-238.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R. S., et al. (1996). Cognitive profiles of difficult to remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88,* 601-638.

Verhelst N. D., & Glas, C. A. W. (1994). The One Parameter Logistic Model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models, foundations, recent developments and applications* (pp. 215-238). New York: Springer-Verlag.

Verhelst, N. D., Glas, C. A. W., & Verstralen, H. H. F. M. (1995). *OPLM: Computer program and manual.* Arnhem: Cito.

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.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., et al. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33,* 468-479.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23,* 159-177.

Ziegler, J. C., & Goswami, U. C. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131,* 3-29.

3

Nature of phonological awareness: An application of IRT[2]

Abstract In the present study, the nature of Dutch children's phonological awareness was examined throughout the elementary school grades. Phonological awareness was assessed using five different sets of items that measured rhyming, phoneme identification, phoneme blending, phoneme segmentation, and phoneme deletion. A sample of 1405 children from kindergarten through fourth grade participated. Results of modified parallel analysis and analyses within the context of item response theory (IRT) showed phonological awareness to be unidimensional across different tasks and grades. Within the unidimensional framework of phonological awareness, the cognitive task requirements for the various tasks were found to differ. In addition to some overlap between the item sets, those for rhyming, phoneme identification, and phoneme blending were easier than those for phoneme segmentation and phoneme deletion. The results lend support to the assumption that phonological awareness is a continuum of availability for phonological representations which can range from partial to full availability (i.e., access).

---

[2] Reference: Vloedgraven, J. M. T., & Verhoeven, L. (2008). The nature of phonological awareness throughout the elementary grades: An item response theory perspective. *Learning and Individual Differences*, doi:10.1016/j.lindif.2008.09.005.

3.1 Introduction

A growing body of research shows phonological awareness to be one of the most important predictors of early reading acquisition (Adams, 1990; Bradley & Bryant, 1978; McBride-Chang & Kail, 2002; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004; Wagner & Torgesen, 1987). In addition, phonological deficits have demonstrated to be the critical factor underlying reading problems (Lyon, Shaywitz, & Shaywitz, 2003; Vellutino, Fletcher, Snowling, & Scanlon, 2004). Finally, intervention studies have found training in phonological awareness to enhance children's literacy skills (Ehri et al., 2001; Gillon, 2000, 2005). A wealth of studies has thus emphasized the critical role of phonological awareness for learning to read. Despite attention to the relations between phonological awareness and reading acquisition, the actual nature of children's phonological awareness has received only scant attention and then mostly among preschool and early elementary school children (e.g., Anthony & Lonigan, 2004; Stahl & Murray, 1994). Sound research on the nature of phonological awareness across a longer period of time is thus generally lacking but nevertheless of major importance for the measurement of children's growth in phonological awareness. Thus it is by no means clear whether the different tasks used to assess phonological awareness across different age groups measure the same underlying ability. Greater insight into the nature of phonological awareness is needed to understand its development and its relations to reading skills. Furthermore, clarification of this issue will give important indications about how to best measure individual differences in phonological awareness.

3.1.1 Nature of phonological awareness

Speaking generally, phonological awareness refers to the sensitivity to the sound structure of a spoken language. A variety of tasks that differ with respect to not only the level of linguistic complexity but also the type of cognitive operation required to successfully perform the task have been used to assess phonological awareness (Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003). With respect to linguistic complexity, the size of the target unit can vary from words, syllables, onsets, and rimes to phonemes. Converging evidence has been provided that children become increasingly sensitive to smaller and smaller linguistic units (e.g., Anthony et al., 2003; Stanovich, 1992; Ziegler & Goswami, 2005). The type of cognitive operation used to assess phonological awareness can be detection, blending, segmentation, and manipulation. Children can *detect* similar and dissimilar phonological units within words before they can *manipulate* the same units within words. Similarly, children acquire the ability to *blend* phonemes into words prior to the ability to *segment* words into phonemes (Wagner, Torgesen, & Rashotte, 1994). The most difficult operations involve the isolation, deletion, and reversal of phonemes (e.g., Yopp, 1988). Children are thus able to perform increasingly complex operations during the course of their development, and the number of operations at their disposal also increases (Anthony et al., 2003).

On a theoretical level, there is considerable dispute about the extent to which phonological awareness constitutes a single unified construct. According to some scholars, different types of phonological abilities reflect distinct abilities. Yopp (1988), for example, provided evidence for a simple phonemic awareness factor and a compound phonemic awareness factor. Blending, sound isolation, and phoneme counting tasks loaded on the first factor while deletion tasks, which involved an extra operation, loaded on the second factor. The findings of Carroll, Snowling, Hulme, and Stevenson (2003) show rhyming and phoneme tasks to tap separate abilities. Høien, Lundberg, Stanovich, and Bjaalid (1995) argue that phoneme awareness, syllable awareness, and rhyming awareness require separate abilities. In a few studies, phoneme blending has been shown to be separate from phoneme segmentation (Wagner, Torgesen, Laughon, Simmons, & Rashotte, 1993; Wagner et al., 1994). The results of these studies showed a two-factor model to provide the best fit for the phonological data although the two factors were strongly related.

In contrast to the conceptualization of phonological awareness as a multidimensional construct, other scholars argue that the various tasks used to measure phonological awareness are simply manifestations of the same underlying ability. According to Stanovich (1992), for example, phonological awareness can be conceptualized as a continuum from "shallow" awareness involving larger phonological units to "deep" awareness involving smaller phonological units. When Stahl and Murray (1994) administered four different phonological awareness tasks that varied in linguistic complexity to kindergartners and first graders, moreover, they found a single underlying factor to best represent the data and this irrespective of whether the data were analyzed by task or level of linguistic complexity. The findings of more recent studies also point in a similar direction. Using confirmatory factor analyses (CFA), Anthony et al. (2002) showed the phonological awareness of 2- to 5-year-old children to best be represented by a single latent ability. In a similar vein, Anthony and Lonigan (2004) concluded that the data from four independent CFA studies showed the different phonological skills measured in children between the ages of 2 and 7 years to represent the same underlying ability.

Related to the question of whether children's developing phonological awareness reflects a single underlying ability or multiple abilities, the question of whether the development of phonological awareness occurs in a discrete stagewise manner or in a more continuous manner can be raised. Anthony et al. (2003) were the first to investigate this question explicitly, and the results of their hierarchical loglinear analyses showed the emergence of phonological awareness skills to occur in a quasi-parallel manner for 2- to 5-year-old children. That is, as children progress on one phonological awareness skill, they also progress on other phonological awareness skills. Determination of the exact development of phonological awareness is of obvious importance for the early identification of reading problems and formulation of interventions to address these problems.

3.1.2 Measurement issues

It is generally acknowledged that the tasks used to assess phonological awareness differ in difficulty (e.g., Adams, 1990; Yopp, 1988). The appropriateness of a particular task appears to depend on the child's level of development (Anthony & Lonigan, 2004; Schatschneider, Francis, Foorman, Fletcher, & Mehta, 1999). Stated differently, different tasks must be administered at different points in time in order to accurately assess a child's phonological awareness. However, a number of measurement problems present themselves in doing this. In kindergarten, a relatively easy task for phonological awareness (e.g., rhyming task) must be administered. In first grade, a more difficult task (e.g., segmentation task) must be administered to accurately assess children's phonological awareness abilities. All of the tasks for phonological awareness are assumed to measure the same underlying ability. The problem is that it is still not clear that different phonological awareness tasks administered at different points in time indeed measure the same underlying ability. If the tasks do not measure the same underlying ability, then the monitoring of phonological awareness over time is not possible. Even if the different measures are found to measure a single underlying ability, moreover, there is still the problem of outcome comparability. Imagine a child who obtains a score of 10 on the rhyming task and another child who obtains a score of 10 on the segmentation task (i.e., the child's score is the number of items answered correctly). Even if it is known that both tasks measure phonological awareness, the two task scores cannot be directly compared because the tasks do not have a shared underlying metric. The aforementioned measurement problems generally apply to the tasks used to measure children's phonological awareness today. Different tasks are used for different grades (Torgesen & Bryant, 1994; Wagner, Torgesen, & Rashotte, 1999); the scores on the kindergarten version of the Test of Phonological Awareness (TOPA) from Torgesen and Bryant, for example, cannot be compared to the scores on the elementary version. It is thus crucial to investigate whether the different tasks measure the same underlying ability across different grades and can be placed along a single ability scale to be able to monitor children's growth in phonological awareness.

In addition, there is the problem of the unreliability of individual task scores. The reliability of the individual assessment of phonological awareness largely depends on the extent to which the task matches the child's level of ability. Floor effects, for example, tell us that the child's phonological abilities are low but not how low; ceiling effects similarly tell us that the child's phonological abilities are high but not how high. In order to provide a reliable measurement of phonological awareness, thus, the level of the assessment task must roughly match the child's level of development on the ability measured and thereby allow the child to answer some items correctly and some items incorrectly.

The aforementioned problems with the measurement of phonological awareness may explain the lack of consensus on the nature of phonological awareness at least in part. Another reason might be that previous researchers have utilized different techniques to pinpoint the nature of phonological awareness. For example, certain researchers (e.g., Høien et al., 1995) have based their conclusions on exploratory factor analyses (EFA) while others have based their conclusions upon the relations between phonological awareness tasks and children's reading scores (e.g., Bryant, MacLean, Bradley, & Crossland, 1990). Still other researchers have applied CFA and structural equation modeling to disentangle the question of dimensionality (Anthony & Lonigan, 2004).

Yet another approach is to draw upon item response theory (IRT) or what is often referred to as latent trait theory (Hambleton, Swaminathan, & Rogers, 1991; Lord & Novick, 1968). In IRT, items with certain characteristics are related to the ability that is measured. This means that the difficulty of the items and the ability of persons are scaled on the same metric. The relations between the probability of a correct response on a particular item and the latent ability can be represented graphically by means of an item characteristic curve (ICC). IRT has a number of advantages, which opens up new perspectives for the measurement of children's phonological awareness. First, the person parameters are independent of the specific tasks administered. Second, the item parameters obtained from a well-fitting IRT model are independent of the sample used to generate the parameters. Another advantage is that IRT allows a broad range of interpretations at the item level. The contributions that items make to the measurement precision can be established, and this is possible for each point of the ability scale. As mentioned before, there are some problems with the measurement of phonological awareness and, in particular with the monitoring of children's development. Use of IRT offers solutions to deal with these measurement problems. By the use of IRT, all items can be placed onto a common measurement scale. Therefore, the scores for different tasks administered on different occasions can be compared, which means that it is possible to measure the development of children's phonological awareness.

3.1.3 The present study

In the present study, it is attempted to gain greater insight into the nature of phonological awareness using IRT. In a pilot study (Vloedgraven & Verhoeven, 2007), the nature of children's phonological awareness was explored via the administration of the same tasks without phoneme deletion to children in kindergarten and first grade. The results indicated a single underlying ability for the different tasks. The present study extends this pilot study with regard to the number of tasks, the number of grades involved, and the use of larger samples per grade. A cross-sectional design of five elementary grade levels is adopted (i.e., kindergarten, first, second, third, and fourth grades), in which the different tasks were administered on two or three occasions throughout the school year.

If different tasks for phonological awareness are used, the question arises of whether the different indicators relate to the same underlying ability in the same manner over time (Horn & McArdle, 1992). IRT is highly appropriate to investigate such an issue. The first question to be addressed is therefore whether the different tasks used to measure phonological awareness including rhyming, phoneme identification, phoneme blending, phoneme deletion, and phoneme segmentation appear to reflect a single underlying ability across the different grades studied or more than one ability. For example, do the tasks used to measure phonological awareness of kindergartners measure the same underlying ability as the tasks used to measure phonological awareness of beginning readers? The same question is valid for the measurement of phonological awareness of a random sample of children and a group of poor readers: do the phonological awareness tasks measure the same underlying ability in both groups of children? With respect to the use of IRT to investigate the dimensionality of phonological awareness, a first study in this direction has already been performed by Schatschneider et al. (1999) who found different tasks to reflect the same underlying ability. The present research extends this research with the inclusion of a rhyming task, which the former did not include, and with sampling also older elementary school children.

If we find a single underlying ability, additional issues will be further investigated using the IRT calibrated scale for phonological awareness. The second question to be addressed is whether the development of children's phonological awareness appears to occur in stages or continuously. Anthony et al. (2003) examined this issue among preschoolers and found evidence for a continuous development, but they did not include a segmentation task because this task was not developmentally appropriate for this age group. In the present study, this issue will therefore be explored from a broader perspective with the inclusion of school aged children and a segmentation task as well. In addition and in contrast to Anthony et al., the children's phonological awareness will be examined at the level of the item, which can give us greater insight into specific patterns of acquisition across the early elementary school grades.

The third and final question to be addressed concerns the suitability of particular tasks for the assessment of children's phonological awareness at specific points in their development (i.e., for specific levels of ability). When the difficulty of a task does not fit a child's level of ability, assessment will be inaccurate. That is, we need to know which tasks can be administered at which points in a child's development in order to obtain the most accurate measures of children's phonological awareness possible.

3.2 Method

3.2.1 Participants

The participants were 1405 children from 57 elementary schools in the Netherlands. This included 291 kindergarten children (KG), 299 grade one children (G1), 251 grade two children (G2), 269 grade three children (G3), and 295 grade four children (G4). The total sample included 778 boys and 627 girls. The schools were randomly selected using the stratified random sampling method to be sure that the schools were representative with regard to school weight (i.e., low, medium, or high). School weight is a combination of ethnic origin and parental level of education applied in the Netherlands to establish resources for the schools. In addition, all of the provinces in the Netherlands were represented in the sample of schools.

In the kindergarten sample, all children were randomly selected for inclusion. In the group of the first- through fourth-grade children, 50% of the sample was randomly selected for inclusion and the other 50% selected on the basis of a score below the 25th percentile on a standardized word-reading test. We have chosen this sampling procedure in order to better investigate the nature of phonological awareness; a question of the present study was whether the phonological awareness tasks administered to a random sample of children, measured the same underlying ability as the phonological awareness tasks administered to a sample of poor readers.

The children were studied for one school year. At the time of initial testing, the kindergartners' mean age was 5.7 years (*SD* = 0.39), the first graders' mean age was 6.8 years (*SD* = 0.74), the second graders' mean age was 7.8 years (*SD* = 0.56), the third graders' mean age was 8.8 years (*SD* = 0.50), and the fourth graders' mean age was 9.8 years (*SD* = 0.55). Between initial and final testing, the total sample decreased by about 6% due to mainly illness or the removal of some children to other schools and/or other places. Given that we wanted to use as much of the data available for each measurement occasion, the sample sizes for different analyses were allowed to vary.

3.2.2 Materials

In the Netherlands, children enter school starting on their fourth birthday and spend the next 2 years in kindergarten. In kindergarten, literacy education is usually limited to storybook reading along with incidental language games to stimulate phonological awareness and beginning literacy. After kindergarten, the children enter first grade and receive formal reading and spelling instruction. Given that Dutch is considered as a relatively transparent orthography with consistent mappings between letters and phonemes, the instruction is phonics-oriented.

For the selection of the most appropriate tasks to measure phonological awareness, the predictive value of the tasks for reading performance was considered. We also selected tasks that are known in the literature to differ in difficulty in order to accurately assess a broad range of phonological awareness abilities (Chard & Dickson, 1999; Høien et al., 1995; Vellutino & Scanlon, 1987; Yopp, 1988). In the end, the following five tasks were selected: rhyming, phoneme identification, phoneme blending, phoneme deletion, and phoneme segmentation. The five tasks represent different aspects of phonological awareness (i.e., size of phonological unit and type of cognitive operation needed) preceding reading acquisition (Adams, 1990). We constructed the tasks for the purpose of the present study, and modeled these measures after tasks described by other researchers (e.g., Høien et al., 1995; Wagner et al., 1993). The tasks contained high frequency monosyllabic words which were drawn from current Dutch word frequency lists (Schaerlaekens, Kohnstamm, & Lejaegere, 1999; Schrooten & Vermeer, 1994). The CV (i.e., consonant-vowel) structure of the target words varied.

For all of the tasks with the exception of the phoneme segmentation task, three response alternatives were first presented both auditorily and visually in the form of pictures in order to reduce memory demands. The target word was then presented auditorily. The child's task was to select the correct picture.

Rhyming. After the presentation of the three response alternatives, the target word was presented and the child was asked to select the response which rhymed with the target word. An example of an item is 'raam, wijn, kus – lijn' [meaning: window, wine, kiss – line]. The task consisted of only CVC words.

Phoneme identification. After the presentation of the three response alternatives, the target word was pronounced along with the initial phoneme from the target word. The child was then asked to select the response which began with the same sound as the target word began with. An example of an item is 'boom, mat, kaal – de m van muis' [meaning: tree, mat, bald – the m of mouse]. All of the initial phonemes were consonants. The task consisted of only CVC words.

Phoneme blending. After the presentation of the three response alternatives, the individual phonemes in the target word were pronounced in the correct order. The child was asked to select the response which corresponded to the target word. An example of an item is 'trein, groen, prijs – t-r-ei-n' [meaning: train, green, prize]. The CV structure of the target words varied; the words had a length of three, four, or five phonemes.

Phoneme deletion. After the presentation of the three response alternatives, the target word was pronounced along with a phoneme to be deleted. The child was asked to select the word that remained after deletion of the pronounced phoneme. An example of an item is 'ster, step, kerk – sterk, laat de k weg' [meaning: star, scooter, church – strong, leave the k]. Only consonants had to be deleted. The phoneme deletion process addressed all possible positions in the word; the target words consisted of three, four, or five phonemes with different CV structures.

Phoneme segmentation. In contrast to the other tasks, the phoneme segmentation task required an oral response. The target word was presented both visually and auditorily. The child was asked to pronounce each of the phonemes in the word in the proper order. An example of an item is 'vuur' [meaning: fire]. The target words contained two to five phonemes.

3.2.3 Procedure

The five tasks were presented on a computer and administered individually by trained research assistants. Each task took approximately 5 to 10 minutes to complete. Every task started with three practice items in which the experimenter provided immediate feedback to facilitate understanding of the task and each task contained 20 test items. All of the items were scored as right or wrong. All of the children with the exception of the first graders were tested on two occasions: at the beginning of the school year (November) and at the end (April/May). The first graders were tested on three occasions, which meant in the middle of the school year (February) in addition to the beginning and the end.

As stated earlier, the measurement precision for a particular task depends on the child's exact level of ability. In previous studies, for example, ceiling effects for rhyming were found for kindergarten children (Stanovich, Cunningham, & Cramer, 1984; van Bon & van Leeuwe, 2003). Schatschneider et al. (1999) showed a phoneme identification task to provide a poor estimate of phonological awareness ability for first-grade children. Furthermore, our own previous research (Vloedgraven & Verhoeven, 2007) showed children to obtain quite high scores for phoneme blending and segmentation at the end of first grade. Phonological awareness tasks are thus not developmentally appropriate for all ability levels as, at certain moments in development, the tasks simply do not distinguish high versus low performers (Chard & Dickson, 1999). Therefore, it is not sensible to administer all of the tasks to all groups of children. We have selected tasks per grade that according to the literature and our previous study suit the child's level of ability (for an overview, see Figure 3.1). It should be noted that the same tasks were administered throughout a single grade (i.e., on the different measurement occasions within a grade).

	Rhyming/Identification					Blending/Segmentation					Deletion				
Grade	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
KG	X	X	X	X	X	X	X	X	X	X					
G1						X	X	X	X	X	X	X	X	X	X
G2											X	X	X	X	X
G3											X	X	X	X	X
G4											X	X	X	X	X

Figure 3.1 *Administration design for the different tasks for each grade*

In addition, not all of the items for a single task were administered to the children on one occasion, in order to avoid an excessive burden on the child. A large item bank was created to encompass various difficulty levels along the continuum of phonological awareness consisting of five tasks of 50 items each. To be able to place all the phonological awareness items onto a common measurement scale, a specific design needs to be used. In this study, an incomplete anchor-test design was employed (Petersen, Kolen, & Hoover, 1989). All of the items of a task were divided in modules of 10 items (see Figure 3.1) and we have tried to make the different modules of a task comparable concerning the difficulty. According to Schreuder and van Bon (1989), the CV structure of words is an important determinant of the difficulty of phonological awareness tasks. In addition to the word structure, the manner of articulation of consonants in a word appears to influence children's achievements on phonological awareness tasks. In Dutch, five manners of articulation are identified (Rietveld & van Heuven, 1997): plosives (b, d, k, p, t), fricatives (f, g, s, v, z), liquids (l, r), nasals (m, n), and glides (h, j, w). In an attempt to construct modules of equal difficulty, we have taken care of a proportional distribution of the different CV structures and articulation manners across the different modules of a task as much as possible.

Within a task, there were fixed combinations of two modules, the so-called booklets. Each child was given one booklet of a particular task (consisting of 20 items) on each occasion and a booklet was never administered twice to the same child. Linking procedures were used to cause overlap between the booklets. The booklets thus differ and are administered to different groups of children but are linked by 10 common items (i.e., anchor items). For each of the tasks, a comparable anchor-test design was followed as illustrated for phoneme blending and phoneme segmentation in Figure 3.2.

	Blending/Segmentation				
Grade	1	2	3	4	5
KG-begin	X	X			
KG-end		X	X		
G1-begin			X	X	
G1-middle				X	X
G1-end	X				X

Figure 3.2 *Design specified for phoneme blending and phoneme segmentation*

Summarizing, different sets of items are administered in the different grades, but because of the overlap in the design, it is possible to present all of the items at one scale of measurement. IRT is exceedingly suitable in doing this (Skaggs & Lissitz, 1986). Furthermore, IRT provides an excellent basis to predict how children, or a group of children may perform on particular items, even when these items have not been administered.

3.2.4 Statistical analyses

The nature of phonological awareness was analyzed in two different manners. First, modified parallel analyses were undertaken. As suggested by Drasgow and Lislak (1983), the eigenvalues for real data sets were compared to the eigenvalues for simulated data sets. That is, a principal-axis factor analysis was performed on the tetrachoric correlation matrix for the different tasks used to measure phonological awareness on each measurement occasion. Simulated data sets with the same number of items and children were constructed, under the assumption of a unidimensional model. That is, the discrimination, difficulty, and ability parameters from a well-fitting IRT model were used to create a simulated data set. The tetrachoric correlations were computed for the simulated data sets and the eigenvalues thus obtained. Scree plots were next created for the real and simulated data sets. Highly similar scree plots indicate unidimensionality. When the second eigenvalue for the real data is considerably larger than the second eigenvalue for the simulated data, however, the underlying structure of phonological awareness is better described using more than one dimension.

The second manner of investigation used to explore the nature of phonological awareness involved several analyses within the context of IRT. Various IRT models differ with regard to the number of item properties, represented by parameters in the model, that they take into account. In the one-parameter model, only item difficulty is estimated; in the two-parameter model, item difficulty and item discrimination are estimated; and in the three-parameter model, item difficulty, discrimination, and guessing are estimated. We adopted the one-parameter logistic model (OPLM), which is an integration of the one- and two-parameter model and links the theoretical advantage of the one- parameter model (i.e., the use of conditional maximum likelihood to estimate item parameters and the sampling independence implied by it) with the flexibility of the two-parameter model (Verhelst & Glas, 1994). The OPLM adjusts for item difficulty and, in order to deal with the dissimilar discriminative power of items, discrimination indices are handled as known constants (i.e., discrete values). The discrimination parameters were selected such that their geometric mean equaled 3. The difficulty parameter indicates the difficulty of an item and the discrimination parameter indicates the extent to which an item distinguishes between various ability levels. The data were analyzed using a computer program for the analysis of dichotomously scored items within the framework of the OPLM (Verhelst, Glas, & Verstralen, 1995).

Various steps within the framework of IRT were taken to investigate the dimensionality of phonological awareness. First, the fit of the OPLM for each of the five tasks was examined. As a next step, the fit of a unidimensional IRT model for all of the phonological awareness items was investigated. To demonstrate item parameter and ability invariance, the item parameters were estimated using the conditional maximum likelihood (CML) method and the ability parameters were estimated using the weighted maximum likelihood (WML) method. The population parameters for the various grades were assessed using the marginal maximum likelihood (MML) method. This method was applied by fixing the item parameters and estimating a mean and a standard deviation in a normal distribution. The MML method is preferred over the WML method for the estimation of means and standard deviations in the population because the individual WML estimators tend to overestimate standard deviations of the ability due to measurement error. For the estimation of the population parameters, only those children who were randomly selected for inclusion in the sample were taken into account.

3.3 Results

3.3.1 Nature of phonological awareness

Given that an incomplete test design was used, it was not possible to compute the correlations between all of the phonological awareness items. Items that were administered, differed for each measurement occasion. The tetrachoric correlations and eigenvalues were therefore computed for the real and simulated data sets for each measurement occasion separately in kindergarten and first grade. We randomly selected ten items from each task that was administered on an occasion to facilitate the interpretation of the factor analysis. For each measurement occasion, the scree plots for the 14 largest eigenvalues for the real and simulated data sets were drawn and can be viewed in Figure 3.3. Inspection of the plots clearly shows a single latent factor, which accounted for 80% to 90% of the variance. Apparently, the breaking point of the scree plots occurred after the first factor. The most convincing evidence for one factor was provided by the highly similar scree plots for the real and simulated datasets. The simulated eigenvalues are computed from a correlation matrix, that is known to be unidimensional, and must therefore demonstrate one factor, every deviant solution is completely due to measurement error. The comparable pattern of the real and simulated eigenvalues indicates that performance across the phonological awareness tasks can largely be accounted for by one factor.

