

SOCIAL SCIENCES

Stability of core language skill from infancy to adolescence in typical and atypical development

Marc H. Bornstein^{1,2*}, Chun-Shin Hahn¹, Diane L. Putnick¹, Rebecca M. Pearson³

Command of language is a fundamental life skill, a cornerstone of cognitive and socioemotional development, and a necessary ingredient for successful functioning in society. We used 15-year prospective longitudinal data from the Avon Longitudinal Study of Parents and Children to evaluate two types of stability of core language skill in 5036 typically developing and 1056 atypically developing (preterm, dyslexic, autistic, and hearing impaired) children in a multiage, multidomain, multimeasure, multi-reporter framework. A single core language skill was extracted from multiple measures at multiple ages, and this skill proved stable from infancy to adolescence in all groups, even accounting for child nonverbal intelligence and sociability and maternal age and education. Language skill is a highly conserved and robust individual-differences characteristic. Lagging language skills, a risk factor in child development, would profitably be addressed early in life.

INTRODUCTION

Early language skills merge into higher-order verbal and mental functioning (1) and so have predictive validity for the development of speech, grammar, reading, academic achievement, and intelligence (2–4). Language skills also predict behavioral adjustment in children (5, 6), even after controls for prior levels of behavior problems and taking into consideration children’s nonverbal intellectual functioning and performance, gender, and ethnicity, as well as their mothers’ verbal intelligence, education, parenting knowledge, and social desirability bias, and their families’ socioeconomic status (7, 8). Achievements in language and literacy open doors to education, occupation, income, and health (9).

Individual differences are a central and manifest characteristic of child language (10–12), as children of the same chronological age vary dramatically in terms of their language skills. One fundamental conceptual issue that has framed debates about individual differences in theory and research across the history of language study and developmental science is their stability (13). Stability is consistency in individual differences over time. Stability in language therefore occurs when some children display relatively high levels of language at one point in time vis-à-vis their peers and continue to display high levels at a later point in time, while other children display consistently lower levels. Language is among the most complex skills a child must master, and so understanding individual differences in language and their developmental stability is of compelling interest to professionals, practitioners, and parents.

Here, we distinguish and study two kinds of stability in child language. One is homotypic stability, maintaining individual rank order in the same characteristic measured in the same metric over time. In language, vocabulary size exemplifies a characteristic that might be indexed in the same way at different ages and show homotypic stability. The other is heterotypic stability, maintaining individual rank order on different manifest characteristics over time, where the different characteristics are theoretically related and pre-

sumed to share the same underlying construct. In language, the shared constructs, vocabulary size at one age and reading comprehension at a later age, might show heterotypic stability.

The present study aims to advance our understanding of child language and its homotypic and heterotypic stability in several novel and substantial ways. (i) This study assesses stabilities of individual variation in language development in large and independent samples. (ii) The study begins earlier in life (6 months) and extends later in life (15 years) than ever before. (iii) The study follows a 13-wave granular longitudinal design that is unprecedented in developmental science. (iv) The study uses a wide range of age-appropriate language domains (from general communication and vocabulary in infancy and early childhood to spelling, reading, and narrative in later childhood and adolescence) broadly construed across a diversity of methods and measures. (v) The study evaluates homotypic and heterotypic stabilities in typically developing children as well as children in four at-risk groups: preterm children and children diagnosed with dyslexia, autism spectrum disorders (ASDs), and hearing impairment. Improved survival and diagnosis have meant markedly increasing numbers of very preterm children and children with ASD, respectively, in the community.

To accomplish these several aims, we identified four major challenges to the study of language stability in children and reached solutions to each. The first two challenges were to capture the multiple domain, method, and measurement approaches to language and their changing age appropriateness. Here, we define “language” broadly to comprise many domains (including phonology, lexicon, grammar, pragmatics, reading, spelling, and narrative), and each can be assessed in multiple ways. Moreover, language differs phenotypically at different ages. The organizational perspective on development posits that the proper way to study development over time is to examine age-appropriate and therefore different, yet conceptually related, measures of the same underlying construct (14). In consequence, no single representation of language development across childhood is best, and no single approach to measurement can predominate. Rather, assessment selection must be guided by tradition, tractability, goal, convenience, and age appropriateness. Developmental scientists today advocate the wisdom of applying multiple assessments and using converging operations of different strategies targeted to the same phenomenon.

Copyright © 2018
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S.
Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA. ²Institute for Fiscal Studies, London, UK. ³School of Social and Community Medicine, University of Bristol, Bristol, UK.

*Corresponding author. Email: marc_h_bornstein@nih.gov

Thus, the first two challenges are to identify sensitive, reliable measures of language with different contents derivable from varying methods and sources that track child age appropriately. We met these challenges, first, by implementing caregiver reports and direct assessments of multiple different aspects of children's language and, second, by extracting shared variance among different measures using latent variables (15). Latent variables constitute a solution to the time-varying requirement of language development because they accommodate multiple age-appropriate indicators and different loadings for the same indicators on child language across age. Latent variables capture empirical covariation among indicators that may manifest differently at each age. Latent variables thereby identify what we call here core language skill. As stated, language assessment across a prolonged developmental timescale perforce entails dynamically changing (age-appropriate) measures. Procedurally, some methodologies can evaluate children directly (as in testing), but others must (or even best) rely on parental report because of the very young age of the child, the possibility of child reactivity to testing or observation, or the experienced and knowledgeable posture of the child's caregiver. Substantively, some measures (grammar or literacy) may be appropriate only at certain ages, whereas others (vocabulary) may be applicable across multiple ages.

The third challenge was to pinpoint stability "in" the child. Stability is often readily ascribed to temporal consistency of a characteristic in the individual. However, valid attribution necessitates simultaneous examination of factors that pervasively influence stability or confound its interpretation. To assess whether core language skill is stable in itself or if any of several third variables that covary with child language underlie stability in core language skill, we assessed and accounted for multiple candidate endogenous (child nonverbal intelligence and sociability) and exogenous (maternal age and education) variables.

The fourth challenge was to evaluate the robustness of stability. Language is a sensitive and demonstrable indicator of human development. Biological risks of many kinds are known to perturb the normal acquisition of language. As reviewed in greater depth in the Supplementary Materials, mean-level differences in language are common in atypically developing preterm children and children with dyslexia, autism, and hearing impairment when compared to typically developing children. However, less is known about how biological risk alters language stability. It is possible that the mechanisms that produce mean-level group differences in language also generate variability in stability. For example, stability of language in children with a language disability, such as dyslexia, may be higher than stability in children without a disability because the processes that restrict language skills also maintain language-disabled children's fixed order relative to one another. Here, we explored several biological and health moderators of language stability in children, including preterm birth, dyslexia, autism, and hearing impairment. Fuller discussions and justifications of the significance of stability, latent variables, and the moderators appear in the Supplementary Materials.

RESULTS

Study 1: Long-term language stability in typical development

Study 1 evaluated long-term language stability in typical development in a 13-wave prospective longitudinal study that used data from the

Children in Focus (CiF) group of the Avon Longitudinal Study of Parents and Children (ALSPAC) (16–18). The final study sample consisted of 925 (429 girls, 46.4%) white, term, monolingual singletons (M gestation = 39.75 weeks, SD = 1.29) free of dyslexia, autism, and hearing impairment. Mothers averaged 29.26 years (SD = 4.48; range, 14 to 43) at childbirth. This study sample was very diverse in terms of maternal education and social class. Table 1 presents the sample size and child age at each data collection wave.

Table 1 also presents language measures, scale scores, and sources used at each data collection wave. Table S1 shows the means, SD s, and ranges of the language measures and covariates for the total sample.

We assessed the fit of a structural model to the data to assess the common convergence of multiple measures on single latent variables of the core language skill, where applicable, and the stability between those language variables (measurement models are discussed in the Supplementary Materials). The a priori model fit the data, scaled Yuan-Bentler ($Y-B$) $\chi^2(618) = 2399.16$, $P < 0.001$, robust comparative fit index (CFI) = 1.00, standardized root mean square residual ($SRMR$) = 0.11, and root mean square error of approximation ($RMSEA$) = 0.00. Figure 1 presents the standardized solution of this stability model. Although $SRMR$ was greater than the usual cutoff of 0.09 (19), it is sensitive to estimation technique and sample size (20) and model complexity (21). Given the excellent fit indicated by the CFI and $RMSEA$ as well as the large sample size and complex model, we deemed it acceptable. All indicators of child language loaded significantly on their factors at each age, which indicated that diverse measures of language formed stable, single factors of core language skill at each age. The stabilities of language were large between successive waves except for one medium-sized stability between the 6-month variable and year 1 factor.

On the basis of an extensive body of research on constructs associated with child language (22–25), and to guard against threats to validity, we controlled for four prominent constructs that might affect child language stability: children's nonverbal intelligence and sociability and mothers' age and education. We then re-evaluated the stability model, taking into consideration these covariates. Figure S1 shows the final covariate model; it fit the data well: scaled $Y-B$ $\chi^2(670) = 2014.22$, $P < 0.001$, robust CFI = 1.00, $SRMR$ = 0.09, and $RMSEA$ = 0.00. The attenuation of stability estimates ranged from 0.02 to 0.19 controlling for covariates. Stabilities of language were still medium to large between successive waves over the first 15 years of life.

Study 2: Long-term language stability in typical and atypical development

Study 2 replicated long-term language stability in an independent sample of typically developing children and evaluated long-term language stabilities in five atypically developing samples. The whole study sample consisted of 5167 (2594 girls, 50.2%) white, monolingual singletons. Mothers averaged 29.03 years (SD = 4.51; range, 15 to 44) at childbirth. This sample was also very diverse in terms of maternal education and social class. Of the 5167 children, 4111 were born term (M gestation = 39.78 weeks, SD = 1.29), were reported free of dyslexia and autism, and with tested bilateral normal hearing, served as the typically developing sample. Atypically developing samples included 435 moderate-late preterm (32 to 36 weeks' gestation, M gestation = 35.05 weeks, SD = 1.19; range, 32 to 36) and 51 very preterm (<32 weeks' gestation, M gestation = 28.92 weeks, SD = 1.75;

Table 1. Study 1: Sample size, child age, and language measures at each data collection wave. *N* represents the number of available observations. Except for wave 1 (child age in months), child ages are in years.

