JAMA | US Preventive Services Task Force | EVIDENCE REPORT

Screening for Speech and Language Delay and Disorders in Children 5 Years or Younger

Evidence Report and Systematic Review for the US Preventive Services Task Force

Cynthia Feltner, MD, MPH; Ina F. Wallace, PhD; Sallie W. Nowell, PhD, CCC-SLP; Colin J. Orr, MD, MPH; Brittany Raffa, MD; Jennifer Cook Middleton, PhD; Jessica Vaughan, MPH; Claire Baker; Roger Chou, MD; Leila Kahwati, MD, MPH

IMPORTANCE Children with speech and language difficulties are at risk for learning and behavioral problems.

OBJECTIVE To review the evidence on screening for speech and language delay or disorders in children 5 years or younger to inform the US Preventive Services Task Force.

DATA SOURCES PubMed/MEDLINE, Cochrane Library, PsycInfo, ERIC, Linguistic and Language Behavior Abstracts (ProQuest), and trial registries through January 17, 2023; surveillance through November 24, 2023.

STUDY SELECTION English-language studies of screening test accuracy, trials or cohort studies comparing screening vs no screening; randomized clinical trials (RCTs) of interventions.

DATA EXTRACTION AND SYNTHESIS Dual review of abstracts, full-text articles, study quality, and data extraction; results were narratively summarized.

MAIN OUTCOMES AND MEASURES Screening test accuracy, speech and language outcomes, school performance, function, quality of life, and harms.

RESULTS Thirty-eight studies in 41 articles were included (N = 9006). No study evaluated the direct benefits of screening vs no screening. Twenty-one studies (n = 7489) assessed the accuracy of 23 different screening tools that varied with regard to whether they were designed to be completed by parents vs trained examiners, and to screen for global (any) language problems vs specific skills (eg, expressive language). Three studies assessing parent-reported tools for expressive language skills found consistently high sensitivity (range, 88%-93%) and specificity (range, 88%-85%). The accuracy of other screening tools varied widely. Seventeen RCTs (n = 1517) evaluated interventions for speech and language delay or disorders, although none enrolled children identified by routine screening in primary care. Two RCTs evaluating relatively intensive parental group training interventions (11 sessions) found benefit for different measures of expressive language skills, and 1 evaluating a less intensive intervention (6 sessions) found no difference between groups for any outcome. Two RCTs (n = 76) evaluating the Lidcombe Program of Early Stuttering Intervention delivered by speech-language pathologists featuring parent training found a 2.3% to 3.0% lower proportion of syllables stuttered at 9 months compared with the control group when delivered in clinic and via telehealth, respectively. Evidence on other interventions was limited. No RCTs reported on the harms of interventions.

CONCLUSIONS AND RELEVANCE No studies directly assessed the benefits and harms of screening. Some parent-reported screening tools for expressive language skills had reasonable accuracy for detecting expressive language delay. Group parent training programs for speech delay that provided at least 11 parental training sessions improved expressive language skills, and a stuttering intervention delivered by speech-language pathologists reduced stuttering frequency.

Editorial page 292

Multimedia

Related article page 329 and **JAMA Patient Page** page 368

Supplemental content

CME at jamacmelookup.com

Related article at jamanetworkopen.com

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Cynthia Feltner, MD, MPH, Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, 725 Martin Luther King Jr Blvd, CB#7295, Chapel Hill, NC 27599 (cindy_feltner@med.unc.edu).

JAMA. 2024;331(4):335-351. doi:10.1001/jama.2023.24647

An estimated 8% of US children aged 3 to 17 years have a communication disorder.[1] Boys are almost twice as likely to be affected than girls (9.6% vs 5.7%,) and higher rates are observed among Black children (10%) compared with Hispanic (6.9%) or White (7.8%) children.[1] These data and other nationally representative prevalence estimates are limited in terms of distinguishing children who have a delay vs specific speech and/or language disorder.

A "delay" refers to development of speech and language in the correct sequence but at a slower rate than expected, whereas a "disorder" refers to development of speech and/or language ability that is qualitatively different from typical development. Speech disorders are characterized by difficulty with forming specific sounds or words correctly (articulation or phonological disorders) or making words or sentences flow smoothly (fluency disorders), and language disorders are characterized by difficulty understanding (receptive language) or speaking (expressive language) relative to their peers.[2] The focus of this review is routine screening for developmental (or "primary") speech or language delay and disorders that are not caused by an injury or another condition (acquired or "secondary" disorders) such as hearing loss (eg, secondary to infection or genetic syndrome) or autism. Evaluation of children with known conditions that affect speech or language development would be part of disease management rather than screening; however, in the context of routine screening, some children who screen positive may go on to receive a primary diagnosis for a disorder such as hearing loss following a diagnostic evaluation.

Many children identified with speech or language delay go on to recover without an intervention.[3] However, observational evidence suggests that school-aged children with speech or language delay may be at increased risk of learning and literacy disabilities.[4-6] and social and behavioral problems,[7] some of which may persist through adulthood.[8,9] Screening for speech and language delay is distinct from overall developmental screening recommended by the American Academy of Pediatrics at 18 and 30 months.[10] Children who screen positive require referral for a diagnostic evaluation to confirm the suspected delay or disorder. Once a diagnosis is confirmed, treatment is variable and individualized to the needs of the child based on how the disorder impairs their function in different settings.

In 2015, the US Preventive Services Task Force (USPSTF) concluded that the evidence was insufficient to assess the balance of benefits and harms of screening for speech and language delay and disorders in children 5 years or younger (I statement).[11] The purpose of the current systematic review was to update the previous evidence review on the benefits and harms of screening for speech and language delay and disorders in children to inform the USPSTF in updating its recommendation.

Methods

Scope of Review

Figure 1 shows the analytic framework and key questions (KQs) that guided the review. Detailed methods are available in the full evidence review.[12] In addition to the KQs, this review looked for evidence related to 3 contextual questions that focused on disparities in the prevalence, detection, and provision and utilization of treatment for speech and language delay or disorders among specific populations of children (eContextual Questions in the Supplement).

Data Sources and Searches

PubMed/MEDLINE, the Cochrane Library, APA PsycInfo, ERIC, and Linguistic and Language Behavior Abstracts (ProQuest) were searched for English-language articles published through January 17, 2023 (eMethods in the Supplement). ClinicalTrials.gov was searched for unpublished studies. The searches were supplemented by reviewing reference lists of pertinent articles, studies suggested by peer reviewers, and comments received during public commenting periods. From January 17, 2023, through November 24, 2023, ongoing surveillance was conducted through article alerts and targeted searches of journals to identify major studies published in the interim that may affect the conclusions or understanding of the evidence and the related USPSTF recommendation.

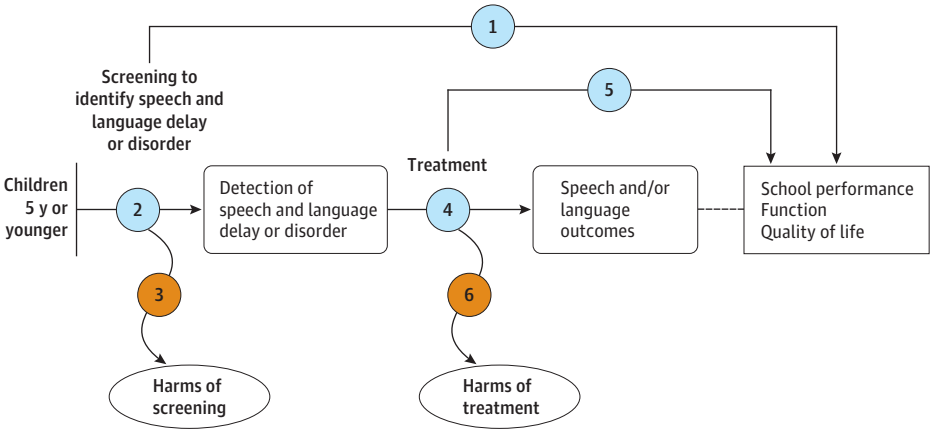
Study Selection

Two investigators independently reviewed titles, abstracts, and full-text articles using prespecified eligibility criteria (eTable 4 in Supplement). Disagreements were resolved by discussion and consensus. For all KQs, English-language studies enrolling unselected children 5 years or younger from primary care or primary care–relevant settings (including childcare, schools, and other education settings) who communicate using any language were eligible. In addition, only studies set in countries categorized as "very high" on the Human Development Index[13] and rated as fair or good quality were included. For studies assessing the benefits and harms of interventions (KQ4, KQ5, and KQ6), those enrolling children referred for treatment or identified by educators or parents as having a possible speech or language problem, and those enrolling children up to age 6 years were also eligible.

For KQ2, studies assessing the accuracy of a screening instrument against a diagnosis reference standard (diagnostic interview, diagnostic questionnaire, or both) were included. Eligible screening instruments had to be feasible for use in primary care and included short questionnaires that could be delivered and interpreted in 10 minutes or less in clinical settings and longer questionnaires completed by parents or teachers outside of a scheduled visit. Studies focusing on the accuracy of general developmental screening tools that did not include a separate component for speech and language skills were excluded.

Randomized clinical trials (RCTs), nonrandomized clinical trials, and controlled cohort studies were eligible for KQ1 and KQ3 (benefit and harms of screening compared with no screening) and KQ6 (harms of interventions compared with an inactive control). For studies reporting on the benefit of interventions to improve speech and language outcomes (KQ4) or academic skills, behavior, function, or quality of life (KQ5), RCTs comparing an intervention with an inactive control were eligible. For KQ4, KQ5, and KQ6, eligible interventions included any treatment designed to improve speech and/or language delay or disorders among eligible populations, regardless of format (eg, individual or group settings, face-to-face, or via telehealth) or delivery personnel (eg, speech-language pathologists [SLPs] or other clinicians, parents, or teachers).

Figure 1. Analytic Framework and Key Questions: Screening for Speech and Language Delay and Disorders in Children 5 Years or Younger



Key questions

1. Does screening for speech and language delay or disorders in children age 5 y or younger improve speech and language outcomes, school performance, function, or quality-of-life outcomes?
2. What is the accuracy of screening tools to detect speech and language delay or disorders in children age 5 y or younger?
3. What are the harms of screening for speech and language delay or disorders in children age 5 y or younger?
4. Do interventions for speech and language delay or disorders in children age 6 y or younger improve speech and language outcomes?
5. Do interventions for speech and language delay or disorders in children age 6 y or younger improve school performance, function, or quality-of-life outcomes?
6. What are the harms of interventions for speech and language delay or disorders?

