

RAPPORT

2024

FORSKNINGSKARTLEGGING

Verktøy for språkkartlegging av barn 0-5 år

En systematisk kartleggingsoversikt

Utgitt av Folkehelseinstituttet
Område for helsetjenester

Tittel Verktøy for språkkartlegging av barn 0-5 år: En systematisk kartleggingsoversikt

English title Language screening tools for children 0-5 years: a systematic scoping review

Ansvarlig Gun Peggy Strømstad Knudsen, konstituert direktør

Forfattere Patricia Sofia Jardim Jacobsen, prosjektleder,
Tiril Cecilie Borge, Forsker
Ingvild Kirkehei, Bibliotekar

ISBN 978-82-8406-425-3

Publikasjonstype Forskningskartlegging

Antall sider 74 (127 inklusiv vedlegg)

Oppdragsgiver Helsedirektoratet

Emneord(MeSH) Mass screening; early diagnosis; routine diagnostic tests; language disorders; child; psychometrics

Sitering Jacobsen PSJ, Borge TC og Kirkehei I. Verktøy for språkkartlegging av barn 0-5 år: en systematisk kartleggingsoversikt. Language screening tools for children 0-5 years: a systematic scoping review–2024. Oslo: Folkehelseinstituttet, 2024.

Innhold

INNHold	3
HOVEDBUdSKAP	5
SAMMENDRAG	6
KEY MESSAGES	9
EXECUTIVE SUMMARY (ENGLISH)	10
FORORD	13
FORKORTELSER	14
INNLEDNING	16
Beskrivelse av problemet/tematikken	16
Kartlegging av barns språkutvikling	16
Språkkartlegging på norske helsestasjoner	17
Hvorfor det er viktig å utføre denne kartleggingsoversikten	17
Mål og problemstilling	17
METODE	18
Beskrivelse av systematisk kartleggingsoversikt	18
Prosjektplan	19
Forsknings spørsmål	19
Overordnet metodebeskrivelse	20
Inklusjonskriterier	21
Litteratursøk	22
Utvelging av litteratur	23
Uthenting av data	23
Kartlegging av kunnskapsgrunnlaget	25
RESULTATER	26
Identifisering av relevante kartleggingsverktøy (forsknings spørsmål 1)	26
Identifisering av forskning som evaluerer kartleggingsverktøy (forsknings spørsmål 2)	26
26	
Beskrivelse av de inkluderte studiene	27
Kartleggingsverktøy vurdert av faggruppen (forsknings spørsmål 1)	28
Kartleggingsverktøy ikke vurdert av faggruppen (identifisert via litteratursøk)	41
DISKUSJON	57

Hovedfunn	57
Er kunnskapsgrunnlaget dekkende, pålitelig og anvendelig?	57
Styrker og svakheter ved kartleggingsoversikten	59
Overensstemmelse med andre litteraturoversikter	60
Resultatenes betydning for praksis	62
Kunnskapshull	63
KONKLUSJON	64
REFERANSER	65
VEDLEGG 1: SØKESTRATEGI	75
VEDLEGG 2: MULIG RELEVANTE STUDIER PUBLISERT PÅ ANDRE SPRÅK ENN ENGELSK, NORSK, DANSK ELLER SVENSK	91
VEDLEGG 3: BRUK AV MASKINLÆRING	93
VEDLEGG 4: EKSKLUDERTE STUDIER LEST I FULLTEKST	95
VEDLEGG 5: PRISMA SJEKKLISTEN	111
VEDLEGG 6: IDENTIFISERTE KARTLEGGINGSVERKTØY (FORSKNINGS-SPØRSMÅL 1)	114
VEDLEGG 7: MÅLEEGENSKAPER EVALUERT PER PUBLIKASJON	116
VEDLEGG 8: DETALJERTE RESULTATER PER PUBLIKASJON	118
Verktøy identifisert i forskningsspørsmål 1	118
Verktøy identifisert i databasesøk	121

