Original Article

Understanding Phoneme Segmentation Performance by Analyzing Abilities and Word Properties

Samantha Bouwmeester[1], Elisabeth H. M. van Rijen[1], and Klaas Sijtsma[2]

[1]Erasmus University Rotterdam, The Netherlands, [2]Tilburg University, The Netherlands

Abstract. Several studies have demonstrated the relationship between phoneme segmentation ability and early reading performance, but so far it is unclear which abilities are involved, and which word properties contribute to the difficulty level of a segmentation task. Using a sample of 596 Dutch children, we investigated the abilities involved in segmenting the phonemes of 45 pseudowords that differed with respect to several properties. First, we found that a combination of short-term memory and speech perception explained variation in segmentation performance. Second, we found that a limited number of word property effects explained the difficulty level of pseudowords rather well. Finally, we constructed a high-reliability scale for measuring segmentation ability.

Keywords: item response theory analysis, linear logistic test model, phoneme segmentation ability, phonological awareness, word segmentation

Introduction

Elementary schools prepare children for acquiring reading ability by practicing phoneme segmentation tasks. A phoneme segmentation task presents a word, for instance, "dark," and children are asked to cut this word into the smallest audible parts possible, in this case "d/a/r/k." By exercising phoneme segmentation tasks, children become sensitive to words consisting of chains of phonemes, and acquire the ability to manipulate these phonemes; that is, they learn that words can be segmented into phonemes and that phonemes can be blended into words. The combination of sensitivity to phonemes and the ability to manipulate them is called phonological awareness. Phonological awareness is assumed to be essential for acquiring initial reading ability (e.g., Bus & IJzendoorn, 1999; Landerl & Wimmer, 2000). Children can only learn how to translate abstract symbols into meaningful language when they are sensitive to the smallest units of oral language, which are the phonemes. Therefore, in kindergarten children start practicing the identification of phonemes, the segmentation of words into phonemes, and the blending of phonemes into words. Segmentation tasks are strong predictors of early reading ability (Geudens & Sandra, 2003).

Researchers do not agree on the skills and abilities required for segmentation (McBride-Chang, Wagner, & Chang, 1997; Sodoro, Allinder, & Rankin-Erickson, 2002). Moreover, they also disagree on which word properties have the strongest influence on phoneme segmentation performance. For example, Schreuder and Van Bon (1989) showed that clusters of consonants are more difficult to segment than vowel-consonant combinations. However, the results of Geudens and Sandra (2003) and Geudens, Sandra, and Van den Broeck (2004) indicate that the cohesion between phonemes interacts with the sonority of consonants.

These disagreements hamper the interpretation of segmentation performance. The purposes of this study were to assess the abilities that are involved in the segmentation of phonemes, and to explain the difficulty level of segmentation tasks using a limited number of word properties. Surprisingly, the extensive research on phonological awareness and early reading processes has failed to stimulate the development of scales for measuring the ability of phonological awareness (for an exception, see Schatschneider, Francis, Foorman, Fletcher, & Mehta, 1999). This study also provides a reliable scale for phonological awareness.

DOI: 10.1027/1015-5759/a000049

Abilities Involved in Segmentation

McBride-Chang et al. (1997) argued that phonological awareness entails at least three abilities: general cognitive ability, verbal short-term memory, and speech perception. General cognitive ability is a prerequisite for mastering a phoneme segmentation task: Children must first understand what is required of them and be capable of carrying out the task (i.e., they must be able to understand what is expected of them and act accordingly). Several studies (e.g., Wagner & Torgesen, 1987; Wagner, Torgesen, Laughton, Simmons, & Rashotte, 1993) have shown that cognitive ability correlates, at least moderately, with performance on phonological awareness tasks. Most of the variation in performance on segmentation tasks caused by individual differences in general cognitive ability may be expected to vanish when a thorough introduction of the task is given and children do exercises before the experiment starts.

Second, to carry out the segmentation task children must memorize the words or pseudowords. This requires verbal short-term memory capacity (Bradley & Bryant, 1985; Wagner et al., 1993). Because pseudowords do not have a semantic meaning, they may be more difficult to recall than meaningful words. Because of greater memory workload, the ability to recall pseudowords is expected to be more important for words containing four or more phonemes than for two- or three-phoneme words. Part of the variation in performance on longer words may be explained by variation in verbal short-term memory capacity but this variation may be absent for short words. Treiman and Weatherston (1992) showed that the more phonemes a syllable contains, the more difficult it was to isolate the initial consonant. McBride-Chang (1995) found that three phonemes were easier to segment than four phonemes, and that four phonemes were easier to segment than five phonemes.