Evidence reviews for the US Preventive Services Task Force (USPSTF) use an analytic framework to visually display the key questions that the review will address to allow the USPSTF to evaluate the effectiveness and safety of a preventive service. The questions are depicted by linkages that relate interventions and outcomes. A dashed line depicts a health outcome that follows an intermediate outcome. For additional information, see the USPSTF Procedure Manual.[16,17]

Data Extraction and Quality Assessment

For each included study, 1 investigator extracted pertinent information about the methods, populations, interventions, comparators, outcomes, timing, settings, and study designs. All data extractions were checked by a second investigator for completeness and accuracy. For newly identified studies, 2 reviewers independently assessed each study's methodological quality using predefined criteria developed by the USPSTF (eMethods in Supplement) and informed by tools designed for various study designs (Cochrane Risk of Bias 2.0 tool for RCTs[14]; Quality Assessment of Diagnostic Accuracy Studies 2 for screening test accuracy).[15] For eligible studies included in the previous update for this topic, quality ratings were spot-checked and carried forward. Disagreements were resolved by discussion.

Data Synthesis and Analysis

Findings for each KQ were summarized in tabular and narrative format. The overall strength of the evidence for each KQ was assessed as high, moderate, low, or insufficient based on the overall quality of the studies, consistency of results between studies, precision of findings, risk of reporting bias, and limitations of the body of evidence using methods developed for the USPSTF (and the Evidence-based Practice Center program).[16,17] Additionally, the applicability of the findings to US primary care populations and settings was assessed. Discrepancies were resolved through consensus discussion.

For studies included for KQ2 (accuracy of screening tools), sensitivity, specificity, likelihood ratios, and predictive values were calculated based on data reported by articles, when sufficient, to compare consistency across similar measures. To determine whether meta-analyses were appropriate, the clinical heterogeneity and methodological heterogeneity of the studies were assessed following established guidance.[18] Due to heterogeneity in populations, outcome measures and other factors, as well as few studies assessing the same screening tool or interventions, meta-analysis was not appropriate.

Results

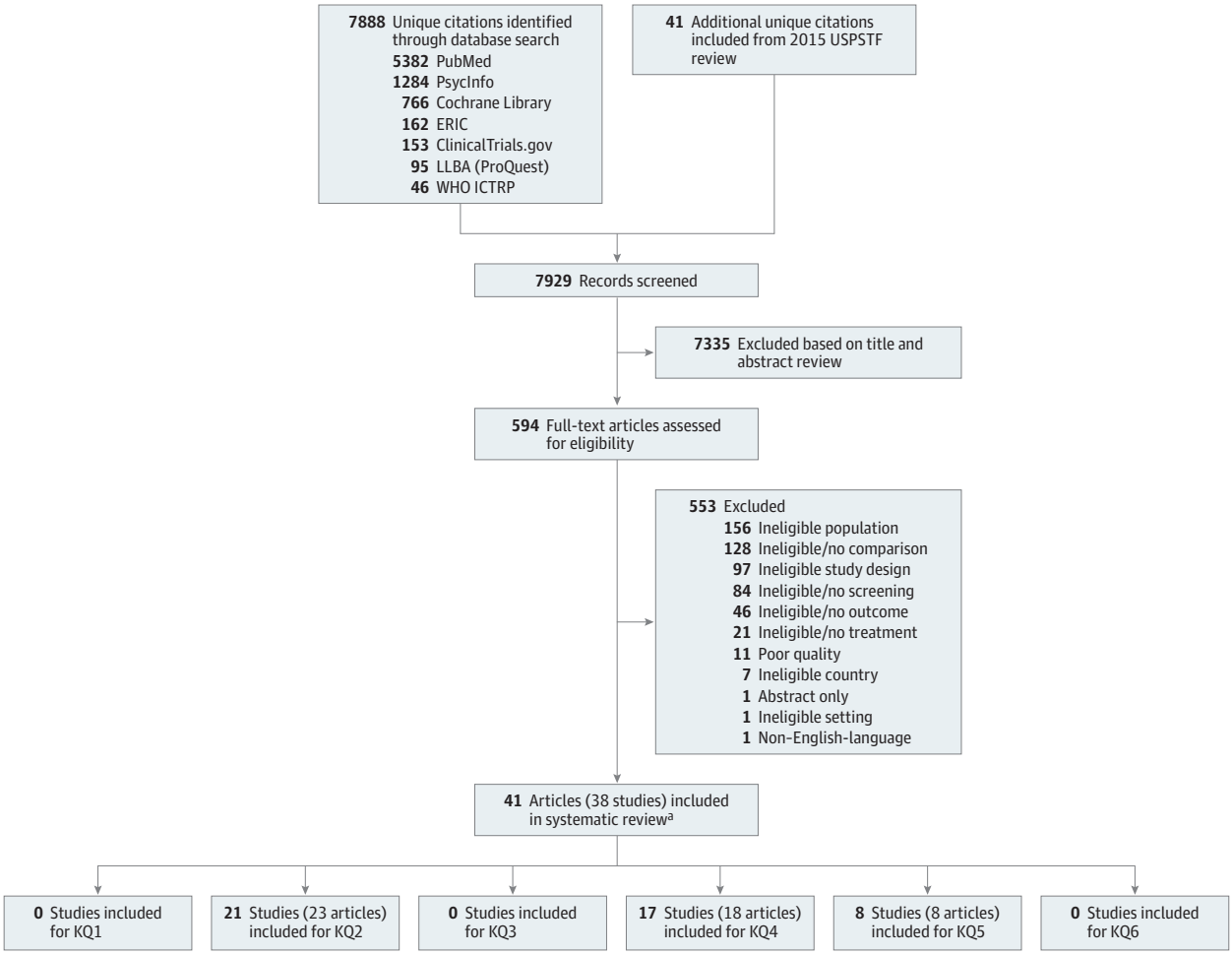
A total of 38 studies (reported in 41 articles) were included (**Figure 2**) in the review. Individual study quality ratings are reported in eTables 5 through 10 in the Supplement.

Benefits of Screening

Key Question 1. Does screening for speech and language delay or disorders in children age 5 years or younger improve speech and language outcomes, school performance, function, or quality-of-life outcomes?

No eligible study addressed this question.

Figure 2. Literature Search Flow Diagram: Screening for Speech and Language Delay and Disorders in Children 5 Years or Younger



ERIC indicates Education Resources Information Center; KQ, key question; LLBA, Linguistics and Language Behavior Abstracts; USPSTF, US Preventive Services Task Force; and WHO ICTRP, World Health Organization International Clinical Trials Registry Platform.

[a] The sum of the number of studies per KQ exceeds the total number of studies because some studies were applicable to multiple KQs.

Accuracy of Screening

Key Question 2. What is the accuracy of screening tools to detect speech and language delay or disorders in children age 5 years or younger?

Twenty-one studies (reported in 23 articles) assessed the accuracy of 23 screening instruments for detecting speech and language delay and disorders in young children against a reference standard (n = 7489) (**Table 1**).[19-41] Seven studies were new to this update.[24,27,30-32,39,41] Of the 23 instruments, 13[19-23,28-32,35,37,38] were designed to be administered to children by a trained examiner, and 10[23-27,33-36,39-41] were parent reports of children's speech or language skills (**Table 2**).

Some screening tools, termed global screening tools, screen for any language problems, while others provide scores for specific aspects of language (eg, expressive communication, receptive language, vocabulary). Twelve global screening tools were evaluated in the studies included the Ages and Stages Questionnaire (ASQ),[23,41] the Davis Observation Checklist for Texas,[19] the Developmental Nurse Screen,[35] the Early Language Scale,[39] the Fluharty Preschool Screening Test (FPST),[20] the General Language Screen,[36] the Hackney Early Language Screening Test/Structured Screening Test (HELST/SST),[28,29] the Infant-Toddler Checklist,[40] the Nurse Screening,[30,31] the Parent Questionnaire,[35] the Screening Kit of Language Development (SKOLD)/Screening Kit of Language Development Black English (SKOLDBE),[21] and the language component of the Sentence Repetition Screening Test (SRST).[38]

Nine other tools provided scores for specific aspects of language, including the Brigance Preschool Screen,[23] the Early Screening Profiles,[23] the Battelle the Elternfragebogen für die Fruberkennung von Riskokindern (ELFRA-2),[33,34] the Sprachentwicklungsscreening (SPES-3) instrument,[24] the Language Development Survey (LDS),[25,26] the Quick Interactive Language Screener (QUILS),[32] the Sure Start Language Measure (SSLM),[41] the Northwestern Syntax Screening Test,[20] and the Battelle Developmental Inventory Screening Test–Communication.[23]

Table 1. Characteristics of Included Studies of Diagnostic Accuracy (KQ2)

