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

Marianne Klem
University of Oslo and Statped, Department of Speech and Language Disorders

Bente Hagtvet
University of Oslo

Charles Hulme
University College London and University of Oslo

Jan-Eric Gustafsson
University of Gothenburg and University of Oslo
Abstract

Purpose: This study investigates 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: We present data from a large-scale two-year longitudinal study, in which 600 children were assessed with a language screening tool (LANGUAGE4) at age 4 years. A sub-sample (N = 206) were assessed on measures of sentence repetition, vocabulary and grammatical knowledge at ages 4, 5 and 6 years. Results: A global latent language factor shows a high degree of longitudinal stability in children from age 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 LANGUAGE4. Importantly, the growth-rates during this two-year time period were parallel for the low- and three higher-performing groups of children.

Conclusions: There is strong stability in children’s language skills between the ages of 4 and 6 years. Our 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.

Key words: screening, latent class analysis, language impairment, growth trajectories
Screening for language delay: Growth Trajectories of Language Ability in Low- and High-Performing Children

Introduction

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 modelling approach with items from a language screening tool (LANGUAGE4 [SPÅK4]; Horn & Dalin, 2008), this paper investigates the stability of children’s language development between 4 to 6 years of age. 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 (see, e.g., Ellis & Thal, 2008 for a brief overview). Different cutoff criteria are used to define language impairments, as no consensus exists on the distinction between impaired, delayed and typical language development (Bishop, 2014; Dollaghan, 2004; Dollaghan, 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 per cent to 19 per cent; 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., 2013).

While recognising 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 surveillance programme 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 minutes to administer, it is efficient and simple while still covering a relatively wide range of language skills that most four-year-olds, according to developmental studies, master (Horn & Dalin, 2008). Recently, 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 paper), that explained a substantial amount of the variance ($R^2=0.61$) 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, 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 language-impaired children may have an initial delay, but that their developmental trajectories are parallel to those of typically developing children (Law et al, 2008; Conti-Ramsden et al, 2012). 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 and across different sub-dimensions 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; Zambrana et al., 2014; Snowling, Duff, Nash & Hulme, in press).

In summary, a major difficulty in screening for language delay is the mismatch between the screening method, primarily developed for categorising 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 (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., 2013), accelerated language growth in early childhood appears to occur in some children who show an initial delay (Conti-Ramsden et al., 2012; Taylor et al., 2013). Conversely, other children may show relatively normal early language
development but go on to develop late emerging language impairments (Dale, et al., 2003; Zambrana et al., 2014; Snowling, Duff, Nash & Hulme, in press).

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 recognised in research (Pennington et al., 2012). Here we emphasize the probabilistic nature of predicting child language outcomes and use a structural equation modelling (SEM) approach to assess the trajectories of language development of sub-groups 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 that assesses 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 allows us to examine whether the equivalent language construct is measured over time. Moreover, using a population-based sample, the current study explores 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 cut-off, Latent Class Analysis allows for a classification based on different performance-profiles and also 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 is on the identification of a stable low-performing group.
Methods

Participants

This study uses a population-based sample of 600 (274 girls, 326 boys) 4-year-old children (mean 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 centre 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 centres in the eastern part of Norway. This subsample consists of 4-year-olds who were assessed with the LANGUAGE4 screening tool at their local health care centres 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

The study reports 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), whilst additional longitudinal data from non-overlapping tests (i.e., measures of vocabulary, grammatical knowledge and sentence repetition) were obtained from sub-sample 2 (N = 206); consequently, data is missing by design.

The LANGUAGE4 assessment procedure was administered as a part of the Norwegian general health surveillance programme for 4 year olds: children were individually assessed at their regional health care centre 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 centres 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 (206 children) were given the additional longitudinal measures. These longitudinal language measures (vocabulary, grammatical knowledge and sentence repetition) were a 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). Our analytical approach to LANGUAGE4 is thus at an item-level. Because 6 items lacked the required variation to be fitted to a model, the structure of LANGUAGE4 has been modelled based on 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 sub-skills at the first-order level; i.e., word definition-, naming-, colour-, comparative adjectives-, preposition-, 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 (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 orally presented 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 (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 (i.e., inflection of nouns, verbs, adjectives etc.). Testing was stopped after 6 consecutive incorrect items at Times 1 and 2, whilst all items were presented at Time 3.