Third, speech perception is hypothesized to be the most important ability involved in segmenting words into phonemes (e.g., Flege, Walley, & Randazza, 1992; Manis et al., 1997). The ability to distinguish speech sounds is implicitly required in all phonological awareness tasks. The influence of general cognitive ability, verbal short-term memory, and speech perception on segmentation is unknown and, as a result, a straightforward interpretation of segmentation performance is impossible.

g
Gsm-MW/MS
Ga-US
Ga-segmentation

Word Properties

Several studies (e.g., Geudens & Sandra, 2003; Treiman, 1984) have shown that properties of words influence segmentation performance. Syllables are made up of smaller subunits called onset and rime (Treiman & Weatherston, 1992). The onset of a syllable is the initial consonant or consonant cluster, and the rime consists of the vowel and the remainder of the syllable. This remainder is usually referred to as the coda (e.g., Treiman & Danis, 1988). According to the onset-rime cohesion hypothesis, phonemes that are within a subunit are adhered more tightly than phonemes that are part of different subunits. Therefore, the segmentation of phonemes within an onset or a rime is expected to be more difficult than the segmentation of phonemes of which one is in the onset and the other in the rime. Thus, it may be more difficult for children to segment the first consonant of the word "tray," because /t/ is part of the onset /tr/, than to segment /r/ from /ay/, because /r/ belongs to the onset and /ay/ belongs to the coda. Schreuder and Van Bon (1989) found that children performed better on stimuli that do not begin with consonant clusters (i.e., consonant (C)/vowel (V) combination, e.g., "vaa" [CV, consonant, vowel]) than on stimuli that begin with these clusters (e.g., "bra" [CCV]).

However, the cohesion between phonemes is not only determined by the onset and rime subunits, but also by the sonority of the consonants. Treiman (1984) suggested that consonants differ with respect to sonority, which determines how closely they adhere to a vowel. According to Treiman (1984), liquids (e.g., /l/ or /r/) tend to adhere more closely to the vowel than nasals (e.g., /m/ or /n/), which in turn adhere more closely to the vowel than obstruents (i.e., plosives [e.g., /p/ or /k/] and fricatives [e.g., /s/ or /g/]). Geudens, Sandra, and Van den Broeck (2004) showed that plosives and fricatives are easier to separate from a vowel than liquids and nasals but that this pattern is characteristic in particular for VCs and not as clear-cut for CVs. This suggests an interaction of sonority and onset-rime cohesion.

McBride-Chang (1995) hypothesized that, since fricatives are pronounced longer (/ffff/, /ssss/) than plosives, which sound short (/k/, /p/), segmenting fricatives from vowels is easier for young children, compared to segmenting plosives from vowels. Contrary to this expectation, McBride-Chang (1995) showed that segmentation performance on fricatives does not differ from performance on plosives. Stahl and Murray (1994) found that children tended to treat certain blends, such as /st/ and /pl/, as units that have a strong cohesion and, thus, may be difficult to segment. However, nasal blends (like /nk/, /nd/ and /mp/) and liquid blends (/ld/) in the coda seemed easier to segment. Schreuder and Van Bon (1989) showed that consonant clusters in the coda are more difficult to segment than vowel-consonant combinations. Children performed better when stimuli end with a vowel-consonant combination (VC, e.g., "aag") than a consonant cluster (VCC, e.g., "urg").

We conclude from these studies that several word properties influence and, thus, explain segmentation performance, and should be incorporated in a set of segmentation tasks.

Choice of Tasks

The segmentation test that was used in this study consisted of 45 one-syllable pseudowords that systematically differed with respect to three word properties. The first property was cluster of consonants. A word contained (1) no cluster of consonants, (2) a cluster of two consonants, or (3) a cluster of three consonants. The second property was location of the cluster. Two possibilities were investigated: onset and coda. The third property was consonant type. Four possibilities were studied, fricatives (v,z,g,f), nasals (m,n), plosives (p,b,k,d,t), and liquids (l,r).

The Appendix shows which word properties were involved for each pseudoword (indicated by 1 scores).

