JSLHR

Research Article

Screening for Language Delay: Growth Trajectories of Language Ability in Low- and High-Performing Children

Marianne Klem,[a,b] Bente Hagtvet,[a] Charles Hulme,[a,c] and Jan-Eric Gustafsson[a,d]

Purpose: This study investigated the stability and growth of preschool language skills and explores latent class analysis as an approach for identifying children at risk of language impairment.
Method: The authors present data from a large-scale 2-year longitudinal study, in which 600 children were assessed with a language-screening tool (LANGUAGE4) at age 4 years. A subsample (n = 206) was assessed on measures of sentence repetition, vocabulary, and grammatical knowledge at ages 4, 5, and 6 years.
Results: A global latent language factor showed a high degree of longitudinal stability in children between the ages of 4 to 6 years. A low-performing group showing a language deficit compared to their age peers at age 4 was identified on the basis of the LANGUAGE4. The growth-rates during this 2-year time period were parallel for the low-performing and 3 higher performing groups of children.
Conclusions: There is strong stability in children's language skills between the ages of 4 and 6 years. The results demonstrate that a simple language screening measure can successfully identify a low-performing group of children who show persistent language weaknesses between the ages of 4 and 6 years.

The aim of screening for language delay is to identify children who are in need of language support (Dockrell, Ricketts, & Lindsay, 2012). Such screening rests on the assumption that "early delays in development predict later delays" (Darrah, Hodge, Magill-Evans, & Kembhavi, 2003, p. 98). Such an assumption is supported by evidence showing strong stability in children's language ability from an early age (Bornstein, Hahn, Putnick, & Suwalsky, 2014; Rice & Hoffman, 2015). Strong stability in this sense refers to a strong degree of consistency in the relative ordering of individuals in a group of children on some characteristic over time (Bornstein & Putnick, 2012). Using a latent variable modeling approach with items from a language-screening tool, LANGUAGE4 (SPRÅK4; Horn & Dalin, 2008), we investigated the stability of children's language development between the ages of 4 and 6 years. A critical question is the extent to which early delays in language development may provide evidence of persistent language impairment (Hayiou-Thomas, Dale, & Plomin, 2013).

The identification of children with language delay is determined on the basis of their language performance levels being "lower than expected relative to their age peers" (Rice, 2013, p. 223). However, deciding whether a language delay in a young child implies the existence of language impairment that will persist is difficult because individual variability is typically large (e.g., see Ellis & Thal, 2008, for a brief overview). Different cutoff criteria are used to define language impairments, as no consensus exists on the distinction among impaired, delayed, and typical language development (Bishop, 2014; Dollaghan, 2004, 2011). This lack of consensus regarding cutoff criteria is confirmed by the wide range of prevalence rates for language impairment that have been reported in different studies (from 2.3% to 19%; see Nelson, Nygren, Walker, & Panoscha, 2006). Moreover, measurement error is likely to be responsible for some of the instability found in classifications of child language impairment over time (Eadie et al., 2014).

Although recognizing such difficulties, the clinical context still requires categorical decisions to ensure that children who need an intervention to ameliorate language-learning difficulties are identified (Bishop, 2014; Coghill & Sonuga-Barke, 2012). In Norway, the general health

[a]University of Oslo, Norway
[b]Statped Sørøst, Norway
[c]University College London, United Kingdom
[d]University of Gothenburg, Sweden

Correspondence to Charles Hulme: c.hulme@ucl.ac.uk

Editor: Rhea Paul
Associate Editor: Shelley Gray
Received August 19, 2015
Revision received November 21, 2015
Accepted December 31, 2015
DOI: 10.1044/2016_JSLHR-L-15-0289

Disclosure: The authors have declared that no competing interests existed at the time of publication.

surveillance program includes a language assessment at age 4 administered by a public health nurse (Sosial-og helsedirektoratet, 2006). The recommended screening procedure, LANGUAGE4 (Horn & Dalin, 2008), is designed to identify 4-year-olds with a language delay who may be at risk of persistent language impairment.

LANGUAGE4 is a simple language assessment procedure specifically developed as a screening tool in the context of a general health check given to all 4-year-olds (Horn & Dalin, 2008). Taking only 10–15 min to administer, it is efficient and simple while still covering a relatively wide range of language skills that most 4-year-olds master, according to developmental studies (Horn & Dalin, 2008). In a recent study, Klem, Gustafsson, and Hagtvet (2015) investigated the construct validity of LANGUAGE4, and their findings suggest that the structure of LANGUAGE4 may be described in terms of a model with one second-order factor, labeled the *LANGUAGE4 factor* in the current article, that explains a substantial amount of the variance (R^2 = .61) in a concurrent latent language criterion factor, defined by four widely used standardized language tests (for further information, see Klem et al., 2015).

Identifying children at risk of persistent language impairments requires an appreciation of the developmental pathways shown by children with and without language impairment (Karmiloff-Smith, 1998; Law, Tomblin, & Zhang, 2008). Results from a recent study by Conti-Ramsden, St. Clair, Pickles, and Durkin (2012), describing language growth over 10 years in a large clinical sample of children diagnosed with language impairment at age 7 years, indicated that the growth of global language abilities could best be accounted for by the *tracking hypothesis* (Law et al., 2008). This hypothesis suggests that children with language impairment may have an initial delay but that their developmental trajectories are parallel to those of typically developing children (Conti-Ramsden et al., 2012; Law et al., 2008). However, studies of younger children suggest greater variability in growth trajectories. Although in some studies parallel growth trajectories in young children with and without language impairment have been documented across different subdimensions of language (Rice, 2013), other evidence suggests that children with delayed language at age 4 years may catch up with their typically developing peers by the age of 5.5 years (Bishop & Edmundson, 1987). This pattern suggests that an initial delay can be followed by an acceleration of language growth relative to typically developing control children, at least in the preschool period (Conti-Ramsden et al., 2012). Finally, there is evidence that, in some children, there is a trajectory of late-onset language delay (e.g., Dale, Price, Bishop, & Plomin, 2003; Snowling, Duff, Nash, & Hulme, 2015; Zambrana, Pons, Eadie, & Ystrom, 2014).

In summary, a major difficulty in screening for language delay is the mismatch between the screening method, primarily developed for categorizing binary conditions, and the nature of a language delay, which is rather a matter of degree (Eriksson, Westerlund, & Miniscalco, 2010). Moreover, although evidence of instability in classification of language impairment before the age of 5 years (e.g., Bishop & Edmundson, 1987) may in part reflect measurement error (Law et al., 2008) and the arbitrary nature of the boundaries defining a language impairment (Eadie et al., 2014), accelerated language growth in early childhood appears to occur in some children who show an initial delay (Conti-Ramsden et al., 2012; Taylor, Christensen, Lawrence, Mitrou, & Zubrick, 2013). Conversely, other children may show relatively normal early language development but go on to develop late emerging language impairments (Dale et al., 2003; Snowling et al., 2015; Zambrana et al., 2014).