**Sentence Repetition.** The 21-item sentence repetition task is an adapted version of the Sentence Memory sub-test 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 sub-test was supplemented with the first five items of the sentence repetition test from the Norwegian version of the 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 Modelling (SEM) with Mplus version 7.11 (Muthén & Muthén, 1998–2012) using the maximum likelihood estimator with robust standard errors (MLR) which is robust to non-normality of data (Muthén & Muthén, 1998–2012).

The data includes both categorical and continuous indicators, i.e., categorical items from LANGUAGE4 and continuous total scores from the longitudinal language tests. Moreover, we assume that two different missing data mechanisms are present. Missing items in LANGUAGE4 are assumed to be MCAR (missing completely at random) since individual children on occasions failed to be 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 MAR (missing at random) 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 item-level) as a means of identifying different sub-groups 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 categorises individuals into classes and also identifies items that best distinguish between classes (Nylund, Asparouhoff, & Muthén, 2007).
To assess the number of latent classes needed, we relied on three fit indexes: the Bayesian Information Criteria (BIC; Schwarz, 1978), where 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 BIC (BIC*; Sclove, 1987) represents noteworthy improvements over BIC (Yang, 2006). Moreover, Nylund et al. (2007) suggest that BIC* also appears to be more consistent when applied to categorical LCA 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, where values close to one indicate good classification accuracy (Geiser, 2013). Because our primary aim was to assess the classification of the low-performing group, we also evaluated “the average latent class probability value for most likely 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 LCA).

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 (.91 and .93).

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 modelled in Klem et al., 2015), because an additional naming item had to be dropped due to 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 and also 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 ($\chi^2 = 826.45$, $p < .001$, $df = 540$, RMSEA = 0.03 CI90 0.026-0.034, CFI=.91, TLI=.90): including additional direct paths from the LANGUAGE4 factor to the latent language factor at ages 5 and 6 produced a non-significant improvement in fit (Satorra-Bentler adjusted chi-square difference = 2.43, $df = 2$), and 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 since 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 non-significant increase in chi square
(Satorra-Bentler adjusted chi-square difference = 1.21, $df = 4$).

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

**Latent Class Analysis**

A latent class analysis based on the LANGUAGE4 items, was used to see if sub-
groups 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, favouring solutions with a larger
number of latent classes. On the other hand, the VLMR favours 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; $p_{VLMR} = 0.21$).

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 0.90) in all these solutions. This low-performing group
included approximately 10 per cent 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.
Based on our interpretation of these LANGUAGE4 item profiles, we labelled 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. Notably, the Low Performing group 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 and Intermediate-Vocabulary-profile) showed item profiles which 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. Lastly, all four sub-groups 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).

Second-Order Growth Mixture Models

A second order growth mixture model (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; Law et al., 2008; Conti-Ramsden et al., 2012).

We first fitted a second-order linear latent growth curve model ($\chi^2=67.28$, $p < 0.001$, $df=36$, RMSEA=0.065 CI90 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, where the mean intercept was constrained to be equal across classes (BIC* = 17875.84) against a model with freely estimated means for the intercept suggests that the model with varying intercepts (BIC* = 17855.00) best accounts for the data (Δ BIC* = 20.84). To provide a 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 favour of High Performing = -3.14; Intermediate-Preposition-profile = -2.20; Intermediate-Vocabulary-profile = -2.10), confirming that the Low Performing group shows considerably lower performance on the latent language factor at the time of identification than the other 3 groups.

Furthermore, the means of the slopes of the growth factor were significant for all sub-groups, showing that all sub-groups, 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* = 17855.00) against a more restricted model with equality constraints on slopes across all classes (BIC* = 17851.37). The difference between these models was non-significant (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 sub-groups. 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 appears to be a common problem in second-order growth mixture modelling (Grimm & Ram, 2011). The non-significant 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 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 are likely to reflect the fact that the parameters of the probabilistic algorithm for class assignment differ slightly between the two models. Importantly, based on probability estimates for their most likely latent class membership there is only a 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 Modelling, we have found that a distinct Low Performing sub-group 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) compared to the three sub-groups of intermediate- and high-performing children. Most importantly, the rates of growth (mean slopes) during this 2-year
period appeared to be parallel across all sub-groups, which in turn implies that the initial gap that the Low Performing group displays 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) and then address the issue of growth in language skills and the Latent Class Analysis approach to identifying a low-performing sub-group 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) and longitudinally (at ages 5 and 6), 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. Expanding on this finding, we found that the longitudinal relation of LANGUAGE4 to later language performance at age 6 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, 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, Weismer,
Miller, Francis, Tomblin, & Kail, 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.

