JSLHR

Review Article

A Systematic Review and Meta-Analysis of Predictors of Expressive-Language Outcomes Among Late Talkers

Evelyn L. Fisher[a]

Purpose: The purpose of this study was to explore the literature on predictors of outcomes among late talkers using systematic review and meta-analysis methods. We sought to answer the question: What factors predict preschool-age expressive-language outcomes among late-talking toddlers?
Method: We entered carefully selected search terms into the following electronic databases: Communication & Mass Media Complete, ERIC, Medline, PsycEXTRA, Psychological and Behavioral Sciences, and PsycINFO. We conducted a separate, random-effects model meta-analysis for each individual predictor that was used in a minimum of 5 studies. We also tested potential moderators of the relationship between predictors and outcomes using metaregression and subgroup analysis. Last, we conducted publication-bias and sensitivity analyses.
Results: We identified 20 samples, comprising 2,134 children, in a systematic review. According to the results of the meta-analyses, significant predictors of expressive-language outcomes included toddlerhood expressive-vocabulary size, receptive language, and socioeconomic status. Nonsignificant predictors included phrase speech, gender, and family history.
Conclusions: To our knowledge this is the first synthesis of the literature on predictors of outcomes among late talkers using meta-analysis. Our findings clarify the contributions of several constructs to outcomes and highlight the importance of early receptive language to expressive-language development.
Supplemental Materials: https://doi.org/10.23641/asha.5313454

Delay in expressive language is the one of the most common reasons for toddlers to be referred for neurodevelopmental evaluation (Whitehurst & Fischel, 1994). A substantial portion of these toddlers are found to be *late talkers,* a term used to describe children under age 3 years with unusually small vocabularies and no concomitant developmental disability or hearing impairment. Overall, the prognosis for late talkers is good, with the majority of children moving into the average range on language measures by preschool (Dollaghan, 2013). However, as a group, late talkers continue to perform lower than children in a control group on language measures in elementary school, and thus they are at heightened risk for language or learning disorders, such as specific language impairment (SLI; Paul, 2000; Rescorla, 2013). SLI is a developmental disorder characterized by persistent, functionally impairing language difficulties, and is typically diagnosed after age 4 years. The present study explores the literature on predictors of expressive-language outcomes among late talkers using systematic review and meta-analysis methods. See Table 1 for a summary of the aim.

Clinical and Theoretical Significance of the Study of Late Talkers

The study of predictors of expressive-language outcomes in late talkers has both clinical and theoretical significance. From a clinical perspective, knowledge about predictors increases the accuracy with which clinicians can distinguish between late bloomers, who will catch up to their peers, and children at high risk for language or learning disorders. The importance of early identification to optimizing outcomes is supported by research indicating that younger children experience greater gains with interventions and that language disorders evident at the age of 5 years remain relatively stable throughout the school years (Aram & Hall, 1989; Long, 2013). Knowledge about predictors also allows clinicians to provide appropriate reassurance to families of children at low risk for language

[a]Department of Psychology, Georgia State University, Atlanta
Correspondence to Evelyn L. Fisher: efisher7@student.gsu.edu.
Editor: Sean Redmond
Associate Editor: Emma Hayiou-Thomas
Received August 2, 2016
Revision received February 1, 2017
Accepted March 31, 2017
https://doi.org/10.1044/2017_JSLHR-L-16-0310

Disclosure: *The author has declared that no competing interests existed at the time of publication.*

Table 1. Study question, population, interventions, comparisons, outcomes, and designs.

Study component	The current study
Overall question	What factors, including child assessment and background information, predict preschool-age expressive-language outcomes among late-talking toddlers?
Population	Late-talking toddlers
Interventions or exposures	Child characteristics with regard to the following predictors: • expressive-vocabulary size • receptive language • phrase speech • socioeconomic status • gender • family history
Comparisons	Comparisons among children of differing levels of the predictors
Outcomes	Expressive language
Study designs	Prospective longitudinal studies

or learning disorders and to reduce the burden of unnecessary interventions on families and service providers.

Several researchers have raised questions about the value of studying late talkers, given the relatively low continuity between late talking and language or learning disorders such as SLI. However, the continued relevance of the study of late talkers is supported by the fact that late talking does have a modest amount of predictive value for future impairment. The estimated proportion of late talkers who go on to display persistent language difficulties varies widely, from 6% to 44% (Dale, Price, Bishop, & Plomin, 2003; Rescorla, 2002). Leonard (2014) points out that lower estimates of continuity between the two conditions come from studies with homogenous samples or stringent inclusion criteria. He argues that these studies may underestimate the predictive value of late talking in the general population because their samples consist of a subset of late talkers with no additional risk factors, such as receptive language delays, below-average nonverbal abilities, and low socioeconomic status (SES). Thus, the true proportion of late talkers that go on to have language or learning disorders such as SLI is probably substantially higher than the prevalence of SLI in the general population. Nonetheless, it is important to emphasize that late talking is a characteristic, not a disorder, and the most likely outcome for individual late talkers is that they will catch up to their peers with typical development.

From a theoretical perspective, the study of predictors enhances our understanding of developmental pathways leading to disordered and nondisordered language. For example, uncertainty remains about the range of normal lexical acquisition patterns and the significance of the fact that some children do not undergo a lexical burst, or acceleration in word learning, at the expected time (Fenson et al., 1994). Clarifying this issue could reveal information about the cumulative impact of adverse circumstances and the changing nature of linguistic competence over the life span. In addition, with regard to measurement, the study of predictors of outcomes in late talkers raises important questions about the reliability of assessment in toddlers and the conceptual distinction between measurements of knowledge and of information processing (Roy & Chiat, 2013).

Predictors of Outcomes Among Late Talkers

Several longitudinal research programs have examined predictors of expressive-language outcomes among late talkers. These predictors include both variables from toddlerhood assessments and background information (Rescorla, 2011). The most frequently examined toddlerhood-assessment predictor is expressive-vocabulary size. It is expected that late talkers with larger expressive vocabularies or milder expressive-vocabulary delays will have better outcomes. Across studies, expressive-vocabulary size has typically shown a statistically significant relationship with later language outcomes among late talkers (Fischel, Whitehurst, Caulfield, & DeBaryshe, 1989; Henrichs et al., 2011; Paul & Fountain, 1999; Westerlund, Berglund, & Eriksson, 2006). This suggests that there is some degree of continuity between toddlerhood and preschool- or school-age language abilities, though authors have been careful to point out that only a small amount of the variance in outcomes is explained by toddlerhood expressive vocabulary.

Several studies have also investigated the relationship between toddlerhood receptive language and language outcomes among late talkers. It is well established that comprehension of language usually precedes production in toddlers, and the profiles of some late talkers are characterized by particularly large comprehension–production gaps (Fenson et al., 1994). Thus, receptive-language assessments may offer an opportunity to examine the integrity of the foundation for language development in late talkers. Receptive language also typically shows a statistically significant relationship to language outcomes among late talkers (Henrichs et al., 2011; Lyytinen, Eklund, & Lyytinen, 2005; Thal, Marchman, & Tomblin, 2013), though this finding has not been observed in all studies (Paul, 2000). It is important to note that samples of late talkers vary in terms of whether or not children with comorbid receptive-language delays were included, so the range of receptive-language abilities was restricted to only average or above average in some studies (Fischel et al., 1989; Rescorla & Schwartz, 1990).

A smaller collection of studies have reported on other aspects of toddlerhood speech and language development as possible predictors of language outcomes, including

indications of phrase speech such as mean length of utterance. The relevance of toddlerhood phrase speech is suspected due to the fact that older children with language disorders tend to show their most severe impairment in the area of grammar. This might suggest that the emergence of early word combinations in toddlerhood is more relevant to early identification of children for long-term difficulties in language than milestones of single-word vocabulary size (Reilly et al., 2010). Some studies examining phrase speech have revealed significant effects, but others have not (Dale et al., 2003; M. J. Moyle, Weismer, Evans, & Lindstrom, 2007).