The limitations of making predictions at the individual level—which, for both substantive and methodological reasons, inevitably will include false positives and false negatives—is well recognized in research (Pennington et al., 2012). Here we emphasize the probabilistic nature of predicting child language outcomes and use a structural equation modeling approach to assess the trajectories of language development of subgroups of children who were classified using a latent class analysis on the basis of the LANGUAGE4 screening tool at age 4 years.

We report a study in which we assessed the extent to which the LANGUAGE4 factor predicts variations in a latent language criterion factor assessed repeatedly between the ages of 4 and 6 years. This latent language factor is defined by the same set of expressive and receptive language measures at each time point, which allowed us to examine whether the equivalent language construct is measured over time. Moreover, using a population-based sample, in the current study we explored the utility of latent class analysis as a means of identifying a group of children with low levels of language performance who may be considered at risk for persisting language impairments. Rather than using an arbitrary cutoff, latent class analysis allows for a classification based on different performance profiles and provides probability estimates for class membership (Coghill & Sonuga-Barke, 2012). Because LANGUAGE4 is a screening procedure designed to identify children in need of language support at age 4 years, our particular focus was on the identification of a stable low-performing group.

Method

Participants

We used a population-based sample of 600 (274 girls, 326 boys) 4-year-old children (*M* age = 52.3 months, *SD* = 2.7 months; age range: 47–60 months), who were enrolled in two different studies of LANGUAGE4. Norwegian was the primary language of all the children in the sample. The children in Subsample 1 (*n* = 394) were recruited from a sample of 4-year-olds who attended their general health check at their local health care center between 1999 and 2002 in four different regions in Norway (i.e., unselected sample). Subsample 2 (*n* = 206) was mainly recruited from Norwegian day care centers in the eastern part of Norway. This subsample consisted of 4-year-olds who were assessed with the LANGUAGE4 screening tool at their local health care centers in 2007 or 2008 and who, at the outset of the study, had no identified condition known to influence language impairment (for further information see Klem et al., 2015).

Design and Procedures

In this article we report data from three assessments made between the ages of 4 to 6 years. The screening tool LANGUAGE4 was administered to the full sample (N = 600), and additional longitudinal data from non-overlapping tests (i.e., measures of vocabulary, grammatical knowledge, and sentence repetition) were obtained from Subsample 2 (n = 206); consequently, data are missing by design. The LANGUAGE4 assessment procedure was administered as a part of the Norwegian general health surveillance program for 4-year-olds: children were individually assessed at their regional health care center by a trained public health nurse.

The longitudinal language tests (see description below) were administered in a fixed order by trained assistants in a separate room in the children's (n = 206) day care centers at Time 1 (M age = 51.1 months, SD = 2.2 months) and Time 2 (M age = 62.8 months, SD = 2.4 months), and the children were usually assessed in school at Time 3 (M age = 75.0 months, SD = 2.3 months).

Tests and Materials

All items in this study were scored dichotomously (i.e., 1 and 0 for *correct* and *incorrect,* respectively). All children at age 4 years were assessed with the LANGUAGE4 screening tool, and Subsample 2 (n = 206) was given the additional longitudinal measures. These longitudinal language measures (vocabulary, grammatical knowledge, and sentence repetition) were part of a comprehensive test battery including nonverbal ability, phonological skills, literacy, and language ability. Only the measures relevant to the current study are presented here, and the total scores for these measures are reported as raw scores (number of items correct).

LANGUAGE4

The assessment materials for this screening test consist of a simple picture folder showing situations that are familiar to 4-year-old children. The assessment involves 33 questions related to three pictures: a living room, a bathroom, and a dog in different positions relative to a kennel. LANGUAGE4 is designed as a systematic observation rather than a formal screening test, and no clear cutoff score determines whether a child is at risk for language impairment (Horn & Dalin, 2008). Thus, our analytical approach to LANGUAGE4 is at an item level. Because six items lacked the required variation to be fitted to a model, the structure of LANGUAGE4 has been modeled on the basis of 27 rather than the full 33 items (see Klem et al., 2015). According to Klem et al. (2015), the internal structure of LANGUAGE4 is best described by a hierarchical model with one second-order factor (LANGUAGE4 factor), which in turn accounts for the variation in a wide range of different language subskills at the first-order level, including aspects of word definition, naming, color, comparative adjectives, prepositions, sentence structure, sentence repetition, and inference skills (see Klem et al., 2015, for a detailed description of these factors and their corresponding items).

Vocabulary

The Norwegian version of the British Picture Vocabulary Scale–Second Edition (BPVS-II; Dunn, Dunn, Whetton, & Burley, 1997) was used to measure receptive vocabulary. The child was asked to match one of four line drawings to a word presented orally by the examiner. The test includes 144 items (Lyster, Horn, & Rygvold, 2010), and testing was stopped after eight incorrect items within a block of 12 items.

Grammatical Knowledge

The Norwegian version of the Grammatic Closure subtest from the Illinois Test of Psycholinguistic Abilities, Revised Edition (ITPA; Kirk, McCarthy, & Kirk, 1968) was used to measure expressive grammatical knowledge. The child was asked to fill in blanks in a series of incomplete sentences presented orally, along with corresponding pictures. The 33-item test covers a wide range of grammatical constructions (such as inflection of nouns, verbs, and adjectives). Testing was stopped after six consecutive incorrect items at Times 1 and 2, and all items were presented at Time 3.

Sentence Repetition

The 21-item sentence repetition task is an adapted version of the Sentence Memory subtest of the Norwegian language screening test, Language 6–16 (Språk 6–16; Ottem & Frost, 2005). In order to avoid floor effects, the original 16-item subtest was supplemented with the first five items of the sentence repetition test from the Norwegian version of the Wechsler Preschool and Primary Scale of Intelligence–Revised (WPPSI-R; Wechsler, 1989). Sentences of increasing length and complexity were spoken by the examiner and the child was required to repeat back exactly what was said. Testing was stopped after three consecutive incorrect repetitions.