The choice of using nonsense pseudowords reflected an attempt to encourage phonological processing and discourage semantic interference. The pseudowords were easy to pronounce. We used the most frequently used vowels in Dutch (i.e., /u/, /i/, /o/, /a/, /e/, /oo/, /aa/, /ie/, /ee/, /oe/). Vowels sounded the same in combination with different consonants.

The following hypotheses were tested with respect to word properties. (1) Clusters of consonants are more difficult to segment than vowel-consonant (CV or VC) combinations (Schreuder & Van Bon, 1989). (2) Liquid combinations are most difficult to segment, followed by nasal, plosive, and fricative combinations, respectively (Geudens et al., 2004; Treiman, 1984). (3) Nasal and liquid consonant clusters (CC) are easier to segment than fricative and plosive clusters (Stahl & Murray, 1994).

Method

Sample

The sample consisted of 596 children from middle class socioeconomic status families. They came from nine Dutch elementary schools, and attended kindergarten (n = 158, M = 73.23, SD = 4.99), Grade 1 (n = 206, M = 87.23, SD = 5.19), and Grade 2 (n = 232, M = 99.42, SD = 5.34).

Instrument and Procedure

Four versions of the segmentation test were constructed, which differed with respect to the presentation order of the words containing the same number of phonemes. Three experimenters administered the test to individual children in a quiet room in the school building. The experimenters were trained master students in psychology. It was explained to the children that they had to cut a word into the smallest parts possible. Some examples were provided, and the child did some exercises to get used to the task format. More exercises were presented when the experimenter thought this was necessary for a particular child, for example, when they tried to cut the word into onset-rhyme subunits instead of phonemes.

The test started after the child understood the task and finished the exercises. The experimenter articulated the word clearly and fluently, and asked the child to cut it into the smallest pieces possible. The words were presented in the order from short (two phonemes) to long (five phonemes). When the child segmented the word correctly, a score of 1 was assigned. An incorrect segmentation received a 0 score. When the child was distracted by some external cause (e.g., telephone, break-bell), the item was scored as a missing.

Data Analysis

The Rasch model (RM; e.g., Fischer & Molenaar, 1995; Rasch, 1960) and Mokken's monotone homogeneity model (MHM; e.g., Mokken, 1971; Sijtsma & Molenaar, 2002) were fitted to the data to assess the ability structure underlying test performance. The linear logistic test model (LLTM; e.g., Fischer, 1973) was fitted to assess the influence of the word properties on segmentation performance.

Rasch Model

Let random variable X_j denote the score (0, 1) on item j, and let i index subjects. The RM assumes that the probability of $X_j = 1$ depends on the subject's latent variable level (often interpreted as ability level) θ_i and the item's difficulty level β_j:

$$P(X_j = 1 \mid \theta_i) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}. \tag{1}$$

This conditional probability is the item response function (IRF). We used the Rasch Scaling Program (RSP; Glas & Ellis, 1994) for estimating and testing the model because RSP offers more methods for diagnosing misfit than most other programs[1]. Parameters were estimated by means of conditional maximum likelihood (CML), which is a common choice because of the availability of sufficient statistics.

RSP uses the asymptotic χ^2 statistic R_1 for testing the null-hypothesis that all IRFs are parallel logistic functions, and the approximate χ^2 statistic Q_2 for testing local independence of the multivariate conditional distribution of the item scores (Glas & Verhelst, 1995). Together, these statistics constitute a full test of the fit of the RM to the data. When the global R_1 test or LR test is significant, for each separate item the local approximate standard normal statistic U_j (Molenaar, 1983) can be used to test the null hypoth-

[1] In addition, we used the eRm package (Mair & Hatzinger, 2007). Both RSP and eRm yielded the same results with respect to parameter estimates and common fit statistics.

esis such that its IRF is logistic with slope 1 against the alternative that it is not. Values of U_j greater than, say, 1.96 (5% significance level), indicate that the IRF is flatter than expected, and values of U_j smaller than, say, –1.96, indicate that the IRF is steeper than expected.

Monotone Homogeneity Model

Misfit of the RM may be the result of multidimensionality (some items measure different abilities than other items) or IRFs that do not have the logistic shape. The MHM (Sijtsma & Molenaar, 2002) may then be used to localize and investigate the misfit. For analyzing data by means of the MHM, we used the program Mokken scale analysis for polytomous items (MSP; Molenaar & Sijtsma, 2000). MSP enables the selection of the items into dimensionally distinct subsets; that is, item subsets measuring different abilities. MSP also estimates the IRFs from the data. The researcher can use visual inspection and statistical testing to assess the shape of the IRFs. In particular, nonmonotone IRFs and relatively flat IRFs do not contribute to a reliable scale score and those items involved may be discarded from the test.