Data collection wave	<i>N</i>	Child age, <i>M</i> (<i>SD</i>)	Measures	Scales	Source
1	859	6.24 (0.45)	Modified DDST (53)	Communication	Caregiver
2	873	1.28 (0.06)	Modified MCDI Words and Gestures (11)	Understand, vocabulary, and social (nonverbal) communication	Caregiver
3	894	1.51 (0.04)	Modified DDST (53)	Communication	Caregiver
4	862	2.01 (0.03)	Modified MCDI Words and Sentences (11)	Vocabulary, grammar, plurals, and tense	Caregiver
5	857	2.07 (0.02)	RDLS (54)	Comprehension	Child assessment
6	852	3.19 (0.09)	Modified MCDI Words and Sentences (11)	Vocabulary, plurals, past tense, and word combination	Caregiver
7	788	4.07 (0.03)	WPPSI (41)	Information, comprehension, vocabulary, and similarity	Child assessment
8	756	5.15 (0.06)	Bus Story Test (55)	Information and sentence length	Child assessment
			RDLS (54)	Comprehension	Child assessment
			Initial Consonant Detection Test (56)	Number of correct responses	Child assessment
9	720	7.44 (0.12)	WORD (57)	Reading and spelling	Child assessment
			Phoneme Detection Task (58)	Number of correct responses	Child assessment
10	661	8.62 (0.18)	WISC (59)	Information, comprehension, vocabulary, and similarity	Child assessment
			WOLD (60)	Listening comprehension and oral expression	Child assessment
11	668	9.84 (0.17)	WORD (57)	Number of correct responses (real and nonreal words)	Child assessment
			NARA II (62)	Accuracy and comprehension	Child assessment
12	502	13.77 (0.15)	TOWRE (63)	Number of correct responses (word and pseudoword)	Child assessment
13	472	15.34 (0.16)	WASI (64)	Vocabulary	Child assessment

range, 25 to 31) children, 322 children with dyslexia (*M* gestation = 39.53 weeks, *SD* = 1.80; range, 27 to 45), 89 children with autism (*M* gestation = 39.47 weeks, *SD* = 2.30; range, 27 to 42), and 221 children who had mild and/or moderate hearing impairment in one ear or in both ears (*M* gestation = 39.31 weeks, *SD* = 2.07; range, 27 to 42). Table S2 shows sample sizes and child ages for each group at each wave.

Study 2 followed the same procedures, language measures, and covariates as those described in study 1, with three small exceptions: Children in study 2 were not tested in the Reynell Developmental Language Scales (RDLS) at 2 years 1 month, and they were not assessed for language at 4 or 5 years. Table S3 shows the means, *SD*s, and ranges of the language measures and covariates by groups.

The a priori model for the whole sample fit the data, scaled *Y*-*B* $\chi^2(344) = 7417.69$, $P < 0.001$, robust CFI = 0.94, SRMR = 0.07, RMSEA = 0.045, and 90% confidence interval (CI) = 0.044 to 0.046. Figure S2 presents the standardized solution of this stability model. All indicators of child language loaded significantly on their factors at each age. The stabilities of language were medium to large between successive waves. Table 2 shows zero-order and partial correlations controlling for covariates between language measures across ages by group, and Table 3 shows point estimates of average stability and their 95% CIs for these correlations. Figure 2 depicts these average stabilities by group. In at-risk groups, all stabilities were medium or large except for the stabilities between 6 months and 1 year after accounting for covariates in moderate-late preterm children

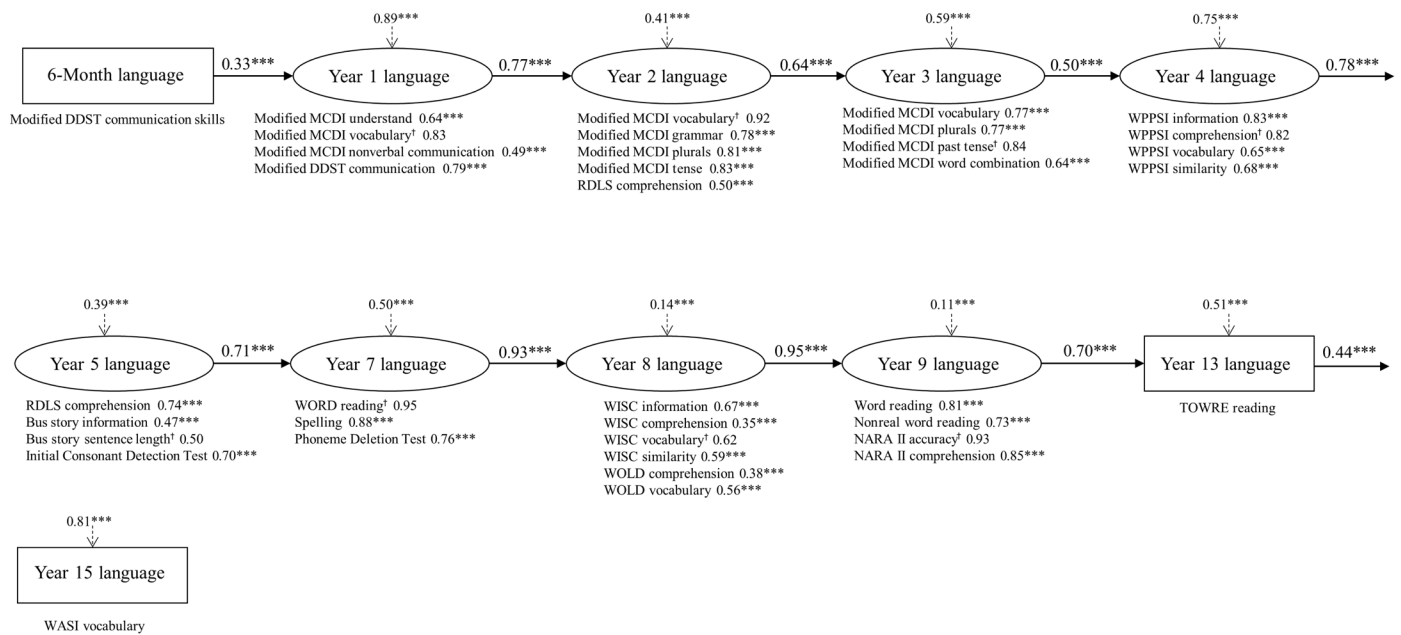


Fig. 1. Study 1. Standardized solution for stability model ($N = 925$). Numbers associated with single-headed arrows are standardized path coefficients; numbers associated with dotted single-headed arrows are error variances or disturbances, the amount of variance not accounted for by paths in the model. Indicators of each latent variable are listed below the latent variable with their factor loadings. [†]Marker indicators of the latent factors (loadings set to 1 to scale and identify the factor). Covariances that were in the model, but not shown in the figure, included year 2 MCDI vocabulary and RDLS comprehension, standardized coefficient = 0.22, $P < 0.001$; year 5 RDLS comprehension and Initial Consonant Detection Test, standardized coefficient = -0.26, $P < 0.001$; year 5 Bus Story information and Bus Story sentence length, standardized coefficient = 0.78, $P < 0.001$; and year 9 word and nonreal word reading, standardized coefficient = 0.39, $P < 0.001$. Correlations of 0.10, 0.30, and 0.50 correspond to small, medium, and large effect sizes, respectively (78).

and in children with autism (Table 2). Most medium-sized stabilities were observed at the earliest ages, between 6 months and 1 year, and between 3 and 7 years across a longer 4-year time span. Atypically developing children’s language performance showed medium-to-large stabilities between successive waves over the span of 15 years, even accounting for child nonverbal intelligence and sociability and maternal age and education.

DISCUSSION

We investigated the longitudinal stabilities of child language from 6 months to 15 years using multiple age-appropriate methods, measures, and reporters, involving a wide variety of different language domains, in relatively large samples of typically developing children, as well as children born preterm and with childhood diagnoses of dyslexia, autism, and hearing impairment, in a prospective long-term microgenetic design. Individual differences tell us about the distribution of language skill, and their stability tells us about the nature and ontogeny of that language skill. We also tested whether a diverse set of controls for third variables and background characteristics accounted for stability in child language.

With respect to the four challenges posed at the outset, clear evidence emerged for individual variation in a core language skill at each of 11 ages, for convergence of multiple indices of language at each age on latent variables representing a core language skill, for the homotypic and heterotypic stability of core language skill over the long term, and for the robustness of long-term stability of core language skill in atypically developing children with several different types of health risk.

As with all developmental constructs, language (and its stability) is a joint product of biology and experience (26, 27). For example, a 2- to 12-year behavior genetics study identified genetic/biological and environmental/experiential sources of individual differences in developing language skills (28). Thus, a consistent personological characteristic, experience, or environment can carry stability. To address this point, we included child and maternal factors known to affect child language as covariates. Long-term stability was obtained separate and apart from both (nonlanguage) endogenous and exogenous covariates. The fact that stability of core language skill across so long a period began so early, was sustained so long, transcended several heterogeneous moderating factors, and was maintained over and above covariates points to a highly conserved and robust individual-differences characteristic in human beings. It further suggests that the search for mechanism(s) underlying stability of core language skill in children is likely to reward basic science as well as applied clinical research.