Hovedbudskap

Språkutvikling hos barn er viktig for barns kognitive og sosiale utvikling. I Norge foregår kartlegging av språk i barnehagealder blant annet på helsestasjoner, vanligvis når barnet er 2 og 4 år. Kartleggingen gjøres ved observasjon, samtaler med foreldre, og ofte ved bruk av språkkartleggingsverktøy. Målet er tidlig oppdagelse av avvik fra normal språkutvikling hos barnet, for å kunne gi nødvendig oppfølging. Anbefalingen om språk i nasjonal faglig retningslinje skal oppdateres i 2024, og det er behov for en oversikt over relevante kartleggingsverktøy, samt en evaluering av disse.

Hensikten med denne oversikten er å identifisere verktøy for å avdekke avvik fra normal språkutvikling hos barn 0-5 år i helsestasjoner. Vi utførte en systematisk kartleggingsoversikt, og inkluderte 47 relevante studier.

- Vi identifiserte totalt 46 verktøy for å kartlegge avvik fra normal forventet språkutvikling hos barn 0-5 år, som vurderes som passende for bruk i en norsk helsestasjonsetting.
- For 24 av de 46 verktøyene identifiserte vi forskning som evaluerte verktøyenes måleegenskaper.
- Det finnes lite forskning om verktøyene som anbefales brukt ved norske helsestasjoner i den nåværende retningslinjen: Vi identifiserte kun én studie beskrevet i to publikasjoner som evaluerte verktøyet SPRÅK 4.
- Det er stor variasjon i hvor mange måleegenskaper som er evaluert for hvert verktøy.
- Flere av studiene fremhever utfordringer ved å utvikle og bruke universelle språkkartleggingverktøy, spesielt for flerspråklige barn.

Resultatene i de inkluderte studiene fremhever kompleksiteten ved å identifisere universelle verktøy for språkkartlegging ment for barn 0-5 år.

Tittel: Verktøy for språkkartlegging av barn 0-5 år: En systematisk kartleggingsoversikt

Hvem står bak denne publikasjonen?

Folkehelseinstituttet, på oppdrag fra Helsedirektoratet

Når ble litteratursøket avsluttet?

Oktober, 2023

Fagfelleevaluering:

Kjetil Gundro Brurberg, Avdelingsdirektør, FHI

Malene Brekke, høyskolelektor/stipendiat, VID

Katrine Giæver, førsteamanuensis, OsloMet

Sammendrag

Innledning

Barns språkutvikling er en kompleks og avgjørende del av deres kognitive og sosiale utvikling, da språk er essensielt for kommunikasjon og overføring av meningsfull informasjon. Avvik fra normal språkutvikling kan påvirke barnets evne til aktiv deltakelse i samspill med jevnaldrende og voksne, og forsinket språkutvikling kan ha betydelige konsekvenser for barnets opplevelse og interaksjon med omverdenen.

Bruk av språkkartleggingsverktøy for å avdekke avvik fra normal språkutvikling på norske helsestasjoner utføres ved 2 og 4 års konsultasjonene. Dette innebærer blant annet observasjon av barnet og samtaler med foreldre/omsorgspersoner. Formålet er tidlig identifikasjon av barn med avvik fra normal språkutvikling. Ved mistanke om forsinket språkutvikling skal barnet henvises til videre språkutredning. Tidlig kartlegging er imidlertid omdiskutert, da det finnes ulike parametere og metoder for kartlegging, samt variasjoner i verktøyformat og hvem som administrerer dem. Nasjonal faglige retningslinje for undersøkelse av språk anbefaler systematisk observasjon av barnets kommunikasjon, språkforståelse, og talespråk. Denne kartleggingsoversikten er nødvendig fordi de nasjonale anbefalingene om språkkartlegging ikke er oppdatert siden 2006, og en oppdatering er planlagt for 2024.

Hensikt

Hensikten med kartleggingen er å identifisere verktøy for å avdekke avvik fra normal språkutvikling hos barn 0-5 år som kan benyttes i en norsk helsestasjonssetting, samt å oppsummere egenskapene til disse verktøyene.