Linear Logistic Test Model

We estimated the effect of combinations of word properties on the difficulty of words by means of the LLTM (Fischer, 1973, 1974; Kubinger, 2009; Scheiblechner, 1972; also, see "The use of LLTM: Cognitive modeling and item-technology analyses," special issue of *Psychology Science Quarterly*, 2008). Let $q_{jt} = 1$ indicate that word *j* has word property *t*, and let $q_{jt} = 0$ indicate absence of this property. The q_{jt} weights are collected in matrix *Q* (Appendix). Let ν_t be the parameter for the effect of word property *t* on the item difficulty β_j. The difficulty level β_j of word *j* can be estimated from a linear combination of the effects ν_t of the *T* word properties *t*: that is, $\beta_j = \sum_{t=1}^{T} q_{jt} \times \nu_t$. The probability that child *i* produces the correct segmentation of word *j* can be estimated by the function in Equation 1, in which we replaced β_j by $\sum_{t=1}^{T} q_{jt} \times \nu_t$, so that

$$P(X_j = 1 \mid \theta_i) = \frac{\exp(\theta_i - \sum_{t=1}^{T} q_{jt} \times \nu_t)}{1 + \exp(\theta_i - \sum_{t=1}^{T} q_{jt} \times \nu_t)}.$$

This is the LLTM. The more similar the βs estimated from the RM and the βs estimated from the LLTM, the better the LLTM explains the data, and the better the difficulty level of the words can be explained by the difficulty levels of the word properties. The model was estimated using the eRm package (Mair & Hatzinger, 2007). We compared the estimated βs of the RM and the βs derived from the LLTM on the basis of estimated parameters for the word properties. Because the LLTM is nested in the Rasch model, we used the likelihood ratio (LR) to test the difference in the fit of the two models.

Results

Thirteen out of the 596 children (0.059%) had missing values on one or more items. Their data were removed from further analysis. To verify whether the presentation order of the words influenced item performance, an ANOVA was done with the total number of correct segmentations (Table 1 shows proportions of correct segmentations for each item) as dependent variable and the four different word orderings as between-subject factor. Word ordering did not have an effect, $F(4, 591) = 0.51$, $p = .73$.

Rasch Model Analysis

For the 45 items, $R_1 = 421$, $df = 44$, $p < .001$, and $LR = 391$, $df = 44$, $p < .001$, which rejects the null hypothesis of 45 logistic IRFs with equal slopes; and $Q_2 = 7742$, $df = 1890$, and $p < .001$, which rejects the null hypothesis of local independence. The item fit statistic U_j showed that the two-phoneme items (in particular: oeg, roo, vaa, ool, which had $U > 3$, indicating that the IRFs were too flat) mainly caused the significance of the R_1 statistic.

Monotone Homogeneity Model Analysis

MSP selected all 45 items in one set, suggesting that rejection of local independence under the Rasch model was the result of the model's restrictiveness and that a single dominant ability drives performance on all 45 items. Plots of the estimated IRFs (not shown here) of the eight 2-phoneme pseudowords identified by the RM analysis showed that they were not logistic, and also that they were rather flat. These eight items were not used for testing the LLTM.

Linear Logistic Test Model Analysis

Technically, the fit of the Rasch model is a prerequisite for the LLTM. Our combined RM and MHM analyses showed that 37 items were approximately unidimensional and had equally-sloped logistic IRFs; hence, the RM fitted this item subset by approximation. Since the purpose of our LLTM analysis was to learn more about the processes leading to item responses, we think it is reasonable to accept an approximately fitting Rasch model and concentrate on the LLTM, interpreting LLTM results with caution.

The estimated βs according to the RM and the LLTM correlated .93 (such values are commonly found; see Holling, Blank, Kuchenbäcker, & Kuhn, 2008; Sonnleitner, 2008), showing good predictability of item difficulty by combinations of word properties. As expected, the Rasch model fitted significantly better than the LLTM; $LR = 444.38$, $df = 25$, $p < .0001$ (this also is a common result in LLTM analyses; Holling et al., 2008; Sonnleitner, 2008).