Limitations to these study results include, among others, the heavy (if necessary) reliance on caregiver report in the early years and the limited (by necessity) number of language domains actually assessed relative to the possible number (see the Supplementary Materials). At three ages, only single language measures were collected; more varied early language measures, or having the same language measure assessed by multiple reporters, would strengthen the study. We did not measure (and so did not eliminate) all possible endogenous factors in children (brain function, motivation, and persistence), but we did measure and so controlled child age, nonverbal intelligence, and sociability as factors in stability. It is challenging to assess many aspects of language in very young children,

Table 2. Study 2: Stability of language across age by groups. Numbers before the slashes represent correlations controlling for child age only (to control age variation within waves); numbers after the slashes represent correlations controlling for child age, nonverbal intelligence, sociability, and maternal age and education. Correlations of 0.10, 0.30, and 0.50 correspond to small, medium, and large effect sizes, respectively (78).

	6 months → year 1	Year 1 → year 2	Year 2 → year 3	Year 3 → year 7	Year 7 → year 8	Year 8 → year 9	Year 9 → year 13	Year 13 → year 15
Typical (<i>n</i> = 4111)	0.36/0.25	0.68/0.57	0.55/0.48	0.23/0.21	0.60/0.44	0.62/0.52	0.79/0.75	0.48/0.30
Moderate-late preterm (<i>n</i> = 435)	0.33/0.16	0.69/0.55	0.59/0.47	0.34/0.31	0.64/0.49	0.65/0.56	0.83/0.80	0.48/0.35
Very preterm (<i>n</i> = 51)	0.48/0.40	0.81/0.62	0.66/0.54	0.27/0.28	0.55/0.38	0.66/0.53	0.64/0.66	0.50/0.25
Dyslexia (<i>n</i> = 322)	0.37/0.20	0.68/0.48	0.54/0.46	0.27/0.24	0.51/0.39	0.50/0.44	0.82/0.82	0.26/0.24
Autism (<i>n</i> = 89)	0.32/0.12	0.70/0.48	0.65/0.58	0.53/0.49	0.70/0.62	0.71/0.66	0.85/0.81	0.53/0.40
Hearing impairment (<i>n</i> = 221)	0.38/0.26	0.72/0.55	0.58/0.50	0.27/0.27	0.54/0.42	0.58/0.54	0.77/0.80	0.37/0.25

Table 3. Study 2: Average stability of language by groups. Average stability represents the mean of correlation coefficients controlled for child age only. Average stability controlled for covariates represents the mean of partial correlations controlled for child age, nonverbal intelligence, sociability, and maternal age and education. The relatively small sample sizes in the very preterm and autism groups contributed to somewhat diminished precision in the point estimates of average correlation and, thus, wider 95% CIs.

	Average stability		Average stability controlled for covariates	
	Point estimate	95% CI	Point estimate	95% CI
Typical (<i>n</i> = 4111)	0.56	0.54–0.58	0.46	0.43–0.48
Moderate-late preterm (<i>n</i> = 435)	0.59	0.53–0.65	0.49	0.41–0.55
Very preterm (<i>n</i> = 51)	0.59	0.37–0.74	0.47	0.22–0.66
Dyslexia (<i>n</i> = 322)	0.52	0.44–0.60	0.44	0.34–0.52
Autism (<i>n</i> = 89)	0.65	0.50–0.75	0.54	0.38–0.68
Hearing impairment (<i>n</i> = 221)	0.55	0.45–0.63	0.47	0.36–0.57

and measurement of language at an early stage perform cannot include all components of language (e.g., grammar). These data were also collected beginning in the 1990s; since then, the treatment of preterm and other at-risk children has changed. Except for hearing impairment, diagnoses of other atypicalities relied on maternal report. Because hearing impairment was measured only once, we do not know whether hearing loss persisted or whether it originated at birth or later in development.

Nonetheless, these results prompt several notable considerations. First, a corollary of the prevailing multidimensional and componential conceptualization of language might be that phenotypically distinguishable language domains are independent of one another. Here, we confirmed that diverse indices of language deriving from different language domains, measures, methods, sources, and contexts, each of which showed individual variation, were positively associated across different ages (23, 29–34). On this basis, we could compute single latent variables of a core language skill at diverse ages. The significant amounts of variance accounted for by each latent variable at each age tested add to the validity of the stability model.

Second, individual differences in core language skill were present from the first years of life, and so relatively stable individual differ-

ences in child language seem to be established early. However, the lowest observed stability coefficient occurred between 6 months and 1 year. As children aged past 1 year, there was more stability (less inconsistency) in language; that is, stabilities from 1 to 13 years were large. A characteristic may not be stable at one age in the life course but may stabilize at a later age. Generally, infancy and early childhood are thought to be less stable (or predictive) periods in life (35), and people are thought to become increasingly consistent in relation to one another as they age (36, 37). Strong stability after 1 year implies that changes among children in their relative rank in core language skill later in development are rare. By contrast, the smaller stability coefficient between 6 months and 1 year indicates that nearly 90% of the variance in 1-year core language skill is not explained by 6-month language. This difference suggests that core language skill is relatively more malleable in early life. It is also possible that lower stability early in life reflects the difficulty in validly assessing language in preverbal infants. In general, however, our findings underscore the importance of identifying lagging language skills early in life and promoting the child's language environment well before formal schooling as a means to enhancing language skill.

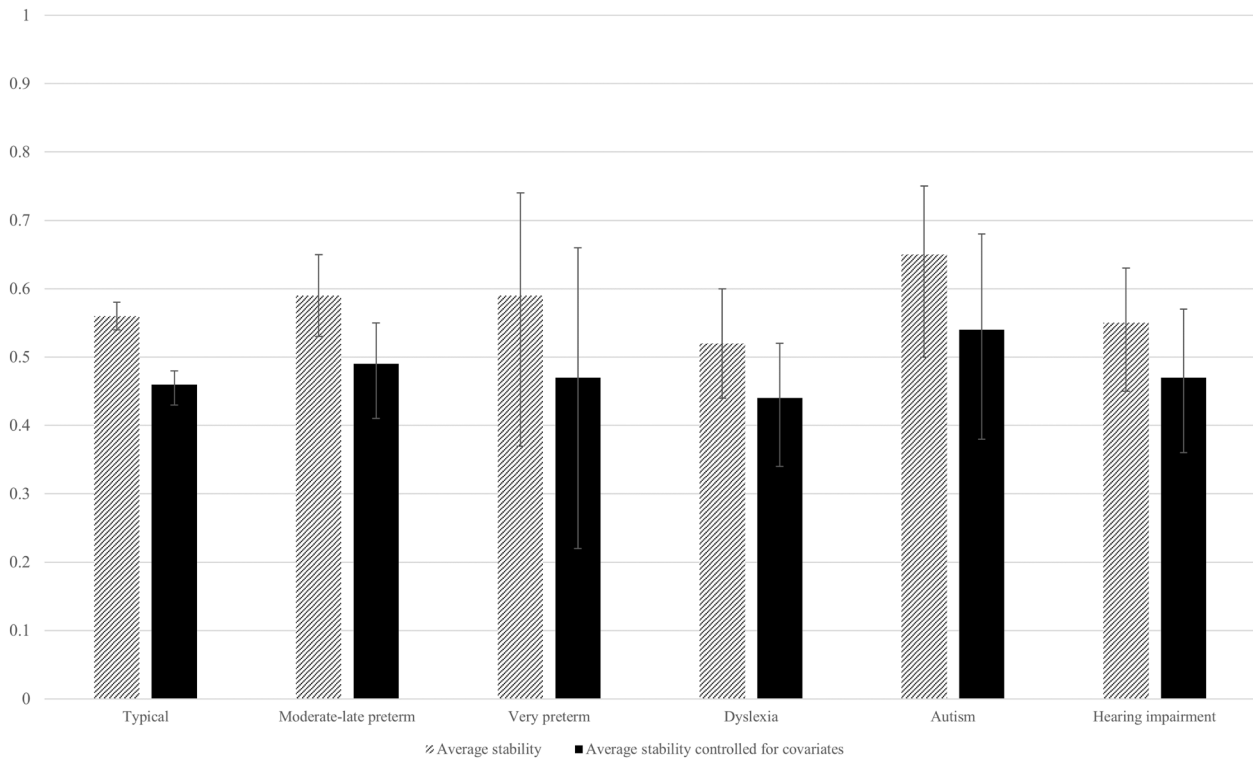


Fig. 2. Study 2. Average stabilities and their 95% CIs of language by group. Average stability represents the mean of correlation coefficients between language measures controlled for child age only. Average stability controlled for covariates represents the mean of partial correlations controlled for child age, nonverbal intelligence, sociability, and maternal age and education.

Large stability coefficients can mislead researchers and practitioners to conclude that language skill in children is set in infancy or toddlerhood. This is not necessarily the case. Stability is a key developmental barometer, but to be stable does not mean to be immutable or impervious to change or intervention. Focusing solely on stability in language overlooks or minimizes dramatic and normative developmental changes in mean level of language. The life-span perspective in developmental science specifies that human beings are open systems, and the plastic nature of psychological functioning ensures both consistency and change across the life course (38). The language skills of individual children (relative to their peers) still shift across time, and even large relative stability leaves significant amounts of common variance unaccounted for. Language is ultimately modifiable by experience or intervention. In language acquisition, development appears to balance the advantages of stability with the adaptive value of early susceptibility to experience.