Source, setting	Study design (No. of participants)	Recruitment setting	Screening tool	Age, mean (range), mo	% Female	Study quality
Alberts et al,[19] 1995 United States	Cross-sectional (n = 59)	Head Start centers in Central Texas	DOCT	48 (52-67)	51	Fair
Allen and Bliss,[20] 1987 United States	Cross-sectional (n = 182)	Childcare centers in suburban Dallas	FPST, NSST	36-47	NR	Fair
Bliss and Allen,[21] 1984 United States	Cross-sectional (n = 602)	Childcare centers in metropolitan Detroit	SKOLD, SKOLDBE	40 (30-48)	48	Fair
Drumwright et al,[22] 1973 United States	Prospective cohort (n = 150)	Head Start, public and private childcare centers, schools, and pediatric clinics in Denver	DASE	(30-72)	NR	Fair
Frisk et al,[23] 2009 Canada	Prospective cohort (n = 110)	Programs providing early intervention services to at-risk children in Ontario	ASQ-CD, BDIST-CD, BPS, ESP	54	32	Fair
Holzinger et al,[24] 2021 Austria	Prospective cohort (n = 2044[a])	Pediatric medical practices in Upper Austria	SPES-3	36 (34-38)[b]	49	Fair
Klee et al,[25] 1998 (study 2) Klee et al,[26] 2000 United States	Prospective cohort (n = 64)	Birth announcements, and local physicians, health departments, and WIC offices in Laramie and Casper, Wyoming	LDS	25 (24-26)	39	Fair
Kok and To,[27] 2019 Hong Kong	Cross-sectional (n = 789)	11 community kindergartens in Hong Kong	ICS-TC	53 (28-81)	47	Fair
Laing et al,[28] 2002 United Kingdom	Cross-sectional (n = 458)	Health center in London	SST	30	44	Good
Law,[29] 1994 United Kingdom	Prospective cohort (n = 189)	Pediatric practice in London	HELST	30	NR	Good
Nayeb et al,[30] 2019 Sweden	Prospective cohort (n = 105[b])	Child health centers in Gävle, Sweden	Nurse screening (Swedish and maternal language)	30	47	Fair
Nayeb et al,[31] 2021 Sweden	Prospective cohort (n = 111[c])	Child health centers in Gävle, Sweden	Nurse screening	30 (29-33)	51	Fair
Pace et al,[32] 2022 (study 2 only) United States	Cross-sectional (n = 126)	University speech and hearing clinic; inclusive public preschool and kindergarten classrooms; Head Start centers	QUILS	56 (38-70)	50	Fair
Sachse et al,[33] 2008 Sachse et al,[34] 2009 Germany	Prospective cohort (n = 117)	Birth announcements in Germany	ELFRA-2 (German version of CDI Words and Sentences)	25 (24-26)	33	Good
Stokes,[35] 1997 Australia	Prospective cohort (n = 398)	Child Health Centres in metropolitan Perth	DNS, parent questionnaire	37 (34-40)	51	Good
Stott et al,[36] 2002 United Kingdom	Prospective cohort (n = 596)	Mailed invitations to children born within Cambridge Health Authority	GLS	36	NR	Fair
Sturner et al,[37] 1993 United States	Prospective cohort (n = 51 [study 1]; n = 147 [study 2])	Schools in a rural county in North Carolina	FPSLST	Study 1: 61 (53-68) Study 2: 62 (55-69)	Study 1: 54 Study 2: 48	Fair
Sturner et al,[38] 1996 United States	Prospective cohort (n = 337[d])	Schools in a rural county in North Carolina	SRST	60 (54-66)	52	Fair
Visser-Bochane et al,[39]2021 The Netherlands	Prospective cohort (n = 265)	Well-child clinics, kindergartens, and schools in the Netherlands	ELS	44 (15-72)	51	Fair
Wetherby et al,[40] 2003 (study 1) United States	Prospective cohort (n = 232)	Public announcements, health care professionals, childcare personnel, and a public health care agency	ITC from CSBS	12-24	NR	Fair
Wilson et al et al,[41]2022 United Kingdom	Prospective cohort (n = 357)	Mailed invitations to parents of children due to receive their universal developmental assessment	ASQ, SSLM	26 (23-30)	47	Fair

Abbreviations: ASQ, Ages and Stages Questionnaire; ASQ-CD, ASQ–Communication Domain; BDIST-CD, Battelle Developmental Inventory Screening Test–Communication Domain; BPS, Brigance Preschool Screen; CDI, MacArthur-Bates Communicative Development Inventory; CSBS, Communication and Symbolic Behavior Scales; DASE, Denver Articulation Screening Exam; DNS, Developmental Nurse Screen; DOCT, Davis Observation Checklist for Texas; ELFRA-2, Elternfragebogen für die Fruberkennung von Riskokindern; ELS, Early Language Scale; ESP, Early Screening Profiles; FPSLST, Fluharty Preschool Speech and Language Screening Test; FPST, Fluharty Preschool Screening Test; GLS, General Language Screen; HELST, Hackney Early Language Screening Test; ICS-TC, Intelligibility in Context Scale-Traditional Chinese; ITC, Infant-Toddler Checklist; KQ, key question; LDS, Language Development Survey; NR, not reported; NSST, Northwestern Syntax Screening Test; QUILS, Quick Interactive Language Screening; SKOLD, Screening Kit of Language Development; SKOLDBE, Screening Kit of Language Development Black English; SPES-3, Sprachentwicklungsscreening; SRST, Sentence Repetition Screening Test; SSLM, Sure Start Language Measure; SST, Structured Screening Test; WIC, Women, Infants, and Children.

[a] Full sample size, based on multiple imputation.

[b] Includes 11 children (10.5%) who did not cooperate during screening and were considered screen positive.

[c] Includes 11 children who were noncooperative during screening. For Model 4, parents of 10 children did not complete parental information.

[d] Based on full sample.

Three of the trained examiner tools specifically screened for articulation skills—the Denver Articulation Screening Exam[22] and the articulation portion of both the Fluharty Preschool Speech and Language Screening Test (FPSLST)[37] and the SRST[38]—and 1

Table 2. Instruments Examined in KQ2 Studies

Instrument	Screening source	Appropriate ages	Domains/skills assessed	Summary scores	No. of items
Ages and Stages Questionnaire -Communication Domain[23,41]	Parent-reported	4 to 60 mo	Broad communication skills	Communication	6 at each age level
Battelle Developmental Inventory Screening Test-Communication Domain[23]	Trained examiner	1 to 8 y	Receptive and expressive language skills[a]	Receptive language Expressive language	9 per each subtest
Brigance Preschool Screen[23]	Trained examiner	45 to 56 mo	Receptive and expressive language skills	Understanding reading (ie, receptive language) Expressive language	Receptive: 2 Expressive: 4
Davis Observation Checklist for Texas[19]	Trained examiner	4 to 5 y	Speaking, understanding, speech fluency, voice, and hearing	Communication	2-5 behaviors in each of 6 areas
Denver Articulation Screening Exam[22]	Trained examiner	2.5 to 7 y	Articulation skills	Articulation	34 sound elements
Developmental Nurse Screen[35]	Trained examiner	34 to 40 mo	Broad language skills	Global language	NR
Early Language Scale[39]	Parent-reported	1 to 6 y	Vocabulary, syntax, morphology, and pragmatics	Global language	26
Early Screening Profiles[23]	Trained examiner	2 y 0 mo to 6 y 11 mo	Word comprehension and production	Verbal concepts	25
ELFRA-2; German version of CDI Words and Sentences[33,34]	Parent-reported	16 to 30 mo	German expressive vocabulary, morphology, and grammar	Expressive language	Vocabulary: 260 Syntax: 25 Morphology: 11
Fluharty Preschool Screening Test[20]/Fluharty Preschool Speech and Language Screening Test[37]	Trained examiner	2 to 5 y	Articulation, and expressive and receptive language skills	Articulation Language	35
General Language Screen[36]	Parent-reported	36 mo	Comprehension, expression, articulation, and pragmatics	Global language	11
Hackney Early Language Screening Test/Structured Screening Test[28,29]	Trained examiner	30 mo	Expressive and receptive language skills	Global language	20
Infant-Toddler Checklist from CSBS[40]	Parent-reported	6 to 24 mo	Emotion and use of eye gaze, communication, gestures, sound use, word use, word understanding, and object use	Social, speech, and symbolic composites Total score	24
Intelligibility in Context Scale-Traditional Chinese[27]	Parent-reported	28 to 71 mo	Functional intelligibility	Articulation	7
Language Development Survey[25,26]	Parent-reported	18 to 35 mo	Expressive vocabulary and word combinations	Expressive language	310
Northwestern Syntax Screening Test[20]	Trained examiner	3 to 8 y	Expressive and receptive knowledge of syntactic forms	Syntactic expression Syntactic comprehension	20 per each subtest
Nurse Screening[30,31]	Trained examiner	2.5 y	Language comprehension and language production	Global language	5 and observation
Parent Questionnaire[35]	Parent-reported	34 to 40 mo	Sentence use, comprehension, articulation, and global problems	Global language	4
Quick Interactive Language Screener[32]	Trained examiner	3 y through 6 y and 11 mo	Comprehension of vocabulary (nouns, verbs, prepositions, conjunctions), syntax (WH questions, past tense, prepositional phrases, embedded clauses), and language learning (noun learning, adjective learning, verb learning, converting active to passive)	Vocabulary, syntax, process, and overall (composite) scores	48
Screening Kit of Language Development/Screening Kit of Language Development Black English[21]	Trained examiner	54 to 66 mo	Vocabulary comprehension, story completion, sentence completion, paired sentence repetition, individual sentence repetition with and without pictures, and comprehension of commands	Global language	20-50 items per each of 7 subtests
Sentence Repetition Screening Test[38]	Trained examiner	54 to 66 mo	Expressive morphology and articulation	Global language articulation	15
SPES-3[24]	Parent-reported[b]	3 y	Expressive vocabulary, expressive grammar	Expressive language	113
Sure Start Language Measure[41]	Parent-reported (to examiner)	2 to 2.5 y	Expressive vocabulary	Expressive vocabulary	50

Abbreviations: CDI, MacArthur-Bates Communicative Development Inventory; CSBS, Communication and Symbolic Behavior Scales; ELFRA-2, Elternfragebogen für die Fruberkennung von Riskokindern; KQ, key question; NR, not reported; SPES-3, Sprachentwicklungsscreening; WH questions, who, when, where, why, what, and how.

[a] Only the Battelle Developmental Inventory Test Receptive Language Scale is included in accuracy analyses.

[b] Although the SPES-3 was designed as both a parent-reported and trained examiner instrument, the authors recommended that only the parent-reported subscales be included as a screen for language delay; therefore, the SPES-3 was classified as a parent-reported instrument.

parent-administered instrument measured articulation.[27] The articulation instruments were considered separately from specific language instruments. All but 3 instruments (ie, ASQ,[23,41] HELST/SST,[28,29] and Nurse Screening[30,31]) were examined in only 1 study each. In addition, 2 studies examined the FPST[20] and a later version with a language component, the FPSLST.[37]

Excluding 2 studies[33,40] that enrolled all children who screened positive and a random sample of children who screened negative, the prevalence of speech and language disorders based on reference standards ranged from 4% to 33% (**Table 3**).

Accuracy of Instruments

As shown in Table 3, the sensitivity of instruments for detecting speech and language disorders and delay ranged from 17% and 100% (median, 86%), and specificity ranged between 32% and 98% (median, 87%). To further examine accuracy, the source of the information (parent report vs trained examiner) and whether the instrument was designed as a global index of speech or language, a specific language skill (eg, word knowledge), or a measure of articulation were considered.

Parent Reported

Sensitivity and specificity of 14 parent-reported tools varied widely (Table 3). Sensitivity ranged from 55% to 93% (median, 84%) and specificity ranged from 32% to 96% (median, 84%).