Given that different tasks were administered in the different grades, clear conclusions could not be drawn on the basis of the factor analyses. Within the framework of IRT, the dimensionality can be investigated for all grades and for all the phonological awareness items together. Thus, the nature of phonological awareness, as measured by five different tasks, was next investigated by the use of IRT for kindergarten through fourth grade.

The model fit of each of the five tasks was first established (by computing item-oriented statistics and an overall statistic). In order to assess whether the individual items for each task fit the OPLM, the p values for all of the items were investigated. A uniform distribution of the frequencies of the p values at the interval [0, 1] favours model fit (Verhelst et al., 1995). For each task, the p values appeared to be rectangularly distributed at the interval, indicating model fit for each of the five tasks.

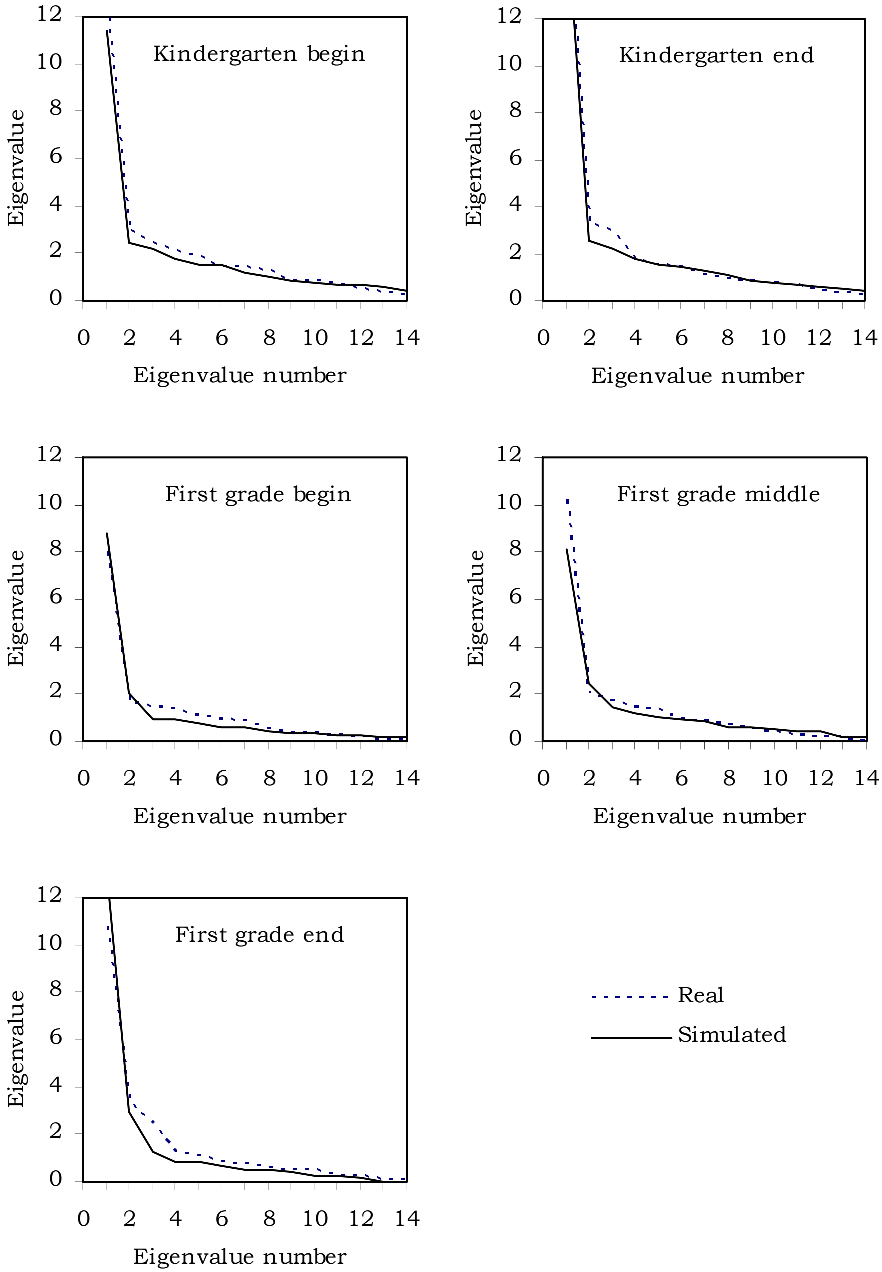

Figure 3.3 *Scree plots for the 14 largest eigenvalues for the real and simulated data sets from kindergarten and first grade*

Next, additional information about the model fit can be provided by the overall R1c test (R1c, with degrees of freedom and *p* value). The *p* values turned out to be rather small, however, it is generally acknowledged that an incomplete design complicates the interpretation of the results of the R1c statistical test. Therefore, the following rule of thumb is commonly used: a model can be viewed to fit the data acceptably when the ratio between the overall R1c to the degrees of freedom is found to be smaller than 1.5:1 (van Weerden, Thijssen, & Verhelst, 2007). For all of the tasks, this ratio appeared to be of the desired order of magnitude (Rhyming: R1c = 182.56, df = 180, p = .44; Phoneme identification: R1c = 239.30, df = 205, p < .05; Phoneme blending: R1c = 235.07, df = 199, p < .05; Phoneme segmentation: R1c = 225.47, df = 153, p < .01; Phoneme deletion: R1c = 512.585, df = 409, p < .01). Both the item-oriented and overall statistical tests showed the OPLM to fit the data for each phonological awareness task, indicating that each task anyhow can best be determined by a single latent factor.

Given that the results of the factor analyses indicate a single latent ability, the next step in the analyses was to investigate the fit of the OPLM for all of the items considered together. The question now is thus whether it is possible to subsume the five tasks administered to groups of different ability levels within the same scale. Also for the total data set, both the item-oriented and overall statistical tests showed satisfactory model fit. The *p* values for the individual items were reasonably distributed at the interval [0, 1] and the R1c was 2128.01 (df = 1538, p < .01).

When an IRT model fits the data, the invariance of item parameter and ability estimates can be demonstrated. The first property claims that item parameters are independent of the specific sample from which they are inferred. The second property states that ability estimates are independent of the particular set of items drawn from the item bank (Hambleton et al., 1991). To further establish the model fit, we investigated these two properties. First, we addressed the expected model property of invariance of item parameters. Evidence for model fit is demonstrated when a linear relationship is found to exist between the difficulty parameters estimated from two samples, even if the samples differ in a way (Lord & Novick, 1968). One important question of the present study was whether the different tasks used to assess phonological awareness measured the same latent ability across grades. To investigate item invariance, the difficulty parameter estimates for the kindergartners and first graders were computed. The relation appeared to be linear and the correlation coefficient, generated to establish the strength of the relationship, turned out to be .97, which shows the invariance of the item parameters. A related question of the present study was whether the phonological awareness tasks measured the same latent ability in a random sample of children as in a sample of poor readers. The difficulty parameters were estimated for both samples and then plotted. Again, a strong linear relation was found to hold between the difficulty parameters for the two groups of children (*r* = .96).

To examine the invariance of the ability estimates, the phonological awareness items were split into two random subsets of items: one containing even items and one containing odd items. The ability parameters for each child were then estimated using the two subsets of items. A strong linear relationship appeared to hold between ability parameters estimated on the basis of the even and odd items (*r* = .84), which provides evidence for the invariance of ability. Taken together, these findings show the phonological awareness items administered to children in different elementary school grades to best be represented by a single underlying ability.

3.3.2 Acquisition of phonological awareness

The preceding findings show phonological awareness to be a unidimensional ability. However, the exact nature of the acquisition of the various phonological awareness skills over time remains unclear. To gain greater insight into the relative complexity of the different phonological awareness skills, the difficulty of the five tasks was examined in terms of their mean difficulty parameter estimates. In Table 3.1, the mean difficulty parameters are presented from least to most difficult. As can be seen, rhyming appeared to be the easiest skill and deletion appeared to be the most difficult skill.

Table 3.1
Average difficulty parameters for the five tasks measuring phonological awareness

	Difficulty parameter	
Task	*M*	*SD*
Rhyming	-0.55	0.26
Phoneme blending	-0.24	0.17
Phoneme identification	-0.17	0.24
Phoneme segmentation	0.53	0.29
Phoneme deletion	0.77	0.22

The question of whether the different phonological skills are acquired in discrete stages or on a more continuous basis still remains to be answered. A graphical representation of the relative difficulties of the items in the different tasks might shed more light on this issue. In Figure 3.4, the relative positions of the difficulty parameters for the items in each task on the unidimensional phonological awareness scale are shown. As can be noticed, the difficulties of the rhyming (RH), phoneme blending (PB), phoneme identification (PI), phoneme segmentation (PS), and phoneme deletion (PD) items clearly differ but also show considerable overlap in terms of difficulty.

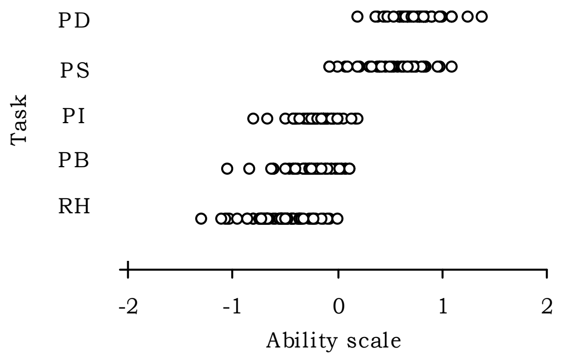

Figure 3.4 *Difficulty parameters for all of the items from the five tasks on the phonological awareness scale*

Before interpreting these results in terms of development, it should be mentioned that some caution needs to be observed in interpreting these findings, because the data are cross-sectional. Despite this, results are expected to give important indications about the development of phonological awareness over time. From Figure 3.4, it can be deduced that a strict stage theory, which requires each specific skill to be completely mastered before progression to the next specific skill, does not apply. A strict stage theory seems thus untenable for the description of the development of phonological awareness. However, a description in terms of continuous development seems largely untenable as well. The results indicate a kind of separation between rhyming, phoneme identification, and phoneme blending — on the one hand — and phoneme segmentation and phoneme deletion — on the other hand. Within these two groups, however, the development appears to be fairly continuous with the different skills thus being acquired more or less simultaneously.

3.3.3 Measurement precision for the different grades

An important advantage of IRT is that the utility of various tasks for different ability levels can be assessed via the computation of test information functions. The test information functions show the contributions of a task to the precision of measurement across the ability continuum. On the basis of the test information function, it is subsequently possible to compute the standard error for an individual ability estimate. The function of the standard error of measurement in IRT is the same as in CTT. But whereas the standard error in CTT is assumed to be constant for all test scores, the standard error in IRT is allowed to vary with ability. Information functions thus vary across the ability continuum, that is, a task might be more informative for some ability levels than for others. Given that the ability distributions move with the grades, we decided to show the (same) information functions for the five tasks for each grade separately in order to indicate the value of the various tasks across the grades (see Figure 3.5). The ability distributions of the different grades are presented below the *x*-axis. The higher the maximum of the information function, the greater the amount of information provided by that particular task at that particular point within the ability distribution.

It is important to note that, given that the items and ability are situated on the same scale, IRT can provide consideration of the measurement precision of tasks in a particular grade without the administration of these tasks in that grade (e.g., Lord, 1980). It is clear that the utility of the five tasks differs somewhat across the different grades. In kindergarten, the rhyming, phoneme identification, and phoneme blending tasks appear to provide the most precise estimates of ability and seem to be particularly informative for the lower scoring kindergartners. In contrast, the phoneme segmentation task provides more information about the abilities of the higher scoring kindergartners. The phoneme deletion task appears to be less suitable to accurately assess phonological awareness of kindergartners. In first and second grade, the phoneme segmentation task appears to be most accurate for the assessment of phonological awareness. The phoneme deletion task provides a little information on the phonological awareness abilities of first and second graders. Examination of the various information functions for the third and fourth grades shows that none of the tasks accurately assesses the phonological awareness abilities of all third and fourth graders. In general, the measurement precision for the various phonological awareness measures decreases as the children's awareness increases. This means that at least some of the measures may still provide accurate estimates of the phonological awareness abilities of at least lower scoring third and fourth graders. The information functions in relation to the ability distributions for the third and fourth graders indeed showed the phoneme segmentation and phoneme deletion tasks to still provide accurate estimates for the lower scoring children in third and fourth grade.

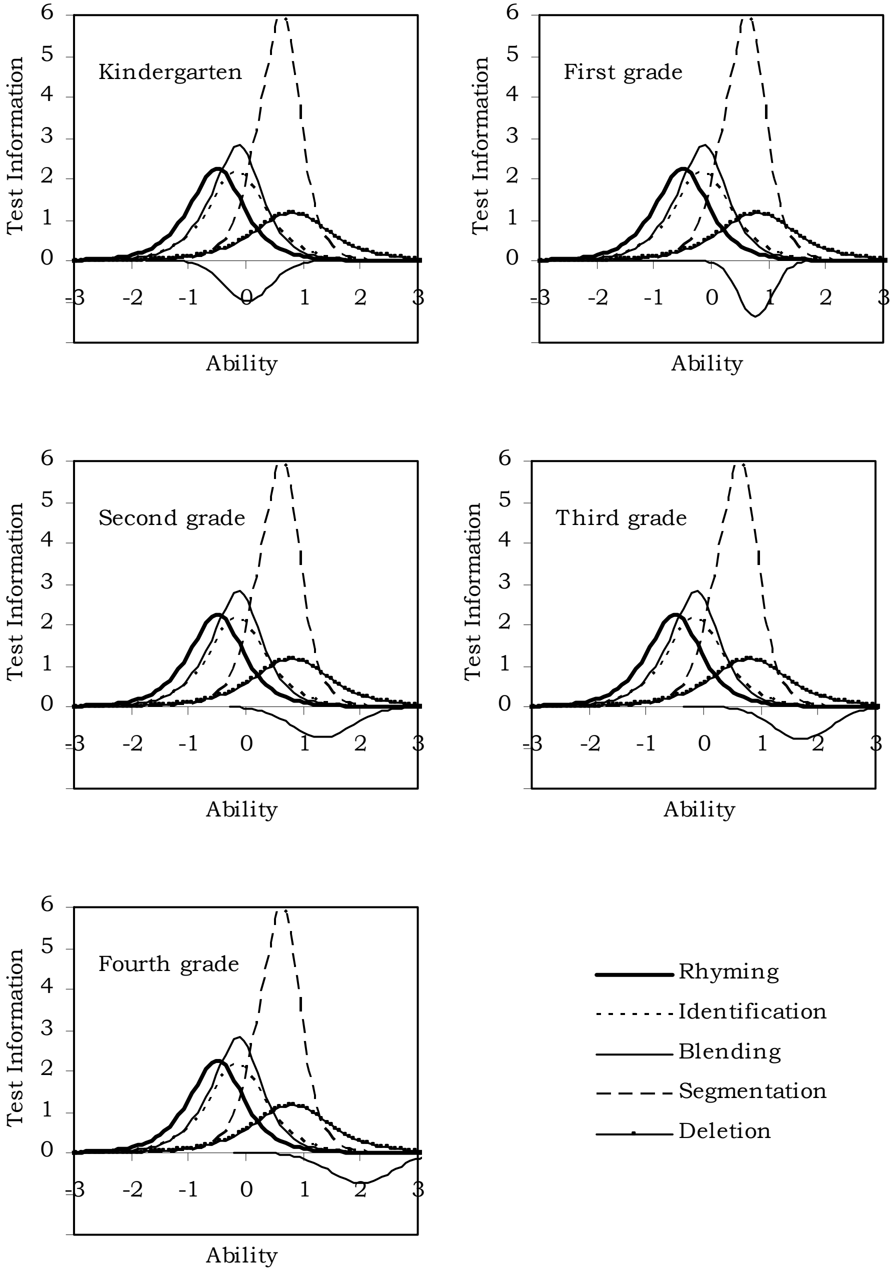

Figure 3.5 *Test information functions for the five tasks measuring phonological awareness for each grade separately*

3.4 Conclusions and discussion

This study extends our knowledge of the nature of phonological awareness throughout the school years. The results of the factor analyses indicated that the different tasks used to measure phonological awareness can be considered as manifestations of the same underlying ability. The various phonological awareness tasks turned out to be highly dominated by a single underlying factor, which accounted for a large percentage of the variance. In addition, the scree plots for the real and simulated data sets were found to be quite similar. The exact nature of the children's phonological awareness across a number of grade levels was further investigated from the perspective of IRT. Both the item-oriented and overall statistical tests showed the fit of the OPLM for all of the items considered together to be quite good. That is, the unidimensionality of phonological awareness has been demonstrated again. Other evidence for this conclusion was provided by the finding of item parameter invariance. In other words, a strong linear relationship was found between the difficulty parameter estimates for the kindergartners and first graders and between those estimates for a random sample of children and a group of poor readers. In a similar vein, ability invariance was also demonstrated.

The assumption of a single underlying dimension to characterize the nature of children's phonological awareness is in agreement with the outcomes of a number of other studies (Anthony & Lonigan, 2004; Anthony et al., 2002; Schatschneider et al., 1999; Stahl & Murray, 1994; Stanovich, 1992). However, this result contradicts claims in the literature that rhyming awareness should be distinguished from phoneme awareness (Carroll et al., 2003; Muter, Hulme, Snowling, & Taylor, 1997). Nevertheless, Anthony and Lonigan (2004) have suggested that rhyming ceiling effects may have artificially separated rhyming from phoneme awareness. They argue that this issue should be further investigated from the perspective of IRT precisely because IRT is less influenced by measurement problems. When Schatschneider et al. employed IRT to investigate the dimensionality of phonological awareness, they indeed found children's phonological awareness to be well-represented by a single underlying ability although rhyming tasks were not included in this study. A variety of phonological awareness tasks including rhyming were therefore administered in the present study and strong support is thus provided for a unified conceptualization of phonological awareness. Given that a much broader range of ability levels was investigated in the present study than in previous studies, moreover, the present findings show the unidimensionality of phonological awareness to also hold for kindergarten through fourth-grade children.

Comparison of the difficulty parameters for the five tasks showed rhyming to be easiest. This was followed by phoneme blending, phoneme identification, phoneme segmentation, and — finally — phoneme deletion. The relative ease of the rhyming task and relative difficulty of the phoneme deletion task is in keeping with the findings of previous studies (Stanovich et al., 1984; Yopp, 1988). The finding that phoneme blending is easier than phoneme segmentation is also consistent with the findings of other studies (e.g., van Bon & van Leeuwe, 2003; Wagner et al., 1994). With respect to the difficulty of the individual items, two separate groups of phonological awareness items appear to emerge. That is, when phonological awareness starts to develop, children seem to acquire rhyming, phoneme identification, and phoneme blending skills; as phonological awareness continues to develop, children seem to acquire phoneme segmentation and phoneme deletion skills. This result corresponds only in part to the conclusions of Anthony et al. (2003) who explicitly investigated this issue and found evidence for a quasi-parallel development of phonological awareness skills in overlapping stages. Our results suggest a distinction between two groups of items in addition to continuous, quasi-parallel development. These partly inconsistent findings can possibly be explained by the use of different methodologies or by the fact that both the tasks and the samples used in the two studies clearly differed. In contrast to Anthony et al., we included a segmentation task and also tested school-age children instead of just kindergartners.

The possible existence of two separate groups of phonological awareness items can be explained in terms of the fact that phonemes are acoustically evanescent (Liberman, Liberman, Mattingly, & Shankweiler, 1980). That is, the ability to segment a word into individual phonemes is inherently difficult because the individual phonemes are not audibly distinguishable. Phonemes overlap and influence each other, which is often referred to as coarticulation. Given this phenomenon, the cognitive operation needed to perform phoneme segmentation and phoneme deletion tasks is far more complex than the cognitive operation needed to perform rhyming, phoneme identification, and phoneme blending tasks (Adams, 1990). The relative difficulty of segmentation and deletion is also in correspondence with the spelling literature. Beginning and poor readers experience greater difficulties with the spelling of rhotic vowels and nasals in clusters as these children have the tendency to treat two consonants in a cluster as a single sound unit. That is, it is hard to separate the individual phonemes acoustically (Snowling, 1994; Treiman, 1993).

The present findings are also consistent with Ziegler and Goswami's grain size theory of reading acquisition (Ziegler & Goswami, 2005). According to this theory, the development of reading depends upon the precision of the underlying phonological representations of words. As regards phonological awareness, it can similarly be assumed that children acquire a partial availability to such representations in kindergarten resulting in progress on the relatively easy phonological awareness tasks. In first grade, when formal reading education starts and explicit instruction in phonics is offered, the children will be able to gain full availability to phonemes and improvement will then be shown on the more complex phonological awareness tasks. Within the unidimensional framework of phonological awareness, the development of phonological awareness seems to proceed from partial towards full access to phonemes with the phonological representations becoming more and more specific. Further research and preferably longitudinal instead of cross-sectional research is still needed, however, to gain greater insight into the acquisition of various phonological awareness skills across the elementary grades.

The third and final question addressed in this research concerned the suitability of the different tasks to assess the phonological awareness abilities of kindergartners through fourth graders. Strong evidence was provided for the assumption that the appropriateness of a specific task depends upon the child's level of ability which is in keeping with the results of previous studies on this issue (Anthony & Lonigan, 2004; Schatschneider et al., 1999). In general, the accuracy of the various tasks measuring phonological awareness diminished as the children's abilities increased. Phonological awareness can thus best be measured in kindergarten and first grade. An interesting finding of the present study is that the phonological awareness abilities of lower scoring third and fourth graders can also be accurately estimated. This finding is in accordance with the results of a study of the development of phonological awareness during the elementary school years showing phonological deficits to not disappear completely over time (de Jong & van der Leij, 2003).

The results of the present study have some significant implications for the early identification of children with reading problems and dyslexia. First of all, it can be concluded that phonological awareness can be conceived as a unidimensional ability across different tasks and grades. The different tasks could be placed along an IRT calibrated scale for phonological awareness which means that it is possible to compare scores from different tasks, and thus to monitor children's growth in phonological awareness throughout the elementary school years. This is a significant result as it is well-known that growth in phonological awareness can account for additional variance in reading ability (Byrne, Fielding-Barnsley, & Ashley, 2000; Hindson et al., 2005). In addition, the present study has shown phonological awareness measures to still provide accurate estimates of the abilities of lower scoring third and fourth graders. The assessment of phonological awareness in the higher grades may be valuable for intervention purposes (cf. Hogan, Catts, & Little, 2005).

Furthermore, the present study makes it clear that the measures selected to assess the phonological awareness of a child must be suited to the child's level of development. Individual differences in phonological awareness can thus best be measured by an IRT based measure consisting of a variety of tasks with the results of the present study providing insight into which measures may be most useful at which levels of ability. These findings can be applied in practice by using methods for optimal test assembly (e.g., Adema, Boekkooi-Timminga, & van der Linden, 1991; van der Linden, 2000). A common technique to construct optimal tests is to maximize the test information at certain ability levels. The combination of items that is optimal will be selected for inclusion in the task. With respect to the measurement of phonological awareness, this means for example that in kindergarten mainly rhyming, phoneme identification, and phoneme blending items will be administered, while in first grade phoneme segmentation items will be overrepresented. An additional advantage of an IRT based measure is the possibility for adaptive testing, in which each child is likely to get different items adjusted to the child's ability level (Hambleton, Zaal, & Pieters, 1991). As a result, the child's phonological awareness ability can be more accurately estimated and the prediction of subsequent reading might be expected to improve as well.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Adema, J. J., Boekkooi-Timminga, E., & van der Linden, W. J. (1991). Achievement test construction using 0-1 linear programming. *European Journal of Operational Research, 55,* 103-111.

Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology, 96,* 43-55.

Anthony, J. L., Lonigan, C. J., Burgess, S. R., Driscoll, K., Phillips, B. M., & Cantor, B. G. (2002). Structure of preschool phonological sensitivity: Overlapping sensitivity to rhyme, words, syllables, and phonemes. *Journal of Experimental Child Psychology, 82,* 65-92.

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly, 38,* 470-487.

Bradley, L., & Bryant, P. E. (1978). Difficulties in auditory organization as a possible cause of reading backwardness. *Nature, 271,* 746-747.

Bryant, P. E., MacLean, M., Bradley, L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology, 26,* 429-438.

Byrne, B., Fielding-Barnsley, R., & Ashley, L. (2000). Effects of preschool phoneme identity training after six years: Outcome level distinguished from rate response. *Journal of Educational Psychology, 92,* 659-667.

Carroll, J. M., Snowling, M. J., Hulme, C., & Stevenson, J. (2003). The development of phonological awareness in preschool children. *Developmental Psychology, 39,* 913-923.

Chard, D. J., & Dickson, S. V. (1999). Phonological awareness: Instructional and assessment guidelines. *Intervention in School and Clinic, 34,* 261-270.

de Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology, 95,* 22-40.

Drasgow, F., & Lissak, R. I. (1983). Modified parallel analysis: A procedure for examining the latent dimensionality of dichotomously scored item responses. *Journal of Applied Psychology, 68,* 363-373.

Ehri, L. C., Nunes, S., Willows, D., Schuster, B., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36,* 250-287.