Background-information variables that are frequently investigated in studies of predictors of outcomes among late talkers include SES, gender, and family history of language or learning disorders or delays. The extant literature on children with typical development indicates that children of lower SES experience slower vocabulary growth from infancy through preschool age (Hart & Risley, 2003). Late talkers of low SES may be more vulnerable to long-term difficulties in language because they are less likely to experience an acceleration in vocabulary growth. Possible mechanisms for the negative impact of low SES on language development include differences in parent–child interaction, such as less child-directed speech, and greater material deprivation and child and parent health problems (Roy & Chiat, 2013). Results of studies of late talkers have been equivocal about whether SES predicts language outcome, with some studies identifying significant effects and others not (Dale et al., 2003; Fischel et al., 1989; Peyre et al., 2014).

With regard to the question of whether there is a gender difference in risk for long-term difficulties among late talkers, studies have found no differences (Armstrong, 2007; Dale et al., 2003; Peyre et al., 2014). The interpretation of gender differences among late talkers is complicated by the fact that many studies use gender-specific norms to determine whether a child is a late talker. Due to this, boys must show smaller expressive-vocabulary sizes than girls in order to be classified as late talkers. However, differences in the relationship between toddlerhood expressive-vocabulary size and language outcomes across genders are not well understood, so this could distort the relationship between gender and language outcomes among late talkers.

A few studies have also compared outcomes among late talkers who do and do not have a family history of language or learning delays or disorders. The majority of these studies have revealed no significant effects (Fischel et al., 1989; Petinou & Spanoudis, 2014; Peyre et al., 2014), with the exception of Lyytinen et al. (2005). However, the studies also vary in terms of which family member was affected, which symptoms were inquired about, and whether the symptoms were persistent or transient. These differences may relate to several genetic substrates with distinct implications for language development.

In summary, there is support for the proposition that expressive-vocabulary size and receptive language explain a small amount of variance in language outcomes among late talkers. The majority of studies have revealed nonsignificant effects for other predictors. However, with regard to some of these predictors, results have been highly inconsistent across studies, leading to further questions about differences in methodology. For example, significant effect sizes in opposite directions have been observed in studies of SES as a predictor (Dale et al., 2003; Levickis, Reilly, Girolametto, Ukoumunne, & Wake, 2014; Peyre et al., 2014). In addition, some predictors were examined predominantly in studies with relatively small sample sizes, suggesting the possibility that insufficient power may have obscured true effects. Of course, it is also possible that there are simply no effects.

Systematic Review and Meta-Analysis

Authors have offered explanations for the divergent and unexpected findings, including differences in sample characteristics, measurement of predictors or outcomes, and statistical power (Desmarais, Sylvestre, Meyer, Bairati, & Rouleau, 2008; Kelly, 1998; Thal & Katich, 1996; Whitehurst, Fischel, Arnold, & Lonigan, 1992). Desmarais et al. (2008), in particular, provided a detailed systematic review of the characteristics of late talkers in toddlerhood, which highlighted methodological heterogeneity among studies of late talkers. In addition, Dollaghan (2013) applied meta-analytic methods to evaluate the degree of continuity between late talking and language impairment. However, no study to date has quantitatively combined the effect sizes of predictors of language outcomes among late talkers using systematic review and meta-analysis methods.

The use of systematic review methods to identify predictors has several advantages. First, the research on late talkers has been cross-disciplinary, and articles have been published in a variety of journals related to speech-language pathology, developmental psychology, and pediatrics. A well-designed systematic review can assist in organizing information that might otherwise be overlooked by researchers in each respective area. Second, the application of consistent systematic review methods to identify studies on several different predictors facilitates comparisons of effect sizes across predictors. Third, efforts to gather gray literature, including unpublished statistics and dissertations, reduce the impact of publication bias on the findings of systematic reviews, creating a more balanced and accurate representation of the extant findings.

Meta-analytic methods also have several advantages. First, they can overcome problems of small sample sizes by increasing the statistical power used to address a research question. In the case of the literature on predictors of expressive-language outcomes for late talkers, meta-analysis has the potential to both clarify whether or not specific variables have any predictive power and maximize the precision of the estimate of variance explained by a predictor. Second, they allow for the quantitative exploration of sources of heterogeneity. Thus, meta-analysis enables hypothesis testing regarding the impact of methodological differences highlighted by study authors as possible sources of heterogeneity.

Aim and Hypotheses of the Present Study

The aim of this project is to apply systematic review and meta-analytic methods to synthesize the studies of predictors of expressive-language outcomes among late talkers. We seek to answer the question: What factors predict preschool-age expressive-language outcomes among late talkers? We hypothesized that toddlerhood expressive-vocabulary size and receptive language would be positively correlated with expressive-language outcomes. We further hypothesized that phrase speech, SES, gender, and family history would not be associated with expressive-language outcomes. (See Table 2 for the hypotheses.)

In order to address the research question, we followed these steps:

1. Identify and describe the relevant studies.
2. Assess the risk of bias in the studies.
3. Determine the main effects.
4. Conduct heterogeneity analyses to explain interstudy variability.
5. Evaluate the robustness of the findings using publication-bias and sensitivity analyses.
6. Identify strengths and weaknesses of the literature.
7. Provide recommendations for future directions.

Method

Systematic Review

Search Protocol

The methods adhere to PRISMA guidelines for systematic reviews and meta-analyses (Liberati et al., 2009). In July 2015, we conducted a literature search using the following electronic databases: Communication & Mass Media Complete, ERIC, Medline, PsycEXTRA, Psychological and Behavioral Sciences, and PsycINFO. We placed no limit on the year of publication. Our search terms were developed after preliminary review of the literature and consultation with a research librarian, and consisted of the following: ("language delay*" OR "late language emergence" OR "late talk*" OR "slow expressive language development") AND (outcome* OR longitudinal OR follow-up). In addition to the electronic-database search, we reviewed citations of studies that met our inclusion criteria, as well as reviews and clinical-topics pieces about late talkers. Last, we contacted authors of relevant records for more information, including unpublished data, additional statistics, and clarification regarding sample overlap across multiple records. In order to minimize the effect of confirmation bias on our analyses, we requested data or effect sizes for all predictors that were likely available to the authors, even if the predictors were unrelated to the research questions of the study.

Inclusion and Exclusion Criteria

We included studies that met the following criteria: (a) Participants were late-talking toddlers. Late-talking status was determined by parent report of expressive vocabulary below a specified cutoff when the child was between ages 18 and 35 months. (b) The study was prospective in design. A follow-up assessment of expressive language was administered at least 5 months after determination of late-talking status and before the age of 5 years. (c) Data were collected for at least one predictor. Statistics included in the article or supplied via e-mail correspondence enabled the calculation of an effect size for the relationship between the predictor and expressive-language outcome. (d) The article was available in English.

We excluded studies for the following reasons: (a) Participants were children with medical conditions or disabilities. (b) Participants were children affected by specific circumstances (e.g., international adoption). (c) The study included fewer than five participants. (d) A primary aim of the study was to evaluate the effectiveness of an intervention. (e) The predictor was unsuitable for meta-analysis due to too few studies (less than five) including that predictor. (f) The sample and predictor overlapped with another record.

Data Extraction

We created and used a data-extraction form to organize bibliographic, participant, and methodological variables from each study, in addition to effect sizes. Participant

Table 2. Operational definitions and hypotheses for predictors of outcome among late talkers.