Global Language vs Specific Language vs Articulation | Limiting analysis to global language instruments based on parent reports, median sensitivity was 74%, ranging between 55% and 89%. Specificity was less variable, ranging between 73% and 95% (median, 79%). In contrast, both sensitivity and specificity of the 3 parent-reported instruments of specific skills (all emerging expressive language skills) were fairly consistent and high (median sensitivity, 91% [range, 83%-93%]; median specificity, 88% [range, 81%-96%]). The 1 parent-rated measure of articulation had a reasonably high sensitivity (86%) but low specificity (32%).

Trained Examiners

The median sensitivity of the 13 screening tools that trained examiners administered to children was 87% (range, 17%-100%), and the median specificity was 88% (range, 58% to 98%). Similar to parent-reported instruments, there is substantial variability in the accuracy of examiner-administered tools.

Global Language vs Specific Language vs Articulation | Restricting the accuracy summary to trained examiner screenings of global language resulted in median sensitivity of 88% (range, 17%-100%) and median specificity of 89% (69%-98%). The median sensitivity of trained examiner instruments for specific language skills was 86% (range, 56%-94%) and median specificity was 70% (range, 58%-90%). Across the 3 trained examiner tools for assessing articulation, the median sensitivity was only 66% (range, 43%-92%); however, median specificity was 96% (range, 93%-97%).

Harms of Screening

Key Question 3. What are the harms of screening for speech and language delay or disorders in children age 5 years or younger?

No eligible study addressed this question.

Benefits of Treatment

Key Question 4. Do interventions for speech and language delay or disorders in children age 6 years or younger improve speech and language outcomes?

Seventeen RCTs (18 articles) compared an intervention for speech and language delay or disorders with an inactive control (no treatment or wait-list control/delayed treatment).[42-59] Study characteristics are shown in eTable 11 in the Supplement. No studies enrolled children identified by routine screening in primary care. Most recruited participants from referrals to speech and language treatment centers (6 studies),[42,47,49,50,53,54] schools or early childhood education centers (4 studies),[43,46,48,56] or via advertisements or a mix of advertisements and outreach to schools, clinical settings, or community-based programs.[44,45,55,57] The mean age of enrolled populations ranged from 18.1 months to 67.8 months, with most (10 studies) enrolling a sample with a mean age of 48 months or older. The proportion of participants who were female ranged from 10% to 49%. Few studies reported on race or ethnicity; in 3 studies set in the US, populations were described as 100% Latino,[45] 100% White,[57] and 1 was inclusive of different groups (2% American Indian, 3% Asian, 2% Black, 26% Hispanic, 12% multiracial, 54% White).[48] Interventions evaluated were heterogeneous and varied in terms of the range of disorders targeted, delivery personnel, intensity/duration, settings, and other factors (eTable 11 in Supplement).

Eight RCTs assessed interventions specific to children with delayed expressive language ("late talkers") and no obvious fluency or speech-sound impairment.[44,45,50-52,56-59] Of these, 3 RCTs evaluated parent-group training interventions focused on strategies to promote their child's language development; training approaches and specific content varied, but all focused on naturalistic strategies (eg, expanding on child utterances, following the child's interests, repeating what the child says, setting up the environment to encourage communication). Of these, 2 RCTs assessed modifications of the Hanen Program for Parents curriculum (featuring a combination of group training sessions composed of a small group of parents and a trained SLP or other trained facilitator, and individual consultations with the SLP while the child is present),[51,58] and 1 evaluated a similar group training program focused on improving child linguistic complexity.[50] Results varied by duration of the intervention and mean age of enrolled populations. In 2 RCTs in which the intervention was delivered to children with a mean age of 27 to 30 months over a longer duration (11 bimonthly 60- to 75-minute sessions in one of the trials[50] and 11 weekly 2.5-hour sessions plus 3 weekly home visits in the other trial[51]), there was consistent benefit across different measures of expressive language outcomes (eTable 12 in the Supplement). The RCT delivering the parent group training to children with a mean age of 18 months over a shorter duration (6 weekly 2-hour sessions) found no significant difference between groups on any measure of receptive or expressive language outcomes.[58]

Five other RCTs assessed different interventions for children with language delay and varied in terms of setting, delivery personnel, and other factors.[44,45,56,57,59] In general, results were inconsistent, with some studies showing improvement on some measures of receptive or expressive language but others not. Results are further summarized in the eResults and eTable 12 in the Supplement.

Table 3. Accuracy of Screening Instruments to Detect Speech and Language Disorders

Instruments (cut point)	Screening subtest	Reference standard	No.	Prevalence, %	% (95% CI)		%			
					Sensitivity	Specificity	PPV	NPV	LR+	LR−
Parent-reported										
Global language instruments										
ASQ-CD[23] ("recommended cutoff")		PLS-4-C	110	4	67 (45-88)[a]	73 (64-82)[a]	32[a]	92[a]	2.4[a]	0.46[a]
		PLS-4-E	110	7	73 (54-91)[a]	76 (67-85)[a]	43[a]	92[a]	3.0[a]	0.36[a]
ASQ-CD[41] full sample (37.5)[b]		PLS-5 Total Language	357	23	55 (44-66)	95 (91-97)	53	95	10.0	0.48
English-only sample (47.5)[b]		PLS-5 Total Language	248	NR[c]	85 (70-94)	84 (78-88)	37	98	5.2	0.18
ELS[39] (15)		Composite based on LS, CCC-2, LLC, LLP, SLC, SWP, SSP	265	11	62 (44-77)[a]	93 (89-96)[a]	53	95	9.2	0.41
GLS[36] (≥2 failures)		DP-II	596	18[d]	75 (67-83)[a]	81 (77-84)[a]	47	94	3.9	0.31[a]
ITC (study 1)[40] (NR)	Aged 12 to 17 mo version	CSBS behavior sample	151	35	89 (80-97)[a]	74 (66-83)[a]	65	92	3.5[a]	0.15[a]
	Aged 19 to 24 mo version	CSBS behavior sample	81	52	86 (75-96)[a]	77 (64-90)[a]	80	83	3.7[a]	0.19[a]
Parent questionnaire[35] (≥1 abnormal response)		SLP rating using language sample, RDLS, Comprehension Scale	381	13	78 (66-89)[b]	91 (88-94)[a]	56	96	8.3[a]	0.24[a]
Specific language instruments										
ELFRA-2 (CDI Words and Sentences)[33,34] (<50 words or 50-80 words and scores for syntax <7 and morphology <2)		SETK-2	117	59	93 (87-99)[a]	88 (78-97)[a]	91	89	7.3[a]	0.08[a]
LDS[25] (study 2); (<50 words or no word combinations)		Clinical judgment on infant MSEL language scales, MLU	64	17	91 (74-100)[a]	87 (78-96)[a]	59	98	6.9[a]	0.10[a]
LDS[26] (>28 screening score)			64		91 (74-100)[a]	96 (91-100)[a]	83	98	24.1[a]	0.09[a]
SPES-3[24] (<41.69)		Composite of SETK-3, AWST-R, language sample	2044[e]	10[f]	88 (77-98)	88 (86-90)	44	98	7.1	0.14
SSLM[41]										
Full sample (19.5)[b]		PLS-5	357	23	83 (74-91)	81 (76-85)	33	98	4.4	0.21
English-only sample (16.5)[b]		PLS-5	248	NR[c]	80 (64-91)	87 (82-91)	41	98	6.2	0.23
Articulation										
ICS-TC[27] (4.29)		HKCAT	789	19[a]	86 (79-90)[a]	32 (28-36)[a]	22[a]	91[a]	1.3[a]	0.45[a]

(continued)

Table 3. Accuracy of Screening Instruments to Detect Speech and Language Disorders (continued)

Instruments (cut point)	Screening subtest	Reference standard	No.	Prevalence, %	% (95% CI)		%			
					Sensitivity	Specificity	PPV	NPV	LR+	LR−
Trained examiner										
Global language instruments										
DOCT[19] (NR)		Composite of MSCA, GFTA, informal language sample	59	17	80 (55-100)[a]	98 (94-100)[a]	89[a]	96[a]	39.2[a]	0.20[a]
DNS[35] (NR)		SLP rating using language sample and RDLS, Comprehension Scale	378	NR	76	97	80	96	NR	NR
FPST[20] (≥1 subtest)		SICD	182	14	60 (41-79)[a]	81 (75-87)[a]	33[a]	93[a]	3.1[a]	0.49[a]
FPSLST[37] (NR)	Language Study 1	TACL-R	51	17[f]	38	85	42	NR	NR	NR
	Language Study 2	TOLD-P	147	22[f]	17	97	50	NR	NR	NR
HELST[29] (≤10)		RDLS	189	26	98 (94-100)[a]	69 (61-77)[a]	53	98	3.1[a]	0.03[a]
SST[28] (<10)		RDLS	282	23	66 (53-76)[a]	89 (85-93)[a]	65[a]	90[a]	6.2[a]	0.38[a]
Nurse screening										
<3 Words[30]		RDLS, Comprehension Scale and spontaneous language observation	105[g]	10	100 (72-100)	81 (71-88)	38	100	5.2	0
≥3 Comprehension questions and ≥2 word combinations[30]		RDLS, Comprehension Scale and spontaneous language observation	105[g]	10	91 (71-88)	91 (59-100)	56	99	19.7	0.1
≥3 Comprehension questions and ≥2 word combinations[31]	Model 3—screening in Swedish and maternal language	RDLS, Comprehension Scale and spontaneous language observation	111[g]	29	88 (71-96)	82 (72-90)	67	94	4.9	0.15
SKOLD/SKOLDBE[21] (<11)	S30	SICD	47	6	100 (100-100)[a]	98 (93-100)[a]	75[a]	100[a]	44.0[a]	0[a]
(<10)	S37	SICD	93	11	100 (100-100)[a]	91 (85-97)[a]	33[a]	100[a]	11.1[a]	0
(<19)	S43	SICD	100	9	100 (100-100)[a]	93 (88-98)[a]	60[a]	100[a]	15.2[a]	0[a]
(<9)	B30	SICD	75	12	89 (68-100)[a]	86 (78-95)[a]	47[a]	98[a]	6.5[a]	0.13[a]
(<14)	B27	SICD	91	9	88 (65-100)[a]	86 (78-92)[a]	37[a]	99[a]	6.0[a]	0.15[a]
(<19)	B43	SICD	54	33	94 (84-100)[a]	78 (64-91)[a]	68[a]	97[a]	4.2[a]	0.07[a]
SRST[38] (<20th percentile)	SRST language	ITPA/BLST	323[h]	11	62 (45-78)[a]	91 (87-94)[a]	44	95[a]	6.6[a]	0.42[a]