Analyses

All analyses were performed using structural equation modeling with Mplus, Version 7.11 (Muthén & Muthén, 1998–2012) using the maximum likelihood estimation with robust standard errors (MLR), which is robust to non-normality of data (Muthén & Muthén, 1998–2012). The data include both categorical and continuous indicators, that is, categorical items from LANGUAGE4 and continuous total scores from the longitudinal language tests. Moreover, we assumed that two different missing data mechanisms are present. Missing items in LANGUAGE4 are assumed to be missing completely at random (MCAR) because, on occasion, individual children were not given all items and no patterns could be detected in the missingness. In contrast, as a part of the study design, for Subsample 1, children were not given the longitudinal language tests. For these data, the weaker missing at random (MAR) assumption appears to be appropriate. The recommended maximum likelihood procedure for missing data under MAR conditions is MLR, which is the default in Mplus (Muthén & Muthén, 1998–2012).

We used different but complementary analytical approaches. First, using an autoregressive approach, we examined the longitudinal relation between the LANGUAGE4 factor at age 4 years and a global criterion latent language construct (latent language factor), defined by vocabulary, grammatical knowledge, and sentence repetition at ages 4, 5, and 6 years. We then explored the use of a latent class analysis on LANGUAGE4 (at the item level) as a means of identifying different subgroups of 4-year-olds, with a particular focus on trying to identify a stable, low-performing group. Latent class analysis is an exploratory procedure used to identify a categorical latent class variable measured by a number of observed response variables. The method categorizes individuals into classes and identifies items that best distinguish between classes (Nylund, Asparouhov, & Muthén, 2007).

To assess the number of latent classes needed, we relied on three fit indexes: the Bayesian information criterion (BIC; Schwarz, 1978), in which the model with the lowest BIC value shows the best fit. Results from simulation studies for latent class analysis models indicate that the sample-size adjusted the BIC (BIC*; Sclove, 1987) represents noteworthy improvements over the BIC (Yang, 2006). Moreover, Nylund et al. (2007) suggest that BIC* also appears to be more consistent when applied to categorical latent class analysis models with unequal classes. Furthermore, we used the Vuong-Lo-Mendell-Rubin likelihood ratio test (VLMR) and the parametric bootstrapped likelihood ratio test (BLRT) in the class enumeration process. These indices are both likelihood ratio tests, and they test whether a model fits better than a corresponding model with fewer classes (Nylund et al., 2007). In addition, the overall quality of classification in each class model was inspected through the entropy values (i.e., a summary of posterior probabilities, in which values close to 1 indicate good classification accuracy; Geiser, 2013). Because our primary aim was to assess the classification of the low-performing group, we also evaluated the probability estimates for classification of individuals based on their most likely latent class membership, for this group separately (see Nylund et al., 2007, for a discussion of different fit indices in relation to how to determine the number of classes in a latent class analysis).

Finally, to assess the utility of this classification approach, we specified a second-order growth mixture model (SOGMM) based on LANGUAGE4 (at the item level) and the latent language factor (defined by vocabulary, grammatical knowledge, and sentence repetition), which allowed us to investigate and compare the developmental trajectories across different classes. A SOGMM combines a longitudinal common factor model (with measurement invariance constraints) with a second-order growth model, in addition to a growth mixture model (see Grimm & Ram, 2009, for a detailed description of a SOGMM).

Results

Estimated means, standard deviations, and bivariate correlations for all longitudinal variables at all three time points are shown in Tables 1 and 2.

Longitudinal Regression Model

To investigate the longitudinal relationship between the LANGUAGE4 factor (i.e., the general language factor derived from the LANGUAGE4 screening tool) and the latent language factor (i.e., the global latent language criterion factor derived from the other longitudinal language measures) at ages 4, 5, and 6 years, we used the model shown in Figure 1. In this model, we used a latent variable simplex model to represent the relationships between our three language outcome measures (i.e., vocabulary, grammatical knowledge, and sentence repetition, reflecting a latent language factor) at ages 4, 5, and 6 years. This simplex model assesses the longitudinal stability of the latent language factor. The stability of this factor is high between consecutive times of measurement (.93 and .91).

The LANGUAGE4 measurement model, which forms the basis for the LANGUAGE4 predictor factor in the current study, is based on 26 items (not 27 as modeled in Klem et al., 2015) because an additional naming item had to be dropped because of lack of variation. Full details of the LANGUAGE4 second-order measurement model are described elsewhere (Klem et al., 2015). The model depicted in Figure 1 shows the LANGUAGE4 factor to be a powerful concurrent predictor of the latent language criterion factor at age 4 years and shows a highly significant indirect path ($\beta = .70$, $p = .00$) to language performance two years later. This model shows a good fit to the data ($\chi^2 = 826.45$, $p < .001$, $df = 540$, root-mean-square error of approximation [RMSEA] = 0.03, 90% confidence interval [CI] [0.026, 0.034], comparative fit index [CFI] = .91, Tucker–Lewis Index [TLI] = .90). Additional direct paths from the LANGUAGE4 factor to the latent language factor at ages 5 and 6 years produced a nonsignificant improvement in fit (Satorra-Bentler adjusted chi-square difference = 2.43, $df = 2$); hence, we adopted the more restrictive model shown in Figure 1. There is also support for the assumption of weak factorial invariance for the latent language factor at ages 4, 5, and 6 years because constraining the unstandardized factor loadings between the latent language factor and each observed language measure to be equal at each time point resulted in a nonsignificant increase in chi square (Satorra-Bentler adjusted chi-square difference = 1.21, $df = 4$).

To summarize, these results show that LANGUAGE4 explains a considerable amount of the individual variation in a latent language criterion factor measured concurrently and two years later. This unidimensional latent language

Table 1. Estimated means, standard deviations, and reliabilities for repeated language measures at all time points.

Measure	Time 1 (age 4)			Time 2 (age 5)			Time 3 (age 6)		
	Mean (*SD*)	α	*n*	Mean (*SD*)	α	*n*	Mean (*SD*)	α	*n*
BPVS-II	41.77 (11.22)	.91	200	56.80 (11.35)	.91	202	72.59 (11.79)	.91	199
GramClos	11.03 (3.91)	.74	204	13.94 (3.71)	.69	202	18.37 (3.79)	.70	198
Sentence Rep	6.53 (1.92)	.63	202	7.98 (2.15)	.70	201	9.12 (2.05)	.70	200

Note. *n* = number of cases with complete data sets for each repeated language measure; BPVS-II = Norwegian version of the British Picture Vocabulary Scale–Second Edition; GramClos = grammatic closure; Sentence Rep = sentence repetition.

criterion factor also shows a high degree of longitudinal stability from ages 4 to 6 years.

Latent Class Analysis

A latent class analysis based on the LANGUAGE4 items was used to see if subgroups could be identified. Table 3 depicts the fit statistics for the class enumeration process. As can be seen, the BIC* is lowered for each added class, favoring solutions with a larger number of latent classes. On the other hand, the VLMR favors the four-class solution over the five-class solution (i.e., the five-class solution does not have a significantly better fit than the four-class solution; pVLMR = 0.21).