Based on 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 per cent 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 compared to
their higher-performing peers. Importantly, the estimated growth trajectories were parallel to
those of their peers, which suggests that the Low Performing children at age 4 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 who 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 cost effectively, and would be likely to be effective in identifying a group of low performing children who would benefit from language enrichment programmes in preschool and the early school years.

An advantage of using Latent Class Analysis over a cut-off 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 the Low Performing group show overall low performance, their performance varies across items, implying that the latent class analysis uses the available information more efficiently than a simple unit-weighted total score. Moreover, any set cut-off score would be arbitrarily set (i.e. the lowest 10 %), while the latent class approach to identification is based on explicit principles and algorithms. Notably, 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 cut-off 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). Importantly, the characteristics of this Low Performing profile is also in accordance with the description of children who should be considered as 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, colours), 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). 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.

Clearly, assessing the utility of such a classification involves questions related to prediction of later outcomes. Our findings from the Second-Order Growth Mixture Model 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 SLI (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 (Law et al., 2008; Conti-Ramsden et al., 2012) as the most appropriate description of growth trajectories across the latent classes. The results also indicate that this pattern of growth is rather homogenous 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.
Finally, 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 & Putnick, 2012; Bornstein et al. 2014). 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 two-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 have not been sufficient to substantially alter their language skills. Moreover, the Low Performing group is likely to include 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. Notably, the findings of parallel growth trajectories also holds for clinical samples (i.e., Conti-Ramsden et al., 2012; Rice, 2013), where 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 sub-populations (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 show 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 show that these children show enduring language difficulties between the ages of 4-6.
years. However, strikingly, our growth curve models show 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. Clearly further research is 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 et al., 2013; for a study that describes a promising approach to intervention).

**Acknowledgements**

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


assessments of motor and communication abilities in typically developing infants—
Implications for screening. *Early Human Development*, 72(2), 97–110. doi:
http://dx.doi.org/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: Department for
Education.

Dollaghan, C. A. (2004). Taxometric analyses of specific language impairment in 3- and 4-
475.

Dollaghan, C. A. (2011). Taxometric analyses of specific language impairment in 6-year-old
10.1044/1092-4388(2011/10-0187)


language performance at 4 and 5 years: Measurement and participant variability.
*International Journal of Language & Communication Disorders, Advanced Online


Leonard, L.B., Weismer, S.E., Miller, C.A., Francis, D.J., Tomblin, J.B., Kail, R.V.


Note: WD = Word Definition, Prep = Prepositions, Inf = Inferences, SR = Sentence Repetition, NM = Naming, Cad = Comparative Adjectives, Col = Colour, 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, CI90 = 90 per cent confidence interval, ** = p<0.01, * = p<0.05
Figure 2

Note: x-axis=LANGUAGE4 items by item number labeled with the corresponding item factor label, i.e., WD = Word definition, NM = Naming, Prep = Prepositions, Sstr = Sentence structure, CAd = Comparative Adjectives, Inf = Inference, Col = Colour, 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 1

Estimated means, standard deviations and reliability for repeated language measures at all time-points

<table>
<thead>
<tr>
<th>Measure</th>
<th>Time 1 (age 4)</th>
<th>Time 2 (age 5)</th>
<th>Time 3 (age 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>α</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>α</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>α</td>
<td>N</td>
</tr>
<tr>
<td>BPVS II</td>
<td>41.77 (11.22)</td>
<td>.91</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>56.80 (11.35)</td>
<td>.91</td>
<td>202</td>
</tr>
<tr>
<td></td>
<td>72.59 (11.79)</td>
<td>.91</td>
<td>199</td>
</tr>
<tr>
<td>GramClos</td>
<td>11.03 (3.91)</td>
<td>.74</td>
<td>204</td>
</tr>
<tr>
<td></td>
<td>13.94 (3.71)</td>
<td>.69</td>
<td>202</td>
</tr>
<tr>
<td></td>
<td>18.37 (3.79)</td>
<td>.70</td>
<td>198</td>
</tr>
<tr>
<td>Sentence Rep</td>
<td>6.53 (1.92)</td>
<td>.63</td>
<td>202</td>
</tr>
<tr>
<td></td>
<td>7.98 (2.15)</td>
<td>.70</td>
<td>201</td>
</tr>
<tr>
<td></td>
<td>9.12 (2.05)</td>
<td>.70</td>
<td>200</td>
</tr>
</tbody>
</table>