Table 1. Proportions-correct (*P*) of the 45 pseudowords in descending order (from easy to difficult)

Word	*P*	Word	*P*	Word	*P*	Word	*P*	Word	*P*	Word	*P*
eep	.92	ool	.80	ner	.75	knif	.66	limt	.58	tern	.43
oeg	.90	voog	.80	soer	.73	snig	.66	strog	.57	morks	.40
noe	.88	neek	.80	gon	.72	ral	.63	biel	.57	galfs	.34
roo	.88	zep	.80	slup	.69	tren	.62	spran	.53	larnt	.33
vaa	.86	meef	.78	din	.69	stir	.61	zelk	.48	telms	.32
aam	.86	tup	.77	mun	.68	nopt	.60	lenst	.48		
daa	.84	lif	.77	rieft	.68	pans	.59	malg	.46		
liek	.81	pug	.77	lom	.68	fops	.59	durpt	.44		

Table 2. Estimated LLTM word property parameters ($\hat{v}$), their standard errors (*SE*), and .95 confidence intervals

Parameter*	$\hat{v}$	*SE*	lower CI	upper CI
2-consonants cluster	–1.797	0.142	–2.075	–1.519
3-consonants cluster	–2.844	0.198	–3.233	–2.455
Cluster location	0.842	0.057	0.730	0.953
Plosive	0.449	0.090	0.273	0.625
Fricative	0.340	0.083	0.178	0.503
Liquid	0.086	0.093	–0.097	0.269
Nasal	–0.157	0.085	–0.324	0.010
Plosive in cluster	0.679	0.083	0.517	0.841
Fricative in cluster	0.439	0.079	0.283	0.595
Liquid in cluster	–0.487	0.084	–0.652	–0.322
Nasal in cluster	–0.239	0.069	–0.376	–0.103

Note. *A singular value decomposition of the Q-matrix showed that the vectors of weights were independent. This indicates a matrix with full rank.

Table 3. Percentile scores of segmentation scale based on 37 pseudowords

Percentile	10	20	30	40	50	60	70	80	90
Number-correct score	1	12	18	24	27	29	32	33	35

Table 2 shows the estimated effect parameters ($\hat{v}$) for the word properties, their standard errors (*SE*), and the 95% confidence intervals (CI). Hypothesis 1 was supported. There was a strong positive effect for clusters of consonants. Thus, a cluster of consonants rendered the segmentation of a word more difficult. A cluster of three consonants was significantly more difficult to segment than a cluster of two consonants. Moreover, a cluster in the coda was more difficult to segment than a cluster in the onset.

Hypothesis 2 was partly supported. Words containing a liquid or a nasal were more difficult to segment than words containing a nasal or a fricative. However, segmenting words containing a liquid was not more difficult than segmenting words containing a nasal, and segmenting words containing a fricative was not more difficult than segmenting words containing a plosive. Nasal- and liquid-consonant clusters were more difficult to segment than plosive- and fricative-consonant clusters. This result contradicts Hypothesis 3.

Scale Construction

Because of their relatively flat IRFs and their easiness, the eight 2-phoneme words were expected to contribute little to a reliable scale score; hence, they were excluded from the scale. Table 3 shows the percentile scores corresponding to the number of correct scores based on the remaining 37 items; 56 children segmented none of the words correctly. However, these children often correctly segmented several two-phoneme pseudowords that had been removed from the scale. Cronbach's α was 0.97. It may be concluded that children in the second decile could segment combinations of vowels and consonants, but had difficulty segmenting clusters of consonants. Children from the sixth decile onward could segment both two and three consonant clusters.

Discussion

The MHM fitted the data well, meaning that one mathematical dimension was enough to explain item performance. With respect to psychological interpretation, we conclude that this dimension most likely represents both short-term memory and speech perception as the simultaneous driving forces of item performance. This justifies the conclusion that none of the items are driven primarily by short-term memory whereas other items are driven primarily by speech perception; these cognitive features are active in combination, not alone, and this is what unidimensionality suggests. The unidimensionality of the segmentation test is convenient because it allows measuring low, intermediate, and high segmentation ability levels on the same scale, using the same interpretation everywhere. The large range of difficulty level of the pseudowords indicates that the test covers a broad ability range and can be used for diagnosing children who differ widely in segmentation ability, or to follow the development of children's segmentation ability for a longer time period.