Gillon, G. (2000). The efficacy of phonological awareness intervention for children with spoken language impairment. *Language, Speech and Hearing Services in Schools, 31,* 126-141.

Gillon, G. (2005). Facilitating phoneme awareness development in 3- and 4-year-old children with speech impairment. *Language, Speech, and Hearing Services in Schools, 36,* 308-324.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Hambleton, R. K., Zaal, N. J., & Pieters, J. P. M. (1991). Computerized adaptive testing: Theory, applications, and standards. In R. K. Hambleton & N. J. Zaal (Eds.), *Advances in educational and psychological testing: Theory and applications* (pp. 341-366). Boston: Kluwer Academic.

Hindson, B. A., Byrne, B., Fielding-Barnsley, R., Newman, C., Hine, D. W., & Shankweiler, D. (2005). Assessment and early instruction of preschool children at risk for reading disability. *Journal of Educational Psychology, 97,* 687-704.

Hogan, T. P., Catts, H. W., & Little, T. D. (2005). The relationship between phonological awareness and reading: Implications for the assessment of phonological awareness. *Language, Speech, and Hearing services in Schools, 36,* 285-293.

Høien, T., Lundberg, I., Stanovich, K. E., & Bjaalid, I. (1995). Components of phonological awareness. *Reading and Writing: An Interdisciplinary Journal, 7,* 171-188.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Journal of Experimental Aging Research, 18,* 117-144.

Liberman, I. Y., Liberman, A. M., Mattingly, I., & Shankweiler, D. (1980). Orthography and the beginning reader. In J. F. Kavanagh & R. L. Venezky (Eds.), *Orthography, reading, and dyslexia* (pp. 137-153). Baltimore: University Park Press.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Mahwah, NJ: Erlbaum.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of Dyslexia, 53,* 1-14.

McBride-Chang, C., & Kail, R. V. (2002). Cross-cultural similarities in the predictors of reading acquisition. *Child Development, 73,* 1392-1407.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24,* 45-64.

Muter, V., Hulme, C., Snowling, M. J., & Taylor, S. (1997). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of Experimental Child Psychology, 65,* 370-396.

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

Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 221-262). New York: American Council on Education.

Rietveld, A. C. M., & van Heuven, V. J. (1997). *Algemene fonetiek* [General phonetics]. Bussum: Coutinho.

Schaerlaekens, A., Kohnstamm, D., & Lejaegere, M. (1999). *Streeflijst woordenschat voor zesjarigen* [Strived vocabulary list for six-year olds]. Lisse: Swets & Zeitlinger.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96,* 265-282.

Schatschneider, C., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology, 91,* 467-478.

Schreuder, R., & van Bon, W. H. J. (1989). Phonemic analysis: Effects of word properties. *Journal of Research in Reading, 12,* 59-78.

Schrooten, W., & Vermeer, A. (1994). Woorden in het basisonderwijs. 15.000 woorden aangeboden aan leerlingen [Words in primary education. 15.000 words offered to pupils]. *Studies in meertaligheid, 6,* 1-44.

Skaggs, G., & Lissitz, R. W. (1986). IRT test equating: Relevant issues and a review of recent research. *Review of Educational Research, 56,* 495-529.

Snowling, M. J. (1994). Towards a model of spelling acquisition: The development of some component skills. In G. D. A. Brown & N. C. Ellis (Eds.), *Handbook of spelling: Theory, process and intervention* (pp. 111-128). Chicester: Wiley.

Stahl, S. A., & Murray, B. A. (1994). Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology, 86,* 221-234.

Stanovich, K. E. (1992). Speculations on the causes and consequences of individual differences in early reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 307-342). Hillsdale, NJ: Erlbaum.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38,* 175-190.

Torgesen, J. K., & Bryant, B. R. (1994). *Test of Phonological Awareness, TOPA.* Austin, Texas: PRO-ED.

Treiman, R. (1993). *Beginning to spell: A study of first-grade children.* New York: Oxford University Press.

van Bon, W. H. J., & van Leeuwe, J. F. J. (2003). Assessing phonemic awareness in kindergarten: The case for the phonemic recognition task. *Applied Psycholinguistics, 24,* 195-219.

van der Linden, W. J. (2000). Optimal assembly of tests with item sets. *Applied Psychological Measurement, 24,* 225-240.

van Weerden, J. J., Thijssen, J. M. W., & Verhelst, N. D. (2007). *Verantwoording van de theorietoetsen techniek* [Account of the theory tests technique]. Arnhem: Cito.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What we have learned in the past four decades? *Journal of Child Psychology and Psychiatry, 45,* 2-40.

Vellutino, F. R., & Scanlon, D. M. (1987). Phonological coding, phonological awareness and reading ability: Evidence from a longitudinal and experimental study. *Merrill–Palmer Quarterly, 33,* 321-363.

Verhelst, N. D., & Glas, C. A. W. (1994). The One Parameter Logistic Model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models, foundations, recent developments and applications* (pp. 215-238). New-York: Springer-Verlag.

Verhelst, N. D., Glas, C. A. W., & Verstralen, H. H. F. M. (1995). *OPLM: Computer program and manual.* Arnhem: Cito.

Vloedgraven, J. M. T., & Verhoeven, L. (2007). Screening of phonological awareness in the early elementary grades: An IRT approach. *Annals of Dyslexia, 57,* 33-50.

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., Laughon, P., Simmons, K., & Rashotte, C. A. (1993). Development of young readers' phonological processing abilities. *Journal of Educational Psychology, 85,* 83-103.

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.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing, CTOPP.* Austin, Texas: PRO-ED.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23,* 159-177.

Ziegler, J. C., & Goswami, U. C. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131,* 3-29.

4

Development of phonological awareness: The case of Dutch[3]

Abstract The aim of the present study was to explore the development of phonological awareness from kindergarten through fourth grade. A univariate regression analysis for latent variables showed a decreased rate of growth with age. For adequate model specification, the effects of receptive letter knowledge, word reading ability, gender, social-economic background, and linguistic diversity were also investigated. Significant main effects of letter knowledge, word reading ability, gender, and socioeconomic status (SES) on phonological awareness were found. However, none of these variables appeared to influence the rate of development for phonological awareness over time, which shows the development of phonological awareness to be largely similar across various groups of children. The results further showed the development of phonological awareness to best be characterized by a quadratic growth curve.

[3] Reference: Vloedgraven, J. M. T., Verhoeven, L., Keuning, J., & Eggen, T. J. H. M. (submitted). Development of phonological awareness throughout the elementary grades: Evidence from a latent variable study.

4.1 Introduction

The strong relationship between phonological awareness and reading ability has been extensively demonstrated by many researchers (e.g., McBride-Chang & Kail, 2002; Stanovich, Cunningham, & Cramer, 1984; Wagner et al., 1997). There is also ample evidence that inadequate phonological awareness is the most important manifestation of the underlying phonological deficit for children with reading problems and dyslexia. The development of phonological awareness has itself received less attention although it is known that growth in phonological awareness explains variance in reading ability over and beyond that accounted for by the children's actual level of phonological awareness (Byrne, Fielding-Barnsley, & Ashley, 2000; Hindson et al., 2005; Spector, 1992). Taken together, these findings emphasize the importance of monitoring the development of phonological awareness throughout the elementary school grades.

4.1.1 Development of phonological awareness

Most of the small number of studies that have investigated the development of phonological awareness have been limited to the examination of pre-readers. For example, Lonigan, Burgess, Anthony, and Barker (1998) examined the development of phonological awareness in 2- to 5-year-old children and found — as expected — the children's phonological abilities to increase with age. However, the course of the children's phonological development was not linear: the rate of development was slower at a younger age and faster at an older age.

The majority of the studies of children's phonological awareness has been concentrated on the identification of differences between particular subgroups of children such as poor readers or dyslexics versus normal readers on a single measurement occasion (e.g., Bruck, 1992). Nevertheless, a few studies that have monitored the development of children's phonological awareness from preschool through the early elementary school grades, when formal literacy instruction starts, do exist. In a study by McBride-Chang, Wagner, and Chang (1997), for example, the association between initial status and growth in phonological awareness was explored from kindergarten through first grade; the phonological awareness of children who had a higher level of phonological awareness to start with was found to develop more quickly than the phonological awareness of the children with a lower level of phonological awareness to start with. When Wagner, Torgesen, and Rashotte (1994) examined the rates of normal development for five different phonological processing abilities from kindergarten through second grade, the results of a repeated measures analysis of variance with a linear and a quadratic contrast for the grade factor showed both a linear and a quadratic effect of grade. In addition, the phonological abilities of the children were still developing in second grade.

No information was thus provided in the preceding studies on the phonological awareness of children in upper elementary school. When de Jong and van der Leij (2003) examined the development of phonological awareness in dyslexic, weak, and normal elementary school readers in a longitudinal Dutch study, they did not describe the general pattern of development for phonological awareness because they were primarily interested in the different manifestations of a phonological deficit. Insight into the development of phonological awareness across a longer period of time is thus lacking. The purpose of the present study was therefore to attempt to fill this gap at least in part and document the development of children's phonological awareness throughout the elementary school grades.

4.1.2 Phonological awareness in relation to literacy

In the study of phonological awareness and the modeling of its development, it is critical that those variables that can possibly influence the development of phonological awareness also be considered. After all, various subgroups of children can certainly develop differently, which means that the identification of a single growth curve may not adequately describe the development of phonological awareness. It is important, for instance, to investigate the implications of children's ability on a related skill for the development of phonological awareness. Converging evidence exists, for example, that — in addition to phonological awareness — letter knowledge constitutes an important predictor of reading (Elbro, Børstrom, & Petersen, 1998; Lonigan, Burgess, & Anthony, 2000). Moreover, letter knowledge and phonological awareness are generally assumed to be strongly related to each other (Bowey, 1994; Burgess & Lonigan, 1998; Johnston, Anderson, & Holligan, 1996).

Little is known about the development of phonological awareness in children with varying levels of letter knowledge or reading ability. According to the reading research literature, different developmental patterns are possible (McBride-Chang et al., 1997). It is possible, for example, that good readers draw upon their already developed reading abilities and therefore progress more quickly than poor readers. It is conversely possible in keeping with a compensatory model that children who show initially lower reading levels eventually catch up to their normal reading peers; that is, the poor readers actually develop at a faster rate than the good readers.

4.1.3 Role of background variables

Numerous studies have been designed to investigate the growth of early reading skills. A general weakness in most of these studies, however, is that relevant background variables have not been included in the analyses (Roth, Speece, Cooper, & DeLaPaz, 1996). The result is inaccurate interpretation of the detected findings. It is thus important that the effects of relevant background variables on the development of phonological awareness be examined in addition to the effects of such related abilities as letter knowledge and word reading skill.

To start with, the effects of gender should be considered as there is evidence that elementary school girls read better than elementary school boys (e.g., Holbrook, 1988; Sammons & Smees, 1998). Given that phonological awareness is one of the most important predictors of reading skill, it can thus be hypothesized that girls will also outperform boys on phonological awareness tasks. However, existing investigations of the effects of gender on phonological awareness have yielded mixed results. When Witcher (2001) examined the effects of gender on the phonological awareness of kindergartners, for instance, girls were found to score higher on average than boys, but the observed difference was not statistically significant. When a literacy task was administered to children in both kindergarten and first grade (Wehry, 2003), a higher initial status was demonstrated for girls, which shows gender differences in the emergent literacy abilities of children to exist. In contrast, Savage and Carless (2004) found no gender differences in the phonological awareness of 5-year-olds.

In other research, socioeconomic status (SES) has been found to account for a significant amount of variance in the reading achievement of children (e.g., Bowey, 1995; Noble, Farah, & McCandliss, 2006). In studies that have specifically investigated the effects of SES on phonological awareness in order to better understand the relations between SES and reading, Raz and Bryant (1990) found significant differences on rhyme oddity tasks for 5- and 6-year-old children: the children from the higher SES sample performed better than the children from the lower SES sample. In regression analyses, both Lonigan et al. (1998) and Fernandez-Fein and Baker (1997) showed SES to be significantly related to the phonological awareness of children: children from middle-income families attained substantially higher scores than children from lower-income families. SES was also found to influence the pattern of development for phonological awareness (Lonigan et al.). That is, social-class differences only played a role in the phonological awareness of the 4- and 5-year old groups and not in the 2- and 3-year-old groups with the children in the middle-income groups scoring higher than the children in the lower-income groups, which shows significantly different developmental patterns for the groups of children distinguished on the basis of SES.

Yet another background variable of possible importance for the study of the development of phonological awareness is linguistic diversity. While linguistic diversity in the form of bilingualism has been hypothesized to exert a facilitative effect on children's phonological awareness, the research to date has yielded only mixed results. Bialystok, Majumder, and Martin (2003), for example, found no significant effects of bilingualism on the phonological awareness development of children in kindergarten through second grade. Rubin and Turner (1989) reported an advantage for bilingual first graders and, in contrast, Bruck and Genesee (1995) found an advantage for bilingual kindergartners that disappeared by the end of first grade.

4.1.4 The present study

The present study was designed to examine the developmental course of phonological awareness during elementary school. However, in order to investigate growth in phonological awareness, the different measures of phonological awareness for the children at different ages must be shown to reflect the same latent ability. To avoid the violation of the implicit assumption of unidimensionality, the growth analyses in the present study were performed using an existing item response theory (IRT) calibrated scale for phonological awareness. In a previous study (Vloedgraven & Verhoeven, 2008), the fit of a unidimensional model for different tasks measuring children's phonological awareness was examined within the framework of IRT. The results showed that phonological awareness can indeed be conceptualized as a single, unidimensional ability during the preschool and elementary school years. The distinctive feature of IRT is that the difficulty of the items and the ability of persons are scaled on the same scale. In IRT, the item and ability parameters are invariant, which implies that different sets of items measure the same construct across different grades. This characteristic makes IRT a sophisticated approach for the measurement of growth. For a detailed description of IRT, see for example Hambleton, Swaminathan, and Rogers (1991) and Loyd (1988).

In the present study, performance on a phonological awareness test that included a variety of tasks was investigated cross-sectionally using relatively large participant samples ($N > 250$) for the kindergarten, first-, second-, third-, and fourth-grade years of elementary school. The first aim of the study was to characterize the development of phonological awareness from kindergarten to fourth grade. In doing this, whether the children progress at a constant rate across time (i.e., linearly) or at a decreasing rate (i.e., quadratically) was examined. For proper model specification, the effects of a number of possibly important other variables were also taken into account. The second aim of the present study was thus to investigate the effects of the children's abilities on a number of related skills on the development of their phonological awareness. In kindergarten, the effects of letter knowledge on the development of phonological awareness were examined. In first through fourth grade, the effects of word reading ability were explored. A related question was whether those children found to have reading problems overcome any concomitant impairments of their phonological awareness or still experience these problems at the end of elementary school. The third aim was to examine the effects of the background variables of gender, SES, and linguistic diversity on the children's phonological awareness and the development of their phonological awareness from kindergarten through fourth grade.

4.2 Method

4.2.1 Participants

Data from our previous study (Vloedgraven & Verhoeven, 2008) were also used in this study. A total of 1405 kindergarten through fourth-grade children coming from 57 regular elementary schools in the Netherlands participated (291 kindergarten children, 299 first-grade children, 251 second-grade children, 269 third-grade children, and 295 fourth-grade children). The total sample consisted of 778 boys and 627 girls. The schools were selected using a stratified random sampling method to assure a distribution of social-economic backgrounds representative of the population. In addition, all of the provinces in the Netherlands were represented within the sample of schools. All of the kindergarten children in the sample were randomly selected for inclusion while 50% of the first- through fourth-grade children in the sample was randomly selected for inclusion and the other 50% selected on the basis of a score below the 25^{th} percentile on a standardized word-reading test.

In the Netherlands, children attend school from the time that they are 4 years of age. They spend 2 years in kindergarten, where no formal literacy instruction is provided but beginning literacy is promoted via the playing of language games, for example. After 2 years of kindergarten, the children enter first grade where they receive formal reading and spelling instruction. Given that Dutch is assumed to have a relatively transparent orthography with consistent mappings between letters and phonemes, the instruction children receive is phonics based.

The children were studied for one school year. At the time of initial testing, the mean age of the kindergartners was 5.7 years (*SD* = 0.39), the mean age of the first graders was 6.8 years (*SD* = 0.74), the mean age of the second graders was 7.8 years (*SD* = 0.56), the mean age of the third graders was 8.8 years (*SD* = 0.50), and the mean age of the fourth graders was 9.8 years (*SD* = 0.55). Between initial and final testing, the total sample decreased by about 6% due to mainly illness or the transfer of children to other schools or other places.

4.2.2 Materials

Phonological awareness. For the selection of the most suitable measures to assess phonological awareness, we have looked at the extent to which a measure represents phonological awareness ability. We also selected measures that are known in the literature to assess a broad range of phonological awareness skills (e.g., Chard & Dickson, 1999; Yopp, 1988). In the end, the following five tasks were selected in order to measure phonological awareness: rhyming, phoneme identification, phoneme blending, phoneme deletion, and phoneme segmentation. These tasks have also been used and described in more detail in our previous study (Vloedgraven & Verhoeven, 2008). The different tasks were presented on a computer. The tasks contained high frequency monosyllabic words with different CV (i.e., consonant-vowel) structures. Each phonological awareness item consisted of three response alternatives which were presented both auditorily and visually in the form of pictures in order to reduce memory demands. Next, the target word was presented auditorily. The child was asked to select the correct picture.

Receptive letter knowledge. This task was also digitally administered. Four letters were presented visually to the children. One of the letters was pronounced, and the child then had to select the correct letter. All of the Dutch letters were presented with the exception of the infrequently occurring letters of [c], [x], [q], and [y]. In addition to these letters, digraphs were included, which meant that the test consisted of 34 items. All of the letters were printed in lower case, and the child's score was the number of items answered correctly.

Word decoding. This paper and pencil task consisted of a list of words (from monosyllabic words containing two to five phonemes to bisyllabic words). The children were asked to read the words aloud as quickly and as accurately as possible in two minutes. The score was the number of words read correctly.

4.2.3 Procedure

Our previous study showed rhyming, phoneme identification, phoneme blending, phoneme deletion, and phoneme segmentation to reflect the same underlying ability (Vloedgraven & Verhoeven, 2008). Earlier research has nevertheless shown not all of the different tasks for phonological awareness to be appropriate for all ability levels; that is, at certain points in development, certain tasks do not distinguish high versus low performers (e.g., Anthony & Lonigan, 2004). Floor and ceiling effects can lead to less accurate measurements of children's ability. For example, a rhyming task is generally able to accurately measure the phonological awareness abilities of kindergartners while a phoneme deletion task is known to be more appropriate to measure the phonological awareness abilities of first graders (e.g., Chard & Dickson, 1999). Administration of all tasks to all grades is not particularly useful, thus. We therefore selected tasks that — according to the literature — fit the level of ability for children in a particular grade (Schatschneider, Francis, Foorman, Fletcher, & Mehta, 1999; Stanovich et al., 1984; van Bon & van Leeuwe, 2003). In kindergarten, items measuring rhyming, phoneme identification, phoneme blending, and phoneme segmentation were administered. In first grade, items measuring phoneme blending, phoneme segmentation, and phoneme deletion were administered. In the second through fourth grades, items measuring phoneme deletion were administered.

The calibrated item bank, yielded by our previous study, thus included five different tasks with each task including 50 items for a total of 250 items. On each measurement occasion, 20 items per presented task in a certain grade were selected and administered to all of the children in that grade. Given that all of the children were thus administered items from the same underlying phonological awareness scale, each child could be assigned a single score for phonological awareness. The use of such an IRT calibrated scale enabled comparison of the phonological awareness scores across children and investigation of the development of the children's phonological awareness over time.

The phonological awareness test was administered on two occasions during the school year for the kindergartners and second, third, and fourth graders: at the beginning of the school year (i.e., in November) and at the end of the school year (i.e., in April). The phonological awareness of the first graders was tested on three occasions during the school year: at the beginning (i.e., in November), in the middle (i.e., in February), and at the end (i.e., in May). The receptive letter knowledge task was only administered to the kindergartners and the word decoding task was administered to children in first through fourth grade. The tasks for receptive letter knowledge and word decoding were administered on only one occasion during each school year to establish the child's level of this ability.

All of the tasks were administered individually in a quiet room outside the classroom. Each task for phonological awareness and the tasks for receptive letter knowledge and word decoding as well took approximately 5 to 10 minutes to complete. The tasks were administered in two sessions to ensure optimal performance on all of the tasks. Each of the tasks for phonological awareness and the receptive letter knowledge task were preceded by three practice items that could either be followed by correction and explanation when the child responded incorrectly or confirmation when the child responded correctly.

4.2.4 Statistical analyses

To investigate the development of phonological awareness, a univariate regression analysis was performed. Given that the phonological awareness items were calibrated using an IRT model, a regression analysis for latent variables was conducted rather than the more common regression analysis for observed variables (e.g., Verhelst & Verstralen, 2002). The reason to choose for a latent regression model is that the residual in this model is not contaminated by estimation errors of the individual observed IRT ability estimates. In addition, a latent model can still be used when different item sets are administered to different groups of children.

In order to measure the development of phonological awareness, several steps were undertaken. The regression model for kindergartners through fourth graders was initially tested. In doing this, we first investigated the main effects of measurement occasion. Thereafter, we examined the effects of abilities closely related to phonological awareness. For this purpose, the children were divided into three ability groups. The kindergartners were classified as having a poor letter knowledge, average letter knowledge, or good letter knowledge. The first through fourth graders were classified as poor readers, average readers, or good readers. A low ability group of children was then established on the basis of letter knowledge or word reading scores below the 25^{th} percentile. The percentiles were determined for each grade separately and based upon only the scores for the random sample groups in grades one through four. The average ability group was established on the basis of scores falling between the 25^{th} and 75^{th} percentiles. The high ability group was defined as children with scores above the 75^{th} percentile. We then performed another regression analysis including measurement occasion and ability group for both kindergartners and first through fourth graders. To determine if the various ability groups develop differently, the interaction between measurement occasion and ability group was examined in particular.

It should be noted that each model also contained the variable “sample” (i.e., random versus non-random sample) to avoid specification errors that may be caused by the unbalanced sampling design. The regression parameter estimates were interpreted in terms of the z scores and effect sizes (κ). The test statistic z is defined as the effect (β) divided by its standard error of measurement and indicates the significance of the effect. To establish the relevance of a particular effect, the effect size (κ) was computed and defined as the effect divided by the residual standard deviation. This definition of effect size is comparable to the one applied by Cohen (1988). An effect size of 0.2 was judged as small, 0.5 as medium, and 0.8 as large.

It is also important to note that the statistics obtained in the analyses outlined above are conditional averages, based on the sample of this study. To estimate averages in the population, we had to correct back for non-random selection from the population (Verhelst & Verstralen, 2002). For the estimation of the population parameters, we re-weighted the data taking this into account. To gain greater insight into the development of phonological awareness, the marginal averages for the three ability groups were presented graphically.

Next, we examined the main effects of the background variables of gender, SES, and linguistic diversity. Whether or not significant interactions occurred between measurement occasion and each of the currently significant background variables was then examined. To determine the best fitting regression model, all of the background variables that were found to be significant in the previous analyses were simultaneously included in a regression analysis. After the final regression model was estimated, it was attempted to approximate the development of phonological awareness in terms of a growth curve. The marginal averages for each of the measurement occasions in each of the grades were estimated and presented graphically. The fit of the growth curve was assessed by means of the total amount of variance explained.

4.3 Results

4.3.1 Development of phonological awareness

To examine the development of phonological awareness from kindergarten through fourth grade, a univariate regression analysis for latent variables was conducted. First, we investigated the main effect of measurement occasion for the entire sample. The first measurement occasion in kindergarten (i.e., measurement occasion KGb) was treated as the reference category and its effect was therefore set to zero. Pairwise contrasts were then computed for all of the successive measurement occasions in order to gain insight into the development of phonological awareness. The results of these analyses are presented in Table 4.1. As can be seen, the effect size (i.e., effect size of the difference between successive occasions) in kindergarten can be judged as moderate; the effect size in first grade can be judged as large; the effect sizes in the second and third grades can be judged as moderate; and the effect size in fourth grade was rather small. These results thus show a declining rate of development. That is, the rate of growth appears to decrease as the children get older.

Table 4.1

Pairwise contrasts for phonological awareness for the successive measurement occasions in kindergarten through fourth grade

Contrasts	β	*SE*	*z*	κ
KGb	0			
KGe – KGb	0.482	0.094	5.110	0.452
G1b – KGe	2.028	0.102	19.964	1.902
G1m – G1b	1.004	0.096	10.457	0.942
G1e – G1m	0.634	0.101	6.299	0.594
G2b – G1e	0.212	0.116	1.837	0.199
G2e – G2b	0.574	0.127	4.524	0.538
G3b – G2e	0.438	0.131	3.359	0.411
G3e – G3b	0.604	0.132	4.569	0.567
G4b – G3e	0.122	0.134	0.916	0.115
G4e – G4b	0.277	0.139	2.000	0.260

Note. KGb = kindergarten begin; KGe = kindergarten end; G1b = first-grade begin; G1m = first-grade middle; G1e = first-grade end; G2b = second-grade begin; and so forth.

4.3.2 Phonological awareness in relation to literacy

The general question to be addressed here was whether the observed pattern of development for phonological awareness would remain when other literacy-related abilities were taken into consideration. The issue was addressed by examining the kindergartner's receptive letter knowledge in connection with their phonological awareness to start with and then the word reading abilities of the children in first through fourth grade in relation to their phonological awareness.