(continued)

Table 3. Accuracy of Screening Instruments to Detect Speech and Language Disorders (continued)

Instruments (cut point)	Screening subtest	Reference standard	No.	Prevalence, %	% (95% CI)		%			
					Sensitivity	Specificity	PPV	NPV	LR+	LR−
Specific language instruments										
BDIST-CD[23] (ROC optimal cutoff)	Receptive[i]	PLS-4-C	110	4	56 (33-78)[a]	70 (60-79)[a]	26[a]	89[a]	1.8[a]	0.89[a]
BPS[23] (ROC optimal cutoff)	Receptive	PLS-4-C	110	4	61 (39-84)[a]	60 (50-70)[a]	23[a]	89[a]	1.5[a]	0.65[a]
	Expressive	PLS-4-E	110	7	91 (79-100)[a]	78 (70-87)[a]	51[a]	97[a]	4.2[a]	0.12[a]
ESP[23] (>1 SD below mean)	Verbal concepts	PLS-4-C	110	4	94 (84-100)[a]	68 (59-78)[a]	40[a]	98[a]	3.0[a]	0.08[a]
	Verbal concepts	PLS-4-E	110	7	86 (72-100)[a]	81 (72-89)[a]	53[a]	96[a]	4.5[a]	0.17[a]
NSST[20] (failure ≥1 subtest)		SICD	182	14	92 (81-100)[a]	48 (41-56)[a]	22[a]	97[a]	1.8[a]	0.16[a]
QUILS[32] (study 2 only) (<25th percentile)	Composite	PLS-5 Auditory Comprehension	126	20	60 (51-69)[a]	90 (70-96)[a]	95[a]	35[a]	6.0	0.66
Articulation instruments										
DASE[22] (<15th percentile)		HAT	150	NR	92	97	NR	NR	NR	NR
FPSLST[37] (NR)	Articulation study 1	AAPS-R	51	4[f]	74	96	50	NR	NR	NR
	Articulation study 2	TD	147	5[f]	43	93	26	NR	NR	NR
SRST[38] (<20th percentile)	SRST Articulation	AAPS-R	325[h]	19	57 (45-69)[a]	95 (93-98)[a]	75	90[a]	12.5[a]	.045[a]

Abbreviations: AAPS-R, Arizona Articulation Proficiency Scale–Revised; ASQ-CD, Ages and Stages Questionnaire–Communication Domain; AWST-R, Aktiver Wortschatztest für 3-bis 5-jährige Kinder; BDIST-CD, Battelle Developmental Inventory Screening Test–Communication Domain; BLST, Bankson Language Screening Test; BPS, Brigance Preschool Screen; CCC-2, Children's Communication Checklist, 2nd Edition–Netherlands; CDI, MacArthur-Bates Communicative Development Inventory; CSBS, Communication and Symbolic Behavior Scales; DASE, Denver Articulation Screening Exam; DNS, Developmental Nurse Screen; DOCT, Davis Observational Checklist for Texas; DP-II, Developmental Profile II; ELFRA-2, Elternfragebogen für die Fruberkennung von Riskokindern; ELS, Early Language Scale; ESP, Early Screening Profiles; FPSLST, Fluharty Preschool Speech and Language Screening Test; FPST, Fluharty Preschool Screening Test; GFTA, Goldman-Fristoe Test of Articulation; GLS, General Language Screen; HAT, Henja Articulation Test; HELST, Hackney Early Language Screening Test; HKCAT, Hong Kong Cantonese Articulation Test; ICS-TC, Intelligibility in Context Scale–Traditional Chinese; ITC, Infant-Toddler Checklist; ITPA, Illinois Test of Psycholinguistic Abilities; LDS, Language Development Survey; LLC, Lexilist Comprehension; LLP, Lexilist Production; LR+, positive likelihood ratio; LR−, negative likelihood ratio; LS, Language Standard; MLU, mean length of utterance; MSCA, McCarthy Scales of Children's Abilities; MSEL, Mullen Scales of Early Learning; NPV, negative predictive value; NR, not reported; NSST, Northwestern Syntax Screening Test; PLS-4-C, Preschool Language Scale, Fourth Edition–Comprehension, PLS-4-E, Preschool Language Scale, Fourth Edition–Expression; PLS-5, Preschool Language Scale, Fifth Edition; PPV, positive predictive value; QUILS, Quick Interactive Language Screener; RDLS, Reynell Developmental Language Scales; ROC, receiver operating characteristic; SETK-2, Sprachentwicklungstest für zweijahrige Kinder; SETK-3, Sprachentwicklungstest für zweijahrige Kinder; SICD, Sequenced Inventory of Communication Development; SKOLD, Screening Kit of Language Development; SKOLDBE, Screening Kit of Language Development Black English; SLC, Schlichting Tests for Language Comprehension; SLP, speech-language pathologist; SPES-3, Sprachentwicklungsscreening; SSP, Schlichting Tests for Sentence Production; SRST, Sentence Repetition Screening Test; SSLM, Sure Start Language Measure; SST, Structured Screening Test; SWP, Schlichting Tests for Word Production; TACL-R, Test for Auditory Comprehension of Language–Revised; TD, Templin-Darley Tests of Articulation Consonant Singles Subtest; TOLD-P, Test of Language Development Primary.

[a] Calculated by the Evidence-based Practice Center.

[b] Optimal cut point using Youden index.

[c] Prevalence not reported for this subsample. Median for sensitivity/specificity includes full sample only and not the English-speaking subsample.

[d] Prevalence for screen failures more than 1.5 SD below the mean is 18%; study calculated accuracy using this value as well as prevalence using cut point of more than 2 SDs below the mean, which was 6%. Data were included for only the former prevalence.

[e] Sample size and prevalence based on imputed sample, which corrected for oversampling of children with positive screening results.

[f] Prevalence data provided by study authors.

[g] Includes 11 children who were noncooperative during screening.

[h] The study investigators weighted the ns based on a stratified sample of 69.

[i] Only the BDIST-CD Receptive Scale is included in accuracy analyses.

Two RCTs assessed fluency treatment for young children. Both focused on the Lidcombe Program of Early Stuttering Intervention.[54,55] This intervention is led by an SLP who trains parents to provide verbal contingencies for stutter-free speech (eg, "that was smooth talking") and stuttering (eg, "that was a bit bumpy") and requests for self-evaluation and self-correction (eg, "can you say that again"). In one of these RCTs, the intervention was delivered in a face-to-face format in a clinical setting[54] and in the other it was delivered via telehealth.[55] Results were consistent in showing a statistically significant improvement in stuttering fluency associated with the intervention. In the face-to-face intervention, children in the intervention group had a 2.3% (95% CI, 0.8-3.9) lower proportion of syllables stuttered than children in the control group at 9 months. Per the authors, this is above the minimum clinically important difference of 1.0% of syllables stuttered (the minimum difference that a listener would be able to distinguish).[54] However, no reference or clear rationale was provided to support this threshold. In the RCT using telehealth delivery of the intervention, the difference between the intervention and control group in change from baseline mean number of syllables stuttered was −3.0% (P = .02) at 9 months.[55]

Evidence on other intervention types targeting specific speech or language problems was limited and is further described in the eResults in the Supplement.

Key Question 5. Do interventions for speech and language delay or disorders in children age 6 years or younger improve school performance, function, or quality-of-life outcomes?

Eight RCTs reported on 1 or more outcomes specific to school performance, function, or quality of life using heterogeneous measures.[42,43,47,48,53,57-59] Characteristics are described above in KQ4 and detailed results are shown in eTable 15 in the Supplement. No RCTs assessing a similar intervention type reported on the same outcome domain, and most studies reporting on similar domains (eg, early literacy) used different outcome measures. In 4 RCTs reporting on a measure of early or emergent literacy skills, 3 found no significant difference between groups.[42,43,48] In contrast, 1 RCT assessing a home-based language delay intervention delivered by trained assistants found benefit for improving letter knowledge associated with the intervention.[59] Two RCTs reported on 1 or more measures of functional communication[42,47] and quality of life/well-being in children[43,53] and found no difference between groups, while 1 RCT evaluating an individual intervention for language delay found significant improvement favoring the intervention for improving child socialization skills and parental stress levels.[57]

Harms of Treatment

Key Question 6. What are the harms of interventions for speech and language delay or disorders?

No eligible study addressed this question.

Discussion

This systematic review synthesized evidence relevant to screening for speech and language delay or disorders in children 5 years or younger. **Table 4** summarizes the main findings of the evidence review. There was no direct evidence on the benefits and harms of screening (KQ1). Potential harms of screening (KQ3) include false-positive results that can lead to unnecessary referrals (and the associated time and economic burden), labeling or stigma, parent anxiety, and other psychosocial harms. Other harms of screening are likely to be minimal because screening is noninvasive.

The studies of screening test accuracy (KQ2) included in this review assessed 23 different tools that varied in terms of whether they were completed by parents vs trained examiners and whether they were designed to detect global speech or language problems vs problems related to specific language skills or articulation. Some screening tools usable in clinical practice may identify children who have a speech or language disorder with reasonable sensitivity and specificity. However, overall evidence was mixed and few screening tools were assessed by more than 1 study each, limiting the ability to make stronger conclusions about the accuracy of specific tools. Parent-reported screening instruments designed to assess expressive language skills displayed consistently high sensitivity and specificity, although precision varied by instrument. In contrast, the accuracy of the parent-reported instruments for global language skill assessment was inconsistent, and precision varied across instruments. The accuracy of examiner-administered screening instruments varied, particularly for instruments designed to assess specific language skills.

Few studies of interventions for speech and language delay or disorder enrolled similar populations and evaluated similar types of interventions (KQ4). Although 2 RCTs of treatment enrolled children newly referred from primary care, it is not clear whether the children were identified via routine screening vs case finding. Other included studies enrolled children referred or recruited via advertisements, and most focused on a specific type of speech delay or disorder. Given these factors, the body of evidence on treatment available for inclusion in this review may not be applicable to the type and severity of disorders that would be detected via routine screening in primary care settings.