Predictor	Definition	Hypothesis
Expressive-vocabulary size	Estimate of the number of different intelligible words produced by the child	Toddlerhood vocabulary size will be positively correlated with expressive-language outcome
Receptive language	Degree of comprehension of language produced by adults	Toddlerhood receptive language will be positively correlated with expressive-language outcome
Phrase speech	Evidence of word combinations, most often measured by mean length of utterances	Toddlerhood phrase speech will not be associated with expressive-language outcome
Socioeconomic status	Economic and social position of the parent(s) relative to others, most often measured by maternal education	Family socioeconomic status will not be associated with expressive-language outcome
Family history	Biological relative(s) having a current or past delay or disorder of speech, language, or learning	Expressive-language outcome will not differ between children with and without a family history
Gender	Parent report of whether the child is male or female	Expressive-language outcome will not differ between boys and girls

variables included mean participant age at intake assessment, expressive-vocabulary size, gender, SES, primary language, and proportion of participants who received intervention. Methodological variables included sample size, attrition, participant inclusion and exclusion criteria, measure and cutoff for determination of late-talking status, duration between intake and follow-up assessments, and measure for each predictor and outcome variable. Language assessments administered at follow-up were coded by type—lexical, grammatical, or combined—according to the language components they emphasized.

Reliability

We examined reliability of both study inclusion and classification of measures. Following the implementation of the search protocol, two coders independently reviewed all records in order to determine eligibility. Following data extraction, two coders independently classified measures of predictors according to which predictor they measured. They independently classified outcome measures according whether they were lexical, grammatical, or combined. With regard to both reliability analyses, the coders discussed instances in which their conclusions differed, and then reached a consensus.

Risk of Bias

We developed a list of study risk-of-bias variables after consulting relevant resources and considering the impact of certain methodological issues on the quality of the information derived from a study with regard to the research question of this systematic review and meta-analysis (The EQUATOR Network, n.d.; Thompson, Diamond, McWilliam, Snyder, & Snyder, 2005). The following study characteristics were desirable: (a) detailed specification of inclusion and exclusion criteria; (b) use of standardized measures with acceptable psychometric properties; (c) use of a direct assessment measure at the follow-up assessment; (d) no conversion of continuous data into a nominal format; (e) no evidence of partial reporting of relationships between predictors and outcomes; and (f) low (< 20%) attrition.

Note that we did not exclude studies on the basis of risk of bias. Rather, we took an inclusive approach to study selection. This allowed us to maximize the comprehensiveness of our systematic review and meta-analysis and still collect and present data related to risk of bias.

Data Processing

Details regarding data processing for each individual study are contained in the footnotes of Table 3. We calculated effect sizes using original data files whenever possible. Additional data-processing steps that were occasionally necessary included pooling subgroups that were not relevant to a particular analysis, eliminating individual cases that did not meet inclusion criteria, and calculating raw score improvement over time as an outcome measure. In addition, if it was possible to calculate effect sizes for multiple outcome time points, the time point closest to age 36 months was selected in order to maximize consistency across studies.

Meta-Analyses

We conducted a separate, random-effects model meta-analysis for each individual predictor using Comprehensive Meta-Analysis software (V. 2; Biostat, Englewood, NJ). We acknowledged significance at $p < .05$. We followed Cohen's (1988) conventions for the interpretation of effects sizes as small (Cohen's $d = 0.2$, Pearson's $r = .1$), medium (Cohen's $d = 0.5$, Pearson's $r = .3$), or large (Cohen's $d = 0.8$, Pearson's $r = .5$).

Main Effects

We reported the main effects of each meta-analysis using one of two statistics: For relationships between continuous predictors and outcomes that were typically presented in the form of correlations, we used Pearson's *r*; for relationships between nominal predictors and outcomes that were typically presented in the form of group differences, we used Cohen's *d*.

Heterogeneity

We determined the statistical significance of heterogeneity using a chi-square test of the *Q* value and estimated the magnitude of heterogeneity using the I^2 value (Borenstein, Hedges, Higgins, & Rothstein, 2009). Higgins et al. (2003) provide guidelines for the interpretation of I^2 as low (< .25), moderate (.50), or high (.75).

Moderation Analyses

We tested potential moderators of the relationship between predictors and outcomes using metaregression (method-of-moments model) for continuous variables and subgroup analysis for nominal variables. Guidelines indicate that moderation analyses are appropriate when the following conditions are met: presence of moderate–high heterogeneity ($I^2 > .25$) and a minimum of eight studies for metaregression or three per subgroup for subgroup analysis (Borenstein et al., 2009).

In all meta-analyses meeting the conditions, we tested duration between intake and follow-up assessments and type of language assessment at follow-up. In addition, we tested several moderators that we hypothesized were relevant to one specific predictor. With regard to expressive-vocabulary size, we used metaregression to determine whether the main effect differed as a function of mean expressive-vocabulary size of the sample. With regard to SES, we used subgroup analysis to determine whether the main effect differed among studies with a wide range of SES, operationally defined as ≥ 10% of parents who did not complete high school or secondary school. With regard to gender, we used subgroup analysis to determine whether the main effect differed among studies that used gender-specific percentile cutoffs to determine late-talker status.

Publication-Bias and Sensitivity Analyses

We conducted publication-bias and sensitivity analyses to evaluate the validity and robustness of meta-analysis findings. We assessed publication bias in meta-analyses

Table 3. Participant and methodological characteristics of included records and unpublished reports.

Study	*n*	Location and primary language	Age at intake (months)	Time to follow-up (months)	% M/F	Socioeconomic status/parent education	Intake vocabulary	% received intervention	Outcome measure(s)
Armstrong (2007)[a] & Lee (2011)[a]	131 63	United States/ English United States/ English	24 24	12 12	51/49[b]	Mean maternal education = 14.18 years; > 10% of mothers did not complete high school[c]	—	—	RDLS Expressive subtests[d,e]
Bishop et al. (2012)[f]	22	United Kingdom/ English	20	28	59/41	Mean age at which mothers stopped education = 21.6 years	< 10	16[g]	(BAS Verbal subtests, Bus Story Test, CCC-2, TEGI)[d,e]
Carson, Klee, Carson, & Hime (2003)[f]	6	United States/ English	24.9	12	83/17	Mean maternal education was between high school diploma and associate's degree	61.17	0	(MLU, MSEL Language subtests)[d,e]
Dale et al. (2003)	740	United Kingdom/ English	24	12	65/35	Mean parent graduated secondary school and qualified for college; approximately 13% did not complete secondary school[c]	< 15	—	CDI Vocabulary, Grammar, and Abstract Language[d,e]
Feldman et al. (2005)[f]	29	United States/ English	24	12	76/24[b]	24% of mothers did not complete high school; 69% high school graduates; 7% college graduates[c]	—	—	MLU,[h] MSCA Verbal subtests,[e] NDW[i]
Fernald & Marchman (2012)[f]	36	United States/ English	18	12	56/44	Mean maternal education = 16.67 years	20	—	CDI Vocabulary[d,i]
Fischel et al. (1989) & Whitehurst, Smith, Fischel, Arnold, & Lonigan (1991)	26 37	United States/ English United States/ English	27.08 27.62	5 5	85/15 92/8	Mean maternal education = 13.96 years "Middle class families"	17.54 18.35	35	(EOWPVT, ITPA Verbal subtests)[e]
Hadley & Short (2005)[f,j] & Hadley & Holt (2006)[f,j]	15	United States/ English	26.07	10	67/33[b]	Mean maternal education = 14.88 years	73.27	40	(CDI Vocabulary,[i] IPSyn,[h] MLU,[h] NDW[i])
Henrichs et al. (2011)	328	Netherlands/ Dutch	18	12	47/53	16% of mothers had < 3 years of secondary school; 27% > 3 years of secondary school or some vocational training; 21% higher vocational training or a bachelor's degree; 36% master's or doctoral degree[c]	—	—	LDS Vocabulary[d,e]
Moyle et al. (2007)[k]	30	United States/ English	24	6	67/33	Mean maternal education = 15.75 years	46.62	37	(CDI Vocabulary,[i] NDW,[i] and PLS Semantic items[i]) and (CDI Grammar,[h] MLU,[h] PLS-3 Syntax items[h])
Levickis et al. (2014)[f,l]	226	Australia/English	18	12	52/48	18% of mothers did not complete high school; 37% high school graduates; 49% college graduates[c]	—	52[m]	PLS-4 Expressive subtests[e]
Lyytinen et al. (2005)[n]	32	Finland/Finnish	24	18	66/34	Mean parent education was vocational school	—	—	(BNT, Inflectional Morphology Test)[e]