The kindergarten children were first divided into three ability groups on the basis of their receptive letter knowledge; recall that receptive letter knowledge was only tested for the kindergartners and this task consisted of 34 items. To be sure that the letter knowledge scores of the high, average, and low ability kindergarten groups differed significantly from each other, a one-way ANOVA with letter knowledge score as the dependent variable and kindergarten ability group as the independent variable was conducted. The letter knowledge means and standard deviations for the three ability groups are presented in Table 4.2. The results show the kindergarten letter knowledge scores to indeed vary significantly across the different kindergarten ability groups, $F\ (2, 283) = 760.41$, $p < .001$. Using a post-hoc Tukey procedure, it was further shown that all of the kindergarten ability groups differed significantly from each other.

Table 4.2
Kindergarten receptive letter knowledge means and standard deviations for three ability groups

Letter knowledge ability group	*N*	*M*	*SD*
Low	81	8.05	1.76
Average	143	15.24	3.32
High	62	27.77	3.52

A latent regression analysis was next conducted with the variables measurement occasion and kindergarten ability group included. The results revealed a main effect of kindergarten ability group, which indicates differences in phonological awareness across ability groups (i.e., differences in phonological awareness depending on the children's receptive letter knowledge). The children in the high ability letter knowledge group scored higher on phonological awareness than the children in the average ability letter knowledge group ($\beta = 1.47$, $z = 13.43$, $\kappa = 1.57$) while the children in the average ability letter knowledge group scored higher on phonological awareness than the children in the low ability group ($\beta = 0.52$, $z = 5.16$, $\kappa = 0.55$).

In order to determine if the phonological awareness of the three kindergarten ability groups developed differently, the interaction between measurement occasion and letter knowledge ability group was examined. A significant interaction did not occur, which shows the pattern of development for phonological awareness to be the same for the three kindergarten ability groups.

In the next set of analyses, the effects of the inclusion of word reading ability in the analyses of the development of phonological awareness for the first through fourth graders were considered. According to their word reading performance, the children were divided into three ability groups. For each grade separately, a one-way ANOVA was then conducted on their word reading scores with ability group as the independent variable. This was done to establish that the word reading scores for the three ability groups clearly differed from each other. The word reading means and standard deviations for the three ability groups per grade are presented in Table 4.3. In each grade, the children's word reading scores differed significantly across the ability groups (G1: F (2, 287) = 476.26, $p < .001$; G2: F (2, 243) = 701.58, $p < .001$; G3: F (2, 264) = 722.15, $p < .001$; G4: F (2, 283) = 494.58, $p < .001$). Post hoc analyses using Tukey's procedure further showed all of the ability groups in each of the grades to significantly differ from each other.

Table 4.3
Word reading means and standard deviations for three ability groups in first through fourth grade

Word reading ability group	*N*	*M*	*SD*
Decoding G1			
Low	88	19.80	7.51
Average	154	47.18	13.23
High	48	117.94	35.61
Decoding G2			
Low	113	32.29	11.08
Average	95	81.75	20.50
High	38	144.66	19.07
Decoding G3			
Low	137	56.85	15.30
Average	94	107.40	18.12
High	36	169.44	18.44
Decoding G4			
Low	139	79.84	22.29
Average	110	132.32	17.20
High	37	188.11	19.93

The next step involved testing the regression model that included measurement occasion and word reading ability (i.e., number of words read correctly in two minutes). The results showed a significant main effect of word reading ability. The children in the high ability word reading group performed significantly better on the phonological awareness measures than the children in the average ability group ($\beta = 0.73$, $z = 8.04$, $\kappa = 0.66$) while the children in the average ability group outperformed the children in the low ability word reading group on the measures of phonological awareness ($\beta = 0.68$, $z = 10.67$, $\kappa = 0.61$). These results thus suggest a moderate influence of word reading ability on phonological awareness.

The interaction between measurement occasion and word reading ability was next examined to determine if the ability groups developed differently or not. The interaction was not significant, which indicates similar growth curves. In other words, those children in grades one through four with higher reading skills showed better phonological awareness than those children in grades one through four with lower reading skills across the board (i.e., no matter what measurement occasion).

For the remainder of the first through fourth grade analyses, we had to correct for the variable “sample” because the samples in these grades were not drawn randomly from the population. The marginal averages were therefore estimated on the basis of the conditional averages provided by the regression model for first through fourth graders on the original ability scale. The marginal averages for phonological awareness were then plotted for the three reading ability groups as in Figure 4.1. As can be seen, the developmental rate was fastest in first grade and leveled off as the children got older (i.e., declining rate of development).

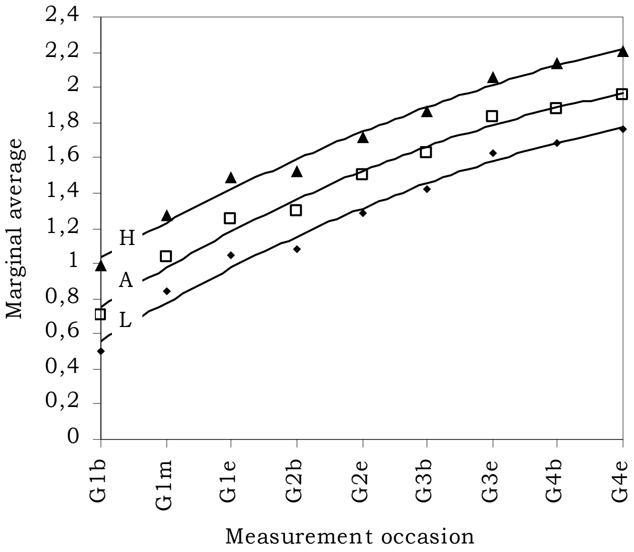

Figure 4.1 *Growth curves for phonological awareness for good (H), average (A), and poor (L) word readers from first through fourth grade*

4.3.3 Role of background variables

Once again, the question of whether all the children studied here showed a similar developmental pattern for phonological awareness or different subgroups could perhaps be distinguished but now on the basis of a number of background variables was raised here. Separate regression analyses on the children's phonological awareness scores were therefore conducted with the variables gender, SES, and linguistic diversity included for this purpose.

The results of the regression analysis showed gender to exert an effect with girls achieving higher phonological awareness scores than boys ($\beta = 0.54$, $z = 6.05$, $\kappa = 0.24$). SES also exerted an effect with the children from a higher SES background producing higher phonological awareness scores than the children from a lower SES background ($\beta = 0.53$, $z = 4.77$, $\kappa = 0.25$). In contrast, no significant effect of linguistic diversity was detected (i.e., the phonological awareness of children with Dutch as the only language used at home versus Dutch in combination with a second language used at home did not differ significantly). The measurement occasion x gender and measurement occasion x SES interactions were next examined but found to be nonsignificant. This shows the developmental patterns for phonological awareness across measurement occasions to be similar for boys and girls and also for children from higher versus lower SES backgrounds.

In order to identify the best fitting regression model for the kindergarten through fourth-grade children, all of the relevant background variables (i.e., measurement occasion, gender, and SES) were entered into the regression analysis. When the pairwise contrasts were computed for all of the successive measurement occasions to gain greater insight into the development of phonological awareness, the influence of measurement occasion proved significant for each adjacent measurement occasion until the end of third grade; that is, the children's phonological awareness scores increased significantly for all of the measurement occasions in kindergarten through third grade ($z > 1.96$). The increases in the children's phonological awareness scores after the end of third grade were not statistically significant. Furthermore, the additional effects of gender and SES continued to exist. The developmental pattern thus seems to remain the same after the inclusion of the relevant background variables.

In the next set of analyses, the fact that not all of the children were randomly selected was controlled for. The marginal averages for phonological awareness were then plotted for each measurement occasion as in Figure 4.2. It should be recalled that the girls generally showed higher phonological awareness scores than the boys and that the children from higher SES backgrounds generally outperformed the children from lower SES backgrounds but otherwise showed similar patterns of development. The growth curve in Figure 4.2 shows more rapid growth on earlier measurement occasions followed by a decreased rate of development on later measurement occasions. Using a least-squares approach to the regression analysis, a quadratic growth curve was indeed found to provide a better fit for the data ($\beta_1 = 0.39$, $t\,(8) = 10.28$, $p < .01$; $\beta_2 = -\ 0.02$, $t\,(8) = -5.26$, $p < .01$; $R^2 = .99$, $F\,(2, 8) = 272.73$, $p < .01$) than a linear growth curve ($\beta_1 = 0.20$, $t\,(9) = 11.44$, $p < .01$; $R^2 = .94$, $F\,(1, 9) = 130.86$, $p < .01$).

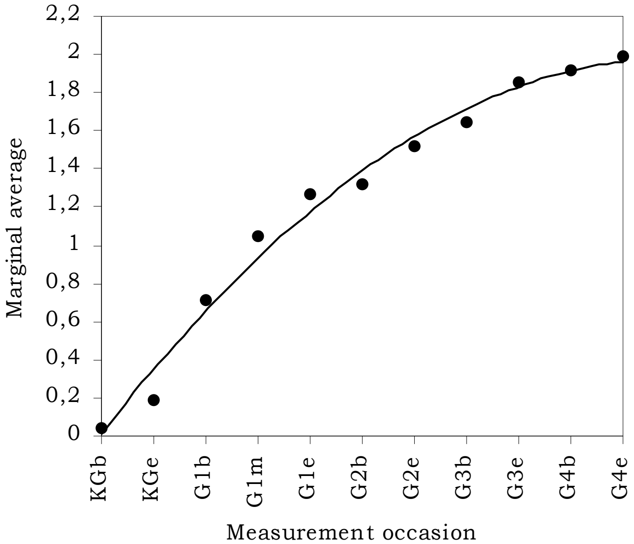

Figure 4.2 *Growth curve for phonological awareness from kindergarten through fourth grade*

4.4 Conclusions and discussion

The first aim in the present research was to describe the development of phonological awareness for children in kindergarten through fourth grade. The results showed the children's phonological awareness to generally increase as they grow older. The results also revealed a declining rate of development over time; that is, the rate of development was fastest in first grade and decelerated as the children grew older. Before interpreting these findings in terms of the general development for a wide variety of children, it is of great relevance to first consider the influence of other variables on the development of phonological awareness.

The second aim in the present study was therefore to examine the possible influence of other related skills on the development of the children's phonological awareness skills. In kindergarten, we explored the effects of the children's receptive letter knowledge. In grade 1 through 4, we explored the effects of the children's word reading ability. The results showed a strong influence of receptive letter knowledge on the phonological awareness of the kindergarten children: the greater the child's receptive letter knowledge, the greater the child's phonological awareness. This association between young children's receptive letter knowledge and their phonological awareness at the same time is in accordance with the results of earlier research (Bowey, 1994; Burgess & Lonigan, 1998; de Jong & van der Leij, 1999; Johnston et al., 1996; Lonigan et al., 2000). The finding that children with greater letter knowledge show higher levels of phonological awareness is also logical as the names of most letters in Dutch (and English, for that matter) simultaneously supply information on their sounds (Adams, 1990; Foorman, Francis, Novy, & Liberman, 1991). The present results further showed the kindergarten children's receptive letter knowledge (i.e., ability group) to not influence the exact course of their development in the domain of phonological awareness in kindergarten. That is, the different ability groups showed dissimilar levels of phonological awareness but similar patterns of phonological awareness development as indicated by the absence of a significant interaction between measurement occasion and ability group in this grade.

In the first through fourth grades of elementary school, the results consistently showed a significant contribution of word reading ability to the level of phonological awareness: the greater the child's reading ability, the greater the child's level of phonological awareness. This strong association between reading ability and level of phonological awareness is consistent with the results of several previous studies (Adams, 1990; Caravolas, Hulme, & Snowling, 2001; Wagner et al., 1994). However, reading ability group did not influence the actual pattern of development for phonological awareness, which was quite uniform for the different reading ability groups in grades one through four (i.e., declining rate of development).

A related question within the context of how children's word reading abilities influence their phonological awareness is whether those children with reading problems in the present study still showed phonological awareness problems at the end of elementary school as found for Dutch dyslexic children by de Jong and van der Leij (2003). The phonological awareness of the children with poor to average word reading abilities clearly lagged behind the phonological awareness of those children with high word reading abilities in the present study and in all of the elementary grades that we studied. In light of the fact that the patterns of development were found to be otherwise quite similar for the different ability groups in the present study, it is unlikely that the poor or average readers will ever catch up to the good readers. Stated differently, both the poor and average readers can be expected to attain a lower maximum on phonological awareness than the good readers in the end which means that their development has not yet reached maturation. These findings provide support for the phonological deficit model of reading disability (e.g., Metsala, Stanovich, & Brown, 1998; Snowling, 2000).

The third aim of the present study was to investigate the effects of some important background variables such as the gender of the child, the social-economic background of the child's family, and linguistic diversity on phonological awareness and its development. The results showed the girls to perform better in general than the boys and the children from a higher social-economic background to attain higher scores than the children from a lower social-economic background although the effects of these variables were small. Linguistic diversity did not appear to influence the children's phonological awareness. Gender and SES did not otherwise influence the growth curves for phonological awareness. That is, a decreased rate of development for phonological awareness after an initial increase between kindergarten and first grade appeared to hold for all of the subgroups of children established on the basis of gender and SES. The findings concerning the effects of background variables are in keeping with the results of previous research. The finding of better performance on the part of girls corresponds to the findings of Wehry (2003). The finding that the children with a higher social-economic background attained higher scores than the children with a lower social-economic background corresponds to the conclusions of several other studies (Fernandez-Fein & Baker, 1997; Lonigan et al., 1998; Raz & Bryant, 1990). Finally, the finding of no influence for linguistic diversity in the present study is in accordance with the outcomes of a recent comprehensive study by Bialystok et al. (2003).

It is interesting to note how our results about the developmental course of phonological awareness relate to previous studies on this issue. The present findings clearly show the development of phonological awareness from kindergarten through the fourth grade of elementary school to be characterized by a declining rate of development. Because none of the relevant other variables appeared to affect the pattern of development for phonological awareness, a quadratic growth curve appeared to characterize the developmental pattern for a wide variety of children. In a previous study, Lonigan et al. (1998) showed the growth in phonological awareness of children between the ages of 2 and 5 years to accelerate in particularly the older groups of children. The present results showed the children's development of phonological awareness to accelerate considerably from kindergarten through first grade (i.e., between the ages of 6 and 7 years). Combining the results of both studies, we can conclude that the development of phonological awareness can best be described by a slow start when children are 2 to 4 years old, followed by a strong increase at the age of 5 to 7 and will eventually decelerate in subsequent years. According to Lonigan et al., the strong increase in 4- to 5- year old children might be a function of the cognitive maturation of children or of the changing quality of children's home literacy environments. The finding from the present study that growth was fastest in first grade can be explained by the explicit phonics instruction children go through in this grade (cf. van Bon & van Leeuwe, 2003). The decreasing rate of growth in second grade was also found by Wagner et al. (1994), but they did not investigate development beyond second grade. The present results show children's phonological awareness to continue to develop even after second grade and, in fact, significant performance increases to occur until the end of third grade. A plateau appears to be reached at that point. To summarize, our findings about the development of phonological awareness over the grades are in concordance with previous studies that investigated development over shorter time lags. The present study has advanced knowledge about the development of phonological awareness by examining development across a longer period of time.

Some possible limitations on the present study and recommendations for future research can now be mentioned. First, the study had a cross-sectional as opposed to longitudinal design. This means that, while the present results provide some important insights into the specific course of development for phonological awareness during early elementary school, future studies should consider the development of phonological awareness from a longitudinal perspective for replication purposes. The use of longitudinal data will also enable detailed analyses of the stability of individual children's phonological awareness over time and of the predictive value of phonological awareness for learning to read. Second, the present findings showed no significant changes in phonological awareness after the end of third grade, which might depend upon the particular measure of phonological awareness used in the present study. According to the relevant research literature, the different sets of items used to assess the phonological awareness of the children in the present study accurately assess a broad range of abilities. Though, an interesting issue for future research is whether phonological awareness might be found to continue to develop even after third grade, if the speed of the phonological awareness measure is taken into consideration.

The present findings have some clear implications for practitioners. The results show that phonological awareness continues to develop in the second and third grades of elementary school, which is after initial reading instruction, and that poor readers continue to underperform on the phonological awareness measures as well. It is thus important that the phonological awareness of poor readers be monitored in the second and third grades in addition to the earlier grades. Consequently, phonological awareness assessment may help to attune treatment programs to the specific needs of these children. Given the strong relation between phonological awareness and reading and the fact that poor readers have not reached their ceiling in phonological awareness, it might be expected that a reading intervention with extended phonics instruction yields positive effects for the literacy skills of these older children. Indeed, previous studies have already provided evidence for the effectiveness of phonological awareness intervention in promoting the reading skills of older children with reading difficulties (Gillon & Dodd, 1995; Swanson, Hodson, & Schommer-Aikins, 2005). In closing, the relatively low phonological awareness of the poor readers in all of the elementary grades examined in the present study underlines the importance of early identification and intervention for children who are clearly at risk for reading failure.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology, 96,* 43-55.

Bialystok, E., Majumder, S., & Martin, M. M. (2003). Developing phonological awareness: Is there a bilingual advantage? *Applied Psycholinguistics, 24,* 27-44.

Bowey, J. A. (1994). Phonological sensitivity in novice readers and nonreaders. *Journal of Experimental Child Psychology, 58,* 134-159.

Bowey, J. A. (1995). Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement. *Journal of Educational Psychology, 8,* 476-487.

Bruck, M. (1992). Persistence of dyslexics' phonological deficits. *Developmental Psychology, 28,* 874-886.

Bruck, M., & Genesee, F. (1995). Phonological awareness in young second language learners. *Journal of Child Language, 22,* 307-324.

Burgess, S. R., & Lonigan, C. J. (1998). Bidirectional relations of phonological sensitivity and prereading abilities: Evidence from a preschool sample. *Journal of Experimental Child Psychology, 70,* 117-141.

Byrne, B., Fielding-Barnsley, R., & Ashley, L. (2000). Effects of preschool phoneme identity training after six years: Outcome level distinguished from rate response. *Journal of Educational Psychology, 92,* 659-667.

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language, 45,* 751-774.

Chard, D. J., & Dickson, S. V. (1999). Phonological awareness: Instructional and assessment guidelines. *Intervention in School and Clinic, 34,* 261-270.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology, 91,* 450-476.

de Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology, 95,* 22-40.

Elbro, C., Borstrøm, I., & Petersen, D. K. (1998). Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly, 33,* 36-60.

Fernandez-Fein, S., & Baker, L. (1997). Rhyme and alliteration sensitivity and relevant experiences among preschoolers from diverse backgrounds. *Journal of Literacy Research, 29,* 433-459.

Foorman, B. R. Francis, D. J., Novy, D. M., & Liberman, D. (1991). How letter-sound instruction mediates progress in first-grade reading and spelling. *Journal of Educational Psychology, 83,* 456-469.

Gillon, G., & Dodd, B. (1995). The effects of training phonological, semantic and syntactic processing skills in spoken language on reading ability. *Language, Speech, and Hearing Services in Schools, 26,* 58-68.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals *of item response theory.* Newbury Park, CA: Sage.

Hindson, B. A., Byrne, B., Fielding-Barnsley, R., Newman, C., Hine, D. W., & Shankweiler, D. (2005). Assessment and early instruction of preschool children at risk for reading disability. *Journal of Educational Psychology, 97,* 687-704.

Holbrook, H. T. (1988). Sex differences in reading: Nature or nurture? *Journal of Reading, 2,* 574-576.

Johnston, R. S., Anderson, M., & Holligan, C. (1996). Knowledge of the alphabet and explicit awareness of phonemes in prereaders: The nature of the relationship. *Reading and Writing: An Interdisciplinary Journal, 8,* 217-234.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent variable longitudinal study. *Developmental Psychology, 36,* 596-613.

Lonigan, C. J., Burgess, S. R., Anthony, J. L., & Barker, T. A. (1998). Development of phonological sensitivity in two- to five-year-old children. *Journal of Educational Psychology, 90,* 294-311.

Loyd, B. H. (1988). Implications of item response theory for the measurement practitioner. *Applied Measurement in Education, 1,* 135-143.

McBride-Chang, C., & Kail, R. V. (2002). Cross-cultural similarities in the predictors of reading acquisition. *Child Development, 73,* 1392-1407.

McBride-Chang, C., Wagner, R., & Chang, L. (1997). Growth modeling of phonological awareness. *Journal of Educational Psychology, 89,* 621-630.

Metsala, J. L., Stanovich, K. E., & Brown, G. D. A. (1998). Regularity effects and the phonological deficit model of reading disabilities: A meta-analytic review. *Journal of Educational Psychology, 90,* 279-293.

Noble, K. G., Farah, M. J., & McCandliss, B. D. (2006). Socioeconomic background modulates cognition–achievement relationships in reading. *Cognitive Development, 21,* 349-368.

Raz, I. S., & Bryant, P. (1990). Social background, phonological awareness and children's reading. *British Journal of Developmental Psychology, 8,* 209-225.

Roth, F., Speece, D. L., Cooper, D. H., & DeLaPaz, S. (1996). Unresolved mysteries: How do metalinguistic and narrative skills connect with early reading? *Journal of Special Education, 30,* 257-277.

Rubin, H., & Turner, A. (1989). Linguistic awareness skills in grade one children in a French immersion setting. *Reading and Writing: An Interdisciplinary Journal, 1,* 73-86.

Sammons, P., & Smees, R. (1998). Measuring pupil progress at key stage 1: Using baseline assessment to investigate value added. *School Leadership & Management, 18,* 389-407.

Savage, R., & Carless, S. (2004). Predicting curriculum and test performance at age 7 years from pupil background, baseline skills and phonological awareness at age 5. *British Journal of Educational Psychology, 74,* 155-171.

Schatschneider, C., Francis, D. J., Foorman, B. F., Fletcher, J. M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology, 91,* 467-478.

Snowling, M. J. (2000). *Dyslexia.* Oxford: Blackwell.

Spector, J. E. (1992). Predicting progress in beginning reading: Dynamic assessment of phonemic awareness. *Journal of Educational Psychology, 84,* 353-363.

Stanovich, K. E., Cunningham, A. E., & Cramer, B. B. (1984). Assessing phonological awareness in kindergarten children: Issues of task comparability. *Journal of Experimental Child Psychology, 38,* 175-190.

Swanson, T. J., Hodson, B. W., & Schommer-Aikins, M. (2005). An examination of phonological awareness treatment outcomes for seventh-grade poor readers from a bilingual community. *Language, Speech, and Hearing Services in Schools, 36,* 336-345.

van Bon, W. H. J., & van Leeuwe, J. F. J. (2003). Assessing phonemic awareness in kindergarten: The case for the phonemic recognition task. *Applied Psycholinguistics, 24,* 195-219.

Verhelst, N. D., & Verstralen, H. H. F. M. (2002). *Structural Analysis of a Univariate Latent Variable (SAUL). Theory and a computer program.* Arnhem: Cito.

Vloedgraven, J. M. T., & Verhoeven, L. (2008). The nature of phonological awareness throughout the elementary grades: An item response theory perspective. *Learning and Individual Differences,* doi:10.1016/j.lindif.2008.09.005.

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.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., et al. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33,* 468-479.

Wehry, S. (2003). *The Early Literacy and Learning Model (ELLM): 2002/2003.* University of North Florida: Florida.

Witcher, S. H. (2001). *Effects of educational kinesiology, previous performance, gender, and socioeconomic status on phonological awareness literacy screening scores of kindergarten students.* Unpublished doctoral dissertation. Virginia Polytechnic Institute and State University: Blacksburg.

Yopp, H. K. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23,* 159-177.

5

Relations between phonological awareness and word reading: A latent variable approach[4]

Abstract The purpose of the present study was to examine the longitudinal relations between Dutch children's phonological awareness and word reading abilities from kindergarten through second grade. Both phonological awareness and word reading were found to be quite stable throughout the early elementary grades. The results further showed initially significant effects of phonological awareness on word reading but once the autoregressive effects of word reading were taken into account, subsequent effects of phonological awareness on word reading were not significant any more. Word reading significantly predicted phonological awareness at times, which lends support to the view that the relations between phonological awareness and reading are bidirectional. Finally, the early effect of phonological awareness on subsequent word reading diminished after the inclusion of letter knowledge although phonological awareness still explained significant additional variance in word reading. Both phonological awareness and letter knowledge were thus found to be unique predictors of reading.

[4] Reference: Vloedgraven, J. M. T., Verhoeven, L., & Eggen, T. J. H. M. (submitted). Relations between phonological awareness and word reading: A latent variable approach.

5.1 Introduction

In learning to read, children must learn to match distinctive visual symbols to units of sound. The process of acquiring these mappings between symbol and sound is often referred to as phonological recoding (Perfetti, 1991; Ziegler & Goswami, 2005). Given that this process of print-to-sound translation can function as a self-teaching device that enables children to decode novel words (Share, 1995), phonological recoding is often assumed to be necessary to learn to read. The quality of the representations of phonological structures in the lexicon is typically assessed using various measures of phonological awareness. In the past 25 years, numerous studies have provided evidence for a relation between phonological awareness and reading (e.g., Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004; Wagner, Torgesen, & Rashotte, 1994). The association is robust and appears to be present even after variance due to IQ, vocabulary, memory, and social class is taken into consideration (Bryant, MacLean, Bradley, & Crossland, 1990). There is nevertheless controversy regarding the direction, magnitude, and processes underlying the associations between phonological awareness and reading. In addition, the development of each of these variables has been given relatively little attention.