Studies of children referred for language delay without obvious speech-sound or fluency disorder suggested that group training interventions offering at least 11 parent training sessions improved expressive language outcomes. For children identified with stuttering, the Lidcombe Program of Early Stuttering Intervention delivered by SLPs improved stuttering fluency at 9 months when delivered either in person or via telehealth. Although 8 RCTs reported on 1 or more outcomes specific to school performance or early literacy, health-related quality of life, function, behavior, or socialization (KQ5), the interventions and populations evaluated were heterogeneous, which limited the ability to assess consistency; most studies found no difference between groups for measures of early literacy, function, and quality of life. However, most trials may not have followed up children for a long enough duration to detect an improvement in quality of life or function that could result from early treatment of a speech and language delay or disorder. No RCTs reported on the harms of interventions; however, given the nature of the interventions, serious harms are unlikely.

Trials are needed that enroll asymptomatic or unselected populations from general primary care settings and directly assess the benefit of screening specifically for speech and language problems. The control groups in these trials could receive either no screening or routine screening for general developmental delay, with no separate score for speech and language problems. Studies are also needed on the potential harms of screening, such as labeling, and harms from

Table 4. Summary of Evidence for Screening and Treatment of Speech and Language Delay and Disorders in Children 5 Years or Younger

No. of studies (No. of participants)	Summary of findings	Consistency and precision	Study quality	Limitations (including reporting bias)	Overall strength of evidence	Applicability
KQ1: Benefits of screening						
No eligible study identified	NA	NA	NA	NA	Insufficient	NA
KQ2: Accuracy of screening						
Parent-reported global language 6 Studies (n = 1941)[23,35,36,39-41]	Sensitivity: median, 74% (range, 55%-89%) Specificity: median, 79% (range, 73%-95%) The Infant-Toddler Checklist had the highest sensitivity (89% and 86%) for each of its 2 age groups The ELS and the ASQ with toddlers had the highest specificity (93% and 95%, respectively)	Mostly consistent and imprecise (for both sensitivity and specificity)	1 Good 5 Fair	Only 1 instrument (ASQ) was included in more than 1 study Reference measures differed across studies One study included all screen failures and a random sample of those who passed Not all studies indicated criteria for screen failure Studies had a wide age range	Low	North American and European parents of infants, toddlers, and preschool children
Parent-reported specific language skills 4 Studies (n = 3245)[24,33,34,41]	Sensitivity: median, 91% (range, 83%-93%) Specificity: 88% (range, 81%-96%) The LDS (revised scoring) displayed a large LR+ and a large LR-; the ELFRA-2 had a large LR-	Sensitivity: fairly consistent, imprecise Specificity: fairly consistent (varies by instrument); the SPES-3 is precise	1 Good 3 Fair	Different reference measures used Small sample size in 1 study Three of the studies included all screen failures and a random sample of those who passed	Moderate	American and European parents of 2- and 3-y-old children
Parent-reported articulation 1 Study (n = 780)[27]	Sensitivity: 86% Specificity: 32%	Sensitivity: unknown consistency, imprecise Specificity: unknown consistency, precise	1 Fair	There was only 1 study of Chinese children Studies had a wide age range May only be appropriate for 4-y-old children	Insufficient	Although the study included parents of children who were speakers of traditional Chinese in Hong Kong and was applicable for them, the instrument would not be applicable to English-speaking children
Examiner-reported global language 10 Studies (n = 2287)[19-21,28-31,35,37,38]	Sensitivity: median, 88% (range, 17%-100%) Specificity: median, 89% (range, 69%-98%)	Mostly consistent, with some instruments showing high (>90%) sensitivity and/or specificity and others showing low or moderate values Precision is inconsistent, varying by instrument; the HELST and SKOLD are precise for sensitivity; the DOCT, SST, 2 of the 3 age levels of the SKOLD, and the SRST are precise for specificity	2 Good 8 Fair	Three instruments were examined in 1 study each; 3 instruments were examined in 2 studies The reference measure varied Criteria for screening failure was not always indicated	Low	Children seen in medical practices in the UK, Sweden, and Australia and in schools in the US One instrument was used with bilingual children
Examiner-reported specific language 3 Studies (n = 418)[20,23,32,a]	Sensitivity: median, 86% (range, 56%-94%) Specificity: median, 70% (range, 58%-90%)	Unclear; both sensitivity and specificity are inconsistent and imprecise; however, tools assess different types of language problems across heterogeneous populations	3 Fair	1 study included 3 instruments, accounting for 5 of the 7 accuracy indices	Insufficient	Children at risk for developmental delays in Canada and childcare centers in the US

(continued)

Table 4. Summary of Evidence for Screening and Treatment of Speech and Language Delay and Disorders in Children 5 Years or Younger (continued)

No. of studies (No. of participants)	Summary of findings	Consistency and precision	Study quality	Limitations (including reporting bias)	Overall strength of evidence	Applicability
Examiner-reported articulation 3 Studies (n = 673)[22,37,38]	Sensitivity: median, 66% (range, 43%-92%) Specificity: median, 96% (range, 93%-97%)	Sensitivity: inconsistent Specificity: consistent Precision unknown (2 studies do not report CIs)	3 Fair	Studies had a wide age range	Low	Children in schools in the US
KQ3: Harms of screening						
No eligible study identified	NA	NA	NA	NA	Insufficient	NA
KQ4: Speech and language outcomes of intervention						
Language delay (parent-delivered) 4 RCTs (n = 378)[45,50,51,58]	Parent-delivered, group training interventions: 2 RCTs assessing interventions delivered over a longer duration (11 bimonthly 60- to 75-min sessions[50] and 11 weekly 2.5-hour sessions plus 3 weekly home visits[51]) found benefit in expressive language outcomes; 1 shorter intervention (6 weekly 2-hour sessions) found no significant difference between groups[58] One RCT of individual home-based parental training intervention found mixed results	Parent-delivered, group training interventions: inconsistent; precise Individual home-based parent training: unknown consistency; imprecise	2 Good 2 Fair	Studies of parental group training differed in duration, intensity, content, and timing of outcome assessment	Parent-delivered, group training interventions: Low Parent-delivered individual training: Insufficient	Parent-group-based training trials that showed benefit enrolled children and parents in the 1990s, results may not be applicable to current practice
Language delay (SLP- or trained staff-delivered) 4 RCTs (n = 270)[44,56,57,59]	One RCT enrolling toddlers (mean age, 21-30 mo) found benefit associated with an individual intervention delivered by an SLP over 12 weeks on multiple measures of expressive language;[57] 3 other RCTs assessing different interventions among older children (mean age, 49.5-59.6 mo) found inconsistent results[44,56,59]	Unknown consistency; mostly imprecise	4 Fair	All studies focused on children with language delay and interventions delivered by an SLP or trained staff; however, populations, settings, and outcome measures were heterogeneous	Insufficient	Children with language delay, who were identified via referrals or advertisements
School-based (tier 1) interventions 2 Cluster RCTs (n = 339)[46,48]	Both found improved receptive and expressive language outcomes associated with the intervention over 52 wk; however, 1 found benefit in some measures (receptive and expressive 1-word picture vocabulary tests focused on vocabulary) but not others (no improvement on standardized measures of oral language)[48]	Mostly consistent; imprecise	2 Fair	One RCT reported only *F* statistics from ANOVA analyses and *P* values, limiting the ability to determine the magnitude of effect; 1 RCT found benefit in some measures of oral language and literacy but not others	Low	Unclear applicability to current preschool curricula in the US; 1 study was set in Spain and 1 in the US
Community-based speech and language disorders 2 RCTs (n = 260 participants)[42,53]	Studies found mixed results with improvement on some domains of speech and language but not others, and no consistent benefit on similar measures or outcome domains	Inconsistent; imprecise	1 Good 1 Fair	Studies both focus on children newly referred from primary care for any speech and language disorder, but differ in country setting (UK and Australia), mean age of enrolled children (34 vs 53 mo), and outcome measures reported	Insufficient	Children newly referred from primary care to existing community-based treatment for speech and language problems in the UK and Australia

(continued)

Table 4. Summary of Evidence for Screening and Treatment of Speech and Language Delay and Disorders in Children 5 Years or Younger (continued)

No. of studies (No. of participants)	Summary of findings	Consistency and precision	Study quality	Limitations (including reporting bias)	Overall strength of evidence	Applicability
Fluency disorders (Lidcombe Program of Early Stuttering Intervention) 2 RCTs (n = 76)[54,55]	Both RCTs found benefit for stuttering fluency associated with the intervention at 9 mo; 1 found a 2.3% reduction in the percentage of syllables stuttered among the intervention vs control group, and the second found the mean number of syllables in the intervention group was significantly lower than in the control group (−3.0; *P* = .02)	Consistent; precise	2 Fair	One RCT delivered the intervention via face-to-face visits, and 1 delivered the intervention via telehealth	Moderate	Children aged 42-56 mo identified with stuttering
Speech-sound disorders 3 RCTs (n = 194)[43,47,49]	One RCT enrolling children with a severe phonological disorder but normal receptive language function found improvement associated with an individual SLP intervention at 16 wk for multiple speech and sound outcomes; 1 RCT assessing an intervention for children with speech motor delay found mixed results; 1 RCT assessing a software-based intervention set in schools for children identified with a speech-sound disorder found no improvement on measures of speech production and speech intelligibility	Unknown; imprecise	3 Fair	Studies focus on children with different types of speech-sound disorders and assess different interventions	Insufficient	Unclear; RCTs are set in different countries and enroll heterogeneous populations of children who differ in age, spoken language, and type of speech disorder
KQ5: Health outcomes of intervention (school performance, function, or quality-of-life outcomes)						
8 RCTs (n = 1239) reported on ≥1 outcomes specific to school performance (or early literacy), function, and QOL[42,43,47,48,53,57-59]	No 2 studies assessing a similar intervention type reported on the same outcome domain; in 4 RCTs assessing a measure of early literacy, 3 found no significant difference between groups and 1 RCT assessing a home-based language-delay intervention delivered by trained assistants found benefit for improving letter knowledge associated with the intervention[59] No study reported benefit for improving function or QOL; 1 individual intervention for language delay found significant improvement favoring the intervention for improving socialization and parental stress level[57]	Unknown; imprecise	2 Good 6 Fair	No 2 studies assessing the same type of intervention reported on a similar outcome measure, limiting the ability to assess consistency of findings	Insufficient	Unclear; RCTs are set in different countries and assess different outcomes among different groups of children, who vary in terms of setting and type of speech and language disorder
KQ6: Harms of intervention						
No eligible study identified	NA	NA	NA	NA	Insufficient	NA

Abbreviations: ANOVA, analysis of variance; ASQ, Ages and Stages Questionnaire; DOCT, Davis Observational Checklist for Texas; ELFRA-2, Elternfragebogen für die Fruberkennung von Riskokindern; ELS, Early Language Scale; HELST, Hackney Early Language Screening Test; KQ, key question; LDS, Language Development Survey; LR−, negative likelihood ratio; LR+, positive likelihood ratio; NA, not applicable; QOL, quality of life; RCT, randomized clinical trial; SKOLD, Screening Kit of Language Development; SLP, speech-language pathologist; SPES-3, Sprachentwicklungsscreening; SRST, Sentence Repetition Screening Test; SST, Structured Screening Test.