(*table continues*)

Table 3. *(Continued).*

Study	*n*	Location and primary language	Age at intake (months)	Time to follow-up (months)	% M/F	Socioeconomic status/parent education	Intake vocabulary	% received intervention	Outcome measure(s)
Paul, Looney, & Dahm (1991)[o] & Paul (1993)[o]	21 37	United States/ English United States/ English	25.6 26.7	12.2 12	71/29 83/17	Participants described as "a middle-class sample"	26.2 22	30	VABS Expressive subdomain,[d,e] DSS[h]
Petinou & Spanoudis (2014)[f]	23	Greece/Greek (Cypriot dialect)	28	8	57/43	—	< 70	—	PLS-3 Expressive subtests[e]
Peyre et al. (2014)	100	France/French	24.26	13.79	64/36[b]	Mean parental education = 13.93 years	12.04	—	(ELOLA, NEPSY subtests)[d,e]
Reilly et al. (2010)	277	Australia/English	24	24	51/49[b,p]	21% of mothers had < 12 years education; 41% ≥ 13 years education; 38% college degree[c,p]	Boys: 39 Girls: 65	—	CELF-P2 Expressive subtests[d,e]
Rescorla & Schwartz (1990)[f]	25	United States/ English	26.32	12.44	100/0	Participants were from "primarily middle- and upper-middle-class families"	20.62	33	IPSyn,[h] MLU,[h] RDLS-R Expressive[e]
Thal, Tobias, & Morrison (1991)	10	United States/ English	22.4	12.1	80/20	All participants were "from middle-class backgrounds"	16.8	—	ELI,[i] MLU[h]
Vuksanovic (2015)	25	Serbia/Serbian	26	5	64/36	All parents had at least completed secondary school	40.39	—	RDLS-R Expressive subtests[e]
Williams & Elbert (2003)[q]	5	United States/ English	27.2	10.6	60/40	—	< 50	0	MLU,[h] NDW[i]

Note. Two citations connected by "&" indicate that the samples overlap, but both records were used to calculate effect sizes. Vocabulary at intake is reported in means or in maximum vocabulary (indicated by the "<" symbol preceding the value). Parenthetical enclosure of more than one outcome measure indicates that the authors combined the measures into a single score. The em dash indicates that information was not available.

M = male; F = female; RDLS = Reynell Developmental Language Scales; RDLS-R = Reynell Developmental Language Scales–Revised; BAS = British Ability Scales; CCC-2 = Children's Communication Checklist–Second Edition; TEGI = Test of Early Grammatical Impairment; MLU = mean length of utterance in a language sample; MSEL = Mullen Scales of Early Learning; CDI = MacArthur Communicative Development Inventories; MSCA = McCarthy Scales of Children's Abilities; NDW = number of different words in a language sample; EOWPVT = Expressive One-Word Picture Vocabulary Test; ITPA = Illinois Test of Psycholinguistic Abilities; IPSyn = index of productive syntax; LDS = Language Development Survey; PLS = Preschool Language Scale; PLS-3 = Preschool Language Scale–Third Edition; PLS-4 = Preschool Language Scale–Fourth Edition; BNT = Boston Naming Test; VABS = Vineland Adaptive Behavior Scales; DSS = Developmental Sentence Scoring; ELOLA = Évaluation du langage oral de l'enfant aphasique; NEPSY = Developmental Neuropsychological Assessment; CELF-P2 = Clinical Evaluation of Language Fundamentals–Preschool–Second Edition; ELI = Early Language Inventory.

[a]Pooled "later late bloomers" and "expressive language delayed" subgroups in order to compare with "late bloomers" groups on outcomes at age 3 years. [b]Study used gender-specific norms and was included in that moderation subgroup in meta-analysis for gender. [c]Study included a substantial portion of parents of low socioeconomic status or education and was included in that moderation subgroup in meta-analysis for socioeconomic status. [d]Outcome was reported in a nominal format, using a specified cutoff score to indicate impaired or unimpaired language. [e]Outcome was coded as combined. [f]Statistics used were from files provided by authors via e-mail correspondence. [g]Percentage of participants either who were receiving intervention or whose parents were concerned about language development. [h]Outcome was coded as grammatical. [i]Outcome was coded as lexical. [j]One case was removed due to normal-range CDI at intake. [k]Available outcomes differed across predictors: Cross-lagged panel analysis included in the article allowed for calculation of relationship between lexical development at age 2 years and grammatical development at age 3 years, and vice versa. [l]Late-talker status was determined at age 18 months, but evaluation of predictors was conducted at 24 months. [m]Percentage of participants who were randomly assigned to intervention condition, which was found to have no effect. [n]Pooled late talkers with dyslexia risk and expressive delay only ("LT1DRa") and late talkers with dyslexia risk and both expressive and receptive delays ("LT2b") subgroups in order to compare with late talkers without a family history of dyslexia ("LTNRa"); pooled "LT1DRa" and "LTNRa" subgroups in order to compare with late talkers with comorbid receptive language delay ("LT2b"). [o]Available outcomes differed across predictors: Relationship was calculated between receptive language at age 2 years and VABS Expressive at age 3 years, as was the relationship between gender and MLU at age 3 years. [p]Data are based on statistics reported for the entire sample, including late talkers and typically developing children. [q]Improvement over time was calculated and used as outcome due to variability in follow-up ages.

with eight or more studies by examining funnel plots for asymmetry and applying Duval and Tweedie's trim-and-fill correction (Borenstein et al., 2009). We assessed sensitivity by exploring the effects of removing each individual study on main effects.

Vote Counting

In order to examine the contribution of the meta-analyses to the literature, we also conducted vote-counting analyses (Higgins & Green, 2008). This approach involves adding up and comparing the number of studies that did and did not find significant results with regard to a research question.

Results

Systematic-Review Results

See Supplemental Material S5 for the PRISMA flowchart detailing search results and numbers of records that we excluded for various reasons. We used multiple sources of data on a single sample in instances in which effect sizes for different predictors of expressive-language outcomes were located in different places. The final sources of data were 23 peer-reviewed articles, one unpublished dissertation, and 11 files received via e-mail correspondence with their authors. In total, we included data from 20 distinct samples of late talkers. For clarity, we will hereafter discuss the data in terms of the 20 samples.

Sample Characteristics

Participant and methodological characteristics are in Table 3. Inclusion and exclusion criteria are in Supplemental Material S1. The studies took place in eight countries and included speakers of six primary languages. There were a total of 2,134 participants in the 20 samples (59% boys, 41% girls). Age at intake ranged from 18 to 28 months, and duration between intake and follow-up assessments ranged from 5 to 28 months. Mean maternal education ranged from 13 to 17 years, with six samples including a wide range of SES. The proportion of children who received intervention varied from 0% to 52%.

Authors most often used parent-completed inventories to determine late-talker status. Five samples used the gender-specific percentile norms of the MacArthur–Bates Communicative Development Inventories, Second Edition (Fenson et al., 2007), which resulted in a lower raw-score cutoff for boys than girls. Mean expressive-vocabulary size ranged from fewer than 10 words to 73 words. Two studies excluded children with below-average receptive language. Supplemental Material S2 contains the methods used to measure the predictors.

Reliability

Interrater reliability for study inclusion and exclusion was 97% (1,764 out of 1,818 records). Reliability for predictor-measure classification was 95% (21 out of 22 measures). Reliability for outcome-measure classification was 90% (18 out of 20 measures). Disagreements were resolved via discussion.