Metode

Vi utførte en systematisk kartleggingsoversikt (engelsk: scoping review). Det er en type kunnskapsoppsummering som kartlegger og narrativt beskriver eksisterende forskningsgrunnlag på et bestemt temaområde. For å identifisere relevante studier gjennomførte en bibliotekar søk i litteraturlidatabaser som MEDLINE og Embase. Vi inkluderte empiriske primærstudier fra år 2000-2023 som evaluerte verktøy/måleinstrumenter for å avdekke avvik fra normal språkutvikling hos barn 0-5 år egnet for bruk ved norske helsestasjoner. Prosjektmedarbeiderne utførte utvelgelse av studier i henhold til inklusjonskriteriene. For hver av de inkluderte studiene hentet vi ut beskrivende informasjon om studien, om verktøyet og hvilke måleegenskaper som ble evaluert. For beskrivelse av måleegenskaper benyttet vi Consensus-based Standards for the

Selection of Health Measurement Instruments (COSMIN). Måleegenskaper iht. COSMIN kan deles opp i ni elementer. Resultatene er fremstilt med enkle narrative presentasjoner av resultater og konklusjoner slik de er rapportert i de inkluderte studiene.

Resultater

Vi identifiserte totalt 46 verktøy for å kartlegge avvik fra normal forventet språkutvikling hos barn 0-5 år, som vurderes som passende for bruk i en norsk helsestasjonsetting. For 24 av disse verktøyene identifiserte vi forskning som evaluerte måleegenskaper.

For de verktøyene som anbefales i de gjeldende retningslinjene for norske helsestasjoner (Screening Av Toåringers Språk (SATS) og SPRÅK 4), fant vi kun en studie beskrevet i to publikasjoner. Begge evaluerte SPRÅK 4. Måleegenskapene ved verktøyet som ble vurdert var intern konsistens (hvor godt alle elementene eller spørsmålene innad et instrument måler det samme konseptet), begrepsvaliditet (i hvilken grad måleinstrumentet måler det konseptet eller fenomenet det er ment å måle) og kriterievaliditet (i hvilken grad resultatene fra instrumentet er korrelert med resultatet fra et «gullstandardinstrument» eller allerede fastsatte diagnoser).

Verktøyet som var mest evaluert – evaluert mht. syv måleegenskaper – var C-BiLLT (Computer-Based Instrument for Low Motor Language Testing). Deretter var det fire verktøy som var evaluert mht. fire måleegenskaper: BITS (Brigance Infant and Toddler Screen), GAPS (Grammar and Phonology Screening), NRT (Nonword Repetition Test), PSRep (Preschool repetition test). De resterende 19 verktøyene var evaluert på tre eller færre måleegenskaper, hvorav ni av verktøyene var evaluert kun på én måleegenskap.

Den måleegenskapen som var mest hyppig evaluert i studiene – målt i 31 studier – var kriterievaliditet (diagnostisk nøyaktighet): hvor godt verktøyet skilte mellom barn med og uten språkvansker, enten samtidig eller longitudinelt. Videre var intern konsistens målt i 16 studier. Seks studier undersøkte kulturell validitet ved vurdering av barns språkferdigheter. To studier som evaluerte PPVT (Peabody Picture Vocabulary Test) rapporterte om kulturell skjevhet i vurderingen av barna når PPVT-IV ble brukt, mens tre studier rapporterte ingen kulturell skjevhet når PPVT-III ble brukt.

For tospråklige barn konkluderte en studie med at screening på begge språkene barna bruker, var den mest nøyaktige metoden for å identifisere barn med språkvansker.