[a] Frisk et al[23] examined 3 instruments and included separate accuracy calculations for the expressive and receptive PLS-4 reference measure. Accuracy outcomes were omitted for the Battelle Developmental Inventory Screening Test with the PLS-4 Expressive Communication Scale due to a possible reporting error in the study.

false-positive results, such as burden on parents due to unnecessary referrals. Such studies would also inform the potential for overdiagnosis associated with routine screening, given that many children who have a speech delay may recover without intervention.[3]

Similarly, studies assessing the accuracy of screening tools among unselected populations, who are ideally recruited through primary care settings, are needed because the prevalence of speech and language problems may vary compared with populations enrolled via advertisements or specialty settings. Specifically, studies that assess the accuracy of existing tools, compared with similar reference standards, would help determine the consistency of findings; because few included studies evaluated the same instrument, our ability to make a strong conclusion about accuracy was limited. Trials of treatment enrolling populations recruited from US primary care settings would help inform the potential benefit of screening because the range of severity and conditions is likely different compared with trials that enroll referred populations. Last, studies that follow up children for a sufficiently long duration to detect improvement in academic performance, function, and quality of life would help in the understanding of whether immediate changes in speech and language outcomes (eg, short-term expansion of vocabulary words) translate into benefit for health and social outcomes.

Limitations

This review excluded studies in children who had a condition known to cause a speech or language problem (eg, hearing loss, autism) to improve the applicability of evidence to populations likely to be detected by routine screening. Studies evaluating primary prevention strategies to promote speech and language development (eg, interventions among groups considered "at risk" or school-based curricula emphasizing language development among children with no developmental delay or disorder) were also excluded. The aim was to limit the review to interventions that are relevant to children with screen-detected speech and language problems and that are appropriate to deliver in primary care settings or refer to from primary care.

Conclusion

This review found no eligible studies that reported on direct benefits or harms of screening compared with usual care or no screening. Parent-reported screening tools for expressive language delay had reasonable accuracy. In contrast, parent-reported screening tools for global language delay had inconsistent accuracy. The accuracy of examiner-administered instruments was also variable, especially for examiner-administered instruments of specific language skills. Existing evidence on treatment of speech and language delay is available from referral populations but not from screen-detected populations. This evidence indicates the benefit from group parent-training programs for speech delay that provide at least 11 parental training sessions for improving expressive language skills, as well as the Lidcombe Program of Early Stuttering Intervention delivered by SLPs for reducing stuttering frequency. Few studies reported on outcomes specific to school performance, function, quality of life, or behavior, and none reported on the harms of interventions.

ARTICLE INFORMATION

Accepted for Publication: December 2, 2023.

Author Affiliations: RTI International–University of North Carolina at Chapel Hill Evidence-based Practice Center (Feltner, Wallace, Middleton, Vaughan, Baker, Kahwati); Department of Medicine, University of North Carolina at Chapel Hill (Feltner); Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill (Feltner, Orr, Middleton, Vaughan, Baker); RTI International, Research Triangle Park, North Carolina (Wallace, Kahwati); Frank Porter Graham Child Development Institute, University of North Carolina at Chapel Hill (Nowell); Department of Pediatrics, University of North Carolina at Chapel Hill (Orr, Raffa); The Pacific Northwest Evidence-Based Practice Center, Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University (Chou); School of Medicine, Division of General Internal Medicine, Oregon Health & Science University (Chou).

Author Contributions: Dr. Feltner had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Feltner, Wallace, Nowell, Raffa, Kahwati.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Feltner, Wallace, Nowell, Raffa, Middleton, Vaughan, Baker, Kahwati.
Critical review of the manuscript for important intellectual content: Feltner, Wallace, Nowell, Orr, Chou, Kahwati.
Statistical analysis: Feltner, Orr.
Obtained funding: Feltner, Chou, Kahwati.
Administrative, technical, or material support: Orr, Middleton, Vaughan, Baker.
Supervision: Feltner.

Conflict of Interest Disclosures: None reported.

Funding/Support: This research was funded under contract 75Q80120D00006, Task Order 75Q80121F32009, from the Agency for Healthcare Research and Quality (AHRQ), US Department of Health and Human Services, under a contract to support the US Preventive Services Task Force (USPSTF).

Role of the Funder/Sponsor: Investigators worked with USPSTF members and AHRQ staff to develop the scope, analytic framework, and key questions for this review. AHRQ had no role in study selection, quality assessment, or synthesis. AHRQ staff provided project oversight, reviewed the evidence review to ensure that the analysis met methodological standards, and distributed the draft for peer review. Otherwise, AHRQ had no role in the conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript findings. The opinions expressed in this document are those of the authors and do not reflect the official position of AHRQ or the US Department of Health and Human Services.

Additional Contributions: We gratefully acknowledge the following individuals for their contributions to this project, including AHRQ staff (Justin Mills, MD, MPH; Tracy Wolff, MD, MPH), Scientific Resource Center for the AHRQ Evidence-based Practice Center Program staff (Robin A. Paynter, MLIS), Pacific Northwest Evidence-based Practice Center staff (Christina Bougatsos, MPH), and RTI International–University of North Carolina–Chapel Hill Evidence-based Practice Center staff (Manny Schwimmer, MPH; Christiane E. Voisin, MSLS; Roberta Wines, MPH; Mary Gendron; Sharon Barrell, MA; Alexander Cone; Teyonna Downing; Michelle Bogus). The USPSTF members, expert reviewers, and federal partner reviewers did not receive financial compensation for their contributions. Evidence-based Practice Center personnel received compensation for their roles in this project.

Additional Information: A draft version of the full evidence review underwent external peer review from 3 content experts (Abigail D. Delehanty, PhD, CCC-SLP, Duquesne University; Virginia Moyer, MD, MPH, University of North Carolina at Chapel Hill; Thelma E. Uzonyi, PhD, CCC-SLP, IMH-E, Kennedy Krieger Institute) and 3 federal partner reviewers (Centers for Disease Control and Prevention; Eunice Kennedy Shriver National Institute of Child Health and Human Development; and National Institute on Deafness and Other Communication Disorders). Comments from reviewers were presented to the USPSTF during its deliberation of the evidence and were considered in preparing the final evidence review. USPSTF members and peer reviewers did not receive financial compensation for their contributions.

Editorial Disclaimer: This evidence review is presented as a document in support of the

accompanying USPSTF Recommendation Statement. It did not undergo additional peer review after submission to *JAMA*.

REFERENCES

1. Black LI, Vahratian A, Hoffman HJ. Communication disorders and use of intervention services among children aged 3-17 years: United States, 2012. *NCHS Data Brief*. 2015;(205):1-8.

2. Centers for Disease Control and Prevention. Language and speech disorders in children. Published 2021. Accessed January 24, 2023 https://www.cdc.gov/ncbddd/developmentaldisabilities/language-disorders.html

3. Law J, Boyle J, Harris F, Harkness A, Nye C. Prevalence and natural history of primary speech and language delay: findings from a systematic review of the literature. *Int J Lang Commun Disord*. 2000;35(2):165-188. doi:10.1111/j.1460-6984.2000.tb00001.x

4. Lewis BA, Freebairn L, Tag J, et al. Adolescent outcomes of children with early speech sound disorders with and without language impairment. *Am J Speech Lang Pathol*. 2015;24(2):150-163. doi:10.1044/2014_AJSLP-14-0075

5. Catts HW, Bridges MS, Little TD, Tomblin JB. Reading achievement growth in children with language impairments. *J Speech Lang Hear Res*. 2008;51(6):1569-1579. doi:10.1044/1092-4388(2008/07-0259)

6. Conti-Ramsden G, Mok PL, Pickles A, Durkin K. Adolescents with a history of specific language impairment (SLI): strengths and difficulties in social, emotional and behavioral functioning. *Res Dev Disabil*. 2013;34(11):4161-4169. doi:10.1016/j.ridd.2013.08.043

7. Glogowska M, Roulstone S, Peters TJ, Enderby P. Early speech- and language-impaired children: linguistic, literacy, and social outcomes. *Dev Med Child Neurol*. 2006;48(6):489-494. doi:10.1017/S0012162206001046

8. Dubois P, St-Pierre MC, Desmarais C, Guay F. Young adults with developmental language disorder: a systematic review of education, employment, and independent living outcomes. *J Speech Lang Hear Res*. 2020;63(11):3786-3800. doi:10.1044/2020_JSLHR-20-00127

9. Schoon I, Parsons S, Rush R, Law J. Children's language ability and psychosocial development: a 29-year follow-up study. *Pediatrics*. 2010;126(1):e73-e80. doi:10.1542/peds.2009-3282

10. Lipkin PH, Macias MM; Council on Children With Disabilities, Section on Developmental and Behavioral Pediatrics. Promoting optimal development: identifying infants and young children with developmental disorders through developmental surveillance and screening. *Pediatrics*. 2020;145(1):e20193449. doi:10.1542/peds.2019-3449

11. Siu AL; US Preventive Services Task Force. Screening for speech and language delay and disorders in children aged 5 years or younger: US Preventive Services Task Force recommendation statement. *Pediatrics*. 2015;136(2):e474-e481. doi:10.1542/peds.2015-1711

12. Feltner C, Wallace IF, Nowell S, et al. Screening for Speech and Language Delays and Disorders in Children Age 5 Years or Younger: An Evidence Review for the US Preventive Services Task Force. Evidence Synthesis No. 234. Agency for Healthcare Research and Quality; 2023. AHRQ publication 23-05306-EF-1.