A third contribution of this study distinguishes homotypic stability (as of vocabulary between year 1 and year 15) from heterotypic stability (as between vocabulary in year 1 and literacy in year 13). The measures, reporters, and contexts for language sampling at the different ages perforce differed. From one point of view, this procedural variation attenuates stability. That is, heterotypic stability between different individual indices of child language likely represents lower-bound estimates of stability considering differences in assessment measures and procedures used at different times. Thus, heterotypic stability is conservative and probably underestimates true stability. By contrast, homotypic stability of identical measures and of latent variables (as we used here) may more closely approximate true stability in language development. Nonetheless, the heterotypic approach to

stability assessment is faithful not only to a developmental perspective but also to a systems perspective on the hierarchical integration of lower-order into high-order abilities with development (1). Literacy, the end goal of our assessments, is conventionally understood as the ability to read, write, spell, listen, and speak (39), but encompasses a progression of skills that begins with comprehension and expression of sounds and then words and culminates with grammar and reading and writing. The latent variable solution to the challenge of heterotypicity is therefore valuable to developmental science in general. An additional point with respect to heterotypic stability was the small associations uncovered in both studies between nonverbal measures of intelligence and language. Of course, to be able to perform in an evaluation of even a nonverbal assessment requires some language ability (if only to be able to follow instructions), and the history of general psychological testing and specific intelligence testing tells us that nonverbal and verbal components of cognition are not strictly independent. For example, the Bayley Scales of Infant Development (40) have a mental development index and a psychomotor development index that correlate, and the Wechsler (41) series of intelligence tests have verbal and performance intelligence quotient (IQ) indices that correlate. In our case, as is also typical, the shared variance in language and nonverbal intelligence measures was small (study 1 range, 2 to 25%; study 2 range, 6 to 21%).

The fourth contribution of this study is analyses of stabilities of individual differences in the language of large numbers of children identified as preterm, dyslexic, autistic, and hearing impaired. From very early in development, core language skill was stable in each group. The findings therefore have implications for psycholinguists and psychologists, pediatricians and psychiatrists, practitioners and

professors, and parents and the public. All stakeholders should be aware that very young children who perform poorly relative to their peers are likely to continue to perform poorly at later ages, which reinforces the desirability of early assessment of language performance and the need for early intervention. Our data suggest that core language skill anticipates verbal and literacy achievements as child development unfolds. Through regular well-child checkups, pediatricians could identify children who have lagging language skills and connect them to early intervention services.

Given the increasing importance of replication in science (42, 43), it is noteworthy that the results of the present studies internally replicate and then extend previous studies with single or fewer language measures taken over shorter periods of time (24).

Last, the present empirical findings articulate with clinical practice; we distinguish between language screening and the accuracy of the multiple domain, measure, and source latent variable approach. Clinically, our approach to estimating child language as latent variables is not a quick tool for early diagnosis or screening; most clinicians do not have the benefit of a rich array of measures or the technical support at hand to estimate latent variables. A screening instrument may be practically valuable, but the latent variable provides a more fundamental understanding of the core construct. Nonetheless, a multimeasure approach to child language has been applied productively in the past for predicting continued language delay (44–47). Notably, interventions that enhance language skills also improve behavioral regulation in children (48–50).

MATERIALS AND METHODS

Study 1: Long-term language stability in typical development

Participants

The ALSPAC is a prospective, population-based, longitudinal transgenerational observational study investigating influences on health and development across the life course. All births in the former Avon Health Authority with an expected date of delivery between 1 April 1991 and 31 December 1992 were eligible. Of the initial 14,541 pregnancies, there were a total of 14,676 fetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age. Because only 2.6% of the ALSPAC sample was non-White (of those participants who provided this data point), and this group was heterogeneous (0.9% Asian, 1.0% Black, and 0.7% other), we focused on the majority group (51, 52). From the 12,075 White European participants in the ALSPAC data, the following exclusion criteria were followed (some children might fall into multiple categories): Children who (i) were twins ($n = 306$), (ii) were born preterm (born less than 37 weeks, $n = 677$), (iii) were hearing impaired ($n = 233$), (iv) had dyslexia ($n = 332$), (v) were diagnosed with autism ($n = 91$), and (vi) were bilingual or spoke a language other than English as their main language ($n = 280$) were excluded, resulting in 9794 children. An additional 8869 children were excluded from study 1 because they were not from the CiF cohort and/or did not have the additional CiF assessments used in the current study. Children who did not fall into the exclusion criteria and provided data at any of the data collection waves were included in study 1.

Maternal education, collected at 32 weeks of pregnancy as an ordinal variable according to increasing levels of achievement, was varied: certificate of secondary education (13.0%), vocational (10.8%), O level (35.2%), A level (25.9%), and university degree (15.1%). Ma-

terial social class ranged from unskilled (2.0%), partly skilled (7.8%), skilled manual (7.2%), skilled nonmanual (42.8%), managerial and technical (33.8%), to professional (6.4%).

Procedures

Child language data were derived from caregiver reports and direct child assessments by trained psychologists during research clinics. The ALSPAC study website includes descriptions of measures used and scoring methods. In addition to the language measures detailed below, caregivers completed questionnaires that supplied demographic information about children's health status, family language, and the like.

Language assessments

We used data collected across 13 ALSPAC collection waves (Table 1). However, we aggregated data collected at ages 1 year 3 months and 1 year 6 months (two waves) into year 1 measures, and those collected at ages 2 years and 2 years 1 month (2 waves) into year 2 measures; thus, we studied 13 waves but calculated stability across 11 ages.

Under 1 year. Caregiver report: At age 6 months, caregivers completed an ALSPAC-modified Denver Developmental Screening Test (hereinafter referred to as modified DDST) (53) adapted for caregiver completion. The “communication” scores were used.

Year 1. Caregiver report: At age 1 year 3 months, the understand, vocabulary, and social (nonverbal) communication scores on the ALSPAC-modified MacArthur Communication Development Inventories (hereinafter referred to as modified MCDI) (11) Words and Gestures were used. Twelve of the 28 questions (called “phrases” on the original MCDI) were asked for the understand scale, and the vocabulary checklist was cut by removing entire sections (i.e., toys, small household items, people, action words, words about time, pronouns, question words, prepositions and locations, and quantifiers), as well as some items from other sections from the original MCDI. Furthermore, some American English items were adapted or replaced with the British English equivalents (e.g., lorry instead of truck; sweater or jumper). Ten of the 12 social communication items (called “first communicative gestures” on the original MCDI) were asked. See the Supplementary Materials for more details about the modified MCDI.

Year 2. Caregiver report: At age 2 years, the vocabulary, grammar, plurals, and tense scores on the modified MCDI Words and Sentences were used. The vocabulary checklist was cut by removing sound effects and animal sounds, small household items, and connecting words as well as by removing items from other sections from the original MCDI. The 4 grammar items (called “word endings” on the original MCDI), 5 irregular plurals (called “word forms—nouns” on the original MCDI), and 20 past tense (called “word forms—verbs” on the original MCDI) items were unmodified.

Direct assessment: At age 2 years 1 month, children in the CiF cohort were administered the RDLS (54). The RDLS comprehension scale measures a child's verbal comprehension by administering a series of activities where the child is asked to respond to and carry out a series of spoken tasks. The raw score was used.

Year 3. Caregiver report: At age 3 years 2 months, the vocabulary, plurals, past tense, and word combination scores on the modified MCDI Words and Sentences were used. The vocabulary checklist was the same as the one used at year 2. The 5 irregular plurals (called “word forms—nouns” on the original MCDI) and 20 past tense (called “word forms—verbs” on the original MCDI) items were unmodified. In addition, two items that ask whether the child uses plurals by adding “-s” to the end of words or uses past tense by adding “-ed” to the end of words were included in the plurals and past tense

scales, respectively. The word combination items were modified from the 14 “complexity” items of the original MCDI. Two items were dropped, and some items were adapted to add a third option (e.g., two feet, two foot, two foot) or to change the object (e.g., “that’s my book” versus “that’s my truck”).

Year 4 (CiF cohort only). Direct assessment: At age 4 years 1 month, four verbal subscale scaled scores ($M = 10$ and $SD = 3$) of the Wechsler Preschool and Primary Scale of Intelligence—Revised UK Edition (WPPSI) (41) were used: information, comprehension, vocabulary, and similarity.

Year 5 (CiF cohort only). Direct assessment: At age 5 years 1 month, the Bus Story Test (55), a screening test of verbal expression, was administered. The assessment involves children listening to a spoken narrative about a bus, accompanied by pictures depicting the events that occur in the story. Children then retell the story with the pictures as support. The child’s narrative is recorded orthographically and scored for information content (number of relevant pieces of information given) and sentence length (mean sentence length of the five longest sentences). In addition, the same RDLS comprehension scale that was used at 2 years 1 month was repeated at age 5 years 1 month. Last, the Initial Consonant Detection Test (56) asked children to identify which two of three words illustrated by line drawings began with the same initial consonants. A total of 10 trials were given, and the number of correct responses was recorded.

Year 7. Direct assessment: At age 7 years 6 months, reading was assessed with measures on the basis of the Wechsler Objective Reading Dimensions (WORD) (57). Pictures and words were used to assess decoding and word reading. The child was shown a series of four pictures. Each picture had four short, simple words underneath it. The child was asked to point to the word that had the same beginning or ending sound as the picture. This request was then followed by a series of three pictures, each with four words beneath, each starting with the same letter as the picture. The child was asked to point to the word that correctly named the picture. The child was then asked to read aloud a series of 48 unconnected words that increased in difficulty. Total numbers of correct responses were used. Spelling was assessed by a series of 15 words that were piloted and chosen by the ALSPAC team (e.g., chin, brought, and telephone). Each word was read aloud on its own, within a specific sentence incorporating the word, and lastly read alone again. The child was asked to write down the spelling of the word even if he or she was just guessing. The total number of words spelled correctly was tallied and used. In addition to the WORD, the Phoneme Detection Task (58) comprised 40 test items of increasing difficulty. It involved asking the child to repeat a word and then to say it again, but with some part of the word (a phoneme or number of phonemes) removed. Total numbers of correct responses were used.

Year 8. Direct assessment: At age 8 years 6 months, four verbal subscale raw scores on the Wechsler Intelligence Scale for Children—III UK Edition (WISC) (59) were used: information, comprehension, vocabulary, and similarity. Two subtests of the Wechsler Objective Language Dimensions (WOLD) (60) were used to measure listening comprehension and oral expression. Listening comprehension involves the child listening to the tester reading aloud a paragraph about a displayed picture. The child then answers questions on what was heard. The child has to make inferences about what was read to them and answer the questions verbally. Expressive vocabulary was assessed by a series of 10 pictures. The total num-

bers of correct responses on comprehension and expressive vocabulary were each tallied and used.