Risk of Bias

A checklist of desirable study characteristics is given in Supplemental Material S3. Nineteen studies specified inclusion and exclusion criteria. Seventeen used standardized measures at intake assessment, and 14 included a direct assessment measure at follow-up. Nine did not convert any continuous data into a nominal format. Twelve reported relationships between all predictors and outcomes or were able to provide additional information via e-mail. Attrition was low in 10 samples, high in five, and unknown in five.

Vote-Counting and Meta-Analysis Results: Assessment Variables

Expressive-Vocabulary Size

Table 4 contains a summary of meta-analysis outcomes for all predictors. Supplemental Material S4 contains the results of vote-counting analyses. Twelve samples, involving 1,113 participants, contributed to the analysis of the relationship between expressive-vocabulary size and expressive-language outcome. Measures of expressive-vocabulary size included parent-completed inventories and informal parent report of vocabulary. According to the vote-counting analysis, a significant effect of expressive-vocabulary size on expressive-language outcome was observed in four out of the 12 samples. According to the meta-analysis, there was a small, statistically significant main effect of the correlation between expressive-vocabulary size and expressive-language outcome ($r = .249$, $p < .01$; see Supplemental Material S6). Expressive-vocabulary size accounted for 6% of the variability in expressive-language outcome ($R^2 = .062$). Heterogeneity was moderate ($I^2 = 46.285$, $Q = 20.478$, $p < .05$). However, none of the three hypothesized moderators—outcome type, duration between intake and follow-up, or mean expressive-vocabulary size—explained a significant amount of heterogeneity. Examination of a funnel plot suggested asymmetry (see Supplemental Material S7), and so we applied a Duval–Tweedie trim-and-fill correction. This correction reduced the main effect but did not change the finding ($r = .138$, $p < .01$). Sensitivity analysis indicated that the finding is robust to the removal of any individual sample.

Receptive Language

Ten samples, involving 527 participants, contributed to the analysis of the relationship between toddlerhood receptive language and expressive-language outcome. Measures of receptive language included standardized direct assessments, standardized parent interviews and questionnaires, and informal parent report. According to the vote-counting analysis, a significant effect of receptive language on expressive-language outcome was observed in four out of the 10 samples. According to the meta-analysis, there was a medium, statistically significant main effect of the correlation between receptive language and expressive-language

Table 4. Summary of meta-analysis results for all predictors.

Predictor	Participants	Studies	Main effect		Heterogeneity		
			Pearson's *r*/Cohen's *d*[a]	95% confidence interval	*Q*	*p*	I^2
Continuous							
Expressive-vocabulary size	1,113	12	.249**	[.133, .358]	20.478	.039	46.285
Receptive language	527	10	.340**	[.215, .454]	12.717	.176	29.228
Phrase speech	851	7	.122	[–.022, .261]	7.434	.283	19.295
Socioeconomic status	1,955	12	.111*	[.040, .181]	16.952	.109	35.109
Nominal[b]							
Gender	1,696	11	0.068	[–.109, .245]	15.019	.131	33.419
Family history	447	6	0.341	[–.034, .716]	8.943	.111	44.093

[a]Pearson's *r* for continuous predictors, Cohen's *d* for nominal predictors. [b]Analyses of nominal predictors were designed such that positive values indicate findings that support the hypotheses. For gender, positive *d* indicates better performance among girls. For family history, positive *d* indicates better performance among children with no family history of language disorders/learning disabilities.

*$p < .05$. **$p < .01$.

outcome ($r = .340$, $p < .01$; see Supplemental Material S8). Receptive language accounted for 12% of the variability in expressive-language outcome ($R^2 = .116$). Heterogeneity was moderate ($I^2 = 29.228$, $Q = 12.717$, $p = .176$). However, none of the three hypothesized moderators—outcome type, duration between intake and follow-up, or inclusion of children with below-average receptive language—explained a significant amount of heterogeneity. A funnel plot was approximately symmetric (see Supplemental Material S9). Sensitivity analysis indicated that the finding is robust to the removal of any individual sample.

Phrase Speech

Seven samples, involving 851 participants, contributed to the analysis of the relationship between toddlerhood phrase speech and expressive-language outcome. Measures included utterance length during observation, standardized parent questionnaires, and informal parent report. According to the vote-counting analysis, a significant effect of phrase speech on expressive-language outcome was observed in one out of the seven samples. According to the meta-analysis, the main effect of the correlation between phrase speech and expressive-language outcome was nonsignificant ($r = .122$, $p = .098$; see Supplemental Material S10). Phrase speech accounted for 2% of the variability in expressive-language outcome ($R^2 = .015$). Heterogeneity was low ($I^2 = 19.295$, $Q = 7.434$, $p = .283$). Sensitivity analysis indicated that the nonsignificant finding is robust to the removal of any individual sample.

Vote-Counting and Meta-Analysis Results: Background Variables

SES

Twelve samples, involving 1,955 participants, contributed to the analysis of the relationship between SES, or parental education (10 maternal, one either parent), and expressive-language outcome. According to the vote-counting analysis, a significant effect of SES on expressive-language outcome was observed in four out of the 12 samples. According to the meta-analysis, there was a small, statistically significant main effect of the correlation between SES and expressive-language outcome ($r = .111$, $p < .01$; see Supplemental Material S11). SES accounted for 1% of the variability in expressive-language outcome ($R^2 = .012$). Heterogeneity was moderate ($I^2 = 35.109$, $Q = 16.952$, $p = .109$). Examination of a funnel plot suggested minor asymmetry (see Supplemental Material S12), so we applied a Duval–Tweedie trim-and-fill correction. This correction reduced the main-effect estimate slightly, but the estimate remained statistically significant ($r = .109$, $p < .01$). Sensitivity analysis indicated that the significant finding is robust to the removal of any individual sample.

Neither outcome type nor duration between intake and follow-up explained a significant amount of heterogeneity. However, subgroup analysis indicated that the main effect of SES was significant only for samples that included a wide range of parental education ($r = .145$, $p < .01$) rather than a limited range ($r = -.022$, $p > .05$).

Gender

Eleven samples, involving 1,696 participants, contributed to the analysis of the relationship between gender and expressive-language outcome. According to the vote-counting analysis, a significant effect of gender on expressive-language outcome was observed in none of the samples. According to the meta-analysis, the main effect of the standardized mean difference in expressive-language outcome between boys and girls was nonsignificant ($d = 0.068$, $p = .451$; see Supplemental Material S13). Heterogeneity was moderate ($I^2 = 33.419$, $Q = 15.019$, $p = .131$). However, none of the three hypothesized moderators—outcome type, duration between intake and follow-up, or use of gender-specific norms—explained a significant amount of heterogeneity. Examination of a funnel plot suggested minor asymmetry (see Supplemental Material S14), so we applied a Duval–Tweedie trim-and-fill correction. This correction reduced the main-effect estimate slightly ($d = 0.056$, $p > .05$). Sensitivity analysis indicated that the nonsignificant main-effect

finding is dependent on the inclusion of the study by Reilly et al. (2010).

Family History

Six samples, involving 447 participants, contributed to the analysis of the relationship between family history and expressive-language outcome. Classification of children as having a family history differed across samples in terms of both which relatives were inquired about (sibling, parent, or either) and which conditions were specified (language disorder, speech delay, or learning disability). According to the vote-counting analysis, a significant effect of family history on expressive-language outcome was observed in one out of the six samples. According to the meta-analysis, the main effect of the standardized mean difference in expressive-language outcome between children with and without a family history was nonsignificant ($d = 0.341$, $p = .075$; see Supplemental Material S15). Heterogeneity was moderate ($I^2 = 44.093$, $Q = 8.943$, $p = .111$). Sensitivity analysis indicated that the nonsignificant main-effect finding is dependent on the inclusion of the study by Petinou and Spanoudis (2014).