The entropy value for the four-latent-class model was 0.84, suggesting that the classes are reasonably well defined in this model. The low-performing group also showed high classification probabilities (all over .90) in all these solutions. This low-performing group included approximately 10% of the sample, a figure that is in line with prevalence estimates for language impairment in young children (Law, Boyle, Harris, Harkness, & Nye, 2000; Nelson et al., 2006; Tomblin et al., 1997).

Figure 2 shows the LANGUAGE4 item profiles of the four different latent classes. On the basis of our interpretation of these LANGUAGE4 item profiles, we labeled the four groups as a high-performing group (*n* = 341), an intermediate prepositions profile group (*n* = 107), an intermediate vocabulary profile group (*n* = 91), and a low-performing group (*n* = 61). The high-performing group showed high performance on most items, and the low-performing group showed considerably lower performance on most items. The low-performing group notably also showed a profile with a markedly poor performance on items requiring repetition of spoken sentences (sentence repetition). The two intermediate groups (intermediate prepositions profile group and intermediate vocabulary profile group) showed item profiles that were weaker than those of the high-performing group, but still considerably higher than those of the low-performing group. However, the two intermediate groups were distinguishable in the sense that they had different relative weaknesses. The intermediate prepositions profile group showed a relative weakness on items requiring the understanding of prepositions. On the other hand, the intermediate vocabulary profile group showed a relative weakness on items requiring descriptions of words (word definitions) but also, to some extent, on items involving sentence repetition. All four subgroups showed a high level of performance, close to ceiling, on the naming items, showing that these items are very easy and are not useful in discriminating between children (see Klem et al., 2015, for a full description of the different LANGUAGE4 items and the corresponding first-order factors).

Table 2. Estimated correlations between all repeated measures at all time points.

Measures and time points	1	2	3	4	5	6	7	8	9
1. SR T1	—								
2. BPVS T1	.319	—							
3. GC T1	.351	.380	—						
4. SR T2	.487	.229	.282	—					
5. BPVS T2	.347	.539	.314	.292	—				
6. GC T2	.349	.341	.419	.304	.415	—			
7. SR T3	.344	.206	.332	.553	.277	.314	—		
8. BPVS T3	.207	.465	.333	.246	.565	.324	.232	—	
9. GC T3	.320	.290	.403	.173	.328	.571	.284	.406	—

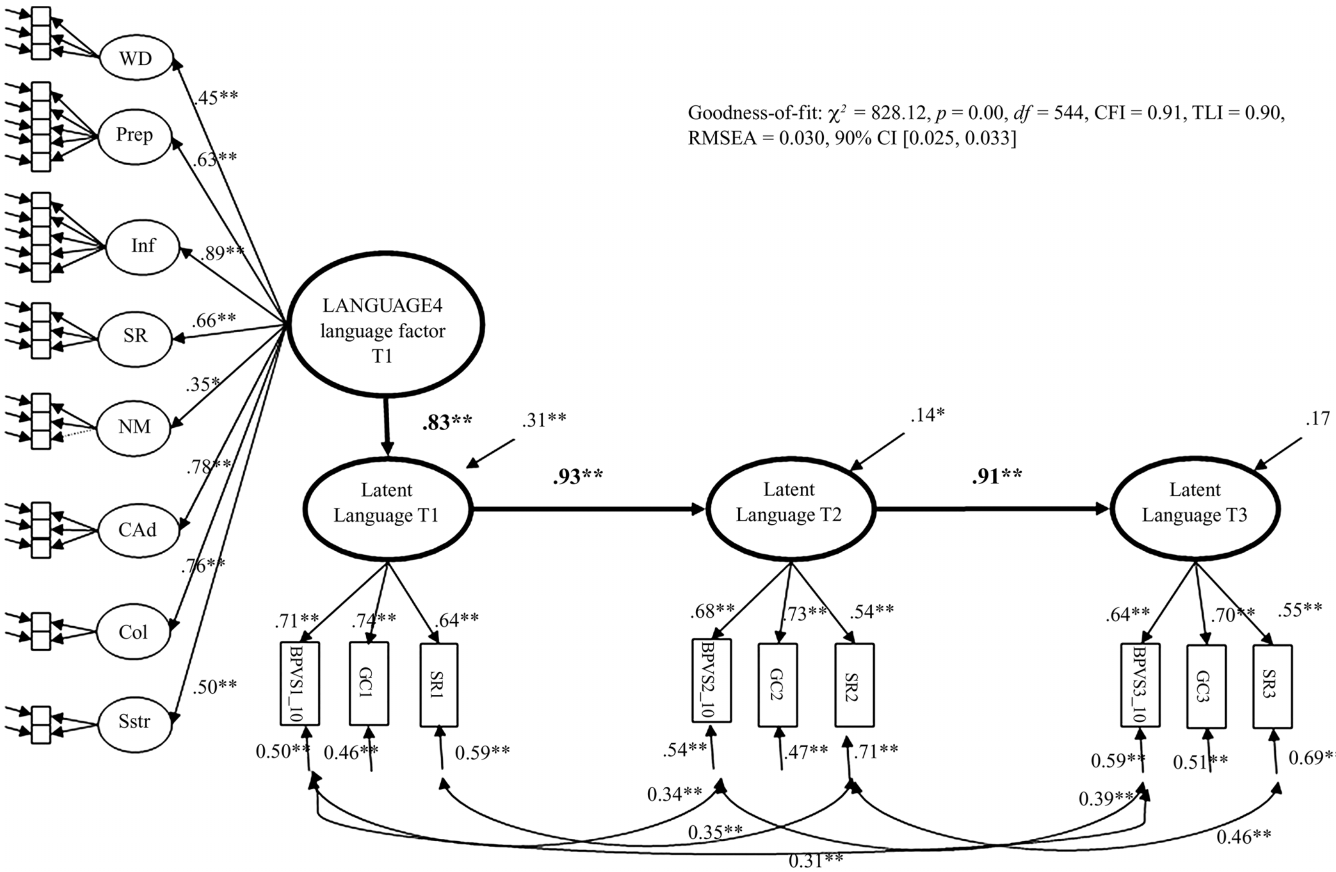
Note. SR = sentence repetition; BPVS = British Picture Vocabulary Scale–Second Edition; GC = grammatic closure; T = time.

Second-Order Growth Mixture Models

A SOGMM provided a means of examining the utility of the latent class approach to classification, as this enabled us to investigate whether the low-performing group would remain at a relative stable disadvantage over time, compared to their peers (cf. the tracking hypothesis; Conti-Ramsden et al., 2012; Law et al., 2008).