Year 9. Direct assessment: At age 9 years 6 months, reading was assessed using the basic reading subtest of the WORD (57). Children were asked to read aloud 10 real words (e.g., huge, union, and smape), followed by 10 nonreal words (e.g., duter, uningest, and smape). Both the real and nonreal words were selected from a larger list of words taken from research conducted by Nunes *et al.* (61). Total numbers of correct responses on real and nonreal words were each tallied and used. The revised Neale Analysis of Reading Ability (NARA II) (62) was used to assess children’s reading skills and comprehension. In this test, children read aloud short passages of stories that resulted in an accuracy score, and their answers to a series of questions about the content of the story resulted in a reading comprehension score.

Year 13. Direct assessment: At age 13 years 6 months, word reading efficiency was assessed by word and pseudoword tests of the Test of Word Reading Efficiency (TOWRE) (63). Children were asked to read out loud 104 real words (e.g., complete and wonderful), followed by a list of 63 nonreal words (e.g., glack and framble). Total numbers of correct responses on real and nonreal words were tallied; because of a very high correlation between the two reading scores, $r = 0.81$, a mean standard score was computed and used in analysis.

Year 15. Direct assessment: At age 15 years 6 months, the vocabulary subscale raw score on the Wechsler Abbreviated Scale of Intelligence (WASI) (64) was used.

Covariates

We assessed the possibility that child nonverbal intelligence (29) and sociability (65), both of which are known to be associated with child language, and mothers’ age and education, both of which are also known to be associated with child language, would account for some of the stability of language competence and performance. Specific covariates (child nonverbal intelligence and sociability) were presumed to be associated with child language variables concurrently or prospectively (but not retrospectively). General covariates (maternal age and education) were presumed to be associated with all child language variables, regardless of the child’s age.

Children’s nonverbal intelligence was assessed three times at clinic visits. At age 4 years 1 month, the performance IQ score of the WPPSI (41) was used. At age 8 years 6 months, the performance IQ score of the WISC (59) was used. At age 15 years 6 months, nonverbal intelligence was measured by the Matrix Reasoning subtest of the WASI (64).

Child sociability was obtained from caregiver reports across data collection waves. At ages 6 months, 1 year 6 months, and 2 years 6 months, the social achievement scores of the adapted DDST (53) were used. At ages 3 years 2 months, 4 years 9 months, and 5 years 9 months, the sociability scores from the Emotionality, Activity, Sociability Temperament questionnaire (66, 67) were used.

Maternal age at childbirth was calculated from the date of delivery and the mother’s date of birth recorded at enrollment. Educational attainment was obtained from a questionnaire sent home at 32 weeks gestation.

Statistical analysis

The SDs and ranges of all language measures (table S1) indicated considerable variation, as is common in the literature and prerequisite to assessments of stability. Variable distributions were examined for univariate normality (68), and transformations were applied to improve distributions. Because of the range of child age at each wave,

we explored correlations of child age with all raw test scores to determine whether age adjustment was warranted. Age-adjusted scores were computed for all language variables that showed significant concurrent correlations with child age and were used in structural equation models (SEMs).

Language stability was evaluated by fitting SEMs using maximum likelihood functions (MLFs) and followed the mathematical models of Bentler and Weeks (69), as implemented in EQS 6.1 (70). SEM is a robust tool for assessing stability because latent variables capture shared variance among their indicators, and so variance uniquely associated with rater bias, random measurement error, or specific error (variance arising from some characteristic unique to a particular indicator that was not accounted for by the factor) is relegated to its error term.

Missing data points (20.4% of the total data) were handled in EQS using full information maximum likelihood with a two-stage Expectation-Maximization estimation of the structured model and the MLF (71). Monte Carlo studies have demonstrated the general superiority of the structured-model EM method implemented in EQS 6.1 compared to other techniques to recover missing data (72, 73). In the course of fitting SEMs, we evaluated Mardia (74) coefficients of multivariate kurtosis and the cases that contributed most to those estimates, as well as the stability of parameter estimates and the cases that contributed disproportionately to parameter estimates. No significant problems with influential cases emerged. Model fit was assessed using scaled $Y-B \chi^2$ statistic, robust CFI, standardized SRMR (75), and RMSEA. Cutoff values ≈ 0.95 for CFI and ≈ 0.09 and ≈ 0.06 for SRMR and RMSEA, respectively, are indicative of a relatively good fit between the hypothesized model and observed data (21). We gave greater weight to the incremental/approximate fit indices than to χ^2 because the χ^2 value is known to be sensitive to sample size (76) and the size of the correlations in the model (77). Standardized path coefficients are presented in text and figures.

For correlations and standardized path coefficients, we adopted conventional magnitudes of r corresponding to small, medium, and large effect sizes as ≈ 0.10 , 0.30, and 0.50, respectively (78, p. 61). All stabilities were large except two medium-sized stabilities between 6 months and year 1 and between the single observed variables at 13 and 15 years.

Next, we explored whether specific covariates were associated with child language measures. We calculated correlations of (i) year 4 WPPSI performance IQ with years 4, 5, 6, and 7 language variables; (ii) year 8 WISC performance IQ with years 8, 9, and 13 language variables; (iii) year 15 WASI matrix reasoning with vocabulary; (iv) child sociability with concurrent language variables from ages 6 months through year 5; and (v) year 5 child sociability with all language variables from years 7 through 15. Children's non-verbal intelligence significantly correlated with all language variables (r values ranged from 0.15 to 0.50, all $P \leq 0.001$); however, children's sociability related to only some language variables, with significant correlations ranging from 0.07 ($P < 0.05$) to 0.46 ($P < 0.001$). To test whether the stability model held controlling for specific covariates and the two general covariates, we re-evaluated the a priori model (Fig. 1) using the adjusted language scores with the shared variance with specific covariates removed and adding the two general covariates as exogenous variables to the SEM. Direct paths from maternal age and education to all eight language-latent variables, and the three observed variables at 6 months and 13 and 15 years, were added to the model.

Study 2: Long-term language stability in typical and atypical development

Participants

Table S2 presents sample sizes and child ages at each data collection wave by groups. Maternal education ranged from secondary education (13.2%), vocational (8.5%), O level (36.3%), A level (26.4%), to university degree (15.6%). Maternal social class ranged from unskilled (1.4%), partly skilled (7.7%), skilled manual (6.3%), skilled nonmanual (42.6%), managerial and technical (35.4%), to professional (6.6%).

At age 7, hearing function was assessed using air conduction pure tone audiometry carried out by audiologists and trained physiology staff during a clinic visit. All measurements were carried out as described in Hall *et al.* (79), and hearing thresholds were measured in both ears according to audiometry procedures recommended by the British Society of Audiology (80). At age 9, children's primary caregivers were asked whether they were ever told that the child has "dyslexia" or "autism," "Asperger's syndrome," or "autistic spectrum disorder."

Procedures

We used data collected across 10 ALSPAC collection waves. However, we aggregated data collected at ages 1 year 3 months and 1 year 6 months (2 waves) into the year 1 measures; thus, we studied language stability across nine ages (fig. S2).

Statistical analysis

The SDs and ranges of all language measures (table S3) again indicate considerable variation. A language stability model was fit on the total sample by using SEM with EQS 6.1 (70). See study 1 for SEM applications. Age-adjusted scores were used in analysis, and missing data points (11.7% of the total data) were handled in EQS using full information maximum likelihood. In the full-sample stability model (fig. S2), all stabilities were medium or large.

Given the complexity of the stability model and the relatively small sizes of the at-risk comparison groups, we generated the generalized least squares factor scores from the SEM and retained them for further analysis. Study 2 language stability of the at-risk subgroups was assessed using Pearson correlation coefficients. Missing data points for the three observed language variables at 6 months and 13 and 15 years were imputed in the control and in each at-risk group (missingness ranged from 7.9 to 17.8% of the total data) separately using the Expectation-Maximization algorithm (81) in SPSS (82). First, zero-order correlations showed language stability before taking specific and general covariates into consideration. Then, language stability was reassessed using partial correlations controlling for general and specific covariates. For specific covariates, we first calculated correlations of (i) year 8 WISC performance IQ with years 8, 9, and 13 language variables; (ii) year 15 WASI matrix reasoning with vocabulary; (iii) child sociability with concurrent language variables from ages 6 months through year 3; and (iv) year 3 child sociability with all language variables from years 7 through 15. Again, child non-verbal intelligence significantly correlated with all language variables (r values ranged from 0.25 to 0.46, all $P < 0.001$), and child sociability related to only some language variables with significant correlations ranging from 0.03 ($P < 0.05$) to 0.51 ($P < 0.001$). Unstandardized residuals of the related language variables controlling for significant covariates, where applicable, were computed before performing specific partial correlations. Last, language stability was reassessed by computing partial correlations of adjusted language scores (with the shared variance with specific covariates removed) controlling for maternal age and education.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/11/eaat7422/DC1>

Study 1 Supplementary Text

Study 1 Materials and Methods

Study 1 Measurement Models

Study 2 Supplementary Text

Table S1. Study 1: Child language measures and covariates: Descriptive statistics.

Table S2. Study 2: Sample size and child age at each data collection wave.

Table S3. Study 2: Child language measures and covariates: Descriptive statistics.

Fig. S1. Study 1.

Fig. S2. Study 2.