Diskusjon og konklusjon

Det er utviklet mange verktøy for å identifisere barn med avvik fra normal forventet språkutvikling. Mange av verktøyene er imidlertid ikke vurdert, eller vurdert i svært liten grad, med hensyn til sine måleegenskaper. Alt i alt fant vi at det er utviklet 24 verktøy som vurderes som passende for bruk i en norsk helsestasjonsetting og der

måleegenskapene er forsket på. Resultatene viser en mangel på studier om måleegenskaper for mange identifiserte verktøy, spesielt for de yngste aldersgruppene.

Samlet sett gir de inkluderte publikasjonene et bilde av kompleksiteten i å vurdere språkutvikling blant yngre barn, inkludert utfordringer knyttet til tidlig identifisering, pålitelighet og validitet av tester, og overførbarhet til ulike populasjoner og aldre. Resultatene våre viser også viktigheten av å vurdere kulturelle og språklige faktorer i utviklingen og administreringen av språktester og behovet for et helhetlig tilnærming i språkkartleggingen.

Resultatene som rapporteres i de inkluderte studiene stemmer i stor grad overens med tendenser rapportert i andre systematiske oversikter. Vår og andre oversikter understreker kompleksiteten ved å identifisere universelle verktøy som skal kunne benyttes for alle barn, utfordringer med å identifisere egnede verktøy for flerspråklige barn, samt behovet for å kombinere ulike verktøy for en mer helhetlig forståelse av barnets språkutvikling.

Denne kartleggingsoversikten danner ikke grunnlag for å gi konkrete anbefalinger for praksis. For å utarbeide et slik kunnskapsgrunnlag trengs det en mer fullstendig systematisk oversikt der resultatene fra de inkluderte studiene vurderes i detalj. I arbeidet vårt har det kommet frem viktig spørsmål som denne kartleggingsoversikten ikke kan svare på. Disse inkluderer hvilke verktøy som er best egnet for språkkartlegging i en norsk helsestasjonssetting og hvordan språkkartleggingen kan best tilpasses flerspråklige barns sine behov. Fagekspertene kan benytte kunnskapsgrunnlaget i oversikten til å vurdere mer inngående måleegenskapene til utvalgte verktøy.

Kunnskapshullene identifisert i oversikten inkluderer manglende evalueringer for flere verktøy, spesielt for den yngste aldersgruppen (0-2 år), få nordiske studier, begrenset antall måleegenskaper evaluert for mange verktøy, og mangel på egnede verktøy for flerspråklige barn. Disse kunnskapshullene gir retning for fremtidig forskning på området.

Key messages

Language development in children is a complex and crucial part of their cognitive and social development. Language screening in preschool children takes place at child health centres in Norway, mainly when the child is 2 and 4 years old. The assessment includes observation, conversations with parents and often with the use of language screening tools. The aim is to detect deviations in language development at an early stage to provide the necessary follow-up. The recommendation on language in the national professional guideline will be updated in 2024, and there is a need for a review of relevant language screening tools and their evaluation.

The purpose of this scoping review is to identify tools for detecting deviations from normal language development in children aged 0-5 years in Norwegian child health centres. We identified 47 relevant studies:

- We identified a total of 46 tools for detecting deviations from normal language development in children 0-5 years, which are considered appropriate for use in Norwegian child health centres.
- We identified research that evaluated measurement properties for 24 of the 46 tools.
- There is little research on the tools recommended for use in Norwegian child health centres in the current guideline: We identified one study described in two publications that evaluated the tool SPRÅK 4.
- There is considerable variation in the number of measurement properties assessed for each tool.
- Several of the included studies emphasise the challenges of developing and using universal language screening tools, especially for multilingual children.

The reported results of the included studies emphasise the complexity of identifying universal language assessment tools for children 0-5 years of age.

Title:

Language screening tools for children 0-5 years: a systematic scoping review

Publisher:

The Norwegian Institute of Public Health conducted the review based on a commission from the Norwegian Directorate of Health

Updated:

Last search for studies: October 2023

Peer review:

Kjetil Gundro Brurberg, Department Director, FHI

Malene Brekke, Assistant Professor /PhD candidate, VID

Katrine Giæver, Associate Professor, OsloMet

Executive summary (English)

Introduction

Children's language development is a complex and crucial part of their cognitive and social development, as language is essential for communication and the transfer of meaningful information. Deviations from normal language development can affect the child's ability to actively participate in interaction with peers and adults, and delayed language development can have significant consequences for the child's experience and interaction with the outside world.