We first fitted a second-order linear latent growth curve model (χ^2 = 67.28, p < .001, df = 36, RMSEA = 0.065, 90% CI [0.040, 0.089], CFI = .94, TLI = .94). The fit indices for this model indicate that a linear model is an acceptable description of the pattern of growth in language skills for children, in the sample as a whole. Next we specified a SOGMM to assess whether there were different growth trajectories across the four latent classes. Testing of a constrained model in which the mean intercept was constrained to be equal across classes (BIC* = 17,875.84) against a model with freely estimated means for the intercept suggests that the model with varying intercepts (BIC* = 17,855.00) best accounts for the data (Δ BIC* = 20.84). To provide a

Figure 1. Model of the longitudinal relationship between LANGUAGE4 and a latent language factor.



Note. WD = word definition, Prep = prepositions, Inf = inferences, SR = sentence repetition, NM = naming, Cad = comparative adjectives, Col = color, Sstr = sentence structure; all from LANGUAGE4. BPVSx_10 = British Picture Vocabulary Scale-II/10 (full scale divided by 10), GCx = grammatic closure, SRx = sentence repetition, T = time, CFI = comparative fit index, TLI = Tucker-Lewis index, RMSEA = root mean square error of approximation, ** $p < 0.01$, * $p < 0.05$.

point of reference when comparing the growth estimates across the different latent classes, we also specified a SOGMM with one class (i.e., the full sample; $N = 600$).

Table 4 shows estimates of the means and variances for the intercepts and slopes for the two specified SOGMMs (i.e., one-class and four-class models). The estimates for mean intercepts across the four classes show that the groups differ in initial level. The effect sizes found for the differences in intercepts between the low-performing group and the other groups are large (Cohen's *d* in favor of high-performing group = −3.14; intermediate preposition profile group = −2.20; intermediate vocabulary profile group = −2.10), confirming that the low-performing group showed considerably lower performance on the latent language factor at the time of identification than the other three groups.

Furthermore, the means of the slopes of the growth factor were significant for all subgroups, showing that all subgroups, on average, demonstrated improvements in language skills over time. We tested whether there were differences in mean slopes across the four classes by comparing the less restricted model (i.e., freely estimated slope means across classes with BIC* = 17,855.00) against a more restricted model with equality constraints on slopes across all classes (BIC* = 17,851.37). The difference between these models was nonsignificant (Δ BIC* = 3.63), indicating that the model with parallel slopes for the different classes provides the most parsimonious description of the data.

The variance of the slope growth factor for the one-class model was not significant (slope variance = 0.02, $p = .56$), suggesting that there is no reason to believe that the slope variance would differ across the subgroups. For this reason, as well as issues related to convergence of the model (i.e., negative residual variances), the four-class SOGMM had an equality constraint on the variances of the intercepts and slopes across all groups. Difficulties with convergence appear to be a common problem in second-order growth mixture modeling (Grimm & Ram, 2009). The nonsignificant slope variance within classes (slope variance = 0.01, $p = .82$) further confirmed that the individual variation of growth is trivial within classes as well.

The second-order growth mixture model was estimated from the combined data from the LANGUAGE4 assessment

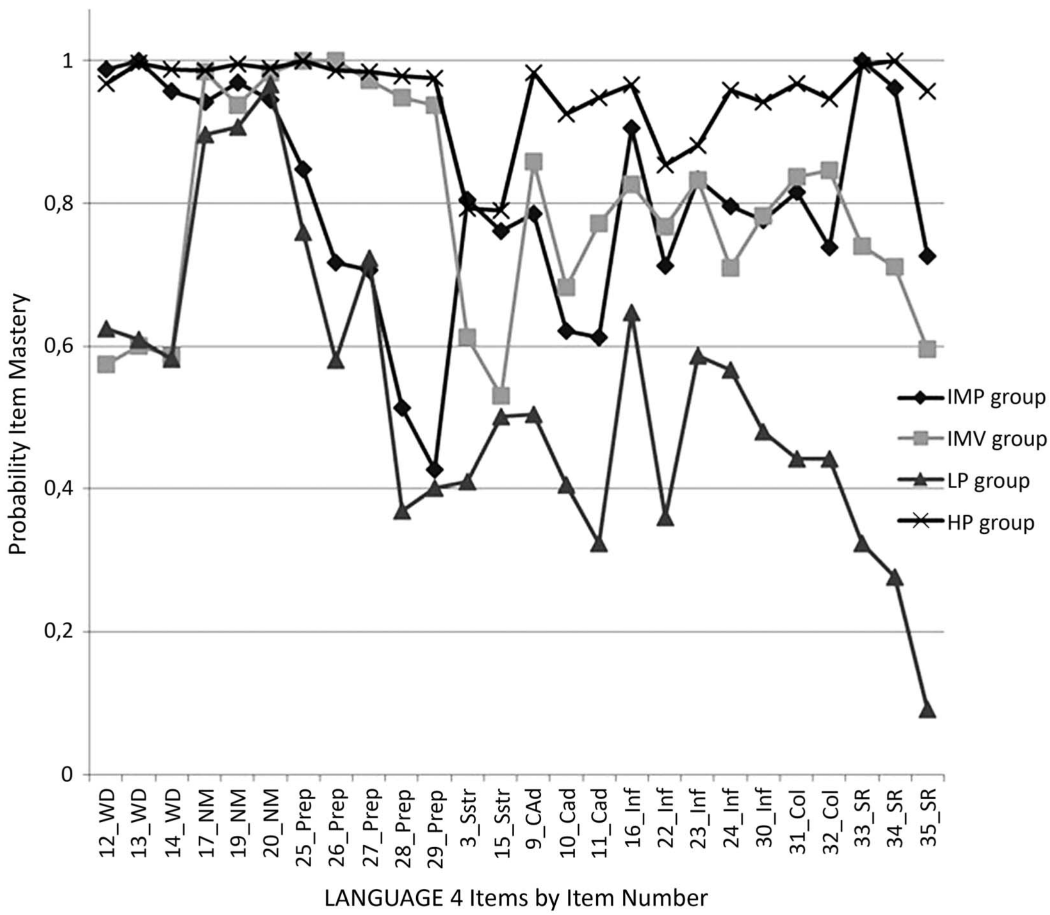
Table 3. Fit statistics for the different latent class models based on the 26 LANGUAGE4 items.