References (83–144)

REFERENCES AND NOTES

- R. Lewontin, *The Triple Helix* (Harvard Univ. Press, 2005).
- R. C. Anderson, P. Freebody, Reading comprehension and the assessment and acquisition of word knowledge. *Adv. Reading/Lang. Res.* **2**, 231–256 (1983).
- G. J. Duncan, C. J. Dowsett, A. Claessens, K. Magnuson, A. C. Huston, P. Klebanov, L. S. Pagani, L. Feinstein, M. Engel, J. Brooks-Gunn, H. Sexton, K. Duckworth, C. Japel, School readiness and later achievement. *Dev. Psychol.* **43**, 1428–1446 (2007).
- J. K. Reynell, M. Huntley, *Reynell Developmental Language Scales* (NFER-Nelson, 1985).
- D. R. Lynam, B. Henry, The role of neuropsychological deficits in conduct disorders, in *Conduct Disorders in Childhood and Adolescence*, J. Hill, B. Maughan, Eds. (Cambridge Univ. Press, 2001), pp. 235–263.
- S. G. K. Yew, R. O'Kearney, Emotional and behavioural outcomes later in childhood and adolescence for children with specific language impairments: Meta-analyses of controlled prospective studies. *J. Child Psychol. Psychiatry* **54**, 516–524 (2013).
- M. H. Bornstein, C.-S. Hahn, J. T. D. Suwalsky, Language and internalizing and externalizing behavioral adjustment: Developmental pathways from childhood to adolescence. *Dev. Psychopathol.* **25**, 857–878 (2013).
- I. T. Petersen, J. E. Bates, B. M. D'Onofrio, C. A. Coyne, J. E. Lansford, K. A. Dodge, G. S. Petit, C. A. Van Hulle, Language ability predicts the development of behavior problems in children. *J. Abnorm. Psychol.* **122**, 542–557 (2013).
- B. MacWhinney, M. H. Bornstein, in *Well-Being: Positive Development Across the Life Course*, M. H. Bornstein, L. Davidson, C. M. Keyes, K. Moore, The Center for Child Well-Being, Eds., (Erlbaum, 2003), pp. 331–339.
- H. M. Feldman, C. A. Dollaghan, T. F. Campbell, M. Kurs-Lasky, J. E. Janosky, J. L. Paradise, Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Dev.* **71**, 310–322 (2000).
- L. Fenson, P. S. Dale, J. S. Reznick, E. Bates, D. J. Thal, S. J. Pethick, M. Tomasello, C. B. Mervis, J. Stiles, Variability in early communicative development. *Monogr. Soc. Res. Child Dev.* **59**, 1–185 (1994).
- C. S. Hammer, P. Morgan, G. Farkas, M. Hillemeier, D. Bitetti, S. Maczuga, Late talkers: A population-based study of risk factors and school readiness consequences. *J. Speech Lang. Hear. Res.* **60**, 607–626 (2017).
- M. H. Bornstein, D. L. Putnick, G. Esposito, Continuity and stability in development. *Child Dev. Perspect.* **11**, 113–119 (2017).
- L. A. Sroufe, The coherence of individual development: Early care, attachment, and subsequent developmental issues. *Am. Psychol.* **34**, 834–841 (1979).
- R. B. Kline, *Principles and Practice of Structural Equation Modeling* (Guilford Press, 2015).
- A. Boyd, J. Golding, J. Macleod, D. A. Lawlor, A. Fraser, J. Henderson, L. Molloy, A. Ness, S. Ring, G. Davey Smith, Cohort Profile: The 'children of the 90s'—The index offspring of the Avon Longitudinal Study of Parents and Children (ALSPAC). *Int. J. Epidemiol.* **42**, 111–127 (2013).
- A. Fraser, C. Macdonald-Wallis, K. Tilling, A. Boyd, J. Golding, G. Davey Smith, J. Henderson, J. Macleod, L. Molloy, A. Ness, S. Ring, S. M. Nelson, D. A. Lawlor, Cohort profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42**, 97–110 (2013).
- J. Golding, M. Pembrey, R. Jones; The Alspac Study Team, ALSPAC—The Avon Longitudinal Study of Parents and Children. *Paediatr. Perinat. Epidemiol.* **15**, 74–87 (2001).
- L. T. Hu, P. M. Bentler, Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct. Equ. Modeling* **6**, 1–55 (1999).
- S. Cangur, I. Ercan, Comparison of model fit indices used in structural equation modeling under multivariate normality. *J. Mod. Appl. Stat. Methods* **14**, 152–167 (2015).
- L.-T. Hu, P. M. Bentler, Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychol. Methods* **3**, 424–453 (1998).
- M. H. Bornstein, C.-S. Hahn, D. L. Putnick, Long-term stability of core language skill in children with contrasting language skills. *Dev. Psychol.* **52**, 704–716 (2016).
- M. H. Bornstein, C.-S. Hahn, D. L. Putnick, Stability of core language skill across the first decade of life in children at biological and social risk. *J. Child Psychol. Psychiatry* **57**, 1434–1443 (2016).
- M. H. Bornstein, C. S. Hahn, D. L. Putnick, J. T. D. Suwalsky, Stability of core language skill from early childhood to adolescence: A latent variable approach. *Child Dev.* **85**, 1346–1356 (2014).
- D. L. Putnick, M. H. Bornstein, S. Eryigit-Madzwamuse, D. Wolke, Long-term stability of language performance in very preterm, moderate-late preterm, and term children. *J. Pediatr.* **181**, 74–79.e3 (2017).
- N. Harlaar, M. Trzakowski, P. S. Dale, R. Plomin, Word reading fluency: Role of genome-wide-single-nucleotide polymorphisms in developmental stability and correlations with print exposure. *Child Dev.* **85**, 1190–1205 (2014).
- M. E. Arterberry, M. H. Bornstein, C. Midgett, D. L. Putnick, M. H. Bornstein, Early attention and literacy experiences predict adaptive communication. *First Lang.* **27**, 175–189 (2007).
- M. E. Hayiou-Thomas, P. S. Dale, R. Plomin, The etiology of variation in language skills changes with development: A longitudinal twin study of language from 2 to 12 years. *Dev. Sci.* **15**, 233–249 (2012).
- M. H. Bornstein, O. M. Haynes, Vocabulary competence in early childhood: Measurement, latent construct, and predictive validity. *Child Dev.* **69**, 654–671 (1998).
- E. Colledge, D. Bishop, G. Koeppen-Schomerus, T. S. Price, F. G. E. Happé, T. C. Eley, P. S. Dale, R. Plomin, The structure of language abilities at 4 years: A twin study. *Dev. Psychol.* **38**, 749–757 (2002).
- P. S. Dale, N. Harlaar, M. E. Hayiou-Thomas, R. Plomin, The etiology of diverse receptive language skills at 12 years. *J. Speech Lang. Hear. Res.* **53**, 982–992 (2010).
- C. J. Johnson, J. H. Beitchman, A. Young, M. Escobar, L. Atkinson, B. Wilson, E. B. Brownlie, L. Douglas, N. Taback, I. Tam, M. Wang, Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *J. Speech Lang. Hear. Res.* **42**, 744–760 (1999).
- J. B. Tomblin, X. Zhang, The dimensionality of language ability in school-age children. *J. Speech Lang. Hear. Res.* **49**, 1193–1208 (2006).
- A. Trouton, F. M. Spinath, R. Plomin, Twins early development study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Res.* **5**, 444–448 (2002).
- N. Bayley, Consistency and variability in the growth of intelligence from birth to eighteen years. *J. Genet. Psychol.* **75**, 165–196 (1949).
- R. J. Sternberg, E. L. Grigorenko, D. A. Bundy, The predictive value of IQ. *Merrill Palmer Q.* **47**, 1–41 (2001).
- B. W. Roberts, W. F. DelVecchio, The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychol. Bull.* **126**, 3–25 (2000).
- R. M. Lerner, *Theories and Methods in Developmental Science* (Routledge, 2018).
- UNESCO, *Education for All: A Global Monitoring Report* (UNESCO, 2006).
- N. Bayley, Bayley scales of infant and toddler development® (3e, Bayley-III®, Pearson, 2005); www.pearsonclinical.com/childhood/products/10000123/bayley-scales-of-infant-and-toddler-development-third-edition-bayley-iii.html
- D. Wechsler, *Wechsler Preschool and Primary Scale of Intelligence* (The Psychological Corporation, Harcourt Brace, ed. Revised, 1990).
- D. G. Bonett, Replication-extension studies. *Curr. Dir. Psychol. Sci.* **21**, 409–412 (2012).
- G. J. Duncan, M. Engel, A. Claessens, C. J. Dowsett, Replication and robustness in developmental research. *Dev. Psychol.* **50**, 2417–2425 (2014).
- M. H. Bornstein, C.-S. Hahn, O. M. Haynes, Specific and general language performance across early childhood: Stability and gender considerations. *First Lang.* **24**, 267–304 (2004).
- L. B. Olswang, B. Rodriguez, G. Timler, Recommending intervention for toddlers with specific language learning difficulties: We may not have all the answers, but we know a lot. *Am. J. Speech Lang. Pathol.* **7**, 23–32 (1998).
- R. Paul, Clinical implications of the natural history of slow expressive language development. *Am. J. Speech Lang. Pathol.* **5**, 5–21 (1996).
- D. Thal, J. Katic, Predicaments in early identification of specific language impairment: Does the early bird always catch the worm?, in *Assessment of Communication and Language*, K. Cole, P. Dale, D. Thal, Eds. (Brookes, 1996), pp. 1–28.
- W. S. Barnett, K. Jung, D. J. Yarosz, J. Thomas, A. Hornbeck, R. Stechuk, S. Burns, Education effects of the Tools of the Mind curriculum: A randomized trial. *Early Child. Res. Q.* **23**, 299–313 (2008).
- A. Diamond, W. S. Barnett, J. Thomas, S. Munro, Preschool program improves cognitive control. *Science* **318**, 1387–1388 (2007).
- A. Winsler, L. Manfra, R. M. Diaz, "Should I let them talk?": Private speech and task performance among preschool children with and without behavior problems. *Early Child. Res. Q.* **22**, 215–231 (2007).
- M. H. Bornstein, J. Jager, D. L. Putnick, Sampling in developmental science: Situations, shortcomings, solutions, and standards. *Dev. Rev.* **33**, 357–370 (2013).