Discussion

Findings of the Meta-Analyses in Context

Findings of the nine meta-analyses can be categorized as medium significant effects, small significant effects, and nonsignificant effects. Toddlerhood receptive language showed a medium significant effect on expressive-language outcomes. Expressive-vocabulary size and SES showed small significant effects. Phrase speech, gender, and family history showed nonsignificant effects.

Of the three child language-development variables, receptive language explained the greatest amount of variance in expressive-language outcome. This is interesting, given that the other variables—expressive-vocabulary size and phrase speech—were expressive. Therefore, our results suggest that how much a toddler understands may be a better predictor of expressive-language outcome than how much he or she says. The finding that children with better comprehension have a better prognosis for language development is consistent with Bishop and Edmundson's (1987) hierarchical model of vulnerability in language components, in which comprehension is foundational and problems in comprehension characterize the most severe language disorders.

The small significant effect of expressive-vocabulary size on language outcomes is consistent with the literature on late talkers in general, which indicates a modest but statistically detectable degree of continuity between toddlerhood expressive language and preschool expressive language. The nonsignificant correlation between phrase speech and expressive-language outcomes suggests that early word combinations may be qualitatively different from the types of grammatical skills that are difficult for older children with language or learning disorders, and that we should not expect difficulties to emerge until more advanced stages, such as the use of function words and bound morphemes.

Our finding that SES was related to expressive-language outcomes for studies that included a wide range of parental education is consistent with research demonstrating a nonlinear relationship between SES and language development, in which only children from low-SES families are negatively affected (Washbrook & Waldfogel, 2010). The finding is also consistent with studies showing that the negative impact of low SES on language development increases over time during childhood (Hart & Risley, 2003). Thus, late talkers of low SES may represent a group that is uniquely vulnerable and in need of specialized intervention that focuses on supporting the needs of the family.

With regard to gender, our finding that expressive-language outcomes did not differ between boys and girls is consistent with the finding of no difference in the prevalence of SLI among kindergarten boys and girls in the largest epidemiological study of SLI to date (Tomblin et al., 1997). Our finding of no relationship between family history and expressive-language outcomes should be interpreted with caution, because we conceptualized family history broadly, and thus the exact predictor varied across studies in terms of which family member was affected, which kinds of disorders or delays were inquired about, and whether the symptoms were persistent or transient. Further studies are needed to determine the relevance of each of these issues to expressive-language outcomes.

Overall Contribution to the Literature

This systematic review and meta-analysis enhances the literature by clarifying relationships between toddlerhood predictors and preschool-age expressive-language outcomes among toddlers. The results of the vote-counting approach indicate that many of the predictors identified as significant in the meta-analysis show ambiguous patterns when one simply compares the number of individual studies that did and did not find significant results. Even among child language variables, in no case did a majority of individual studies report significant findings. This highlights the utility of quantitatively combining study findings using a method that is sensitive to sample size and within-study variance. Meta-analysis is a useful tool for building future research questions and moving a line of research forward.

Strengths, Limitations, and Recommendations

The Late-Talker Literature

Across the studies, the most common risk-of-bias issue was the conversion of continuous data to a nominal format. Authors frequently justify this choice by elevating the importance of the distinction between disordered and nondisordered expressive language in how they frame their research question. This certainly does have some advantages, especially in a clinical context requiring related binary

decisions (i.e., intervention or no intervention). However, exclusively reporting binary outcomes is disadvantageous when it comes to enhancing our understanding of language development for several reasons. First, the approach involves discarding information that could be used in statistical analyses by artificially minimizing variability in outcomes (Kerlinger & Lee, 1999). Second, it distorts our understanding of score distributions and relationships among variables (Thompson et al., 2005). Last, because there is no gold standard for the diagnosis of SLI, binary outcomes complicate efforts to synthesize results across studies.

Additional risk-of-bias issues included use of unstandardized measures, partial reporting, and high or unreported attrition. On average, older studies showed more methodological and reporting weaknesses than recent ones, suggesting that standards may have evolved over time. The variability in use of standardized measures reflects both authors' resources and perceptions of the relative value of different types of measures—parent questionnaires, language-sample analysis, and standardized measures. In general, omnibus standardized direct assessment measures of language development are preferable because they are less susceptible to reporting bias and have well-established psychometric properties. Additional methods may provide supplementary information. Partial reporting was undesirable for the purpose of the meta-analyses but typically appeared to be indicative of differences in the research questions of each study rather than a decision to withhold results. Attrition rates were higher for large population-based studies, which is expected given that more heterogeneous samples would have included greater numbers of families with significant barriers to follow-up. Last, we noted that few studies reported that the staff members administering assessments or transcribing language samples were unaware of the presentation of the children in toddlerhood. Taking this step could reduce confirmation bias in longitudinal studies.

Our survey of the literature on predictors of expressive-language outcomes in late talkers also revealed several areas in which research is sparse. For example, few authors have examined the relationship between the age at which the child qualified as a later talker and outcome. This is surprising, given that the mean child age at intake in the studies ranged from 18 to 28 months, and research indicates that the stability of low language skills increases as children get older (Thal & Katich, 1996). Therefore, it seems reasonable to expect that when a child is showing a delay in expressive vocabulary is an important question in predicting outcomes. Little information was also available on aspects of parenting behavior and parent well-being. This information would add to our understanding of the roles of environmental factors, especially as possible mediators of the relationship between SES and expressive-language outcomes.

A major strength of the literature was the addition of several large, population-based studies of language development within the last 10–15 years. The longitudinal research on late talkers now consists of two types of studies: those on large, population-based samples and those on small, carefully characterized samples (Rescorla, 2011). Researchers have opposing views regarding which of these study types is superior (Desmarais et al., 2008; Reilly et al., 2008). However, it is our view that each type is suitable for addressing distinct sets of questions, and the combination of the two types has far more potential to move the field forward than either type alone. For example, population-based studies have greater statistical power, allow the exploration of lower base-rate child and family characteristics, and suffer less from the problem of selection bias and therefore have more heterogeneous samples. Small samples have the advantages of controlling for unrelated conditions and allowing detailed data collection using methods that are time-consuming and costly. An additional strength of the literature is the gradual addition of studies of toddlers learning languages other than English. Petinou and Spanoudis (2014) note that cross-linguistic examinations of late talkers may shed light on the development of the language disorders by revealing patterns of difficulty in specific linguistic features that are related to language outcomes.

The Present Review

There are several limitations of the present review. First, we excluded both studies that were not available in English and studies for which an effect size could not be calculated. Second, our heterogeneity analyses often failed to explain the significant interstudy variability. This could indicate either that we did not identify the relevant methodological factors or that complex interactions among methodological variables obscured our analyses. Third, the practice of collapsing information across multiple studies in meta-analyses necessarily complicates the interpretation of findings by minimizing the importance of the methodological and theoretical diversity of the studies. For example, the criteria for classifying a child as a late talker varied widely, from approximately the 2nd to the 20th percentile on standardized measures of expressive vocabulary in toddlerhood. This could have affected our analyses due to floor effects in more severely delayed samples.

The inclusiveness of the review may also be viewed as a limitation. A few of the measures of predictors lacked good psychometric properties, which may limit the validity of the data derived from them. Even among well-established measures, diversity in the content may have influenced analyses. For example, expressive-language outcome measures included parent-completed vocabulary inventories, standardized direct assessment measures of vocabulary and grammar, and statistics from language-sample analyses. Rescorla (2011) notes that vocabulary outcomes tend to be better than grammar outcomes for late talkers, and Nippold and Schwarz (1996) suggest that persistent difficulties among late talkers may only be observable in skills that are emerging at the age at which outcomes were assessed. Also, the minimum number of participants for study inclusion was low. In the context of the random-effects model, this may lead to small studies being too heavily represented in the results.