5.1.1 Nature of phonological awareness

Phonological awareness can be defined as the sensitivity to the sound structure of oral language. Converging evidence exists that there is a developmental progression in the acquisition of phonological awareness from larger to smaller phonological units (e.g., Anthony & Francis, 2005; Ziegler & Goswami, 2005). However, there is considerable debate about whether all units have a predictive value for reading. According to Hulme et al. (2002), sensitivity to phonemes is a better predictor of reading than rhyming ability. In contrast, Goswami and Bryant (1990) have provided evidence that rhyming ability is of greater importance for the prediction of reading. Lonigan, Burgess, and Anthony (2000) have argued that there are problems with the use of such a predictor approach, namely that it is generally assumed that various types of phonological awareness exist. In addition, the overlap between the different measures of phonological awareness is not taken into consideration. Finally, the fact that the different measures have different reliabilities is often ignored.

To properly investigate the relations between phonological awareness and reading, thus, we must first know more about the exact nature of phonological awareness. Anthony and Francis (2005, p. 256) reported that "methodologically sound studies using large samples, multiple measures, and advanced statistics support a unified phonological awareness construct." By means of confirmatory factor analysis, it was demonstrated that the different tasks measuring phonological awareness are manifestations of a single underlying ability (Anthony & Lonigan, 2004; Anthony et al., 2002). Moreover, studies using item response theory (IRT) have yielded the same conclusion (e.g., Schatschneider, Francis, Foorman, Fletcher & Mehta, 1999; Vloedgraven & Verhoeven, 2008).

The significance of phonological awareness depends upon not only its predictive ability for reading skills but also on the stability of phonological awareness abilities over time. As suggested by Wagner et al. (1997), the value of early screening for reading problems will increase as individual differences in phonological awareness remain consistent over time. When Wagner et al. examined the development of phonological processing abilities in a latent-variable longitudinal correlation study with kindergartners through fourth graders, the results showed the individual differences in phonological awareness to be rather stable throughout the children's development. In a similar vein, Lonigan et al. (2000) showed the latent phonological awareness variable to be highly stable from late preschool through early elementary school.

5.1.2 Relations between phonological awareness and reading

In general, there are three alternative views on the nature of the relations between phonological awareness and word reading. First, it has been suggested that the development of phonological awareness precedes the acquisition of reading skills. Evidence for this view comes from numerous longitudinal studies in which phonological awareness is found to be an important predictor of later reading skills (for reviews, see Wagner & Torgesen, 1987; Wagner et al., 1994). Second, it has been suggested that the development of reading influences the development of phonological awareness (Morais, Alegria, & Content, 1987). Evidence for this view comes from the low level of phonological awareness found for not only prereaders (Liberman, Shankweiler, Fischer, & Carter, 1974) but also illiterate adults and readers of a nonalphabetic written language (Lukatela, Carello, Shankweiler, & Liberman, 1995; Read, Zhang, Nie, & Ding, 1986). Finally, it has been suggested that the causal relations between phonological awareness and reading are bidirectional (Ehri, 1992; Perfetti, Beck, Bell, & Hughes, 1987; for a review, see Shankweiler & Fowler, 2004). That is, individual differences in phonological awareness have been found to influence the development of reading abilities and, conversely, individual differences in the process of learning to read have been found to influence the development of phonological awareness.

Related to the issue of how the relations between phonological awareness and reading skills can best be characterized is the issue of whether the relations between phonological awareness and reading are relatively stable over time or not. When Wagner et al. (1997) studied children from kindergarten through fourth grade, individual differences in phonological awareness were found to influence subsequent individual differences in reading achievement for all of the different periods investigated. In contrast, de Jong and van der Leij (1999) found evidence for changing relations between phonological awareness and reading skills over time in their study of a Dutch sample of children. In kindergarten, phonological awareness did not relate to subsequent reading skills. After the start of literacy education in first grade, phonological awareness was found to influence later reading. The influence of phonological awareness increased during the course of first grade, but disappeared thereafter. In other words, the association between phonological awareness and subsequent reading skill appears to depend upon the moment of measurement. According to de Jong and van der Leij, a plausible explanation for the time-limited importance of phonological awareness in relation to reading may be due to the relative transparency of the Dutch language.

5.1.3 Methodological issues

When examining the connections between phonological awareness and reading, it is important to formulate a correctly specified model of causal relations by minimizing the effects of common sources of misspecifications. One general source of model misspecification is not permitting bidirectional relations (Wagner et al., 1994). Along these lines, Schult (1999) has emphasized that a strong association between two variables assessed concurrently does not provide information on the causal relations between the variables or the direction of these relations, be it phonological awareness preceding later reading or the other way round. Time (i.e., measurement of an independent variable prior to measurement of a dependent variable) should thus be taken into account in any study of causal relations, and this can be only achieved by means of longitudinal studies. To examine any changes in the relations between two variables and the directionality of the relations between two variables, measurement should be undertaken on several occasions. Unfortunately, many studies have been restricted to a single occasion (e.g., McDougall, Hulme, Ellis, & Monk, 1994). Longitudinal studies in which children's phonological awareness abilities were measured at the time these children might already be able to read (e.g., Juel, Griffith, & Gough, 1986) have to contend with the same problem.

Another common source of model misspecification is the omission of clearly known plausible causes (Macmillan, 2002; Schult, 1999; Wagner et al., 1994). An example of such a plausible cause is the so-called autoregressive effect of a variable measured on a particular occasion on the same variable measured on a later occasion (Gollob & Reichardt, 1987). According to Schult, the demonstration of an association over time is not sufficient evidence to conclude that a causal relation exists. In order to establish causality, the possibility of a spurious association between two variables (i.e., an association that is the result of variation in another variable) must also be ruled out. With respect to the associations between children's kindergarten phonological awareness and later reading, for example, it has simply been assumed that kindergartners cannot read and that it therefore makes no sense to assess their reading and examine the possible autoregressive effects of their early reading skills on later reading skills. As Castles and Coltheart (2004) claim in their recent review, however, many preschoolers already possess some literacy skills, such as letter knowledge.

A large amount of research shows letter knowledge to be one of the single best predictors of later reading skills in an alphabetic language with close correspondences between graphemes and phonemes (e.g., Lonigan et al., 2000; Snowling, Gallagher, & Frith, 2003; Stevenson & Newman, 1986). In addition, several studies have shown that letter knowledge strongly influences phonological awareness (e.g., Burgess & Lonigan, 1998; Caravolas, Hulme, & Snowling, 2001). Along these lines, studies of pre-readers have found evidence that a certain amount of letter knowledge is needed to show some awareness of phonemes (Barlow-Brown & Connelly, 2002; Johnston, Anderson, & Holligan, 1996). It seems logical to assume that phonological awareness may be activated by letter knowledge, as Barron (1994), for example, has suggested that letter knowledge supplies children with phonological representations that include information on the orthographic units in a word. Knowledge of letters can function as a concrete referent and thereby make the phonemes in a language less abstract (see also Wagner et al., 1997). Given the strong associations of letter knowledge with both phonological awareness and reading, thus, it is possible that the relations between phonological awareness and reading may be largely mediated by letter knowledge. Among others, Blaiklock (2004) has indeed provided evidence for a major role of letter knowledge in the relations between phonological awareness and reading. Strong associations were found between phonological awareness and reading, but these relations were greatly affected by letter knowledge. When verbal ability and phonological memory were controlled for, the size of the correlations between phonological awareness and reading remained virtually the same; when letter knowledge was controlled for, however, most of the correlations became nonsignificant. While many researchers have emphasized the importance of controlling for existing and pre-existing literacy skills (e.g., Bowey & Francis, 1991;

Castles & Coltheart, 2004; Elbro, Borstrøm, & Petersen, 1998), relatively few studies have actually done this.

A third common source of model misspecification is measurement error (Wagner et al., 1994). Due to measurement error, estimates of the observed correlations between two variables can generally be lower than the true correlations in the population. Many of the studies of phonological awareness and reading have not paid adequate attention to the just described possible sources of model misspecification.

5.1.4 The present study

The present study was designed to determine the development of children's phonological awareness and word reading abilities and to map the longitudinal relations between these two variables from kindergarten through second grade. Prior to examination of the development of phonological awareness and word reading over time, however, it must first be demonstrated that each ability has the same psychological significance over time. That is, to avoid violation of the implicit assumption of unidimensionality, growth analyses were undertaken using IRT calibrated scales. As previous research conducted within the framework of IRT has shown one latent ability to underlie different sets of items used to measure phonological awareness (Vloedgraven & Verhoeven, 2008), the development of phonological awareness was investigated using an existing IRT based scale for phonological awareness. In order to examine the development of children's word reading abilities, it was attempted to construct an IRT based scale for word reading for the purpose of this study. IRT has some important advantages over classical test theory (CTT) due to the fact that in IRT the items and the ability of individuals can be located along the same scale. One advantage is that the estimated ability is independent of the specific items administered, given that these items are taken from the same calibrated item bank. In addition, when an IRT based scale appears to fit in different grades, it can be assumed that the same underlying ability is being measured in all of the grades. This is why IRT is very appropriate to study development over time. For detailed descriptions of IRT, see for example Hambleton, Swaminathan, and Rogers (1991) or Loyd (1988).

The first aim of the present study was to examine how phonological awareness and word reading develop across the grades. Before investigating the relations between these two variables, we need to know whether the children sufficiently progressed on both phonological awareness and word reading over time. The second aim was to investigate the relations between phonological awareness and word reading. To start with, the autoregressive effects of both phonological awareness and word reading were examined. That is, the extent to which later phonological awareness and word reading can be predicted by earlier measurement of the same variables was investigated. The size and directions of the associations between phonological awareness and reading over time were examined next. Of particular interest was the stability or instability of the relations as the children moved from beginning to skilled reading. The third aim of the present study was to investigate the impact of the children's receptive letter knowledge at the kindergarten level for the associations between phonological awareness and word reading. Given that other studies have shown the effects of phonological awareness on subsequent reading to be largely mediated by children's letter knowledge, it is critical that this also be examined in the present study.

In keeping with the research literature and in recognition of the need to consider relevant background variables that possibly influence the development of phonological awareness and word reading (Bruck & Genesee, 1995; Noble, Farah, & McCandliss, 2006; Sammons & Smees, 1998), it was decided to explore the effects of gender, socioeconomic status (SES), and linguistic diversity (i.e., children with Dutch as the only language used at home versus Dutch in combination with a second language used at home) on the children's phonological awareness and word reading scores. In order to avoid the methodological limitations of some previous studies and minimize possible sources of model misspecification, a longitudinal design was adopted. This allowed us to investigate not only the directionality of the relations between phonological awareness and reading ability but also the autoregressive effects of early reading and letter knowledge on later reading and thereby minimize the possibility of the relation between phonological awareness and reading being spurious. Finally, a covariance matrix was estimated under the assumption of an IRT model in order to make sure that the estimated effects are not attenuated by measurement errors (see statistical analyses for further explanation).

5.2 Method

5.2.1 Participants

Data from our previous study (Vloedgraven & Verhoeven, 2008) were also partly used in this study. A total of 590 children from 39 elementary schools in the Netherlands participated in the present study. Of these, 291 were recruited from kindergarten and 299 from first grade. For the present study, both groups were then followed across a period of two years. At the time of initial testing, the mean age of the kindergartners was 5.7 years (*SD* = 0.39); the mean age of the first graders was 6.8 years (*SD* = 0.74). The younger sample contained 158 boys and

133 girls; the older sample contained 169 boys and 130 girls. The schools were selected using a stratified random sampling method to assure a distribution of social-economic backgrounds representative of the population. In addition, all of the provinces in the Netherlands were represented within the sample of schools. All of the kindergarten children were randomly selected for inclusion in the sample while 50% of the first-grade children was randomly selected for inclusion and the other 50% selected on the basis of a word-reading score below the 25th percentile on a standardized test. About 10% of the children left the longitudinal sample during the research period due to mainly illness or placement in another school, or family move to another place.

In Dutch education, kindergarten starts when a child turns 4 and lasts for 2 years. Beginning literacy is generally stimulated in kindergarten via, for example, language games, but no structural literacy program is used. Children start learning to read at the age of 6 years in first grade when they receive formal literacy instruction. Given the relatively transparent orthography of Dutch with consistent mappings between letters and phonemes, the instruction children receive is phonics based.

5.2.2 Materials

Three areas of abilities were tested: phonological awareness, word reading, and letter knowledge.

Phonological Awareness (PA). To assess phonological awareness, measures that are known in the literature to represent phonological awareness were selected for use. The following five tasks were assumed to provide an adequate measure of phonological awareness: rhyming, phoneme identification, phoneme blending, phoneme deletion, and phoneme segmentation and are known to measure a broad range of phonological awareness skills (e.g., Adams, 1990; Vellutino & Scanlon, 1987). These tasks have also been used and described in more detail in our previous study (Vloedgraven & Verhoeven, 2008). All of the tasks were presented digitally on a computer screen and contained high frequency monosyllabic words with different CV (i.e., consonant-vowel) structures. Each phonological awareness item consisted of three response alternatives which were presented both auditorily and visually in the form of pictures to avoid overloading the children's working memories. The target word was then presented auditorily, and the child's task was to select the correct picture.

The analyses from this previous study yielded an IRT calibrated item bank. Both the item-oriented and overall statistical tests showed satisfactory model fit (Verhelst, Glas, & Verstralen, 1995). A uniform distribution of the *p* values for the individual items at the interval [0, 1] favours model fit. The *p* values for the individual items appeared to be reasonably distributed at this interval. In addition, the ratio between the overall R1c to the degrees of freedom was found to be smaller than 1.5:1 (R1c = 2128.01, df = 1538, $p < .01$). For the conceptualization of phonological awareness as a unidimensional ability, it is crucial that the different measures be shown to relate to the same underlying ability in the same manner over time. The results of previous research have shown the different tasks to indeed reflect a single underlying ability across the different grades studied.

Word Decoding Accuracy (WDA). This computer-based task assessed children's ability to correctly decode isolated words of varying difficulty. The children were required to name words presented individually on a screen. A total of 70 (monosyllabic and bisyllabic) words were administered on each measurement occasion from the middle of first grade through second grade.

The question was whether all the word decoding items administered during the first and second grades could be represented along the same scale. To establish whether the IRT model fits the data, the p values for the individual items were examined and an overall test statistic was computed. The p values for the individual items were fairly distributed at the interval [0, 1]. In addition, the ratio between the overall R1c-statistic and the degrees of freedom was 1.5 or lower (R1c = 2469.19, df = 1834, $p < .01$). Both the item-oriented statistics and an overall statistic showed the IRT model to fit the data.

Receptive Letter Knowledge (RLK). This task was also digitally administered. Four letters were presented visually at the same time. The target letter was then pronounced and the children were required to point to the correct letter. All of the Dutch letters were presented with the exception of the following infrequently occurring letters: [c], [x], [q], and [y]. Digraphs such as [ie], and [au] were also included, which meant that the test consisted of 34 items. All of the letters were printed in lower case.

The task for receptive letter knowledge was only administered at the beginning and the end of kindergarten, because this task does not constitute a reliable measure of letter knowledge in first grade because knowledge of letters at this time largely depends upon the reading methods used. For the purpose of the present study, it was attempted to construct an IRT based scale for receptive letter knowledge in kindergarten. To assess model fit, both the item-oriented statistics and an overall test statistic were computed. The p values for the individual items were rectangularly distributed at the interval [0, 1] and the R1c value was 136.19 (df = 90, $p < .01$), which show the model to fit the data reasonably.

5.2.3 Procedure

Previous research has shown different tasks for phonological awareness to not distinguish high versus low performers at certain points in development (e.g., Anthony & Lonigan, 2004). Floor and ceiling effects have been found to lead to inaccurate measurement of children's ability. For example, a rhyming task can accurately measure the phonological awareness of kindergartners while a phoneme deletion task is known to more accurately measure the phonological awareness of first graders. For this reason, tasks were selected per grade to fit the children's level of ability according to the literature (e.g., Schatschneider et al., 1999; van Bon & van Leeuwe, 2003). In kindergarten, items measuring rhyming, phoneme identification, phoneme blending, and phoneme segmentation were administered. In first grade, items measuring phoneme blending, phoneme segmentation, and phoneme deletion were administered. In second grade, items measuring phoneme deletion were administered.

The calibrated item bank included five different tasks with each task encompassing 50 items for a total of 250 items. On each measurement occasion, 20 items per presented task in a certain grade were selected and presented to all of the children in a particular grade. Because all of the children were administered items coming from the same underlying phonological awareness scale, each child could be assigned a single score for phonological awareness. Given that the scale was an IRT calibrated scale, the children's scores could be compared and the development of phonological awareness over time also investigated.

As already mentioned, two cohorts of children were used. Kindergartners were followed from their last year in kindergarten through first grade; first graders were followed from first grade through second grade. All of the children were assessed on five occasions during the relevant two-year period. In kindergarten, the children were tested at the beginning and end of the school year (i.e., KGb and KGe). In first grade, the children were tested at the beginning, middle, and end of the school year (i.e., G1b, G1m, and G1e). In second grade, the children were tested at the beginning and end of the year (i.e., G2b and G2e). The research design of the study is summarized in Table 5.1. Characteristic of this design is that the two cohorts overlapped each other or were linked via so-called anchors, which allows us to link all of the measurement occasions and describe the development of the children's early literacy from kindergarten through second grade without a fully longitudinal design. Given that the two cohorts were not identically sampled from the population, however, the cohort data was reweighted in such a manner that the two cohorts became exchangeable.

The children were individually tested by trained research assistants. Test administration was conducted in two sessions of about 20 minutes each within a

2-3-week period to ascertain optimal performance on all of the tasks. Each task was preceded by three practice items to teach the children the task. During the practice items, children were also given corrective feedback.

Table 5.1
Research design

	Measurement occasions						
Cohort	KGb	KGe	G1b	G1m	G1e	G2b	G2e
Cohort 1	X	X	X	X	X		
Cohort 2			X	X	X	X	X

5.2.4 Statistical analyses

The development of phonological awareness and word reading in addition to the longitudinal relations between these two variables were investigated over time using the AMOS 6.0 program for structural equation modeling (SEM; Arbuckle, 2005). SEM is a statistical modeling technique that determines relationships among variables. A major advantage of SEM is that each variable can function as both an independent variable and a dependent variable within the same model (Schumacker & Lomax, 1996), which allows the modeling of bidirectional relations.

SEM is typically used to assess both the measurement and structural components of a causal model. The measurement model represents the associations between observed indicators and hypothetical constructs or latent variables. The structural model represents the associations between the latent variables, describes the effects, and assigns the explained and unexplained variance. In the present study, the measurement and structural models were separated. As already mentioned, the phonological awareness, word decoding, and receptive letter knowledge items were calibrated using IRT. In the IRT model that we used, the variable “item responses” is assumed to be explained by a unidimensional latent ability. When the model is considered true, the means and covariance matrix for the latent ability can be computed directly using the observed weighted sum scores and the item parameter estimates (see, for example, Andersen & Madsen, 1977; Mislevy, 1984), and yet only the structural coefficients need to be estimated with SEM software. Separation of the measurement and structural models has the added advantage that the sample used for calibration purposes can differ from the sample used to estimate the structural model. The structural model can also be estimated even when the individuals in the sample have been assessed using different tests.

The computations of the latent covariance matrix and means were conducted using the MULTI computer program (Kamphuis, 1993, 1998). However, in the present study, two problems presented themselves during the estimation of the covariance matrix. The first was that the design was incomplete. The second was that the design was unbalanced. The problem of an incomplete research design could be solved by MULTI as the software can handle missing data, provided it is missing at random (MAR; Little & Rubin, 1987). In order to apply MULTI, it was assumed that the missing data was missing at random. The problem of an unbalanced research design could be solved by reweighting the data. In order to estimate representative means and covariances, non-random selection from the population was corrected for. Children with scores below the 25th percentile on a standardized word-reading test were oversampled in the present study. Data were therefore generated for two groups of children, namely those with scores above the 25th percentile and those with scores below the 25th percentile. The true data sets (i.e., without measurement error) for the two groups of children were then combined, and we corrected for the variable "sample" to establish the same ratio as in the population (i.e., ratio of children with scores above the 25th percentile to children with scores below the 25th percentile should be 3:1). The data were generated in such a manner that the corrected data set contained ability estimates for 2000 children. Next, the new data set was used to estimate the covariance matrix.

To answer the research questions, different SEM analyses were conducted in a stepwise manner. Prior to examination of the longitudinal relations between phonological awareness and word reading, the effects of the background variables of gender, SES, and linguistic diversity on these variables were investigated. Separate regression analyses on the children's phonological awareness and word reading scores for each of the background variables were conducted with a software package for the structural analysis of univariate latent person parameters (SAUL; Verhelst & Verstralen, 2002).

Next, the autoregressive effects of phonological awareness and word reading were investigated by estimating the simplex model for both variables (Jöreskog, 1970). The underlying assumption in this model is that the child's ability on a particular measurement occasion can be explained by the same ability assessed on a previous measurement occasion. The correlation matrix for the simplex model is characterized by a particular structure, moreover. The largest value in each column should be next to the main diagonal for that column, and the correlation coefficients should decline as they deviate from the main diagonal.

The relations between the latent variables in kindergarten through second grade were next investigated by testing several models. A model that only takes the autoregressive relations for phonological awareness and word reading into account constituted the starting point for these analyses. The disturbance variables for phonological awareness and word reading assessed on the same measurement occasions were allowed to correlate. Finally, receptive letter knowledge at the kindergarten level was also included in the model.

To determine the goodness of fit for the estimated models, a chi-square test (x^2, with degrees of freedom and *p* value) — which measures the discrepancy between the covariance matrix implied by the model and the sample covariance matrix — is often used. However, when the sample is large, the x^2 value tends to be significant in all cases. Stated differently, use of a chi-square test with a large sample can lead to a rejection of the model even when the model is largely consistent with the analyzed data. Three other indices were therefore used to evaluate the goodness of fit for the estimated models: the comparative fit index (CFI), the normed fit index (NFI), and the standardized root mean square residual (SRMR). When CFI and NFI exceed .90, the model can be assumed to fit the data acceptably. Furthermore, the SRMR should be lower than .10 to provide a reasonable fit (Hu & Bentler, 1999). The structural parameters (i.e., standardized regression coefficients) were estimated using the maximum likelihood (ML) method.

5.3 Results

5.3.1 Descriptive statistics

The means and standard deviations for phonological awareness and word reading on the different measurement occasions are presented in Tables 5.2 and 5.3, respectively. For each variable, the scores were linearly transformed to fit a scale ranging from 1 to 100. For receptive letter knowledge assessed at the beginning and end of kindergarten, the means were 26.28 and 36.49, respectively; and the standard deviations were 16.50 and 24.27, respectively. It should be noted that these statistics were estimated on the basis of a true data set (N = 2000) under the assumption of an IRT model and thus contain no measurement error. Two findings regarding these statistics are worth mentioning. First, the children clearly improved both their phonological awareness and word reading abilities in the grades that we studied. Second, the results suggest a declining rate of development for phonological awareness. A strong increase is apparent at the beginning of first grade but, as children grow older, the rate of growth decreases. For word reading, the developmental pattern is less clear due to the smaller number of measurement occasions.

In Tables 5.2 and 5.3, the correlations among the phonological awareness abilities and among the word reading abilities assessed on the different occasions are also presented. For both variables, the correlation matrix reflects a simplex structure: the largest value in each column is next to the main diagonal for that column and the correlation coefficients gradually decline as they depart from the main diagonal.

Table 5.2
Means, standard deviations, and correlations for phonological awareness assessed on different occasions in kindergarten through second grade

	KGb	KGe	G1b	G1m	G1e	G2b	G2e
M	25.94	28.83	45.57	54.80	58.15	62.24	70.67
SD	13.68	12.94	10.58	11.60	13.66	13.34	14.93
KGb	-						
KGe	.88	-					
G1b	.58	.62	-				
G1m	.50	.55	.67	-			
G1e	.49	.58	.62	.79	-		
G2b	.33	.31	.39	.55	.68	-	
G2e	.29	.22	.33	.48	.64	.79	-

Table 5.3
Means, standard deviations, and correlations for word reading assessed on different occasions in first and second grade

	G1m	G1e	G2b	G2e
M	31.73	40.80	49.77	54.98
SD	13.75	13.07	10.88	11.17
G1m	-			
G1e	.89	-		
G2b	.69	.80	-	
G2e	.64	.74	.86	-

The question of effects of gender, SES, and linguistic diversity on phonological awareness and word reading was next considered. The results of regression analyses show gender to exert a significant effect on both phonological awareness and word reading with girls achieving not only higher phonological awareness scores than boys (β = 0.12, z = 1.99, κ = 0.12) but also higher word reading scores (β = 0.19, z = 3.35, κ = 0.23). The effects of SES and linguistic diversity were found to be nonsignificant for both phonological awareness (β = 0.11, z = 1.95, κ = 0.11 and β = 0.06, z = 0.86, κ = 0.06, respectively) and word reading (β = 0.13, z = 1.90, κ = 0.16 and β = 0.09, z = 0.91, κ = 0.11, respectively).