52. J. Jager, D. L. Putnick, M. H. Bornstein, II. More than just convenient: The scientific merits of homogeneous convenience samples. *Monogr. Soc. Res. Child Dev.* **82**, 13–30 (2017).
53. W. K. Frankenburg, J. B. Dodds, The Denver developmental screening test. *J. Pediatr.* **71**, 181–191 (1967).
54. J. K. Reynell, *The Reynell Developmental Language Scales* (NFER-Nelson Windsor, ed. revised, 1977).
55. C. E. Renfrew, *Bus Story Test (The Renfrew Language Scales)* (Routledge, 1997).
56. B. Byrne, R. Fielding-Barnsley, Recognition of phoneme invariance by beginning readers. *Read Writ.* **5**, 315–324 (1993).
57. J. Rust, S. Golombok, G. Trickey, *Wechsler Objective Reading Dimensions* (Psychological Corporation, 1993).
58. J. Rosner, D. P. Simon, The auditory analysis test: An initial report. *J. Learn. Disabil.* **4**, 384–392 (1971).
59. D. Weschler, J. Golombok, S. Rust, *Manual for the Weschler Intelligence Scale for Children* (Psychological Corporation, ed. 3, 1999).
60. J. Rust, *The Manual of the Wechsler Objective Language Dimensions (WOLD)* (The Psychological Corporation, 1996).
61. T. Nunes, P. Bryant, J. Olsson, Learning morphological and phonological spelling rules: An intervention study. *Sci. Stud. Read.* **7**, 289–307 (2003).
62. M. D. Neale, *Neale Analysis of Reading Ability—Revised: Manual for Schools* (NFER-Nelson, 1997).
63. R. K. Wagner, J. K. Torgesen, C. A. Rashotte, *Comprehensive Test of Phonological Processes* (PRO-ED Publishing, Inc., 1999).
64. D. Weschler, J. Golombok, S. Rust, *Manual for the Weschler Abbreviated Scale of Intelligence* (Psychological Corporation, 1999).
65. M. H. Bornstein, O. M. Haynes, K. M. Painter, Sources of child vocabulary competence: A multivariate model. *J. Child Lang.* **25**, 367–393 (1998).
66. A. H. Buss, R. Plomin, *Temperament: Early Developing Personality Traits* (Erlbaum, 1984).
67. M. H. Bornstein, C.-S. Hahn, D. L. Putnick, R. Pearson, Stability of child temperament: Multiple moderation by child and mother characteristics. *Br. J. Dev. Psychol.* **10.1111/bjdp.12253** (2018).
68. B. G. Tabachnick, L. S. Fidell, *Using Multivariate Statistics* (Allyn and Bacon, ed. 6, 2012).
69. P. M. Bentler, D. G. Weeks, Linear structural equations with latent variables. *Psychometrika* **45**, 289–308 (1980).
70. P. M. Bentler, *EQS 6 Structural Equations Program Manual* (Multivariate Software, 2006).
71. M. Jamshidian, P. M. Bentler, ML estimation of mean and covariance structures with missing data using complete data routines. *J. Educ. Behav. Stat.* **24**, 21–24 (1999).
72. M. S. Gold, P. M. Bentler, Treatments of missing data: A Monte Carlo comparison of RBHDI, iterative stochastic regression imputation, and expectation-maximization. *Struct. Equ. Modeling* **3**, 319–355 (2009).
73. K.-H. Yuan, P. M. Bentler, Normal theory based test statistics in structural equation modelling. *Br. J. Math. Stat. Psychol.* **51**, 289–309 (1998).
74. K. V. Mardia, Measures of multivariate skewness and kurtosis with applications. *Biometrika* **57**, 519–530 (1970).
75. M. W. Browne, R. Cudeck, in *Testing Structural Equation Models*, K. A. Bollen, J. S. Long, Eds. (Sage, 1993), pp. 136–162.
76. G. W. Cheung, R. B. Rensvold, Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Modeling* **9**, 233–255 (2002).
77. J. Miles, M. Shelvin, A time and a place for incremental fit indices. *Personal. Individ. Differ.* **42**, 869–874 (2007).
78. J. Cohen, P. Cohen, *Applied Multiple Regression/Correlation Analysis for Behavioral Sciences* (Lawrence Erlbaum Associates, 1983).
79. A. Hall, M. Pembrey, M. Lutman, C. Steer, M. Bitner-Glindzicz, Prevalence and audiological features in carriers of GJB2 mutations, c.35delG and c.101T>C (p.M34T), in a UK population study. *BMJ Open* **2**, e001238 (2012).
80. British Society of Audiology, “Pure tone air and bone conduction threshold audiometry with and without masking and determination of uncomfortable loudness levels” (2012); www.thebsa.org.uk/wp-content/uploads/2014/04/BSA_RP_PTA_FINAL_24Sept11_MinorAmend06Feb12.pdf.
81. A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **39**, 1–38 (1977).
82. IBM Corp. IBM SPSS Statistics for Windows, Version 22.0 (IBM Corp., 2003).
83. D. Gentner, A. Özyürek, Ö. Gurcanli, S. Goldin-Meadow, Spatial language facilitates spatial cognition: Evidence from children who lack language input. *Cognition* **127**, 318–330 (2013).
84. S. M. Pruden, S. C. Levine, J. Huttenlocher, Children’s spatial thinking: Does talk about the spatial world matter? *Dev. Sci.* **14**, 1417–1430 (2011).
85. R. Fivush, Constructing narrative, emotion, and self in parent-child conversations about the past, in *The Remembering Self: Construction and Accuracy in the Self-Narrative*, U. Neisser, R. Fivush, Eds. (Cambridge Univ. Press, 1994), pp. 136–157.
86. R. Fivush, C. A. Haden, Parent-child reminiscing and the construction of a subjective self, in *The Development of Social Cognition and Communication*, B. D. Homer, C. S. Tamis-LeMonda, Eds. (Lawrence Erlbaum, 2005), pp. xvii–378.
87. M. Lazaridis, The emergence of a temporally extended self and factors that contribute to its development: From theoretical and empirical perspectives: Abstract. *Monogr. Soc. Res. Child Dev.* **78**, vii (2013).
88. J. W. Astington, J. A. Baird, Eds., *Why Language Matters for Theory of Mind* (Oxford Univ. Press, 2005), pp. xii–355.
89. P. A. de Villiers, The role of language in theory-of-mind development: What deaf children tell us, in *Why Language Matters for Theory of Mind*, J. W. Astington, J. A. Baird, Eds. (Oxford Univ. Press, 2005), pp. xii–355.
90. B. Schick, P. De Villiers, J. De Villiers, R. Hoffmeister, Language and theory of mind: A study of deaf children. *Child Dev.* **78**, 376–396 (2007).
91. A. C. Watson, K. M. Painter, M. H. Bornstein, Longitudinal relations between 2-year-olds’ language and 4-year-olds’ theory of mind. *J. Cogn. Dev.* **2**, 449–457 (2001).
92. M. H. Bornstein, D. L. Putnick, Stability of language in childhood: A multiage, multidomain, multimeasure, and multisource study. *Dev. Psychol.* **48**, 477–491 (2012).
93. M. I. Appelbaum, R. B. McCall, in *Handbook in Child Psychology: Vol. 1. History, Theory, and Methods*, W. Kessen, Ed., P. H. Mussen, Series Ed. (Wiley, 1983), pp. 415–476.
94. R. B. McCall, Nature-nurture and the two realms of development: A proposed integration with respect to mental development. *Child Dev.* **52**, 1–12 (1981).
95. R. M. Lerner, R. M. Hershberg, L. J. Hilliard, S. K. Johnson, in *Developmental Science: An Advanced Textbook*, M. H. Bornstein, M. E. Lamb, Eds. (Psychology Press, ed. 7, 2015), pp. 3–41.
96. C. O. Eckerman, C. C. Davis, S. M. Didow, Toddlers’ emerging ways of achieving social coordinations with a peer. *Child Dev.* **60**, 440–453 (1989).
97. M. J. Guralnick, Social interactions among preschool children. *Except. Child.* **46**, 248–253 (1980).
98. C. Howes, C. C. Matheson, Sequences in the development of competent play with peers: Social and social pretend play. *Dev. Psychol.* **28**, 961–974 (1992).
99. D. Bellinger, Consistency in the pattern of change in mother’s speech: Some discriminant analyses. *J. Child Lang.* **7**, 469–487 (1980).
100. K. A. Clarke-Stewart, L. P. Vanderstoep, G. A. Killian, Analysis and replication of mother-child relations at two years of age. *Child Dev.* **50**, 777–793 (1979).
101. L. Rescorla, T. Fechnay, Mother-child synchrony and communicative reciprocity in late-talking toddlers. *J. Speech Hearing Res.* **39**, 200–208 (1996).
102. G. Cross, in *The Development of Communication*, N. Waterson, C. Snow, Eds. (Wiley, 1978), pp. 199–216.
103. B. McLaughlin, D. White, T. McDevitt, R. Raskin, Mothers’ and fathers’ speech to their young children: Similar or different? *J. Child Lang.* **10**, 245–252 (1983).
104. D. P. Hartmann, C. B. Abbott, K. E. Pelzel, in *Developmental Science: An Advanced Textbook*, M. H. Bornstein, M. E. Lamb, Eds. (Psychology Press, ed. 7, 2015), pp. 113–214.
105. H. Blencowe, S. Cousens, M. Z. Oestergaard, D. Chou, A.-B. Moller, R. Narwal, A. Adler, C. V. Garcia, S. Rohde, L. Say, J. E. Lawn, National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: A systematic analysis and implications. *Lancet* **379**, 2162–2172 (2012).
106. J. D. Eicher, J. R. Gruen, Language impairment and dyslexia genes influence language skills in children with autism spectrum disorders. *Autism Res.* **8**, 229–234 (2015).
107. J. Rispens, P. Been, Subject-verb agreement and phonological processing in developmental dyslexia and specific language impairment (SLI): A closer look. *Int. J. Lang. Commun. Disord.* **42**, 293–305 (2007).
108. E. K. Robertson, M. F. Joannis, A. S. Desroches, A. Terry, Past-tense morphology and phonological deficits in children with dyslexia and children with language impairment. *J. Learn. Disabil.* **46**, 230–240 (2013).
109. M. J. Snowling, M. Melby-Lervåg, Oral language deficits in familial dyslexia: A meta-analysis and review. *Psychol. Bull.* **142**, 498–545 (2016).
110. E. Gerrits, E. de Bree, Early language development of children at familial risk of dyslexia: Speech perception and production. *J. Commun. Disord.* **42**, 180–194 (2009).
111. M. Alt, T. Hogan, S. Green, S. Gray, K. Cabbage, N. Cowan, Word learning deficits in children with dyslexia. *J. Speech Lang. Hear. Res.* **60**, 1012–1028 (2017).
112. M. Duranovic, S. Tinjak, A. Turbic-Hadzagic, Morphological knowledge in children with dyslexia. *J. Psycholinguist. Res.* **43**, 699–713 (2014).
113. R. Lovio, R. Näätänen, T. Kujala, Abnormal pattern of cortical speech feature discrimination in 6-year-old children at risk for dyslexia. *Brain Res.* **1335**, 53–62 (2010).
114. S. van Viersen, E. H. de Bree, M. Verdam, E. Krikhaar, B. Maassen, A. van der Leij, P. F. de Jong, Delayed early vocabulary development in children at family risk of dyslexia. *J. Speech Lang. Hear. Res.* **60**, 937–949 (2017).
115. J. D. Eicher, A. M. Montgomery, N. Akshoomoff, D. G. Amaral, C. S. Bloss, O. Libiger, N. J. Schork, B. F. Darst, B. J. Casey, L. Chang, T. Ernst, J. Frazier, W. E. Kaufmann, B. Keating, T. Kenet, D. Kennedy, S. Mostofsky, S. S. Murray, E. R. Sowell, H. Bartsch, J. M. Kuperman, T. T. Brown, D. J. Hagler Jr., A. M. Dale, T. L. Jernigan, J. R. Gruen; Pediatric Imaging Neurocognition Genetics Study, Dyslexia and language impairment associated genetic