Model	Class count	Log L	BIC*	pVLMR	pBLRT	Number of free parameters	Entropy
Two class	435, 165	–5,317.46	10,805.69	0.000 (0.000)	0.000	53	0.878
Three class	388, 115, 97	–5,212.84	10,683.45	0.037 (0.038)	0.000	80	0.848
Four class	341, 107, 91, **61**	–5,100.74	10,546.25	0.019 (0.020)	0.000	107	0.838
Five class	349, 118, **59**, 42, 32	–5,033.56	10,498.89	0.212 (0.215)	0.000	134	0.881
Six class	356, 83, **58**, 38, 34,31	–4,987.96	10,494.70	0.320 (0.322)	0.000	161	0.901

Note. Class count = number of individuals classified in each class based on their most likely latent class membership (boldface indicates class count for the low-performing class across models); BIC* = sample-size adjusted BIC; pVLMR = probability value for the Vuong-Lo-Mendell-Rubin likelihood ratio test (adjusted values in parentheses); pBLRT = probability value for the parametric bootstrapped likelihood ratio test.

and the three repeated administrations of the three language tests. Table 4 shows that there is a small difference in the number of children in the four classes of the latent class model based on LANGUAGE4 items alone (see Table 3) compared to the four classes in the second-order growth mixture model. These differences likely reflect the fact that the parameters of the probabilistic algorithm for class assignment differ slightly between the two models. It is important to note that, based on probability estimates for their most likely latent class membership, there is only a

Figure 2. LANGUAGE4 item profiles based on a latent class analysis.



Note. x-axis = LANGUAGE4 items by item number labeled with the corresponding item factor label: WD = word definition, NM = naming, Prep = prepositions, Sstr = sentence structure, CAd = comparative adjectives, Inf = inference, Col = Color, SR = sentence repetition, HP group = high-performing group, LP group = low-performing group, IMV group = intermediate-vocabulary-profile group, IMP group = intermediate-prepositions-profile group.

Table 4. Estimated means and variances for intercepts and slopes for the two second-order growth mixture models.

Models	Full		Four-class							
Latent class label	One class (*n* = 600)		HP (*n* = 331)		IMP (*n* = 114)		IMV (*n* = 93)		LP (*n* = 62)	
Latent language factor growth parameters	Means	Variances	Means	Variances	Means	Variances	Means	Variances	Means	Variances
Intercept (*p*)	4.18 (.00)	0.67 (.00)	4.39 (.00)	0.49 (.00)	3.62 (.00)	0.49 (.00)	3.54 (.00)	0.49 (.00)	1.82 (.00)	0.49 (.00)
Slope (*p*)	1.53 (.00)	0.02 (.56)	1.47 (.00)	0.01 (.82)	1.71 (.00)	0.01 (.82)	1.70 (.00)	0.01 (.82)	1.48 (.00)	0.01 (.82)

Note. Full = one-class model (i.e., point of reference with the full sample); HP = high-performing class; IMP = intermediate prepositions class; IMV = intermediate vocabulary class; LP = low-performing class.

minor difference in the number of individuals classified in the low-performing group, showing that the identification is relatively stable across the two models.

In summary, by using latent class analysis and second-order growth mixture modeling we found that a distinct low-performing subgroup of children is identifiable based on the LANGUAGE4 items and that this class shows a large difference in starting level (i.e., mean intercept at age 4 years) compared to the three subgroups of intermediate- and high-performing children. It is important to note that the rates of growth (mean slopes) during this 2-year period appeared to be parallel across all subgroups, which in turn implies that the initial gap that the low-performing group displayed in relation to their peers persists over time.

Discussion

We first discuss the longitudinal relationship between LANGUAGE4 (i.e., the hierarchical model with one second-order LANGUAGE4 factor) and later language abilities (i.e., the longitudinal latent language factor). Then, we address the issue of growth in language skills and the latent class analysis approach to identifying a low-performing subgroup of children.

The LANGUAGE4 Factor as a Predictor of Later Language Abilities

The results from the longitudinal autoregressive model (see Figure 1), in which the LANGUAGE4 factor was related to a latent language factor both concurrently (at age 4 years) and longitudinally (at ages 5 and 6 years), confirm the previous concurrent findings of Klem et al. (2015) that the LANGUAGE4 factor explains a considerable amount of the latent language criterion factor at age 4 years. Expanding on this finding, we found that the longitudinal relation of LANGUAGE4 to later language performance at age 6 years was strong but operated only indirectly via the latent language construct at ages 4 and 5 years (i.e., LANGUAGE4 does not predict later language performance over and above the autoregressive effect).

The indirect effect of LANGUAGE4 on later language performance was expected because the stability of childhood language appears to be strong from an early age (Bornstein et al., 2014; Bornstein & Putnick, 2012; Rice & Hoffman, 2015). The high degree of longitudinal stability in the latent language factor implies that the ranking of individuals within the group is largely preserved over time.

The high stability of childhood language skills demonstrates that language outcomes at a given age are strongly predicted by language skills at an earlier age (Bornstein & Putnick, 2012). From this perspective, the second-order LANGUAGE4 factor appears to be a powerful longitudinal predictor of later language outcomes. However, despite convincing evidence for the strong stability of individual differences in language skills in children from ages 4 to 6 years, considerable common variance remains unaccounted for at ages 5 and 6 years, which shows that individual children can still change their status relative to each other in terms of their language abilities over time (Bornstein & Putnick, 2012). Previous research suggests that a range of cognitive skills not assessed in the current study, including processing speed and working memory skills, may predict variance in children's language skills (Leonard et al., 2007).

Identification and Developmental Trajectories of Low-Performing Children

Our findings from the mixture growth approach complement the results from the autoregressive approach. Our particular focus here was to assess whether a latent class approach would identify a stable, well-defined low-performing group of children on the basis of LANGUAGE4.

On the basis of a latent class analysis of the LANGUAGE4 data, we identified a low-performing group of children whose language skills were considerably poorer than the rest of the sample. These children represented approximately 10% of the full sample ($N = 600$). The results from the second order growth mixture model showed that the low-performing group on average had a large initial gap in language ability at age 4 years compared to their higher performing peers. It is important to note that the estimated growth trajectories were parallel to those of their peers, which suggests that the low-performing children at age 4 years are likely to remain low performing compared to their peers at the age of 6 years in the absence of effective language interventions. On this basis, we argue that the use of latent class analysis from the LANGUAGE4 screening data provides a promising approach for identifying 4-year-olds in need of language support. The use of a computer-based scoring system for screening data would easily enable this analysis to be applicable to clinical practice as the basis for telling clinicians which children may be at risk for language impairment. In short, the current research has implications for clinical practice and suggests that routine screening for language delays could be implemented in a cost-effective manner and would be likely to be effective in identifying a group of low-performing children who would benefit from language enrichment programs in preschool and the early school years.