Implications and Future Directions

In clinical contexts, this review highlights the importance of accurately measuring receptive language when assessing need for intervention in late talkers. Additional considerations include the child's expressive-vocabulary size and SES. Parent resources and reactions to intervention are also especially important in making decisions respecting evaluations of late talkers, due to the fact that late talkers are at relatively low risk for language or learning disorders. Thus, the clinician must be mindful of the possibility that the burden of intervention could outweigh the benefit (Paul, 2000).

Although we identified several statistically significant predictors, it is important to put our findings in perspective. The most powerful predictor, receptive language, explained only 12% of the variability in expressive-language outcome; the lower limit of the confidence interval suggests that it may even explain as little as 5% of the variability. This is consistent with other studies of the stability of low language skills from toddlerhood to preschool, which also provide evidence of low sensitivity and specificity in predicting outcomes among late talkers (Dollaghan, 2013). Moreover, several researchers have pointed out that the number of late talkers that go on to have SLI underestimates the prevalence of SLI in the population (Leonard, 2013). Therefore, the majority of children with SLI are not former late talkers. Overall, the current literature indicates that there are multiple pathways leading to SLI. Being a late talker is a risk factor for language or learning disorders, but it is neither a clinical condition nor a certain sign of a disorder to come. Given the results of this review and the state of our ability to predict outcomes for individual late talkers, the clinical recommendations described by Paul (2000)—that late talkers with comorbid receptive impairment should receive intervention and those with no comorbid receptive impairment should receive only occasional monitoring of language growth—are reasonable.

Future directions that proceed naturally from the findings of this review include the use of multifactorial models to predict expressive-language outcomes. Models would ideally be developed for each increment of 2–3 months in child age, to allow for the fact that the influence of specific predictors over time changes and to increase our understanding of changes in the level of diagnostic accuracy that is feasible over time. These models ultimately may be used to develop screening measures for identifying young children at the highest risk for SLI on a public-health level (Rescorla & Dale, 2013).

The exploration of domain-general information-processing skills that underlie typical language development and may be core deficits leading to SLI is an important next step for the study of predictors of outcomes among late talkers. These skills include memory, processing speed, and statistical learning (J. Moyle, Stokes, & Klee, 2011). The relevance of these constructs to late talkers is supported by studies indicating that verbal working-memory deficits remain apparent in late talkers after they have caught up to their peers in vocabulary and grammar (Bavin & Bretherton, 2013; Thal et al., 2013). Research on information-processing skills could reveal a deeper level of cognitive mechanisms that explain the variability in subsequent language development.

The combination of multifactorial models and increased knowledge about information processing may lead to a shift in focus among researchers investigating toddlerhood risk for language or learning disorders. Studying subgroups of children with weaknesses other than expressive-vocabulary delay may actually be more productive in terms of enhancing our developmental account of language or learning disorders. The results of the present study, for example, suggest the possibility that toddlers with receptive delays constitute a subgroup of children at higher risk. Although late talking is a salient clinical characteristic known to be related to increased risk for language or learning disorders, this apparent relationship may be driven by the fact that late talking is secondary to delays in comprehension for some children.

In conclusion, this systematic review and meta-analysis provides the first quantitative summary of predictors of expressive-language outcomes among late talkers. We identified receptive language, expressive-vocabulary size, and SES as statistically significant predictors. Important future steps for this field include the use of multifactorial models and investigation of domain-general cognitive underpinnings of language development.

Acknowledgments

This review article was a general examination project submitted by Evelyn L. Fisher to the Department of Psychology at Georgia State University, Atlanta. I thank Nicole Lim for her participation in abstract review and measure coding and Gordon Warren for his input on statistical methods. I also thank MaryAnn Romski, Rose Sevcik, and Tricia King for their feedback on the project. Last, I thank the many investigators who provided supplementary information about their studies: Dorothy Bishop, Cecyle Carson, Susan Ellis Weismer, Heidi Feldman, Anne Fernald, Pamela Hadley, Marcia Kurs-Lasky, Joanne Lee, Penny Levickis, Virginia Marchman, Kakia Petinou, Hugo Peyre, Leslie Rescorla, Jasmina Vuksanovic, and the Early Language in Victoria Study team members, especially Fiona Mensah, Sheena Reilly, and Edith Bavin. Without their responsiveness and support, this project would not have been possible.

References

References marked with an asterisk indicate studies included in the meta-analyses.

Aram, D. M., & Hall, N. E. (1989). Longitudinal follow-up of children with preschool communication disorders: Treatment implications. *School Psychology Review, 18,* 487–501.

***Armstrong, E. S.** (2007). School-age cognitive and achievement outcomes for late talkers and late bloomers: Do late bloomers really bloom? Retrieved from ProQuest Information and Learning database. (Accession number 2007-99008-286)

Bavin, E. L., & Bretherton, L. (2013). The Early Language in Victoria Study: Late talkers, predictors, and outcomes. In

L. A. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 3–21). Baltimore, MD: Brookes.

Bishop, D. V. M., & Edmundson, A. (1987). Specific language impairment as a maturational lag: Evidence from longitudinal data on language and motor development. *Developmental Medicine & Child Neurology, 29,* 442–459.

***Bishop, D. V. M., Holt, G., Line, E., McDonald, D., McDonald, S., & Watt, H.** (2012). Parental phonological memory contributes to prediction of outcome of late talkers from 20 months to 4 years: A longitudinal study of precursors of specific language impairment. *Journal of Neurodevelopmental Disorders, 4,* 3. https://doi.org/10.1186/1866-1955-4-3

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, United Kingdom: Wiley.

***Carson, C. P., Klee, T., Carson, D. K., & Hime, L. K.** (2003). Phonological profiles of 2-year-olds with delayed language development: Predicting clinical outcomes at age 3. *American Journal of Speech-Language Pathology, 12,* 28–39.

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

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

Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *International Journal of Language & Communication Disorders, 43,* 361–389.

Dollaghan, C. (2013). Late talker as a clinical category: A critical evaluation. In L. A. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 91–112). Baltimore, MD: Brookes.

The EQUATOR Network. (n.d.). *Enhancing the quality and transparency of health research*. Retrieved from http://www.equator-network.org/

***Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., & Paradise, J. L.** (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development, 76,* 856–868. https://doi.org/10.1111/j.1467-8624.2005.00882.x

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., . . . Stiles, J. (1994). Variability in early communicative development [Monograph]. *Monographs of the Society for Research in Child Development, 59*(5, Serial No. 242), 1–185.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur–Bates Communicative Development Inventories, Second Edition*. Baltimore, MD: Brookes.

***Fernald, A., & Marchman, V. A.** (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development, 83,* 203–222. https://doi.org/10.1111/j.1467-8624.2011.01692.x

***Fischel, J. E., Whitehurst, G. J., Caulfield, M. B., & DeBaryshe, B.** (1989). Language growth in children with expressive language delay. *Pediatrics, 83,* 218–227.

***Hadley, P. A., & Holt, J. K.** (2006). Individual differences in the onset of tense marking: A growth-curve analysis. *Journal of Speech, Language, and Hearing Research, 49,* 984–1000. https://doi.org/10.1044/1092-4388(2006/071)

***Hadley, P. A., & Short, H.** (2005). The onset of tense marking in children at risk for specific language impairment. *Journal of Speech, Language, and Hearing Research, 48,* 1344–1362. https://doi.org/10.1044/1092-4388(2005/094)

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator, 27*(1), 4–9.

***Henrichs, J., Rescorla, L., Schenk, J. J., Schmidt, H. G., Jaddoe, V. W. V., Hofman, A., . . . Tiemeier, H.** (2011). Examining continuity of early expressive vocabulary development: The Generation R Study. *Journal of Speech, Language, and Hearing Research, 54,* 854–869. https://doi.org/10.1044/1092-4388(2010/09-0255)

Higgins, J. P. T., & Green, S. (Eds.). (2008). *Cochrane handbook for systematic reviews of interventions*. Chichester, United Kingdom: Wiley.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ: British Medical Journal, 327*(7414), 557–560.