In general, with respect to verbal short-term memory, performance on long words was worse than on short words. The MHM analysis suggested that two-phoneme items measured the same composite of short-term memory and speech perception as the other items, but that their IRFs were not logistic and also relatively flat. This explains the misfit of the RM. Moreover, these items were very easy. This could mean that verbal short-term memory is hardly needed for the two-phoneme words, which would represent a deviation from unidimensionality.

The results of the LLTM analysis showed that the small number of hypothesized word properties well explained the difficulty level of the pseudowords. Treiman's (1984) hypothesis, that vowel-liquid combinations are most difficult to segment followed by vowel-nasal, vowel-plosive, and vowel-fricative combinations, respectively, was partly confirmed. Vowel-liquid combinations were indeed more difficult to segment than vowel-fricative or vowel-plosive combinations, but they did not differ from vowel-nasal combinations. Our results showed a large difference between obstruents (plosives and fricatives) and sonorants (liquids and nasals) in vowel-consonant combinations. As in English, in Dutch sonorants adhere more closely to preceding vowels than obstruents, which renders them difficult to segment.

In general, a cluster of two or three consonants is more difficult to segment than a vowel-consonant combination. This agrees with Schreuder and Van Bon (1989), who only compared vowel-consonant and cluster consonant combinations in the coda. However, this result should be interpreted which caution because clusters of consonants are confounded with the number of phonemes. Therefore, we cannot decide whether three-consonant cluster words are more difficult to segment than two-consonant cluster words because they consist of more phonemes and therefore require more memory capacity, or because the combination of three consonants requires more speech perception ability.

The segmentation scale is a handy and highly reliable tool for teachers and remedial teachers to assess a child's segmentation ability. A scale score can be used to compare segmentation performance of different children or performance of a single child over time. Moreover, the scale score provides information about whether a child has mastered certain word properties. This may be useful information for diagnosing early problems in reading ability.

Acknowledgments

We thank Serena van IJsselmuide, Deborah Vonk, and Tamara Wally for their assistance in collecting the data.

Ga-US
Gsm-MW/MS

Ga-segmentation

References

Bradley, L., & Bryant, P. E. (1985). *Rhyme and reason in reading and spelling.* Ann Arbor, MI: University of Michigan Press.

Bus, A. G., & IJzendoorn, M. H. van (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91,* 403–414.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica, 37*, 359–374.

Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests* [An explanation of the theory behind psychological tests]. Bern, Switzerland: Huber.

Fischer, G. H., & Molenaar, I. W. (1995). *Rasch models: Foundations, recent developments, and applications.* New York: Springer.

Flege, J. E., Walley, A. C., & Randazza, L. (1992). A developmental study of native and nonnative vowel perception. *Journal of the Acoustical Society of America*, 2415.

Geudens, A., & Sandra, D. (2003). Beyond implicit phonological knowledge: No support for an onset-rime structure in children's explicit phonological awareness. *Journal Of Memory and Language, 49*, 157–182.

Geudens, A., Sandra, D., & Van den Broeck, W. (2004). Segmenting two-phoneme syllables: Developmental differences in relation with early reading skills. *Brain and Language, 90*, 338–352.

Glas, C. A. W., & Ellis, J. L. (1994). *Rasch scaling program*. Groningen, The Netherlands: iecProGAMMA.

Glas, C. A. W., & Verhelst, N. D. (1995). Testing the Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models, foundations, recent developments, and applications* (pp. 69–95). New York: Springer.

Holling, H., Blank, H., Kuchenbäcker, K., & Kuhn, J. T. (2008). Rule-based item design of statistical word problems: A review and first implementation. *Psychology Science Quarterly, 50*, 363–387.

Kubinger, K. D. (2009). Applications of the linear logistic test model in psychometric research. *Educational and Psychological Measurement, 69*, 232–244.

Landerl, K., & Wimmer, H. (2000). Deficits in phoneme segmentation are not the core problem of dyslexia: Evidence from German and English children. *Applied Psycholinguistics, 21,* 243–262.

Mair, P., & Hatzinger, R. (2007). Extended rasch modeling: The ERM package for the application of IRT models in R. *Journal of Statistical Software, 20*, 1–20.

Manis, E. R., McBride-Chang, C. A., Seidenberg, M. S., Keating, P., Doi, L. M., Munson, B., & Petersen, A. (1997). Are speech perception deficits associated with developmental dyslexia? *Journal of the Acoustical Society of America, 66*, 211–235.

McBride-Chang, C. (1995). What is phonological awareness? *Journal of Educational Psychology, 87,* 179–192.