5.3.2 Relations between phonological awareness and reading

To investigate the relations between phonological awareness and word reading, several SEM analyses were conducted in a stepwise manner on the covariance matrix, which was estimated on the basis of a true data set. In the first step, the possibility of autoregressive effects for both phonological awareness and word reading was examined via estimation of the simplex model for the two abilities separately. It should be noted that in light of the significant influence of gender on both phonological awareness and word reading, gender was included as a control variable in each model. The statistics used to evaluate the goodness of fit for the simplex models are reported in Table 5.4. As expected, the x^2 was very large due to the large sample size, so no value is attached to these results. The other statistics show the fit of the simplex model to be satisfactory for both the children's phonological awareness and word reading. That is, the phonological awareness and word reading scores on a particular measurement occasion are adequately predicted by the same abilities measured on the previous occasion.

Table 5.4

Goodness of fit statistics for simplex models of phonological awareness and word reading in kindergarten through second grade

Variable	x^2	*df*	*p*	CFI	NFI	SRMR
PA	834	21	.00	.93	.93	.08
WDA	99	6	.00	.99	.99	.02

To investigate the longitudinal relations between phonological awareness and word reading from kindergarten through second grade, the simplex models for phonological awareness and word reading were next combined into a single model. Those disturbance variables assessed on the same measurement occasion were allowed to correlate. To start with, a model that included all possible cross time-lag effects of phonological awareness on word reading was tested. In other words, phonological awareness on a particular occasion was consistently assumed to predict word reading on the next occasion. The fit of this model, as indicated in

Table 5.5 (model 1) was only just acceptable. A model in which word reading on a particular occasion was assumed to consistently predict phonological awareness on the next occasion was then tested. The fit of this model was found to be quite poor (see Table 5.5, model 2) although some of the effects of word reading on phonological awareness measured one occasion later were significant. This suggests a bidirectional association between the two variables, which was therefore tested next (see Table 5.5, model 3). As can be seen, the fit of the model 3 was satisfactory and better than that of the previous two models. The differences between models 1 and 3 and between models 2 and 3 were both highly significant (x^2 = 401, *df* = 3, $p < .01$ and x^2 = 978, *df* = 4, $p < .01$, respectively).

Table 5.5

Goodness of fit statistics for longitudinal models of phonological awareness and word reading from kindergarten through second grade

Model	x^2	*df*	*p*	CFI	NFI	SRMR
1. PA → WDA	2173	47	.00	.91	.90	.10
2. WDA → PA	2750	48	.00	.88	.88	.22
3. PA ←→ WDA	1772	44	.00	.92	.92	.08
4. PA ←→ WDA, and RLK	2554	64	.00	.91	.91	.09

The nonsignificant relations were next omitted from model 3 to produce the final version of this model. The fit of the final version of the model was similar to the fit of the original version (x^2 = 1778, *df* = 48, $p < .01$, CFI = .92, NFI = .92, SRMR = .08). In Figure 5.1, this model and the standardized regression coefficients are displayed. To keep the figure manageable, the disturbance variables are not presented. Two results are of primary interest. First, as Figure 5.1 shows, phonological awareness is initially a strong predictor of later word reading, but this influence diminishes over time. When the autoregressive effects of word reading are taken into account, phonological awareness does not explain any additional variance in word reading. Second, word reading in both first and second grades is found to exert a significant effect on phonological awareness measured on the next occasion, which provides evidence for bidirectional relations between phonological awareness and word reading.

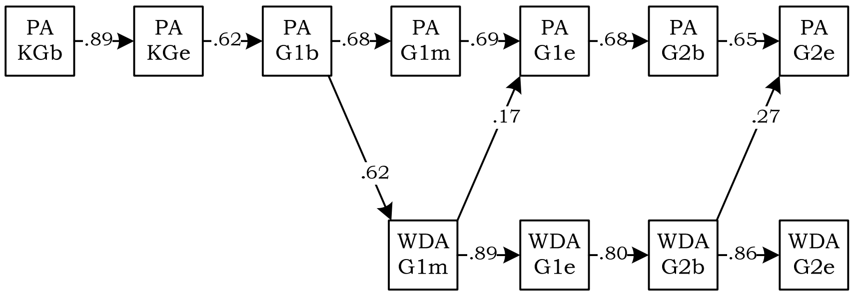

Figure 5.1 *Longitudinal model with bidirectional relations between phonological awareness and word reading*

The influence of kindergarten letter knowledge on the relations between phonological awareness and word reading was examined next. The fit statistics presented in Table 5.5 for model 4 show the fit of the model with bidirectional relations to still be adequate even after receptive letter knowledge has been taken into consideration. The same relations as in the previous model were found to be nonsignificant and therefore omitted from the model. The fit of this final version of model 4 (see Figure 5.2) was comparable to the fit of the original version (x^2 = 2560, *df* = 68, p < .01, CFI = .91, NFI = .91, SRMR = .09). The percentage of the total variance in word reading achievement accounted for by receptive letter knowledge, phonological awareness, and previous word reading skills was substantial on each measurement occasion with a range of 50% to 77%. Interestingly, with the inclusion of receptive letter knowledge, the effects of phonological awareness on later word reading clearly declined. Nevertheless, the results show phonological awareness to account for a significant amount of additional variance in word reading in the middle of first grade after control for receptive letter knowledge. Both phonological awareness and receptive letter knowledge thus appear to be important and unique predictors of reading.

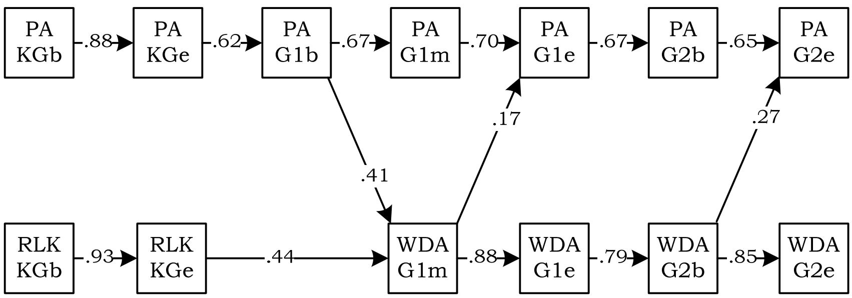

Figure 5.2 *Longitudinal model with bidirectional relations between phonological awareness and word reading after control for receptive letter knowledge*

5.4 Conclusions and discussion

On the basis of the results of this study on the development of children's phonological awareness and word reading and the longitudinal relations between these two variables in kindergarten through second grade, several conclusions can be drawn. First, the children's phonological awareness and word reading skills generally increase throughout the grades studied. A notable result is the substantial growth in phonological awareness at the beginning of first grade, which can be attributed to the start of formal literacy instruction at this time (cf. van Bon & van Leeuwe, 2003). The finding that phonological awareness continues to develop through at least second grade is also consistent with the results of previous studies (e.g., Wagner et al., 1994).

Prior to investigation of the relations between phonological awareness and word reading, the autoregressive effects of the two variables were explored. The results showed both phonological awareness and word reading to be relatively stable underlying abilities. That is, the scores for both phonological awareness and word reading on a particular measurement occasion were found to predict the scores for the same ability measured on the next occasion rather well. The finding of stable individual differences in phonological awareness (Lonigan et al., 2000; Wagner et al., 1997) and word reading (Aarnoutse, van Leeuwe, Voeten, & Oud, 2001; Torgesen & Burgess, 1998) is in accordance with the results of earlier research.

When the longitudinal relations between phonological awareness and word reading were investigated, the results showed phonological awareness to initially predict word reading significantly. However, when the autoregressive effects of word reading were taken into account, subsequent effects of phonological awareness on reading were not significant any more. In contrast, Wagner et al. (1997) found individual differences in phonological awareness to influence subsequent individual differences in word reading for each period they studied (i.e., kindergarten through fourth grade). Nevertheless, the findings of the present study are in keeping with the results of another Dutch study (de Jong & van der Leij, 1999) that also showed the influence of phonological awareness on subsequent reading achievement to be restricted to the first year of formal literacy instruction and thus limited over time. According to these authors, this result can perhaps be explained by the relative transparency of Dutch orthography together with the phonics based instruction that the children receive.

A striking finding in the present study was that in addition to the influence of phonological awareness on subsequent word reading, word reading also influenced subsequent phonological awareness. This finding is consistent with the view that the relations between phonological awareness and reading are bidirectional (Perfetti et al., 1987). However, de Jong and van der Leij did not find such an effect of reading on subsequent phonological awareness in their study that also included Dutch children. The discrepancy in results might be explained by the fact that the measures of phonological awareness used in de Jong and van der Leij study and the present study clearly differed. In the present study, the phonological awareness items measured a broad range of skills that varied in difficulty, including the relatively difficult phoneme segmentation and phoneme deletion skills, which the study by de Jong and van der Leij did not. The finding of an effect of reading on later — more complicated — phonological awareness in the present study and the absence of such an effect in the previous study is thus commensurate with the view that lower levels of phonological awareness can facilitate later reading skills and the acquisition of reading skills can facilitate later — more complicated — phonological awareness (Perfetti et al., 1987; Shankweiler & Fowler, 2004; Stanovich, 1986).

Furthermore, reading skills continued to account for additional variance in phonological awareness even after its autoregressive effects were taken into account while phonological awareness no longer played a role in reading skills after the autoregressive effects of earlier reading skills were taken into account. This finding can perhaps be explained in part by the higher levels of stability for word reading than for phonological awareness; the higher the degree of stability, the less variance to be accounted for by effects other than autoregressive effects. The size of the influence of word reading on subsequent phonological awareness was about a third of the size of the influence of phonological awareness on word reading in first grade, which can also possibly be explained by the inclusion of autoregressive effects in the model. The effect of reading on later phonological awareness was found even after control for the effects of previous scores on phonological awareness whereas the strong influence of phonological awareness on later reading was only found when no autoregressive effects of reading were included. When we changed the model by omitting the autoregressive effects of phonological awareness, the effects of reading on phonological awareness were indeed found to increase substantially for each time period examined.

Another interesting finding is the initial influence of reading in the middle of first grade on later phonological awareness, the disappearance of such an effect at the end of first grade, and then an influence again at the beginning of second grade. This renewed effect might be explained in terms of increased reflection on Dutch orthography as a result of spelling instruction. In Dutch, phoneme-to-grapheme relations are less consistent than grapheme-to-phoneme relations, which causes spelling to be more difficult than reading for Dutch children (Bosman & van Orden, 1997). Starting in second grade, Dutch children receive explicit instruction on important spelling rules (Bosman, de Graaff, & Gijsel, 2006), and it is certainly conceivable that such instruction can stimulate their reading skills and also their phonemic awareness.

In addition to autoregressive effects for phonological awareness and word reading, letter knowledge is another factor that should be taken into account in any investigation of the relations between phonological awareness and reading (e.g., Blaiklock, 2004; de Jong & van der Leij, 1999). The third aim of the present study was thus to examine the impact of children's kindergarten letter knowledge on the relations between phonological awareness and reading. The present results showed the effect of phonological awareness on subsequent word reading to decline after the inclusion of receptive letter knowledge. This finding is in agreement with the results of earlier research (Blaiklock, 2004; Wagner & Torgesen, 1987). The fact that the relation between phonological awareness and reading is largely mediated by the role of letter knowledge supports the common association of phonological awareness and reading with letter knowledge. Indeed, letter knowledge is a critical factor for learning to read (Byrne, 1998) and appears to stimulate phonological awareness by providing children information on the phonological structure of words (Burgess & Lonigan, 1998). Although the effect of phonological awareness on later word reading decreased after letter knowledge was taken into account, phonological awareness continued to have an additional influence on reading and thus remained a significant predictor. The result that both letter knowledge and phonological awareness independently predicted reading is also consistent with the results of the small number of previous studies that took the role of letter knowledge into account when examining the relations between phonological awareness and reading (Elbro et al., 1998; Lonigan et al., 2000).

Despite the importance of the present findings, a few limitations on the present study and recommendations for future research can still be mentioned. First, the results are based on two cohorts, as opposed to one cohort followed from kindergarten through second grade. With the use of the MULTI program, it was possible to estimate a complete covariance matrix, which enabled us to test models for kindergarten through second-grade children. One problem, however, is that the estimated models could be either better or worse approximations of reality than suggested by the fit indices. While the present results provide some valuable indications about the specific relations between phonological awareness and word reading in the early elementary grades, conclusions based upon the results for a single cohort followed longitudinally would be more powerful. It is thus important for future studies to find out whether the present results can be replicated when a single cohort of children is followed from kindergarten through second grade.

A second limitation is that not all plausible causes were included in the models we tested. Some researchers, for example, have argued that vocabulary and

IQ should be controlled for when investigating the relations between phonological awareness and reading. Although previous studies have shown phonological awareness to be related to reading skills even after any variance due to vocabulary and IQ has been taken into account (e.g., Bryant et al., 1990; Lonigan et al., 2000), the present results could nevertheless provide a fuller picture of these relations when the additional variables of vocabulary and IQ were also taken into account. In future research, these variables should thus be considered in addition to the role of letter knowledge and the autoregressive effects of the target variables, as shown by the results of this study.

The present results have important implications for educational practice. First, the finding that both phonological awareness and word reading can be conceived as relatively stable underlying abilities and that these abilities are highly related to each other highlight the importance of early screening and intervention for children possibly at risk for reading problems. Second, future screening instruments should include a measure of letter knowledge in addition to measures of phonological awareness to better predict reading skill. With the application of the knowledge provided by this and other recent studies of the relations between phonological awareness and reading, reading problems in older children and adults can be expected to decrease in the future.

References

Aarnoutse, C., van Leeuwe, J., Voeten, M., & Oud, J. (2001). Development of decoding, reading comprehension, vocabulary and spelling during the elementary school years. *Reading and Writing: An Interdisciplinary Journal, 14,* 61-89.

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Andersen, E. B., & Madsen, M. (1977). Estimating the parameters of the latent population distribution, *Psychometrika, 42,* 357-374.

Anthony, J. L., & Francis, D. J. (2005). Development of phonological awareness. *Current Directions in Psychological Science, 14,* 255-259.

Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology, 96,* 43-55.

Anthony, J. L., Lonigan, C. J., Burgess, S. R., Driscoll, K., Phillips, B. M., & Cantor, B. G. (2002). Structure of preschool phonological sensitivity: Overlapping sensitivity to rhyme, words, syllables, and phonemes. *Journal of Experimental Child Psychology, 82,* 65-92.

Arbuckle, J. L. (2005). *AMOS 6.0 User's guide.* Chicago: SPSS/Erlbaum.

Barlow-Brown, F. & Connelly, V. (2002). The role of letter knowledge and phonological awareness in young Braille readers. *Journal of Research in Reading, 25,* 259-270.

Barron, R.W. (1994). The sound-to-spelling connection: Orthographic activation in auditory word recognition and its implications for the acquisition of phonological awareness and literacy skills. In V.W. Berninger (Ed.), *The varieties of orthographic knowledge 1: Theoretical and developmental issues* (pp. 219-242). Dordrecht: Kluwer Academic.

Blaiklock, K. E. (2004). The importance of letter knowledge in the relationship between phonological awareness and reading. *Journal of Research in Reading, 27,* 36-57.

Bosman, A. M. T., de Graaff, S., & Gijsel, M. A. R. (2006). Double Dutch: The Dutch spelling system and learning to spell in Dutch. In: R. M. Joshi & P. G. Aron (Eds.), *Handbook of orthography and literacy* (pp. 135-150). Mahwah, NJ: Erlbaum.

Bosman, A. M. T., & van Orden, G. C. (1997). Why spelling is more difficult than reading. In C. A. Perfetti, L. Rieben, & M. Fayol, (Eds.), *Learning to spell: Research, theory, and practice across languages* (pp. 173-194). Hillsdale, NJ: Erlbaum.

Bowey, J. A., & Francis, J. (1991). Phonological analysis as a function of age and exposure to reading instruction. *Applied Psycholinguistics, 12,* 91-121.

Bruck, M., & Genesee, F. (1995). Phonological awareness in young second language learners. *Journal of Child Language, 22,* 307-324.

Bryant, P. E., MacLean, M., Bradley, L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology, 26,* 429-438.

Burgess, S. R., & Lonigan, C. J. (1998). Bidirectional relations of phonological sensitivity and prereading abilities: Evidence from a preschool sample. *Journal of Experimental Child Psychology, 70,* 117-141.

Byrne, B. (1998). *The foundation of literacy: The child's acquisition of the alphabetic principle.* Sussex: Psychology Press.

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language, 45,* 751-774.

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91,* 77-111.

de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology, 91,* 450-476.

Ehri, L. C. (1992). Reconceptualizing the development of sight word reading and its relationship to recoding. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 105-143). Hillsdale, NJ: Erlbaum.

Elbro, C., Borstrøm, I., & Petersen, D. K. (1998). Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly, 33,* 36-60.

Gollob, H. F., & Reichardt, C. S. (1987). Taking account of time lags in causal models. *Child Development, 58,* 80-92.

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

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory.* Newbury Park, CA: Sage.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1- 55.

Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002). Phoneme awareness is a better predictor or early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology, 82,* 2-28.

Johnston, R. S., Anderson, M., & Holligan, C. (1996). Knowledge of the alphabet and explicit awareness of phonemes in prereaders: The nature of the relationship. *Reading and Writing: An Interdisciplinary Journal, 8,* 217-234.

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

Juel, C., Griffith, P., & Gough, P. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78,* 243-255.

Kamphuis, F. H. (1993). Estimation and prediction of individual ability in longitudinal studies. In J. H. L. Oud & R. A. W. van Blokland-Vogelesang (Eds.), *Advances in longitudinal and multivariate analysis in the behavioral sciences* (pp. 41-42). Nijmegen: ITS.

Kamphuis, F. H. (1998). *MULTI.* Arnhem: Cito.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Journal of Experimental Child Psychology, 18,* 201-212.

Little, R. J. A. & Rubin, D. B. (1987). *Statistical analysis with missing data.* New York: John Wiley.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent variable longitudinal study. *Developmental Psychology, 36,* 596-613.

Loyd, B. H. (1988). Implications of item response theory for the measurement practitioner. *Applied Measurement in Education, 1,* 135-143.

Lukatela, K., Carello, C., Shankweiler, D., & Liberman, I. Y. (1995). Phonological awareness in illiterates: Observations from Serbo-Croatian. *Applied Psycholinguistics, 16,* 463-487.

Macmillan, B. (2002). Rhyme and reading: a critical review of the research methodology. *Journal of Research in Reading, 25,* 4-42.

McDougall, S., Hulme, C., Ellis, A., & Monk, A. (1994). Learning to read: The role of short-term memory and phonological skills. *Journal of Experimental Psychology, 58,* 112-133.

Mislevy, R. J. (1984). Estimating latent distributions. *Psychometrika, 49,* 359-381.

Morais, J., Alegria, J., & Content, A. (1987). The relationships between segmental analysis and alphabetic literacy: An interactive view. *Cahiers de Psychologie Cognitive, 7,* 415-438.

Noble, K. G., Farah, M. J., & McCandliss, B. D. (2006). Socioeconomic background modulates cognition–achievement relationships in reading. *Cognitive Development, 21,* 349-368.

Perfetti, C. A. (1991). Representations and awareness in the acquisition of reading competence. In L. Rieben & C. A. Perfetti (Eds.), *Learning to read: Basic research and its implications* (pp. 33-44). Hillsdale, NJ: Erlbaum.

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

Read, C., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition, 24,* 31-34.

Sammons, P., & Smees, R. (1998). Measuring pupil progress at key stage 1: Using baseline assessment to investigate value added. *School Leadership & Management, 18,* 389-407.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96,* 265-282.

Schatschneider, C., Francis, D. J., Foorman, B. F., Fletcher, J. M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology, 91,* 467-478.

Schult, R. (1999). *Investigating the social world* (2nd ed.). Pineforge Press.

Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling.* Mahwah, NJ: Erlbaum.

Shankweiler, D., & Fowler, A. E. (2004). Questions people ask about the role of phonological processes in learning to read. *Reading and Writing: An Interdisciplinary Journal, 17,* 483-515.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition, 55,* 151-218.

Snowling, M. J., Gallagher, A., & Frith, U. (2003). Family risk of dyslexia is continuous: Individual differences in the precursors of reading skills. *Child Development, 74,* 358-373.

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

Stevenson, H. W., & Newman, R. S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development, 57,* 646-659.

Torgesen, J. K., & Burgess, S. R. (1998). Consistency of reading related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 161-188). Hillsdale, NJ: Erlbaum.

van Bon, W. H. J., & van Leeuwe, J. F. J. (2003). Assessing phonemic awareness in kindergarten: The case for the phonemic recognition task. *Applied Psycholinguistics, 24,* 195-219.

Vellutino, F. R., & Scanlon, D. M. (1987). Phonological coding, phonological awareness and reading ability: Evidence from a longitudinal and experimental study. *Merrill–Palmer Quarterly, 33,* 321-363.

Verhelst, N. D., Glas, C. A. W., & Verstralen, H. H. F. M. (1995). *OPLM: Computer program and manual.* Arnhem: Cito.

Verhelst, N. D., & Verstralen, H. H. F. M. (2002). *Structural Analysis of a Univariate Latent Variable (SAUL). Theory and a computer program.* Arnhem: Cito.

Vloedgraven, J. M. T., & Verhoeven, L. (2008). The nature of phonological awareness throughout the elementary grades: An item response theory perspective. *Learning and Individual Differences,* doi:10.1016/j.lindif.2008.09.005.

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.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., et al. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33,* 468-479.

Ziegler, J. C., & Goswami, U. C. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131,* 3-29.

6

General discussion

The present thesis reports on the results of four empirical studies on the nature and development of phonological awareness and its relations to literacy throughout the early elementary grades in the Netherlands. The main issues that were the focus of the previous chapters include the nature of phonological awareness, the question how children's phonological awareness abilities develop over the grades, and finally, the relations between phonological awareness and word reading over time. In this chapter, we will provide a summary of the major findings, followed by an integration of these findings into a developmental model of phonological awareness. Subsequently, the limitations of the present thesis will be described. Finally, two examples of how the theoretical findings of the present thesis can be applied in practice will be presented.

6.1 Phonological awareness and learning to read

6.1.1 Nature of phonological awareness

First of all, an attempt was made to arrive at an answer to the question about the nature of phonological awareness. The crucial question in Chapter 2 and 3 was whether the different tasks for phonological awareness reflect a single underlying ability across the different grades studied or separate abilities. We have addressed this issue from an item response theory (IRT) perspective, as suggested by Anthony and Lonigan (2004) because IRT entails less measurement problems than classical test theory (CTT). The present thesis extended previous studies on this issue with regard to the diversity of tasks, and the number of grades involved. Results have provided accumulating evidence for the conceptualization of phonological awareness as a single underlying ability across tasks and grades (i.e., kindergarten through fourth grade). This conclusion is in accordance with the findings of a number of recent other studies (Anthony & Lonigan, 2004; Lonigan, Burgess, & Anthony, 2000; Schatschneider et al., 1999).

A related question concerned the order of acquisition of the various phonological awareness skills over time. This issue was addressed in Chapter 3 as well. Within the unidimensional framework of phonological awareness, the cognitive task requirements for the different tasks appeared to differ. In addition to a considerable overlap between the tasks, results indicate a kind of distinction between two groups of phonological awareness items. Children seem to first master the relatively easy rhyming, phoneme identification, and phoneme blending skills and seem to progress on the more complex phoneme segmentation and phoneme deletion skills later on.

6.1.2 Development of phonological awareness

The next issue to be investigated concerned the question how children's phonological awareness abilities develop over the grades. To gain more insight into this issue, we examined both the developmental pattern of phonological awareness and the stability of phonological awareness abilities over the grades. In the study described in Chapter 4, we examined the developmental pattern of phonological awareness from kindergarten through fourth grade, in which the effects of a number of important other variables were also taken into account. The findings showed the children's phonological awareness abilities to generally increase as they grew older. The results further demonstrated a decreasing rate of development over time; that is, the rate of development was fastest in first grade and decelerated as the children got older. These results resemble those reported by Wagner, Torgesen, and Rashotte (1994), but they did not investigate development beyond second grade. In addition, we found significant main effects of receptive letter knowledge, word reading ability, gender, and SES for phonological awareness but none of these variables appeared to affect the pattern of development for phonological awareness. A quadratic growth curve turned out to describe the development of phonological awareness for a wide variety of children. An important finding of this study was that growth in phonological awareness was no longer significant after the end of third grade. Given that the developmental pattern was similar for children of different reading ability levels, both the poor and average readers are expected to achieve a lower level of phonological awareness than the good readers in the end. This means that the poor and average readers after third grade no longer continue to progress in phonological awareness, while their development has not yet reached maturation. These results provide support for the phonological deficit model of reading disability (e.g., Metsala, Stanovich, & Brown, 1998; Snowling, 2000; Stanovich & Siegel, 1994).

The result that different reading ability groups show dissimilar levels of phonological awareness but similar patterns of phonological awareness development seems to indicate that phonological awareness is a rather stable ability. However, the stability of phonological awareness can better be investigated by monitoring children's phonological awareness abilities within a longitudinal design. In Chapter 5, a longitudinal model was presented addressing the development of phonological awareness from kindergarten through second grade. Results showed that scores for phonological awareness on a particular measurement occasion can predict the performance on the same ability on the next measurement occasion rather well for each interval examined, which provides evidence for phonological awareness as a stable underlying ability. This stability was also demonstrated in other studies (Lonigan et al., 2000; Wagner et al., 1997).