13. United Nations Development Programme. Human Development Report 2020: the next frontier: human development and the Anthropocene. Published 2020. Accessed January 24, 2023 http://report2020.archive.s3-website-us-east-1.amazonaws.com/

14. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898. doi:10.1136/bmj.l4898

15. Whiting PF, Rutjes AW, Westwood ME, et al; QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536. doi:10.7326/0003-4819-155-8-201110180-00009

16. US Preventive Services Task Force. US Preventive Services Task Force Procedure Manual. Published May 2021. Accessed January 24, 2023 https://www.uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes/procedure-manual

17. Agency for Healthcare Research and Quality Effective Health Care Program. Methods guide for effectiveness and comparative effectiveness reviews. Published content last reviewed October 2022. Accessed January 24, 2023 https://effectivehealthcare.ahrq.gov/products/collections/cer-methods-guide

18. West SL, Gartlehner G, Mansfield AJ, et al. *Comparative Effectiveness Review Methods: Clinical Heterogeneity*. Agency for Healthcare Research and Quality; 2010.

19. Alberts FM, Davis BL, Prentice L. Validity of an observation screening instrument in a multicultural population. *J Early Interv*. 1995;19(2):168-177. doi:10.1177/105381519501900209

20. Allen DV, Bliss LS. Concurrent validity of two language screening tests. *J Commun Disord*. 1987;20(4):305-317. doi:10.1016/0021-9924(87)90012-8

21. Bliss LS, Allen DV. Screening Kit of Language Development: a preschool language screening instrument. *J Commun Disord*. 1984;17(2):133-141. doi:10.1016/0021-9924(84)90019-4

22. Drumwright A, Van Natta P, Camp B, Frankenburg W, Drexler H. The Denver articulation screening exam. *J Speech Hear Disord*. 1973;38(1):3-14. doi:10.1044/jshd.3801.03

23. Frisk V, Montgomery L, Boychyn E, et al. Why screening Canadian preschoolers for language delays is more difficult than it should be. *Infants Young Child*. 2009;22(4):290-308. doi:10.1097/IYC.0b013e3181bc4db6

24. Holzinger D, Weber C, Barbaresi W, Beitel C, Fellinger J. Language screening in 3-year-olds: development and validation of a feasible and effective instrument for pediatric primary care. *Front Pediatr*. 2021;9:752141. doi:10.3389/fped.2021.752141

25. Klee T, Carson DK, Gavin WJ, Hall L, Kent A, Reece S. Concurrent and predictive validity of an early language screening program. *J Speech Lang Hear Res*. 1998;41(3):627-641. doi:10.1044/jslhr.4103.627

26. Klee T, Pearce K, Carson DK. Improving the positive predictive value of screening for developmental language disorder. *J Speech Lang Hear Res*. 2000;43(4):821-833. doi:10.1044/jslhr.4304.821

27. Kok ECE, To CKS. Revisiting the cutoff criteria of Intelligibility in Context Scale–Traditional Chinese. *Lang Speech Hear Serv Sch*. 2019;50(4):629-638. doi:10.1044/2019_LSHSS-18-0073

28. Laing GJ, Law J, Levin A, Logan S. Evaluation of a structured test and a parent-led method for screening for speech and language problems: prospective population based study. *BMJ*. 2002;325(7373):1152. doi:10.1136/bmj.325.7373.1152

29. Law J. Early language screening in city and Hackney: the concurrent validity of a measure designed for use with 2 1/2-year-olds. *Child Care Health Dev*. 1994;20(5):295-308. doi:10.1111/j.1365-2214.1994.tb00392.x

30. Nayeb L, Lagerberg D, Westerlund M, Sarkadi A, Lucas S, Eriksson M. Modifying a language screening tool for three-year-old children identified severe language disorders six months earlier. *Acta Paediatr*. 2019;108(9):1642-1648. doi:10.1111/apa.14790

31. Nayeb L, Lagerberg D, Sarkadi A, Salameh EK, Eriksson M. Identifying language disorder in bilingual children aged 2.5 years requires screening in both languages. *Acta Paediatr*. 2021;110(1):265-272. doi:10.1111/apa.15343

32. Pace A, Curran M, Van Horne AO, et al. Classification accuracy of the Quick Interactive Language Screener for preschool children with and without developmental language disorder. *J Commun Disord*. 2022;100:106276. doi:10.1016/j.jcomdis.2022.106276

33. Sachse S, Von Suchodoletz W. Early identification of language delay by direct language assessment or parent report? *J Dev Behav Pediatr*. 2008;29(1):34-41. doi:10.1097/DBP.0b013e318146902a

34. Sachse S, von Suchodoletz W. Response to reader comments on Early identification of language delay by direct language assessment or parent report?. *J Dev Behav Pediatr*. 2009;30(2):176. doi:10.1097/DBP.0b013e31819f1c9f

35. Stokes SF. Secondary prevention of paediatric language disability: a comparison of parents and nurses as screening agents. *Eur J Disord Commun*. 1997;32(2 spec No):139-158. doi:10.1111/j.1460-6984.1997.tb01628.x

36. Stott CM, Merricks MJ, Bolton PF, Goodyer IM. Screening for speech and language disorders: the reliability, validity and accuracy of the General Language Screen. *Int J Lang Commun Disord*. 2002;37(2):133-151. doi:10.1080/13682820110116785

37. Sturner RA, Heller JH, Funk SG, Layton TL. The Fluharty Preschool Speech and Language Screening Test: a population-based validation study using sample-independent decision rules. *J Speech Hear Res*. 1993;36(4):738-745. doi:10.1044/jshr.3604.738

38. Sturner RA, Funk SG, Green JA. Preschool speech and language screening: further validation of the sentence repetition screening test. *J Dev Behav Pediatr*. 1996;17(6):405-413. doi:10.1097/00004703-199612000-00006

39. Visser-Bochane MI, van der Schans CP, Krijnen WP, Reijneveld SA, Luinge MR. Validation of the Early Language Scale. *Eur J Pediatr*. 2021;180(1):63-71. doi:10.1007/s00431-020-03702-8

40. Wetherby AM, Goldstein H, Cleary J, Allen L, Kublin K. Early identification of children with communication disorders. *Infants Young Child*. 2003;16(2):161-174. doi:10.1097/00001163-200304000-00008

41. Wilson P, Rush R, Charlton J, Gilroy V, McKean C, Law J. Universal language development screening: comparative performance of two questionnaires. *BMJ Paediatr Open*. 2022;6(1):e001324. doi:10.1136/bmjpo-2021-001324

42. McLeod S, Davis E, Rohr K, et al. Waiting for speech-language pathology services: a randomised controlled trial comparing therapy, advice and device. *Int J Speech Lang Pathol*. 2020;22(3):372-386. doi:10.1080/17549507.2020.1731600

43. McLeod S, Baker E, McCormack J, et al. Cluster-randomized controlled trial evaluating the effectiveness of computer-assisted intervention delivered by educators for children with speech sound disorders. *J Speech Lang Hear Res*. 2017;60(7):1891-1910. doi:10.1044/2017_JSLHR-S-16-0385

44. Thordardottir E, Cloutier G, Ménard S, Pelland-Blais E, Rvachew S. Monolingual or bilingual intervention for primary language impairment? a randomized control trial. *J Speech Lang Hear Res*. 2015;58(2):287-300. doi:10.1044/2014_JSLHR-L-13-0277

45. Peredo TN, Mancilla-Martinez J, Durkin K, Kaiser A. Teaching Spanish-speaking caregivers to implement *EMT en Español*: a small randomized trial. *Early Child Res Q*. 2022;58:208-219. doi:10.1016/j.ecresq.2021.08.004

46. Acosta-Rodríguez VM, Ramírez-Santana GM, Hernández-Expósito S. Intervention for oral language comprehension skills in preschoolers with developmental language disorder. *Int J Lang Commun Disord*. 2022;57(1):90-102. doi:10.1111/1460-6984.12676

47. Namasivayam AK, Huynh A, Granata F, Law V, van Lieshout P. PROMPT intervention for children with severe speech motor delay: a randomized control trial. *Pediatr Res*. 2021;89(3):613-621. doi:10.1038/s41390-020-0924-4

48. Wilcox MJ, Gray S, Reiser M. Preschoolers with developmental speech and/or language impairment: efficacy of the Teaching Early Literacy and Language (TELL) curriculum. *Early Child Res Q*. 2020;51:124-143. doi:10.1016/j.ecresq.2019.10.005

49. Almost D, Rosenbaum P. Effectiveness of speech intervention for phonological disorders: a randomized controlled trial. *Dev Med Child Neurol*. 1998;40(5):319-325. doi:10.1111/j.1469-8749.1998.tb15383.x

50. Gibbard D. Parental-based intervention with pre-school language-delayed children. *Eur J Disord Commun*. 1994;29(2):131-150. doi:10.3109/13682829409041488

51. Girolametto L, Pearce PS, Weitzman E. Interactive focused stimulation for toddlers with expressive vocabulary delays. *J Speech Hear Res*. 1996;39(6):1274-1283. doi:10.1044/jshr.3906.1274

52. Girolametto L, Pearce PS, Weitzman E. Effects of lexical intervention on the phonology of late talkers. *J Speech Lang Hear Res*. 1997;40(2):338-348. doi:10.1044/jslhr.4002.338

53. Glogowska M, Roulstone S, Enderby P, Peters TJ. Randomised controlled trial of community based speech and language therapy in preschool children. *BMJ*. 2000;321(7266):923-926. doi:10.1136/bmj.321.7266.923

54. Jones M, Onslow M, Packman A, et al. Randomised controlled trial of the Lidcombe Programme of Early Stuttering Intervention. *BMJ*. 2005;331(7518):659. doi:10.1136/bmj.38520.451840.EO

55. Lewis C, Packman A, Onslow M, Simpson JM, Jones M. A phase II trial of telehealth delivery of the Lidcombe Program of Early Stuttering Intervention. *Am J Speech Lang Pathol*. 2008;17(2):139-149. doi:10.1044/1058-0360(2008/014)

56. Robertson SB, Ellis Weismer S. The influence of peer models on the play scripts of children with specific language impairment. *J Speech Lang Hear Res*. 1997;40(1):49-61. doi:10.1044/jslhr.4001.49

57. Robertson SB, Ellis Weismer S. Effects of treatment on linguistic and social skills in toddlers with delayed language development. *J Speech Lang Hear Res*. 1999;42(5):1234-1248. doi:10.1044/jslhr.4205.1234

58. Wake M, Tobin S, Girolametto L, et al. Outcomes of population based language promotion for slow to talk toddlers at ages 2 and 3 years: Let's Learn Language cluster randomised controlled trial. *BMJ*. 2011;343:d4741. doi:10.1136/bmj.d4741

59. Wake M, Tobin S, Levickis P, et al. Randomized trial of a population-based, home-delivered intervention for preschool language delay. *Pediatrics*. 2013;132(4):e895-e904. doi:10.1542/peds.2012-3878