An advantage of using latent class analysis over a cutoff score lies in the former's ability to identify low-performing children based on their observed performance profile being distinctly different from other profiles. Even though children in the low-performing group showed overall low performance, their performance varied across items, implying that the latent class analysis uses the available information more efficiently than a simple unit-weighted total score. Moreover, any set cutoff score would be arbitrarily set (e.g., the lowest 10%), while the latent class approach to identification is based on explicit principles and algorithms. It is interesting to note that latent class analysis also provides probability estimates of class membership (Coghill & Sonuga-Barke, 2012), which would be helpful clinically in the case of children who may fall just short of a typical cutoff criterion.

The LANGUAGE4 item profile of the low-performing group differed from those of the other three latent classes in the sense that they had generally lowered probabilities for correctly answering several items sets (e.g., items related to word definitions, prepositions, inference making, and sentence repetition). The characteristics of this low-performing profile is also in accordance with the description of children who should be considered to be at risk for language impairment according to the LANGUAGE4 manual (Horn & Dalin, 2008), which states that a poor performance on several different tasks (e.g., items of inference, sentence repetition, colors), rather than isolated weaknesses on a few item sets, should lead to a further referral of the child to an appropriate speech and language service (Horn & Dalin, 2008).

From a methodological point of view, the use of latent class analysis in this context has enabled us to identify a low-performing group that is very much in line with the description of who should be identified as at risk for language impairment according to the manual. Still, the low-performing group of children is likely to be the one falling at the lower end of the normal distribution (hence the large difference in intercepts at age 4 years). However, using latent class analysis in this context appears beneficial because it provides an alternative, more probabilistic approach to identifying children who may be at risk of language impairment. The stability of the latent class of low-performing children across different class solutions suggests that the classification of this group is robust.

Assessing the utility of such a classification clearly involves questions related to prediction of later outcomes. Our findings from the SOGMM indicate a substantial similarity in growth trajectories at the second-order level across all the identified latent classes. As such, this finding supplements the reported results from the autoregressive approach by suggesting that the longitudinal stability of language abilities is consistent across the latent classes. In line with previous studies on developmental pathways of language abilities in children with specific language impairment (e.g., Conti-Ramsden et al., 2012; Law et al., 2008; Rice, 2004), the current finding of parallel growth trajectories across all the latent classes suggests that, other things equal, the children in the low-performing group are likely to remain low relative to their age peers over time. This pattern provides support for the tracking hypothesis (Conti-Ramsden et al., 2012; Law et al., 2008) as the most appropriate description of growth trajectories across the latent classes. The results also indicate that this pattern of growth is rather homogeneous both across and within groups. Moreover, this finding suggests that the low-performing group identified on the basis of a low LANGUAGE4 profile may be considered at risk for a persisting language impairment.

Although there appears to be strong stability in the development of childhood language, it is important to note that such stability does not mean that language skills are not amenable to intervention (Bornstein et al., 2014; Bornstein & Putnick, 2012). One limitation of the current study is that we have not accounted for the possible impact that educational interventions may have had on language development during this 2-year time period. It is clear, however, that to the extent to which the low-performing children have received language interventions during this time period, their scope and impact were not sufficient to substantially alter their language skills. Moreover, the low-performing group likely includes children with low language abilities of different etiology, and there is reason to believe that the impact of different interventions would vary in accordance with such heterogeneity as well. It is interesting to note that the findings of parallel growth trajectories also holds for clinical samples (i.e., Conti-Ramsden et al., 2012; Rice, 2013) in which children are known to receive speech and language therapy. This finding highlights the need for further research to examine the different mechanisms that govern growth in language acquisition in childhood in general, as well as across different subpopulations (Rice, 2013). One other limitation of this study is the relatively short time period covered; it would be most valuable to have longer term follow-up data on children whose language skills are deficient in the age range studied here (4 to 6 years).

Conclusion

We have reported data from a language screening procedure administered at age 4 years. Using a latent class analysis approach to categorization, we showed that it is possible to identify a well-defined, low-performing group of children on the basis of this simple screening tool. This low-performing group represents roughly the bottom 10% of children in the sample, and they appear to have substantial and global language weaknesses. Subsequent longitudinal analyses indicated that these children showed enduring language difficulties between the ages of 4 to 6 years. However, it is notable that our growth curve models showed that the rate of language development in these children is comparable to the rate of growth in the rest of sample. We believe our results demonstrate that population-based language screening at age 4 years is viable and can be used to identify children who are likely to show persistent language weaknesses. Further research is clearly necessary to confirm the utility of this approach to screening and to identify the extent to which early language difficulties can be ameliorated by suitable interventions (see Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013, for a study that describes a promising approach to intervention).

Acknowledgments

The study was funded by the Norwegian Research Council (grant number 185459/F10). We greatly appreciate the participation of the children, parents, and public health nurses in this study. We also wish to thank Erna Horn for kindly giving us permission to use the norm data of LANGUAGE4 as a part of our analysis.

References

Bishop, D. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language & Communication Disorders, 49,* 381–415. doi:10.1111/1460-6984.12101

Bishop, D. V. M., & Edmundson, A. (1987). Language-impaired 4-year-olds: Distinguishing transient from persistent impairment. *Journal of Speech and Hearing Disorders, 52,* 156–173.

Bornstein, M. H., Hahn, C.-S., Putnick, D. L., & Suwalsky, J. T. D. (2014). Stability of core language skill from early childhood to adolescence: A latent variable approach. *Child Development, 85,* 1346–1356. doi:10.1111/cdev.12192

Bornstein, M. H., & Putnick, D. L. (2012). Stability of language in childhood: A multiage, multidomain, multimeasure, and multisource study. *Developmental Psychology, 48,* 477–499.

Coghill, D., & Sonuga-Barke, E. J. (2012). Annual research review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders—Implications of recent empirical study. *Journal of Child Psychology and Psychiatry, 53,* 469–489. doi:10.1111/j.1469-7610.2011.02511.x

Conti-Ramsden, G., St. Clair, M. C., Pickles, A., & Durkin, K. (2012). Developmental trajectories of verbal and nonverbal skills in individuals with a history of specific language impairment: From childhood to adolescence. *Journal of Speech, Language, and Hearing Research, 55,* 1716–1735.

Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay 1. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research, 46,* 544–560.

Darrah, J., Hodge, M., Magill-Evans, J., & Kembhavi, G. (2003). Stability of serial assessments of motor and communication abilities in typically developing infants—Implications for screening. *Early Human Development, 72,* 97–110. doi:10.1016/S0378-3782(03)00027-6

Dockrell, J., Ricketts, J., & Lindsay, G. (2012). *Understanding speech, language and communication needs: Profiles of need and provision*. London, United Kingdom: Department for Education.

Dollaghan, C. A. (2004). Taxometric analyses of specific language impairment in 3- and 4-year-old children. *Journal of Speech, Language, and Hearing Research, 47,* 464–475.