Kelly, D. J. (1998). A clinical synthesis of the "late talker" literature: Implications for service delivery. *Language, Speech, and Hearing Services in Schools, 29,* 76–84.

Kerlinger, F. N., & Lee, H. B. (1999). *Foundations of behavioral research* (4th ed.). Fort Worth, TX: Wadsworth.

***Lee, J.** (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics, 32,* 69–92. https://doi.org/10.1017/S0142716410000299

Leonard, L. B. (2013). Alternative routes to language impairment. In L. A. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 363–376). Baltimore, MD: Brookes.

Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). Cambridge, MA: MIT Press.

***Levickis, P., Reilly, S., Girolametto, L., Ukoumunne, O. C., & Wake, M.** (2014). Maternal behaviors promoting language acquisition in slow-to-talk toddlers: Prospective community-based study. *Journal of Developmental & Behavioral Pediatrics, 35,* 274–281. https://doi.org/10.1097/DBP.0000000000000056

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *The BMJ, 339,* b2700. https://doi.org/10.1136/bmj.b2700

Long, T. (2013). Early intervention. In M. L. Batshaw, N. J. Roizen, & G. R. Lotrecchiano (Eds.), *Children with disabilities* (7th ed., pp. 547–557). Baltimore, MD: Brookes.

***Lyytinen, P., Eklund, K., & Lyytinen, H.** (2005). Language development and literacy skills in late-talking toddlers with and without familial risk for dyslexia. *Annals of Dyslexia, 55,* 166–192.

Moyle, J., Stokes, S. F., & Klee, T. (2011). Early language delay and specific language impairment. *Developmental Disabilities Research Reviews, 17,* 160–169. https://doi.org/10.1002/ddrr.1110

***Moyle, M. J., Weismer, S. E., Evans, J. L., & Lindstrom, M. J.** (2007). Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *Journal of Speech, Language, and Hearing Research, 50,* 508–528.

Nippold, M. A., & Schwarz, I. E. (1996). Children with slow expressive language development: What is the forecast for school achievement? *American Journal of Speech-Language Pathology, 5*(2), 22–25. https://doi.org/10.1044/1058-0360.0502.22

***Paul, R.** (1993). Patterns of development in late talkers: Preschool years. *Communication Disorders Quarterly, 15*(1), 7–14. https://doi.org/10.1177/152574019301500103

Paul, R. (2000). Predicting outcomes of early expressive language delay: Ethical implications. In D. V. M. Bishop & L. B. Leonard (Eds.), *Speech and language impairments in children: Causes, characteristics, intervention and outcome* (pp. 195–209). New York, NY: Psychology Press.

Paul, R., & Fountain, R. (1999). Predicting outcomes of early expressive language delay. *Infant-Toddler Intervention: The Transdisciplinary Journal, 9*(2), 123–135.

***Paul, R., Looney, S. S., & Dahm, P. S.** (1991). Communication and socialization skills at ages 2 and 3 in "late-talking" young children. *Journal of Speech and Hearing Research, 34,* 858–865.

***Petinou, K., & Spanoudis, G.** (2014). Early language delay phenotypes and correlation with later linguistic abilities. *Folia Phoniatrica et Logopaedica, 66,* 67–76. https://doi.org/10.1159/000365848

***Peyre, H., Bernard, J. Y., Forhan, A., Charles, M.-A., De Agostini, M., Heude, B., & Ramus, F.** (2014). Predicting changes in language skills between 2 and 3 years in the EDEN mother–child cohort. *PeerJ, 2,* e335. https://doi.org/10.7717/peerj.335

Reilly, S., Wake, M., Bavin, E., Eadie, P., Bretherton, L., & Prior, M. R. (2008). Letter regarding "A systematic review of the literature on characteristics of late-talking toddlers," by Desmarais et al. *International Journal of Language & Communication Disorders, 43,* 473–475. https://doi.org/10.1080/13682820802242825

***Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., … Bretherton, L.** (2010). Predicting language outcomes at 4 years of age: Findings from Early Language in Victoria Study. *Pediatrics, 126,* e1530–e1537. https://doi.org/10.1542/peds.2010-0254

Rescorla, L. A. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 45*(2), 360–371. https://doi.org/10.1044/1092-4388(2002/028)

Rescorla, L. A. (2011). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews, 17,* 141–150. https://doi.org/10.1002/ddrr.1108

Rescorla, L. A. (2013). Late-talking toddlers: A 15-year follow-up. In L. A. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 219–239). Baltimore, MD: Brookes.

Rescorla, L. A., & Dale, P. S. (2013). Where do we stand now? Conclusions and future directions. In L. A. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 377–387). Baltimore, MD: Brookes.

***Rescorla, L. A., & Schwartz, E.** (1990). Outcome of toddlers with specific expressive language delay. *Applied Psycholinguistics, 11,* 393–407. https://doi.org/10.1017/S0142716400009644

Roy, P., & Chiat, S. (2013). Teasing apart disadvantage from disorder: The case of poor language. In C. R. Marshall (Ed.), *Current issues in developmental disorders* (pp. 125–150). New York, NY: Psychology Press.

Thal, D. J., & Katich, J. (1996). Predicaments in early identification of specific language impairment: Does the early bird always catch the worm? In K. N. Cole, P. S. Dale, & D. J. Thal (Eds.), *Assessment of communication and language* (pp. 1–28). Baltimore, MD: Brookes.

Thal, D. J., Marchman, V. A., & Tomblin, J. B. (2013). Late-talking toddlers: Characterization and prediction of continued delay. In L. A. Rescorla & P. S. Dale (Eds.), *Late talkers: Language development, interventions, and outcomes* (pp. 169–201). Baltimore, MD: Brookes.

***Thal, D., Tobias, S., & Morrison, D.** (1991). Language and gesture in late talkers: A 1-year follow-up. *Journal of Speech and Hearing Research, 34,* 604–612.

Thompson, B., Diamond, K. E., McWilliam, R., Snyder, P., & Snyder, S. W. (2005). Evaluating the quality of evidence from correlational research for evidence-based practice. *Exceptional Children, 71,* 181–194.

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

***Vuksanovic, J. R.** (2015). Relationship between social interaction bids and language in late talking children. *International Journal of Speech-Language Pathology, 17,* 527–536.

Washbrook, E., & Waldfogel, J. (2010). *Cognitive gaps in the early years: A summary of findings from the report "Low Income and Early Cognitive Development in the UK."* London, United Kingdom: Sutton Trust.

Westerlund, M., Berglund, E., & Eriksson, M. (2006). Can severely language delayed 3-year-olds be identified at 18 months? Evaluation of a screening version of the MacArthur-Bates Communicative Development Inventories. *Journal of Speech, Language, and Hearing Research, 49*(2), 237–247.

Whitehurst, G. J., & Fischel, J. E. (1994). Early developmental language delay: What, if anything, should the clinician do about it? *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 35,* 613–648.

Whitehurst, G. J., Fischel, J. E., Arnold, D. S., & Lonigan, C. J. (1992). Evaluating outcomes with children with expressive language delay. In S. F. Warren & J. Reichle (Eds.), *Causes and effects in communication and language intervention* (pp. 277–313). Baltimore, MD: Brookes.

***Whitehurst, G. J., Smith, M., Fischel, J. E., Arnold, D. S., & Lonigan, C. J.** (1991). The continuity of babble and speech in children with specific expressive language delay. *Journal of Speech and Hearing Research, 34,* 1121–1129.

***Williams, A. L., & Elbert, M.** (2003). A prospective longitudinal study of phonological development in late talkers. *Language, Speech, and Hearing Services in Schools, 34,* 138–153.