6.1.3 Relations between phonological awareness and reading

Another focal issue in the present thesis concerned the longitudinal relations between phonological awareness and word reading during the early elementary grades. This issue was fully addressed in Chapter 5. The results demonstrated phonological awareness to have a significant influence on reading initially, but once the autoregressive effects of reading were taken into account, subsequent effects of phonological awareness on word reading were not significant any more. In addition, reading appeared to exert significant effects on phonological awareness. These findings support the view that the relations between phonological awareness and reading are bidirectional (Perfetti et al., 1987; Shankweiler & Fowler, 2004).

Furthermore, the early effect of phonological awareness on children's subsequent reading declined when the children's receptive letter knowledge was also included in the model. Nevertheless, both phonological awareness and the children's letter knowledge remained unique predictors of the children's reading development which is in agreement with the outcomes of other studies (Elbro, Børstrom, & Petersen, 1998; Lonigan et al., 2000).

For the prediction of reading skills, several researchers have investigated the effects of different phonological awareness tasks on reading. The finding that different phonological awareness skills have different predictive values for reading is often conceived as evidence for multidimensionality. A number of researchers found that rhyming and phoneme tasks correlate differently with literacy measures and argue therefore, that these tasks reflect separate abilities. Hulme et al. (2002), for example, demonstrated phoneme awareness to be a better predictor of reading, while in contrast, Bryant, MacLean, Bradley, and Crossland (1990) found rhyme awareness most important. Anthony and Lonigan (2004) argue that the inconsistent findings about the different predictive values of the different skills show that the relative superiority of certain skills is unreliable. Measurement artefacts, like ceiling effects, are expected to obscure the relationship between phonological awareness and reading (Anthony & Lonigan, 2004; Ziegler & Goswami, 2005). Given that the present thesis showed phonological awareness to be a unidimensional ability, it was possible to address the issue of prediction from another perspective. That is, this result enabled us to investigate the predictive value of the phonological awareness measure broadly defined (i.e., including items that assessed a broad range of phonological awareness skills). Lonigan et al. (2000) were the first to explicitly investigate this issue, and the results showed the broader construct of phonological awareness to be a strong predictor of children's reading skills. While Lonigan et al. tested English-speaking children, we tested Dutch children and in addition, we enlarged the sample size per grade. The fact that we arrived at the same conclusion can be seen as a strong cross-linguistic support for the claim that the global construct of phonological awareness is an important predictor of children's subsequent reading skills.

6.1.4 Towards a developmental conceptualization of phonological awareness

The present thesis provided convincing evidence that the different tasks for phonological awareness are all manifestations of a single underlying ability in the different grades studied. Furthermore, in addition to a substantial overlap between the tasks, results indicate a kind of separation between rhyming, phoneme identification, and phoneme blending, on the one hand, and phoneme segmentation and phoneme deletion, on the other hand. These findings seem to be fully commensurate with Ziegler and Goswami's (2005) grain size theory of reading acquisition. According to this theory, the development of reading depends upon the precision of the underlying phonological representations of words. With respect to phonological awareness, it can similarly be assumed that children acquire a partial access to such representations in kindergarten, which results in progress on such phonological awareness skills as rhyming, phoneme identification, and phoneme blending. In first grade, when the children receive formal literacy education, they will be able to gain full access to phonemes as the teaching of the orthography may facilitate the identification of phonemes and progress will then be demonstrated with respect to the more complex skills of phoneme segmentation and phoneme deletion. This view is also consistent with the suggestion that phonological awareness is a consequence of learning to read (Morais, Bertelson, Cary, & Alegria, 1986; Perfetti et al., 1987) and that most phonological awareness tasks can develop in the absence of reading skills while segmentation requires reading acquisition for its development (van Bon & van Leeuwe, 2003). Thus, within the unidimensional framework of phonological awareness, the development of phonological awareness seems to proceed from partial towards full access to phonemes with the phonological representations becoming more and more specific.

The research findings altogether seem to be largely in agreement with the developmental model of phonological awareness. According to this model, phonological awareness can be viewed as a single ability that develops continuously across the school years (Anthony, Lonigan, Driscoll, Phillips, & Burgess, 2003; Stanovich, 1992). Anthony and colleagues assert that there are three major predictions for the developmental model of phonological awareness. The first prediction concerns the unidimensionality of phonological awareness as measured by different tasks which has been sufficiently demonstrated in the present thesis. The second prediction concerns the order of acquisition of the different phonological awareness skills. Our results showed a kind of distinction between two groups of items in addition to continuous, quasi-parallel development as reported by Anthony et al. (2003). The third test for the developmental conceptualization of phonological awareness was provided by the longitudinal study that showed phonological awareness to be a stable ability over time. In addition, this study demonstrated the broader construct of phonological awareness to be a significant predictor of children's reading skills. Summarized, the results of the four studies together lend support to the conceptualization of phonological awareness as a unitary construct that develops along a continuum of availability for phonological representations which can range from partial to full access.

6.2 Limitations and future research directions

In the present thesis, we made an attempt to overcome some limitations of other studies. With respect to the nature of phonological awareness, Schatschneider et al. (1999) were the first who investigated this issue within the framework of IRT. We extended this research by also administering a rhyming task and by also measuring third and fourth graders. Furthermore, we attempted to overcome some methodological limitations of previous studies. For example, in the present study, a large sample of children was measured at several occasions in order to investigate bidirectional relationships. In addition, we included the most important other plausible causes as the autoregressive effects of reading and letter knowledge in our analyses to minimize the possibility that the relation between phonological awareness and word reading is spurious. Of course, the present research has also some limitations that need to be mentioned.

An important question of the present thesis was whether phonological awareness can be conceived as a unidimensional ability or consists of separate abilities. We mainly investigated the nature of phonological awareness within the framework of IRT. While the application of IRT has many advantages and is also regularly used for dimensionality assessment, one can also criticize this choice, because the IRT model we used in this thesis a priori assumes unidimensionality. We have partly met such objections by also using other techniques for the assessment of the dimensionality. To examine the nature of phonological awareness, a two-factor analysis using MINRES was performed and modified parallel analyses were conducted. However, these methods have their own limitations. For example, both methods are based on tetrachoric correlations, which have relatively large standard errors (Brown, 1977). Another approach to disentagle the issue of dimensionality is the use of confirmatory factor analysis (CFA). In CFA, it is tested whether an a priori hypothesis about the number of factors, based on theoretical models, can be supported by the data. Anthony and Lonigan (2004) analyzed the data of four different studies by means of CFA to examine the nature of phonological awareness. Findings from these studies converged on the view that phonological awareness can best be conceptualized as a unidimensional ability. While the interpretation of factors is unambiguous in CFA, given that the different models can be statistically compared, this approach has also some limitations. Anthony and Lonigan comment on CFA studies that results can be influenced by measurement artefacts, like floor and ceiling effects and differential reliabilities. Another limitation of this approach is the strong reliance on fit indices, which can result in incorrect conclusions (Marsh, Hau, & Wen, 2004). For future studies, Anthony and Lonigan have recommended the application of IRT to address this issue, because it is less influenced by measurement problems. The present thesis has addressed the dimensionality of phonological awareness from an IRT perspective and results again demonstrated the different tasks for phonological awareness to be manifestations of the same underlying ability. However, it is still better to examine whether similar conclusions can be drawn when both methods are used on the same data sets. Future research should investigate the nature of phonological awareness by using CFA as well as IRT to definitively solve this issue.

A general problem in comparing the results of different studies on phonological awareness is that a wide variety of tasks have been used to measure the construct of phonological awareness. In addition, even within tasks there is much variability. For example, the difficulty of items within tasks is influenced by CV structure, articulatory factors, and position of phonemes. Every researcher uses different tasks and items to measure phonological awareness, thus the results should always be interpreted within the context of the specific tasks and items administered. Bearing this in mind, we have carefully selected the tasks for phonological awareness on the basis of relevant research literature in order to measure a broad range of phonological awareness skills (see Chapter 1 for a rationale for the inclusion of each task). Furthermore, we have attempted to take care of a proportional distribution of the different CV structures and articulation manners within each task as much as possible. Additional research within IRT is needed to find out whether the present findings can be replicated when phonological awareness is measured by other tasks and items. Moreover, validity research for the measurement of phonological awareness is called for. Another interesting issue to investigate is whether the same results will be found if the speed of the phonological awareness tasks is taken into account.

Another limitation concerns the design of the present study. The studies that were described in Chapter 2, 3, and 4 were based on a cross-sectional design. In Chapter 5, the results of a longitudinal study on the relations between phonological awareness and word reading were presented. However, this study was not completely longitudinal, given that the development from kindergarten through second grade was investigated, while the children were monitored for two school years. Although not all of the cells in the covariance matrix were really observed, it was nevertheless possible to estimate a complete covariance matrix by using the software MULTI (Kamphuis, 1998). The problem, however, is that by the estimation of some covariances, it is possible that the estimated models are a less good or better approximation to reality than suggested by the fit indices. Yet, there is no reason to cast doubt on the present results, although a completely longitudinal design is always preferable.

Summarizing, it may be stated that converging evidence has been provided for the developmental conceptualization of phonological awareness, but still more research, especially longitudinal, is needed to consolidate this conclusion.

6.3 Measurement of phonological awareness: From theory to practice

The present thesis has shown the different tasks for phonological awareness to reflect the same underlying ability across the different grades. It was also demonstrated that the suitability of a specific task for the measurement of phonological awareness depends upon the child's ability level. The utility of the various tasks appeared to differ across the different grade levels. The question is thus not which phonological awareness task is most important but which tasks can be administered at which points in a child's development in order to obtain the most accurate measures of children's phonological awareness possible. This means that children's phonological awareness would be best assessed by different tasks that measure different phonological awareness skills and that this broader construct, instead of the specific tasks for phonological awareness, should be used for screening purposes. An interesting possibility for the measurement of children's phonological awareness is the use of computerized adaptive tests (CATs), in which each child takes a unique test that is adapted to the child's ability level (e.g., Straetmans & Eggen, 1998). In the next section, we will tell more about adaptive tests in general, and provide an example of how the phonological awareness measures can be administered adaptively, based on the theoretical findings of the present thesis.

Furthermore, the results of the present study showed both phonological awareness and receptive letter knowledge to be unique predictors of reading in first grade. An important question is how these findings can be applied in practice. It seems a good idea to include measures for both phonological awareness and letter knowledge in a screening instrument for the early identification of children with reading problems. An example will be provided of how a combination of these two variables can contribute to the early screening of reading problems.

6.3.1 Computerized adaptive tests

The finding that the different phonological awareness tasks are all manifestations of the same underlying ability resulted in an IRT calibrated item bank for phonological awareness. All of the items have certain characteristics, such as the content (i.e., rhyming, phoneme identification, phoneme blending, phoneme segmentation, and phoneme deletion) and in addition, some psychometric characteristics (i.e., difficulty and discrimination). The availability of an IRT calibrated item bank offers possibilities for CATs (for an overview, see Wainer, 2000). The reason that IRT is excellently appropriate for application in CATs is that it places the ability and the difficulty of the items on the same scale. This makes it possible to tailor a test to the examinee's ability level. Furthermore, abilities can be estimated using each set of items of the calibrated item bank and scores on different tasks can be compared without any problems. In IRT, the concept of item information function is available for the selection of optimal items (e.g., Eggen & Verschoor, 2006; Straetmans & Eggen, 1998).

Although adaptive testing is more expensive and time-consuming to conduct compared to paper-and-pencil tests, the value of adaptive tests for practice is generally acknowledged. Main advantage of adaptive testing is the efficiency of measurement, given that items are selected that provide the most information about the person's ability level. An adaptive test can be shortened by 50%, while keeping the same measurement precision as a fixed test (Weiss & Kingsbury, 1984). For the purpose of large-scale screening in education, shorter testing times have the preference. A second important advantage of adaptive tests is that examinees will be tested at their own ability level. That is, each person receives items that are challenging without being too difficult or too easy. Still, another advantage is that learning effects are largely prevented due to the fact that at each measurement occasion different items are administered. Furthermore, like any computer-based test, adaptive tests can provide the scores immediately after the test administration and are thus very convenient for educators.

Before we discuss the application of adaptive testing for the measurement of phonological awareness, we will first describe how CATs generally work. The standard computer-adaptive testing procedure is an iterative algorithm (Eggen, 2004), which is visually presented in Figure 6.1. At the start, if information about the examinee's ability is available, it could be used. Otherwise, the CAT will assume that the ability of the person is average, which results in the administration of an item of medium difficulty. After each administered item, the following three steps are taken. First, the item bank is searched for the best item to administer, given the current estimate of the examinee's ability level. The most applied method for item selection is the maximum information approach (Lord, 1970), in which the item that provides the most information at a certain ability level will be selected. This item with the greatest information is an item whose difficulty is closely adapted to the examinee's ability level and that can discriminate well between persons with ability levels around the target level. As a second step, the selected item is administered and the examinee gives a correct or incorrect answer. Third, the new ability estimate is computed, based on all of the previous responses. This iterative process will continue until a stopping criterion has been met. Often, the test is stopped when the standard error of measurement falls below a previously assessed value, which guarantees the accuracy of the ability estimates.

It has been shown that maximum information item selection in CATs results in a 50% probability of answering the items correctly for each individual (Eggen & Verschoor, 2006). In other words, children will answer about half of the items incorrectly. It is possible that CATs are perceived as very difficult, especially by young children, which might lead to fear of failure and as a consequence, possible lower achievements. For young children, it seems a good idea to select items with a higher success probability. Eggen and Verschoor showed that item selection aimed at a 60 or 70% success probability is possible without a great loss in measurement precision which entails that CATs can thus be greatly adapted to young children.

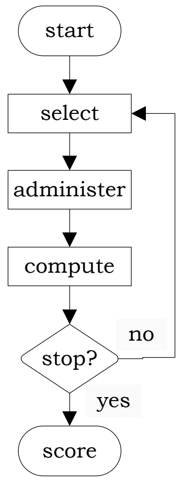

Figure 6.1 *Iterative algorithm in the basic computer-adaptive testing procedure*

Clearly, the potential of adaptive testing is great, though in modern CAT applications many constraints are enforced upon the selection of items. These constraints mainly concern the content of the items to be administered and the item exposure. For example, a test designer can specify that some aspects of the target ability have to occur in a certain proportion (Kingsbury & Zara, 1991). In addition, adaptive tests often have item exposure constraints to prevent over- and underexposure of particular items. With respect to the measurement of phonological awareness, some additional practical issues complicate the construction of such an adaptive test. The different tasks used to assess phonological awareness (e.g., blending and segmentation) have their own instructions, for instance. As a result, the items from the different tasks must be administered consecutively: the items from one task with its own instructions must first be administered followed by the items from a different task with its own instructions.

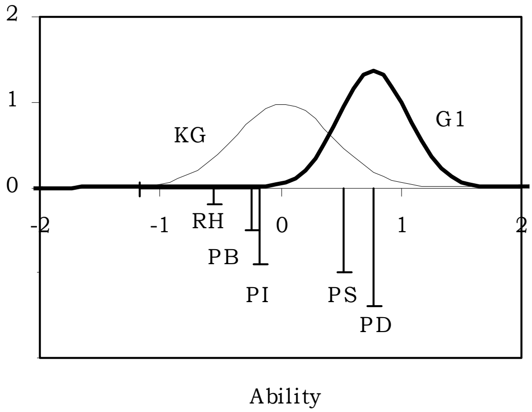

Figure 6.2 *Mean difficulty of the different tasks in relation to ability*

Then the question of the order of administration of the different tasks can be raised, given that the findings of the present thesis have shown the five tasks to differ in difficulty. Results have indicated a distinction between two groups of items showing rhyming (RH), phoneme identification (PI), and phoneme blending (PB) skills to be first acquired, followed by the acquisition of phoneme segmentation (PS) and phoneme deletion (PD) skills. In Figure 6.2, the ability distributions of kindergartners (KG) and first graders (G1) are presented on the y-axis and the mean difficulty of each of the five tasks is put on the x-axis (i.e., ability scale). Do these findings and the results of the information functions (see Chapter 2 and 3) provide us with recommendations for ongoing practice? The difficulty of the items measuring rhyming, phoneme identification, and phoneme blending appears to fit the ability of the lower and average scoring kindergartners. It could be recommended to successively administer a set of rhyming items, blending items, and phoneme identification items to kindergartners. Within each task, items should be administered according to the CAT algorithm. An interesting idea is to administer the tasks themselves in an adaptive way as well. For example, it is recommendable to only present the kindergartner with the segmentation and deletion items if the child obtains a score that falls above a certain value, so that only the higher scoring kindergartners receive the relatively difficult segmentation and deletion items. For the next measurement occasion, the last estimate of the child's ability could determine whether the items of the three relatively easy tasks or the items of the two relatively difficult tasks will be administered. That is, children with lower phonological awareness abilities will be presented with the easy items, and vice versa. Of course, if the child obtains a score above a certain criterion on the easy items, the items of the difficult tasks can still be administered to accurately assess a child's ability. A possibility is to stop the test if the child gets a score above a specified criterion (i.e., gain a sufficient level of phonological awareness) rather than at a specific age or in case of an accurate estimate of ability.

It should be noted that the application of the CAT algorithm for the segmentation items is complicated, given that this task requires an oral response and the answers can thus not be automatically scored. To be able to present the segmentation items in an adaptive way, someone is needed to put responses into the computer immediately after each item has been administered.

6.3.2 Early screening of reading problems

For the early screening of reading problems, even more important than the prediction of reading abilities in general is the prediction of whether a child will develop reading difficulties or not (see also Gijsel, Bosman, & Verhoeven, 2006). Therefore, we have investigated the discriminative power of both phonological awareness and receptive letter knowledge. Can these two variables, as measured in kindergarten, predict which children will fail in reading and which will be typical readers in first grade? To answer this question, the number of valid positives and valid negatives versus false positives and false negatives can be computed. The valid positive rate reflects the number of children who were identified as at risk by the screening measure and indeed became poor readers. The false positive rate refers to the number of children who were identified as at risk but who appeared to be typical readers. The valid negative rate reflects the number of children who were identified as not risk and indeed became typical readers. In contrast, the false negative rate refers to the number of children who were identified as not at risk but who appeared to develop reading problems later. To establish the value of a screening instrument, the sensitivity and specificity indices are often calculated. The sensitivity index (i.e., number of valid positives / (valid positives + false negatives)) reflects the ability of the instrument to correctly identify poor readers. The specificity index (i.e., number of valid negatives / (valid negatives + false positives)) reflects the ability of the instrument to correctly identify typical readers. For screening purposes, it has been suggested that the desired order of magnitude of these indices is .75 (Gredler, 2000; Hammill, Mather, Allen, & Roberts, 2002).

To establish the discriminative power of the tasks that were used in the study described in Chapter 5, for each task (i.e., phonological awareness, receptive letter knowledge, and word reading), children were divided in two groups. Children were classified as having poor abilities (i.e., scoring below the 25th percentile – *P*25) or typical abilities (i.e., scoring above the 25th percentile). Phonological awareness (PA) and receptive letter knowledge (RLK) were measured at the end of kindergarten and reading in the middle of first grade. First, we investigated whether phonological awareness can adequately predict which children will fail in reading and which will be typical readers. As a next step, the discriminative power of receptive letter knowledge was examined. The relevant statistics are presented in Table 6.1.

Table 6.1
Proportions of valid and false positives and negatives, and sensitivity and specificity indices for phonological awareness and receptive letter knowledge

At-risk criterion	Valid positives	False positives	Valid negatives	False negatives	Sensitivity index	Specificity index
PA-*P*25	.54	.46	.81	.19	.47	.85
RLK-*P*25	.53	.47	.81	.19	.45	.85
PA or RLK-*P*25	.47	.53	.83	.17	.57	.76
PA or RLK-*P*10 or PA and RLK-*P*30	.55	.45	.91	.09	.67	.86

For both variables, the specificity index appears to be higher than the sensitivity index, which means that these variables can better predict which children will be typical readers than which children will become poor readers. However, for the early screening of reading problems and for early intervention programs, it is more important to predict which children will develop reading problems. As a next step, in an attempt to increase the ability to correctly identify at-risk children, we have combined the scores on phonological awareness and letter knowledge. Children with a score below the 25th percentile on one of the two tasks were categorized as at risk. As a result, the sensitivity index appeared to increase to .57, which is still too low. Moreover, as a side effect, the number of false positives also increased (see Table 6.1). Next, statistics were computed for the following “at-risk criterion”: scores below the 10th percentile on one of the two tasks or scores below the 30th percentile on both tasks. The sensitivity index increased to .67 and the specificity index to .86 (see Table 6.1). The sensitivity index is still lower than .75, but it might be expected that this value will further increase when the phonological awareness measures are administered adaptively. The number of false positives (false alarm) is rather high, but it is difficult to judge about this proportion, given that no information is available about the number of children who already received an intervention in kindergarten. When some of the at-risk children indeed received an intervention in kindergarten, it might be expected that some of these children will become typical readers in first grade (i.e., false positive). For screening purposes, this combination criterion seems a good possibility, given that the ability to correctly identify children with reading problems increases with about 20% compared to only one predictor as criterion. Almost 70% of the poor readers in first grade can be already predicted in kindergarten.

To conclude, the present thesis has provided some important implications for educational practice. The results yielded valuable information about which tasks can be administered at which moments in a child's development and have set the stage for the adaptive testing of phonological awareness, which is a very convenient possibility for educators. The present thesis has thus demonstrated how individual differences in phonological awareness can be adequately measured throughout the early elementary grades.

Furthermore, it has been shown that it is important to include both a measure for phonological awareness and for letter knowledge in a screening instrument for the early identification of children with reading problems. As a next step, the children who are identified as at risk may receive early interventions before these children develop reading problems. According to the Response to Intervention (RTI) model, those who not respond to intervention, have more severe reading problems and need more intensive instruction (e.g., Klingner & Edwards, 2006). This means that it is very important to monitor children's progress in early literacy skills over time. Given that phonological awareness can be conceived as a unidimensional ability across tasks and grades, children's phonological awareness abilities can be adequately monitored throughout the school years. In this way, the present findings have offered new possibilities for the screening of children who are at risk for reading problems.

References

Anthony, J. L., & Lonigan, C. J. (2004). The nature of phonological awareness: Converging evidence from four studies of preschool and early grade school children. *Journal of Educational Psychology, 96,* 43-55.

Anthony, J. L., Lonigan, C. J., Driscoll, K., Phillips, B. M., & Burgess, S. R. (2003). Phonological sensitivity: A quasi-parallel progression of word structure units and cognitive operations. *Reading Research Quarterly, 38,* 470-487.

Brown, M. B. (1977). Algorithm AS 116: The tetrachoric correlation and its standard error. *Applied Statistics, 26,* 343-351.

Bryant, P. E., MacLean, M., Bradley, L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology, 26,* 429-438.

Carroll, J. M., Snowling, M. J., Hulme, C., & Stevenson, J. (2003). The development of phonological awareness in preschool children. *Developmental Psychology, 39,* 913-923.

Eggen, T. J. H. M. (2004). *Contributions to the theory and practice of computerized adaptive testing.* Twente University, Enschede.

Eggen, T. J. H. M., & Verschoor, A. J. (2006). Optimal testing with easy or difficult items in computerized adaptive testing. *Applied Psychological Measurement, 30,* 379-393.

Elbro, C., Borstrøm, I., & Petersen, D. K. (1998). Predicting dyslexia from kindergarten: The importance of distinctness of phonological representations of lexical items. *Reading Research Quarterly, 33,* 36-60.

Gijsel, M. A. R., Bosman, A .M. T., & Verhoeven, L. (2006). Kindergarten risk factors, cognitive factors, and teacher judgments as predictors of early reading in Dutch. *Journal of Learning Disabilities, 39,* 558-571.

Gredler, G. R. (2000). Early childhood screening for developmental and educational problems. In B. A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (pp. 399-411). Boston: Allyn & Bacon.

Hammill, D. D., Mather, N., Allen, E. A., & Roberts, R. (2002). Using semantics, grammar, phonology, and rapid naming tasks to predict word identification. *Journal of Learning Disabilities, 35,* 121-136.

Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002). Phoneme awareness is a better predictor or early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology, 82,* 2-28.

Kamphuis, F. (1998). *MULTI.* Arnhem: Cito.

Kingsbury, G. G., & Zara, A. R. (1991). A comparison of procedures for content-sensitive item selection in computerized adaptive tests. *Applied Measurement in Education, 4,* 241-261.

Klingner, J. K., & Edwards, P. (2006). Cultural considerations with response to intervention models. *Reading Research Quarterly, 41,* 108-117.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent variable longitudinal study. *Developmental Psychology, 36,* 596-613.

Lord, M. F. (1970). Some test theory for tailored testing. In W. H. Holzman (Ed.), *Computer assisted instruction, testing, and guidance* (pp. 139-183). New York: Harper and Row.

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's findings. *Structural Equation Modeling, 11,* 320-341.

Metsala, J. L., Stanovich, K. E., & Brown, G. D. A. (1998). Regularity effects and the phonological deficit model of reading disabilities: A meta-analytic review. *Journal of Educational Psychology, 90,* 279-293.

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24,* 45-64.

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

Schatschneider, C., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology, 91,* 467-478.

Shankweiler, D., & Fowler, A. E. (2004). Questions people ask about the role of phonological processes in learning to read. *Reading and Writing: An Interdisciplinary Journal, 17,* 483-515.