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

Mokken, R. J. (1971). *A theory and procedure of scale analysis*. Berlin, Germany: De Gruyter.

Molenaar, I. W. (1983). Some improved diagnostics for failure of the Rasch model. *Psychometrika, 48*, 49–72.

Molenaar, I. W., & Sijtsma, K. (2000). *User's manual MSP5 for windows. A program for Mokken Scale analysis for Polytomous items* [software manual]. Groningen, The Netherlands: iecProGAMMA.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests.* Copenhagen, Denmark: Nielsen & Lydiche.

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*, 439–449.

Scheiblechner, H. (1972). Das Lernen und Lösen complexer Denkaufgaben (Learning and solving complex thought problems). *Zeitschrift für experimenteller und angewandte Psychologie, 19,* 481–520.

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

Sijtsma, K., & Molenaar, I. W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage.

Sodoro, J., Allinder, R. M., & Rankin-Erickson, J. L. (2002). Assessment of phonological awareness: Review of methods and tools. *Educational Psychological Review, 14,* 223–259.

Sonnleitner, P. (2008). Using the LLTM to evaluate an item-generating system for reading comprehension. *Psychology Science Quarterly, 50*, 345–362.

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

Treiman, R. (1984). On the status of final consonant clusters in English syllables. *Journal of Verbal Learning and Verbal Behavior, 23,* 343–356.

Treiman, R., & Danis, C. (1988). Are speech perception deficits associated with developmental dyslexia? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 145–152.

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

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

Samantha Bouwmeester

Institute of Psychology
FSW
Erasmus University
P. O. Box 1738
NL-3000 DR Rotterdam
The Netherlands
Tel. +31 10 408-2795
E-mail bouwmeester@fsw.eur.nl

Appendix

Item (*j*) by Property (*t*) Matrix (*Q*).

	2-consonants cluster	3-consonants cluster	Cluster location	Fricative	Nasal	Plosive	Liquid	Fricative in cluster	Nasal in cluster	Plosive in cluster	Liquid in cluster
biel	0	0	0	0	1	1	0	0	0	0	0
din	0	0	0	0	1	1	0	0	0	0	0
pug	0	0	0	1	1	0	0	0	0	0	0
tup	0	0	0	0	0	1	0	0	0	0	0
ner	0	0	0	1	1	0	0	0	0	0	0
meef	0	0	0	1	1	0	0	0	0	0	0
neek	0	0	0	1	0	1	0	0	0	0	0
mun	0	0	0	0	1	0	0	0	0	0	0
lom	0	0	0	0	1	0	1	0	0	0	0
ral	0	0	0	0	0	0	1	0	0	0	0
lif	0	0	0	1	0	0	1	0	0	0	0
liek	0	0	0	0	0	1	1	0	0	0	0
gon	0	0	0	1	1	0	0	0	0	0	0
zep	0	0	0	1	0	1	0	0	0	0	0
voog	0	0	0	1	0	0	0	0	0	0	0
soer	0	0	0	1	0	0	1	0	0	0	0
malg	1	0	0	0	1	0	0	1	0	0	1
tern	1	0	0	0	0	1	0	0	1	0	1
pans	1	0	0	0	0	1	0	1	1	0	0
nopt	1	0	0	0	1	0	0	0	0	1	0
limt	1	0	0	0	0	0	1	0	1	1	0
rieft	1	0	0	0	0	0	1	1	0	1	0
fops	1	0	0	1	0	0	0	1	0	1	0
zelk	1	0	0	1	0	0	0	0	0	1	1
tren	1	0	1	0	1	0	0	0	0	1	1
knif	1	0	1	1	0	0	0	0	1	1	0
stir	1	0	1	0	0	0	1	1	0	0	0
slup	1	0	1	0	0	1	0	1	0	0	1
snig	1	0	1	1	0	0	0	1	1	0	0
telms	0	1	0	0	0	1	0	1	1	0	1
morks	0	1	0	0	1	0	0	1	0	1	1
durpt	0	1	0	0	0	1	0	0	0	1	1
larnt	0	1	0	1	0	0	0	0	1	1	1
lenst	0	1	0	0	0	0	1	1	1	1	0
galfs	0	1	0	1	0	0	0	1	0	0	1
strog	0	1	1	1	0	0	0	1	0	1	1
spran	0	1	1	0	1	0	0	1	0	1	1