Note: N = number of cases with complete datasets for each repeated language measure, BPVS II = Norwegian version of the British Picture Vocabulary Scale, second edition, GramClos = Grammatic Closure, Sentence Rep = Sentence Repetition. SD = Standard deviation, α = Chronbach’s alpha
Table 2

*Estimated Correlations between all repeated measures at all time-points*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>SR T1</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>BPVS T1</td>
<td>.319</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>GC T1</td>
<td>.351</td>
<td>.380</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>SR T2</td>
<td>.487</td>
<td>.229</td>
<td>.282</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>BPVS T2</td>
<td>.347</td>
<td>.539</td>
<td>.314</td>
<td>.292</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6.</td>
<td>GC T2</td>
<td>.349</td>
<td>.341</td>
<td>.419</td>
<td>.304</td>
<td>.415</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>SR T3</td>
<td>.344</td>
<td>.206</td>
<td>.332</td>
<td>.553</td>
<td>.277</td>
<td>.314</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>8.</td>
<td>BPVS T3</td>
<td>.207</td>
<td>.465</td>
<td>.333</td>
<td>.246</td>
<td>.565</td>
<td>.324</td>
<td>.232</td>
<td>-</td>
</tr>
<tr>
<td>9.</td>
<td>GC T3</td>
<td>.320</td>
<td>.290</td>
<td>.403</td>
<td>.173</td>
<td>.328</td>
<td>.571</td>
<td>.284</td>
<td>.406</td>
</tr>
</tbody>
</table>

*Note: SR = Sentence Repetition; BPVS = British Picture Vocabulary Scale, second edition; GC = Grammatic Closure, T = time*
Table 3

*Fit statistics for the different latent class models based on the 26 LANGUAGE4 items*

<table>
<thead>
<tr>
<th>Model</th>
<th>Class count</th>
<th>Log L</th>
<th>BIC*</th>
<th>pVLMR</th>
<th>pBLRT</th>
<th>No. of free parameters</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>2class</td>
<td>435, 165</td>
<td>-5317.46</td>
<td>10805.69</td>
<td>0.000</td>
<td>0.000</td>
<td>53</td>
<td>0.878</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.000)</td>
<td></td>
</tr>
<tr>
<td>3class</td>
<td>388, 115, 97</td>
<td>-5212.84</td>
<td>10683.45</td>
<td>0.037</td>
<td>0.000</td>
<td>80</td>
<td>0.848</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.038)</td>
<td></td>
</tr>
<tr>
<td>4class</td>
<td>341, 107, 91, 61</td>
<td>-5100.74</td>
<td>10546.25</td>
<td>0.019</td>
<td>0.000</td>
<td>107</td>
<td>0.838</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.020)</td>
<td></td>
</tr>
<tr>
<td>5class</td>
<td>349, 118, 59, 42, 32</td>
<td>-5033.56</td>
<td>10498.89</td>
<td>0.212</td>
<td>0.000</td>
<td>134</td>
<td>0.881</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.215)</td>
<td></td>
</tr>
<tr>
<td>6class</td>
<td>356, 83, 58, 38, 34, 31</td>
<td>-4987.96</td>
<td>10494.70</td>
<td>0.320</td>
<td>0.000</td>
<td>161</td>
<td>0.901</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.322)</td>
<td></td>
</tr>
</tbody>
</table>

Note: Class count = number of individuals classified in each class based on their most likely latent class membership (bold fonts indicates class count for the Low Performing class across models), BIC* = sample-size adjusted BIC, p VLMR = probability value for the Vuong-Lo-Mendell-Rubin Likelihood Ratio Test (adjusted values in parentheses), p BLRT = probability value for the parametric Bootstrapped Likelihood Ratio Test.
Table 4  Estimated Means and Variances for Intercepts and Slopes for the two Second-Order Growth Mixture Models

<table>
<thead>
<tr>
<th>Latent Class label</th>
<th>Full</th>
<th>Four-class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>One-class (N=600)</td>
<td>HP (N=331)</td>
</tr>
<tr>
<td>Intercept (p)</td>
<td>4.18 (.00)</td>
<td>0.67 (.00)</td>
</tr>
<tr>
<td>Slope (p)</td>
<td>1.53 (.00)</td>
<td>0.02 (.56)</td>
</tr>
</tbody>
</table>

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, p = p-values
Figure captions

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

Figure 2. LANGUAGE4 item profiles based on a Latent Class Analysis