Snowling, M. J. (2000). *Dyslexia.* Oxford: Blackwell.

Stanovich, K. E. (1992). Speculations on the causes and consequences of individual differences in early reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 307-342). Hillsdale, NJ: Erlbaum.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86,* 24-53.

Straetmans, G. J. J. M., & Eggen, T. J. H. M. (1998). Computerized adaptive testing: What it is and how it works. *Educational Technology, 38,* 1, 45-52.

van Bon, W. H. J., & van Leeuwe, J. F. J. (2003). Assessing phonemic awareness in kindergarten: The case for the phonemic recognition task. *Applied Psycholinguistics, 24,* 195-219.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., et al. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33,* 468-479.

Wainer, H. (Ed.). (2000). *Computerized adaptive testing: A primer* (2nd ed.). Hillsdale, NJ: Erlbaum.

Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21,* 361-375.

Ziegler, J. C., & Goswami, U. C. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131,* 3-29.

Summary

In this thesis, the nature and development of phonological awareness and its relations to literacy were studied throughout the early elementary school grades within the Dutch context. Phonological awareness refers to access to and an understanding of the sound structure of oral language. Phonological awareness is known to be the critical factor in reading development. Growth in phonological awareness has also been found to account for variance in reading beyond that accounted for by the actual level of phonological awareness. Numerous studies have investigated the relations between phonological awareness and reading, but the exact nature of phonological awareness has yet to receive sufficient attention despite its importance for early literacy development. In the present research, data on the development of phonological awareness in relation to reading skills were analyzed from the perspective of item response theory (IRT) as this approach is exceptionally suited for the monitoring of abilities over time.

In Chapter 2, the possibilities for the measurement of growth in phonological awareness were explored in a pilot study. The phonological awareness of 172 kindergartners and 173 first graders was assessed using four tasks: rhyming, phoneme identification, phoneme blending, and phoneme segmentation. The results showed the different tasks to reflect a single underlying ability in the kindergarten and first grades. The tasks appeared to clearly differ in difficulty; rhyming turned out to be the easiest task and phoneme segmentation the most difficult task with phoneme blending and phoneme identification occurring in between. In addition, strong growth in phonological awareness was observed from kindergarten to first grade. The different measures of phonological awareness could thus be placed along a single ability scale and clearly reflect any changes in this ability. In other words, it is possible to monitor the development of Dutch children's phonological awareness in the early elementary grades using the selected measures.

The nature of phonological awareness was further examined in greater detail in the study reported on in Chapter 3. In an extension of the pilot study with regard to the number of tasks, the number of grades involved, and the use of larger samples per grade, a phoneme deletion task was added to the other four tasks. A total of 1405 kindergarten through fourth-grade children was then studied. The results again showed phonological awareness to be unidimensional across tasks and grades. In a next step, the relative difficulty of the various items used to measure phonological awareness was investigated. In addition to some overlap, the

items measuring rhyming, phoneme identification, and phoneme blending were found to be easier than the items measuring phoneme segmentation and phoneme deletion. One problem frequently encountered with the measurement of phonological awareness is the inaccuracy of the measurement, which may be due to the use of a task not suited to a particular child's level of ability. The results of the study reported on in Chapter 3 provide insight into which measures may be most useful at which levels of ability and clearly show the accuracy of the various tasks to decrease as the children's abilities increase with the most accurate estimates of ability found for kindergarten and first grade. In grades two, three, and four, only the abilities of the lower scoring children were accurately estimated. In kindergarten, the rhyming, phoneme identification, phoneme blending, and phoneme segmentation tasks all appeared to accurately measure the children's phonological awareness. In the higher grades, in contrast, primarily the phoneme segmentation and phoneme deletion tasks were found to be informative.

In a shift of focus from the tasks used to assess phonological awareness to the children that were tested, the development of children's phonological awareness from kindergarten through fourth grade was considered in Chapter 4. The study described in the previous chapter provided an IRT calibrated scale for phonological awareness, which constituted the input for the analyses in this study. A declining rate of development for phonological awareness was detected; that is, the rate of growth was fastest in first grade and decelerated as the children got older with no significant progress after the end of third grade. The results further showed significant main effects of receptive letter knowledge, word reading ability, gender, and SES on the children's phonological awareness. Children with greater letter knowledge and higher reading skills performed better on phonological awareness tasks than children with less letter knowledge and lower reading skills. Furthermore, girls produced higher phonological awareness scores than boys and children from a higher SES background outperformed children from a lower SES background. However, none of these variables appeared to affect the developmental pattern for phonological awareness. That is, a quadratic growth curve best characterized the development of phonological awareness for a wide variety of children.

Starting from the assumption that phonological awareness can be conceptualized as a unitary construct, the relations between phonological awareness broadly defined (i.e., measured using a number of different tasks) and word reading were next investigated longitudinally for children in kindergarten through second grade. A first finding from this study described in Chapter 5 was that both phonological awareness and word reading appear to be quite stable abilities in the early elementary grades. The results further showed phonological awareness to initially predict reading significantly but once the autoregressive effects of reading were taken into account, subsequent effects of phonological awareness on word reading were not significant any more. Furthermore, children's reading abilities were found to influence their later phonological awareness, which supports the view that the relations between phonological awareness and reading are bidirectional. In addition, letter knowledge turned out to be another important predictor of reading. Both phonological awareness and letter knowledge were found to be unique predictors of reading performance.

Finally, in Chapter 6, some general conclusions were presented and integrated into a developmental model of phonological awareness and in addition, some practical applications were presented. The findings of the four studies considered together lend support to the conceptualization of phonological awareness as a continuum of availability for phonological representations — a continuum that can range from partial to full phoneme access. For educational practice, the availability of an IRT calibrated item bank offers excellent opportunities for adaptive testing (i.e., testing tailored to the individual child's level of ability). For the early screening for reading problems, use of a combination of phonological awareness and receptive letter knowledge tasks is recommended to determine which kindergarten children are at a risk for first-grade reading failure. For screening purposes, it is also important that the development of children's early literacy skills be adequately monitored as those who do not respond well to early intervention may have more severe reading problems. The results of the present research provided converging evidence that phonological awareness can be conceptualized as a unidimensional ability across tasks and grades, which provides new possibilities for the assessment of the development of phonological awareness over time.

Samenvatting

In dit proefschrift wordt verslag gedaan van een onderzoek naar de ontwikkeling van het fonologisch bewustzijn in relatie tot geletterdheid bij kinderen in Nederland. Fonologisch bewustzijn verwijst naar de toegang tot en het begrip van de klankstructuur van gesproken taal. Het is algemeen bekend dat fonologisch bewustzijn de belangrijkste factor is die ten grondslag ligt aan de ontwikkeling van lezen. Ook is er in voorgaande studies bewijs gevonden voor het feit dat groei in fonologisch bewustzijn kan zorgen voor extra verklaarde variantie in lezen boven het feitelijke niveau van fonologisch bewustzijn. Talrijke studies hebben de relaties tussen het fonologisch bewustzijn en lezen onderzocht, terwijl de onderliggende structuur, oftewel, de dimensionaliteit van het fonologisch bewustzijn, veel minder aandacht heeft gekregen. Gedetailleerde kennis over de onderliggende structuur is echter nodig om de ontwikkeling van het fonologisch bewustzijn en de relaties met lezen beter te kunnen onderzoeken. De data van het huidige onderzoek zijn geanalyseerd binnen het raamwerk van de item respons theorie (IRT), aangezien deze benadering buitengewoon geschikt is voor het volgen van vaardigheden in de tijd.

In hoofdstuk 2 wordt verslag gedaan van een pilotstudie, waarin de mogelijkheden voor het meten van groei in fonologisch bewustzijn worden nagegaan. Het fonologisch bewustzijn werd gemeten bij 172 kinderen in groep 2 en 173 kinderen in groep 3 door de volgende 4 taken: rijm, identificatie van beginfoneem, synthese en analyse. De resultaten lieten zien dat de verschillende taken uitingen waren van dezelfde onderliggende vaardigheid. Verder verschilden de taken in moeilijkheid; rijm bleek de gemakkelijkste taak te zijn, gevolgd door synthese en identificatie van beginfoneem. Analyse bleek het moeilijkst te zijn. Daarnaast heeft dit onderzoek een sterke groei in fonologisch bewustzijn van groep 2 naar groep 3 laten zien. De verschillende taken voor fonologisch bewustzijn konden dus op één onderliggende schaal geplaatst worden en bleken in staat te zijn veranderingen in deze vaardigheid te meten. Met andere woorden, het is met behulp van de gekozen taken mogelijk de ontwikkeling van het fonologisch bewustzijn bij Nederlandse kinderen te volgen gedurende de beginjaren van de basisschool.

De onderliggende structuur van het fonologisch bewustzijn is nader onderzocht in een volgende studie. Deze studie, die beschreven is in hoofdstuk 3, kan worden gezien als een uitbreiding van de pilotstudie met betrekking tot het aantal taken, het aantal jaargroepen en de grootte van de steekproef per jaargroep.

Een taak voor deletie werd toegevoegd aan de overige 4 taken en deze taken werden afgenomen bij 1405 kinderen uit de groepen 2 tot en met 6. Opnieuw werd aangetoond dat fonologisch bewustzijn opgevat kan worden als een eendimensionaal construct wat betreft de verschillende taken gemeten in de verschillende groepen. Vervolgens is de relatieve moeilijkheid van alle items voor fonologisch bewustzijn onderzocht. Naast wat overlap bleken de items bedoeld voor het meten van rijm, identificatie van beginfoneem en synthese eenvoudiger dan de items voor analyse en deletie. Een veelvoorkomend probleem bij het meten van het fonologisch bewustzijn is de onnauwkeurigheid van een meting, wat verklaard kan worden doordat mogelijk een taak gebruikt wordt die niet past bij het vaardigheidsniveau van een bepaalde leerling. Resultaten van deze studie hebben duidelijk laten zien welke taken het beste bij welke vaardigheidsniveaus afgenomen kunnen worden. Daarnaast is aangetoond dat de nauwkeurigheid van de verschillende taken afneemt naarmate de vaardigheid van de kinderen toeneemt. De vaardigheidsschattingen bleken het meest nauwkeurig in de groepen 2 en 3. In de groepen 4, 5 en 6 kon alleen de vaardigheid van de lager scorende kinderen nog nauwkeurig gemeten worden. Aangetoond werd dat in groep 2 de taken voor rijm, identificatie van beginfoneem, synthese en analyse in staat waren het fonologisch bewustzijn van kinderen nauwkeurig te meten. In de hogere groepen, echter, bleken vooral de analyse- en deletietaak informatief te zijn.

Terwijl de studies beschreven in de hoofdstukken 2 en 3 vooral gericht waren op de taken voor het fonologisch bewustzijn, zijn de studies waarvan in de hoofdstukken 4 en 5 verslag wordt gedaan meer gericht op de kinderen die getoetst zijn. Het doel van het onderzoek in hoofdstuk 4 was om de ontwikkeling van kinderen in het fonologisch bewustzijn te beschrijven van groep 2 tot en met groep 6. De studie beschreven in hoofdstuk 3 leverde een IRT-gekalibreerde itembank voor fonologisch bewustzijn, welke als uitgangspunt is genomen voor de analyses van deze studie. De resultaten lieten een afnemende groei in fonologisch bewustzijn zien; dat wil zeggen dat de mate van groei het sterkst was in groep 3 en afnam naarmate kinderen ouder werden. Na groep 5 bleken de fonologisch bewustzijn scores van kinderen niet meer significant vooruit te gaan. Verder toonden de resultaten aan dat er significante effecten waren van receptieve letterkennis, woordleesvaardigheid, geslacht en SES op het fonologisch bewustzijn. Kinderen met meer letterkennis en meer leesvaardigheden presteerden beter op de taken voor het fonologisch bewustzijn dan kinderen met minder letterkennis en minder leesvaardigheden. Bovendien, meisjes haalden hogere fonologisch bewustzijn scores dan jongens en kinderen met een hogere SES achtergrond presteerden beter dan de kinderen met een lagere SES achtergrond. Geen van deze variabelen bleek echter het ontwikkelingsverloop van het fonologisch bewustzijn te beïnvloeden. Dat wil zeggen, een kwadratische groeicurve was het best in staat de ontwikkeling in het fonologisch bewustzijn te beschrijven voor een uiteenlopende groep kinderen.

Uitgaande van de opvatting dat fonologisch bewustzijn het beste kan worden opgevat als een eendimensionaal construct, zijn vervolgens de relaties tussen fonologisch bewustzijn in brede zin (gemeten door verschillende taken) en lezen onderzocht. Hoofdstuk 5 beschrijft de resultaten van deze longitudinale studie bij kinderen in groep 2 tot en met 4. Een eerste uitkomst van dit onderzoek was dat zowel fonologisch bewustzijn als lezen beschouwd kunnen worden als relatief stabiele vaardigheden gedurende de beginjaren van de basisschool. De resultaten lieten verder zien dat fonologisch bewustzijn aanvankelijk een significante voorspeller van lezen was, maar dit effect verdween wanneer de autoregressieve effecten van lezen ook in het model waren opgenomen. Daarnaast werd aangetoond dat fonologisch bewustzijn niet alleen lezen kon voorspellen, maar dat ook het omgekeerde het geval was, namelijk dat lezen een significante voorspeller was van fonologisch bewustzijn. Deze resultaten ondersteunen het bestaan van wederkerige relaties tussen fonologisch bewustzijn en lezen. Tot slot bleek letterkennis een andere belangrijke voorspeller van lezen te zijn. De gegevens in deze studie lieten zien dat zowel fonologisch bewustzijn als letterkennis unieke voorspellers van lezen zijn.

In het laatste hoofdstuk 6 zijn de algemene conclusies weergegeven en geïntegreerd in een ontwikkelingsmodel voor fonologisch bewustzijn. De resultaten van de vier studies samen leiden tot de conceptualisatie van fonologisch bewustzijn als een continuüm van toegang tot fonologische representaties dat loopt van gedeeltelijke tot volledige toegang tot fonemen. Vervolgens is in dit hoofdstuk de vertaalslag gemaakt naar de praktijk. Wat betekenen de theoretische bevindingen uit de verschillende onderzoeken voor het meten van fonologisch bewustzijn in de praktijk? Voor de onderwijspraktijk biedt de beschikbaarheid van een IRT-gekalibreerde itembank uitstekende mogelijkheden voor adaptieve toetsen, waarbij elk kind een andere toets maakt die is aangepast aan het niveau van het kind. Daarnaast is voor de vroege screening van leesproblemen aanbevolen gebruik te maken van een combinatie van taken voor fonologisch bewustzijn en receptieve letterkennis in groep 2 om te voorspellen welke kinderen in groep 3 problemen met lezen zullen ondervinden. Voor het screenen op leesproblemen is het verder van belang om de ontwikkeling van kinderen in de vroege leesvaardigheden nauwkeurig te volgen, aangezien bekend is dat kinderen die ondanks interventie niet vooruitgaan de meest ernstige leesproblemen hebben. De resultaten van dit proefschrift hebben voldoende bewijs geleverd voor de conceptualisatie van fonologisch bewustzijn als een eendimensionale vaardigheid over de verschillende groepen, wat nieuwe mogelijkheden biedt voor het volgen van de ontwikkeling in fonologisch bewustzijn in de tijd.

Appendix A Basic elements item response theory

Several handbooks have been written that address item response theory (IRT; e.g., Lord, 1980; van der Linden & Hambleton, 1997). In this appendix, a short overview of the basic elements of IRT will be presented to provide support in the interpretation of the results of this thesis.

There are two general frameworks for the development and analyses of tests: a) CTT (i.e., classical test theory), and b) IRT. Both theories introduce a number of theoretical concepts and assumptions and specify their relationships in test models.

The focus of the two theories is completely different. The basic notion in CTT is the true score on a certain test. The examinee's ability depends on the items of that particular test. Moreover, the item statistics (e.g., *p* value) depend on the sample from which they are obtained. This dependency reduces their usefulness.

In contrast, in IRT, the concept to be measured is central. This concept is considered as an unobservable or latent variable. Characteristic for IRT is that the difficulty of the items and the ability of persons are positioned on the same scale. Because of the specific characteristics of IRT, this theory has some important advantages compared to CTT:

- ability estimates (scores) are test-independent
- item statistics (e.g., difficulty, discrimination) are group-independent

Because of these characteristics, IRT is very suitable for large-scale investigation with several populations and item banks consisting of many items. Consider an item bank consisting of 200 items to be calibrated for a broad range of ability levels. In this case, it is unfeasible to administer each item to each person. In IRT, this problem can be solved by using incomplete designs. In these designs, there are subsets of overlapping items which are administered to different groups of children. These groups are not necessarily random, enabling to tailor the sets of items to different ability groups. The overlap constitute the link between all of the items. Because of this overlap, it is possible to place all of the items on one scale of measurement. As a result, examinees who take different tests can be compared without problems. Table A.1 provides an overview of the main differences between IRT and CTT.

Table A.1

Main differences between IRT and CTT

Area	IRT	CTT
Level of modeling	Item	Test
Item-ability relationship	Difficulty of the items and the ability of persons are scaled on the same scale	Not specified
Ability	Ability scores are reported on the scale (based on item scores and the IRT model)	Test score (sum of items answered correct on a particular test)
Item statistics	b, a, c (for three-parameter model) plus corresponding item information functions	Item difficulty (*p*) Item discrimination (*r*)

One-Parameter Logistic Model (OPLM)

For the present thesis, the data were analyzed within the framework of IRT. Different IRT models exist. The primary distinction among the most popular IRT models is between the number of parameters used to describe the items. In the present thesis, we have used the OPLM. In the OPLM, difficulty parameters are estimated and discrimination parameters are dealt with as known constants (i.e., discrimination indices can vary, but have discrete values) and the OPLM can therefore be considered as a synthesis between the one- and two-parameter model. The relation between ability (denoted by the Greek letter theta, θ) and examinee's item performance can be represented by an item characteristic curve (ICC). The ICC for the OPLM is given by the equation:

$$P_i(\theta) = \frac{e^{\alpha_i(\theta-\beta_i)}}{1+e^{\alpha_i(\theta-\beta_i)}}$$

$P_i(\theta)$: probability that an examinee with ability θ answers item i correctly:

β_i: difficulty parameter of item i

a_i: discrimination parameter of item i

e: transcendental number with value 2.718, and

$P_i(\theta)$: is an S-shaped curve with values between 0 and 1 over the ability scale. The higher the value of the latent variable θ, the higher the probability of a correct response.

Characteristics ability scale:

- exact values are undetermined (any linear transformation is allowed)
- the scale is fixed by setting the sum of the item difficulties equal to zero
- theoretical range is from negative infinity to positive infinity, while practical considerations usually limit the range of values from -3 to +3

Parameters in the OPLM:

- β_i (difficulty) parameter for an item: point on the ability scale where the probability of a correct response is 0.5. This parameter is a location parameter, which indicates the position of the ICC in relation to the ability scale. The higher the value of the β_i parameter, the more ability is needed for an examinee to have a 50% chance of answering the item correct.
- a_i (discrimination) parameter: slope of the ICC curve at the point β_i on the ability scale. The higher the value of the a_i parameter, the better the item can discriminate between persons of different ability levels.

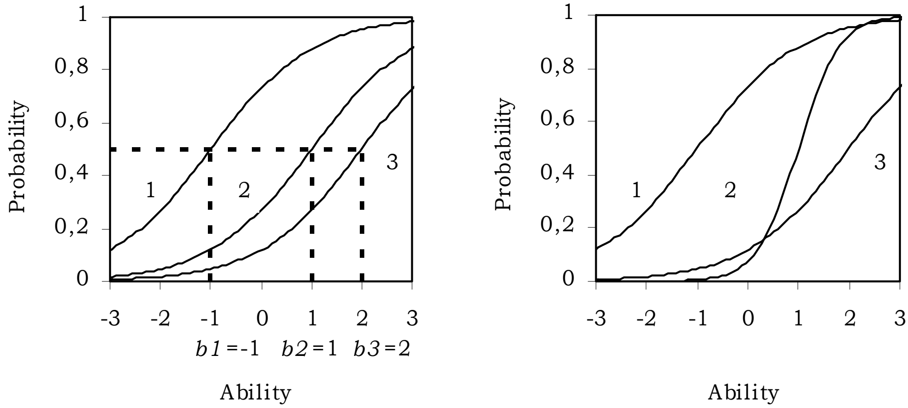

Figure A.1 *Sample ICCs: (a) three items with different difficulty parameters; (b) three items with different difficulty and discrimination parameters*

Figure A.1a shows some sample ICCs. The item parameters are as follows: Item 1, β_1 = -1.0; item 2, β_2 = 1.0; and item 3, β_3 = 2.0. As can be seen, the curves differ only by their location on the ability scale. The most difficult item (3) is located to the right end of the ability scale and the easiest item (1) is located to the left end of the ability scale. Figure A.1b shows the same ICCs as Figure A.1a, with the only difference that the discrimination parameter of item 2 has increased.

Information functions

Another property of IRT is the existence of item information functions. These functions express the precision of measurement at different ability levels. Some relevant features regarding the concept of information include:

- the amount of information provided by an item is largest when the probability of a correct answer is 0.5, thus at its β value on the ability scale
- items with more discriminating power contribute more to measurement precision than items with less discriminating power
- the sum of item information functions in a test results in a test information function
- the estimate of the standard error for an ability estimate is inversely proportional to the square root of the test information evaluated at the estimate
- the measurement precision can be computed for each point of the ability scale in contrast to reliability in CTT, which is assumed to be constant for all test scores
- when the model holds for different populations, the same item information functions apply for these different populations. By presenting the information functions of different item sets and the ability distributions of different groups in the same figure, it is possible to show the utility of different sets of items for different groups of children.

In Figure A.2, the test information function and the standard error of measurement of a certain item set is shown. The ability scale is put on the x-axis and the amount of information provided by this set of items and the corresponding standard error of measurement are shown on the y-axis. As can be seen, the lower the standard error of measurement, the higher the amount of information. In the same figure, the ability distributions of two groups are presented below the x-axis. As can be noticed, children from group 2 generally have higher ability levels than children from group 1. It can be derived from the figure that this item set provides the most accurate estimates for the better performing children from group 1 and for the children at the lower end of the ability distribution in group 2. The information value of this item set decreases for the children at the lower end of the ability distribution in group 1 and for the children at the higher end of the ability distribution in group 2. In other words, these two groups of children can not be accurately measured by this item set. In this way, it can be determined which tasks can best be administered at which ability levels.

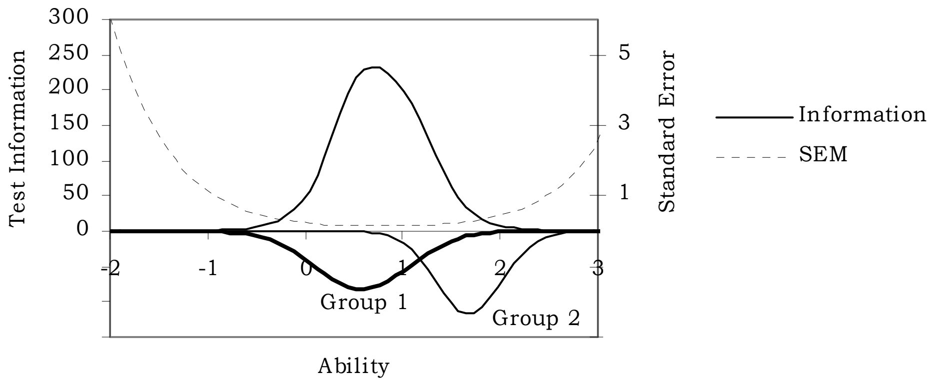

Figure A.2 *Test information function and the standard error of measurement of an item set in relation to the ability of two groups*

References

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Mahwah, NJ: Erlbaum.

van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of modern item response theory.* New York: Springer.

Curriculum Vitae

Judith Vloedgraven is geboren op 11 april 1981 te Heeten. Na het behalen van haar VWO-diploma op het Carmel College Salland te Raalte in 1999, begon ze in datzelfde jaar met de opleiding Pedagogische Wetenschappen aan de Katholieke Universiteit Nijmegen. Tijdens het laatste jaar van de studie volgde ze naast het reguliere programma tevens de lerarenvariant, een cursus in didactische vaardigheden. Gedurende de studie werkte ze als studentassistent bij de afdeling statistiek. In 2003 is ze cum laude afgestudeerd met als specialisatie 'Orthopedagogiek: leren en ontwikkeling'. Na haar studie is ze in 2004 voor enkele maanden naar India vertrokken en heeft daar als vrijwilliger les gegeven op een school voor kansarme kinderen. Bij terugkomst startte ze in september 2004 haar onderzoekswerkzaamheden bij het Expertisecentrum Nederlands dat is verbonden aan de (huidige) Radboud Universiteit. Naast haar werkzaamheden voor haar promotieonderzoek heeft ze in samenwerking met Cito een screeningsinstrument ontwikkeld dat bedoeld is om leerlingen met een risico op leesproblemen vroegtijdig te kunnen signaleren. Dit instrument bevat opgaven voor het meten van fonologisch bewustzijn en letterkennis in de groepen 2 en 3. Verder heeft ze tijdens haar promotietraject studenten begeleid bij hun afstudeerstage en onderwijsgevende taken verricht. In 2008 heeft ze haar proefschrift afgerond. Vanaf januari 2009 zal ze als toetsdeskundige werkzaam zijn bij Cito.