Dollaghan, C. A. (2011). Taxometric analyses of specific language impairment in 6-year-old children. *Journal of Speech, Language, and Hearing Research, 54,* 1361–1371. doi:10.1044/1092-4388(2011/10-0187)

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale* (2nd ed.). London, United Kingdom: GL Assessment.

Eadie, P., Nguyen, C., Carlin, J., Bavin, E., Bretherton, L., & Reilly, S. (2014). Stability of language performance at 4 and 5 years: Measurement and participant variability. *International Journal of Language & Communication Disorders, 49,* 215–227. doi:10.1111/1460-6984.12065

Ellis, E. M., & Thal, D. J. (2008). Early language delay and risk for language impairment. *Perspectives on Language Learning and Education, 15,* 93–100. doi:10.1044/lle15.3.93

Eriksson, M., Westerlund, M., & Miniscalco, C. (2010). Problems and limitations in studies on screening for language delay. *Research in Developmental Disabilities, 31,* 943–950. doi:10.1016/j.ridd.2010.04.019

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry, 54,* 280–290. doi:10.1111/jcpp.12010

Geiser, C. (2013). *Data analysis with Mplus*. New York, NY: Guilford Press.

Grimm, K. J., & Ram, N. (2009). A second-order growth mixture model for developmental research. *Research in Human Development, 6,* 121–143. doi:10.1080/15427600902911221

Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R. (2013). Language impairment from 4 to 12 years: Prediction and etiology. *Journal of Speech, Language, and Hearing Research, 57,* 850–864. doi:10.1044/2013_JSLHR-L-12-0240

Horn, E., & Dalin, A. (2008). *SPRÅK 4: Brukerveiledning* [LANGUAGE4: User's manual] (Revised ed.). Bergen, Norway: Designtrykkeriet.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences, 2,* 389–398.

Kirk, S. A., McCarthy, J. D., & Kirk, W. D. (1968). *Illinois Test of Psycholinguistic Abilities* [Measurement instrument]. Norwegian version by Gjessing, H. J., & Nygaard, H. D., 1975. Urbana, IL: University of Illinois Press.

Klem, M., Gustafsson, J.-E., & Hagtvet, B. (2015). The dimensionality of language ability in 4-year-olds: Construct validation of a language screening tool. *Scandinavian Journal of Educational Research, 59,* 195–213. doi:10.1080/00313831.2014.904416

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). The feasibility of universal screening for primary speech and language delay: Findings from a systematic review of the literature. *Developmental Medicine & Child Neurology, 42,* 190–200. doi:10.1111/j.1469-8749.2000.tb00069.x

Law, J., Tomblin, J. B., & Zhang, X. (2008). Characterizing the growth trajectories of language-impaired children between 7 and 11 years of age. *Journal of Speech, Language, and Hearing Research, 51,* 739–749. doi:10.1044/1092-4388(2008/052)

Leonard, L. B., Weismer, S. E., Miller, C. A., Francis, D. J., Tomblin, J. B., & Kail, R. V. (2007). Speed of processing, working memory, and language impairment in children.

Journal of Speech, Language, and Hearing Research, 50, 408–428. doi:10.1044/1092-4388(2007/029)

Lyster, S.-A. H., Horn, E., & Rygvold, A.-L. (2010). Ordforråd og ordforrådsutvikling hos norske barn og unge. Resultater fra utprøvingen av British Picture Vocabulary Scale II, second edition (BPVS-II). [Vocabulary and vocabulary development among Norwegian children and youth. Results from the trial of the British Picture Vocabulary Scale II, second edition]. *Spesialpedagogikk,* (09), 35–43.

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nelson, H. D., Nygren, P., Walker, M., & Panoscha, R. (2006). Screening for speech and language delay in preschool children: Systematic evidence review for the US Preventive Services Task Force. *Pediatrics, 117,* 298–319. doi:10.1542/peds.2005-1467

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14,* 535–569.

Ottem, E., & Frost, J. (2005). *Språk 6–16: Screening test: Manual* [Language 6–16: Screening test. User's manual]. Oslo, Norway: Bredtvet kompetansesenter.

Pennington, B. F., Santerre-Lemmon, L., Rosenberg, J., MacDonald, B., Boada, R., Friend, A., . . . Olson, R. K. (2012). Individual prediction of dyslexia by single vs. multiple deficit models. *Journal of Abnormal Psychology, 121,* 212–224. doi:10.1037/a0025823

Rice, M. L. (2004). Growth models of developmental language disorders. In M. L. Rice & S. F. Warren (Eds.), *Developmental language disorders: From phenotypes to etiologies* (pp. 207–240). Mahwah, NJ: Erlbaum.

Rice, M. L. (2013). Language growth and genetics of specific language impairment. *International Journal of Speech-Language Pathology, 15,* 223–233. doi:10.3109/17549507.2013.783113

Rice, M. L., & Hoffman, L. (2015). Predicting vocabulary growth in children with and without specific language impairment: A longitudinal study from 2;6 to 21 years of age. *Journal of Speech, Language, and Hearing Research, 58,* 345–359. doi:10.1044/2015_JSLHR-L-14-0150

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461–464. doi:10.1214/aos/1176344136

Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52,* 333–343.

Snowling, M. J., Duff, F., Nash, H., & Hulme, C. (2015). Language profiles and literacy outcomes of children with resolving, emerging or persisting language impairments. *Journal of Child Psychology and Psychiatry*. Advance online publication. doi:10.1111/jcpp.12497

Sosial-og helsedirektoratet [The Norwegian Directorate for Health and Social Affairs]. (2006). *Nasjonale retningslinjer for undersøkelse av syn, hørsel, og språk hos barn.* [National clinical guidelines for examinations of vision, hearing, and language abilities in children]. Oslo, Norway: Author.

Taylor, C. L., Christensen, D., Lawrence, D., Mitrou, F., & Zubrick, S. R. (2013). Risk factors for children's receptive vocabulary development from four to eight years in the Longitudinal Study of Australian Children. *PLoS ONE, 8,* e73046.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40,* 1245–1260.

Wechsler, D. (1989). *WPPSI-R manual: Wechsler Preschool and Primary Scale of Intelligence–Revised*. San Antonio, TX: Harcourt Brace Jovanovich.

Yang, C.-C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistics & Data Analysis, 50,* 1090–1104. doi:10.1016/j.csda.2004.11.004

Zambrana, I. M., Pons, F., Eadie, P., & Ystrom, E. (2014). Trajectories of language delay from age 3 to 5: Persistence, recovery, and late onset. *International Journal of Language & Communication Disorders, 49,* 304–316.