- markers influence cortical thickness and white matter in typically developing children. *Brain Imaging Behav.* **10**, 272–282 (2016).
116. N. M. Raschle, B. L. C. Becker, S. Smith, L. V. Fehlbauer, Y. Wang, N. Gaab, Investigating the influences of language delay and/or familial risk for dyslexia on brain structure in 5-year-olds. *Cereb. Cortex* **27**, 764–776 (2017).
 117. S. L. Rimrodt, D. J. Peterson, M. B. Denckla, W. E. Kaufmann, L. E. Cutting, White matter microstructural differences linked to left perisylvian language network in children with dyslexia. *Cortex* **46**, 739–749 (2010).
 118. U. Richardson, P. Kulju, L. Nieminen, P. Torvelainen, Early signs of dyslexia from the speech and language processing of children. *Int. J. Speech-Lang. Pathol.* **11**, 366–380 (2009).
 119. D. Garrido, D. Petrova, L. R. Watson, R. Garcia-Retamero, G. Carballo, Language and motor skills in siblings of children with autism spectrum disorder: A meta-analytic review. *Autism Res.* **10**, 1737–1750 (2017).
 120. K. W. Sheppard, K. M. Boone, B. Gracious, M. A. Klebanoff, L. K. Rogers, J. Rausch, C. Bartlett, D. L. Coury, S. A. Keim, Effect of omega-3 and -6 supplementation on language in preterm toddlers exhibiting autism spectrum disorder symptoms. *J. Autism Dev. Disord.* **47**, 3358–3369 (2017).
 121. L. J. Hahn, N. C. Brady, L. McCary, L. Rague, J. E. Roberts, Early social communication in infants with fragile X syndrome and infant siblings of children with autism spectrum disorder. *Res. Dev. Disabil.* **71**, 169–180 (2017).
 122. E. Haebig, J. R. Saffran, S. Ellis Weismer, Statistical word learning in children with autism spectrum disorder and specific language impairment. *J. Child Psychol. Psychiatry* **58**, 1251–1263 (2017).
 123. W. A. Helland, T. Helland, Emotional and behavioural needs in children with specific language impairment and in children with autism spectrum disorder: The importance of pragmatic language impairment. *Res. Dev. Disabil.* **70**, 33–39 (2017).
 124. E. M. Whyte, K. E. Nelson, Trajectories of pragmatic and nonliteral language development in children with autism spectrum disorders. *J. Commun. Disord.* **54**, 2–14 (2015).
 125. L. B. Adamson, R. Bakeman, K. Suma, D. L. Robins, An expanded view of joint attention: Skill, engagement, and language in typical development and autism. *Child Dev.* 10.1111/cdev.12973 (2017).
 126. M. M. Davidson, S. E. Weismer, Reading comprehension of ambiguous sentences by school-age children with autism spectrum disorder. *Autism Res.* **10**, 2002–2022 (2017).
 127. M. Jyotishi, D. Fein, L. Naigles, Investigating the grammatical and pragmatic origins of wh-questions in children with autism spectrum disorders. *Front Psychol.* **8**, 319 (2017).
 128. E. L. Coderre, M. Chernenok, B. Gordon, K. Ledoux, Linguistic and non-linguistic semantic processing in individuals with autism spectrum disorders: An ERP study. *J. Autism Dev. Disord.* **47**, 795–812 (2017).
 129. C. Brynkvog, I.-M. Eigsti, M. Jørgensen, S. Lemcke, O.-S. Bohn, P. Krøjgaard, Syntax and morphology in Danish-speaking children with autism spectrum disorder. *J. Autism Dev. Disord.* **47**, 373–383 (2017).
 130. M. F. Westerveld, J. Paynter, D. Trembath, A. A. Webster, A. M. Hodge, J. Roberts, The emergent literacy skills of preschool children with autism spectrum disorder. *J. Autism Dev. Disord.* **47**, 424–438 (2017).
 131. E. L. Bavin, E. K. Baker, in *Innovative Investigations of Language in Autism Spectrum Disorder*, L. R. Naigles, Ed. (Language and the Human Lifespan Series, American Psychological Association, 2017), pp. 35–47.
 132. J. C. Visser, N. N. J. Rommelse, M. Lappenschaar, I. J. Servatius-Oosterling, C. U. Greven, J. K. Buitelaar, Variation in the early trajectories of autism symptoms is related to the development of language, cognition, and behavior problems. *J. Am. Acad. Child Adolesc. Psychiatry* **56**, 659–668 (2017).
 133. K. D. Bopp, P. Mirenda, Prelinguistic predictors of language development in children with autism spectrum disorders over four-five years. *J. Child Lang.* **38**, 485–503 (2011).
 134. S. R. Edmunds, L. V. Ibañez, Z. Warren, D. S. Messinger, W. L. Stone, Longitudinal prediction of language emergence in infants at high and low risk for autism spectrum disorder. *Dev. Psychopathol.* **29**, 319–329 (2017).
 135. K. E. Nelson, J. Welsh, in *The Speech/Language Connection*, R. Paul, Ed. (Paul H. Brookes, 1997), pp. 199–225.
 136. E. Klima, U. Bellugi, *The Signs of Language* (Harvard Univ. Press, 1979).
 137. M. Marschark, *Psychological Development of Deaf Children* (Oxford Univ. Press, 1993).
 138. E. K. Newport, R. P. Meier, in *The Cross-Linguistic Study of Language Acquisition: Vol. 1. The Data*, D. I. Slobin, Ed. (Erlbaum, 1985), pp. 881–938.
 139. P. S. Day, Deaf children's expression of communicative intentions. *J. Commun. Disord.* **19**, 367–385 (1986).
 140. M. Harris, *Language Experience and Early Language Development: From Input to Uptake* (Erlbaum, 1992).
 141. K. P. Meadow, M. T. Greenberg, C. Erting, H. Carmichael, Interactions of deaf mothers and deaf preschool children: Comparisons with three other groups of deaf and hearing dyads. *Am. Ann. Deaf* **126**, 454–468 (1981).
 142. D. Moores, *Educating the Deaf: Psychology, Principles and Practices* (Houghton Mifflin, 1987).
 143. M. Rodda, C. Grove, *Language, Cognition, and Deafness* (Erlbaum, 1987).
 144. M. H. Bornstein, A. M. Selmi, O. M. Haynes, K. M. Painter, E. S. Marx, Representational abilities and the hearing status of child/mother dyads. *Child Dev.* **70**, 833–852 (1999).

Acknowledgments: We are grateful to all the families who took part in this study, the midwives for help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. **Funding:** The UK Medical Research Council and Wellcome (grant ref. 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors, and M.H.B. will serve as guarantor for the contents of this paper. This research was also supported by the Intramural Research Program of the NIH/NICHD, USA, and an International Research Fellowship in collaboration with the Centre for the Evaluation of Development Policies (EdePO) at the Institute for Fiscal Studies (IFS), London, UK, funded by the European Research Council (ERC) under the Horizon 2020 research and innovation programme (grant agreement no. 695300-HKADeC-ERC-2015-AdG). **Author contributions:** M.H.B., C.-S.H., and D.L.P. conceptualized the study. R.M.P. curated the data. C.-S.H. analyzed the data. M.H.B. and C.-S.H. wrote the original draft of the manuscript. D.L.P. and R.M.P. reviewed and edited the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Data used for this submission will be made available on request to the Executive (alspac-exec@bristol.ac.uk). The ALSPAC data management plan describes in detail the policy regarding data sharing, which is through a system of managed open access.

Submitted 29 March 2018

Accepted 24 October 2018

Published 21 November 2018

10.1126/sciadv.aat7422

Citation: M. H. Bornstein, C.-S. Hahn, D. L. Putnick, R. M. Pearson, Stability of core language skill from infancy to adolescence in typical and atypical development. *Sci. Adv.* **4**, eaat7422 (2018).