The use of language screening tools to identify deviations from normal language development at Norwegian child health centres is carried out at the 2- and 4-year consultations. This includes observation of the child and conversations with parents/carers. The purpose is early identification of children with deviations from normal language development. If delayed language development is suspected, the child should be referred for further language assessment. However, early assessment is controversial, as there are different parameters and methods for assessment, as well as variations in tool formats and who administers them. National guidelines recommend systematic observation of the child's communication, language comprehension and spoken language. This scoping review is necessary because the national recommendations on language mapping have not been updated since 2006, and an update is planned for 2024.

Objective

The objective of this scoping review is to identify tools for detecting deviations from normal language development in children 0-5 years that can be used in Norwegian child health centres, and to summarise the characteristics of these tools.

Method

We conducted a systematic scoping review. A scoping review is a type of knowledge synthesis that maps and narratively describes the existing research base in a specific topic area. To identify relevant studies, a librarian conducted searches in literature databases such as MEDLINE and Embase. We included empirical primary studies from 2000-2023 that evaluated tools/measurement instruments to detect deviations from normal language development in children aged 0-5 years suitable for use in Norwegian child health centres. The project staff selected studies according to the inclusion criteria. For each of the included studies, we extracted descriptive information about the study, the tool and the measurement properties that were evaluated according to Con-

sensus-based Standards for the Selection of Health Measurement Instruments (COSMIN). The results are presented with simple narrative presentations of results and conclusions as reported in the studies.

Results

In summary, we identified a total of 46 tools for detecting deviations from normal language development in children 0-5 years, which are considered appropriate for use in Norwegian child health centres. For 24 of these tools, we identified research that evaluated measurement properties.

For the tools that are recommended according to current guidelines for Norwegian child health centres (Screening of Two-Year-Old Language (SATS) and SPRÅK 4), we found only one study described in two publications. Both evaluated SPRÅK 4. The measurement properties of the tool that were assessed were internal consistency (how well all the items or questions within an instrument measure the same concept), construct validity (the extent to which the instrument measures the concept or phenomenon it is intended to measure) and criterion validity (the extent to which the results from the instrument are correlated with the results from a "gold standard" instrument or already established diagnoses).

The most evaluated tool - evaluated in terms of seven measurement properties - was C-BiLLT (Computer-Based Instrument for Low Motor Language Testing). For four tools four measurement properties were evaluated: BITS (Brigance Infant and Toddler Screen), GAPS (Grammar and Phonology Screening), NRT (Nonword Repetition Test), PSRep (Preschool repetition test). The remaining 19 tools were evaluated on three or fewer measurement properties, of which nine of the tools were evaluated on only one measurement property.

The measurement property that was most frequently evaluated in the studies - measured in 31 studies - was criterion validity (diagnostic accuracy): How well the tool distinguished between children with and without language difficulties. Furthermore, internal consistency was measured in 16 studies. Six studies examined cultural validity in the assessment of children's language skills. Two studies evaluating the PPVT (Peabody Picture Vocabulary Test) reported cultural bias in the assessment of children when the PPVT-IV was used, while three studies reported no cultural bias when the PPVT-III was used.

For bilingual children, one study concluded that screening in both languages spoken by the children was the most accurate method for identifying children with language difficulties.

Discussion and Conclusion

Many tools have been developed to identify children with deviations from normal expected language development. However, many of the tools have not been evaluated, or

have been evaluated to a very limited extent, regarding their measurement properties. Overall, we found that 24 tools have been developed for which there is research on their measurement properties, and which are considered appropriate for use in a Norwegian child health centre setting. The results show a lack of studies on the measurement properties of many identified tools, especially for the youngest age groups. Overall, the included publications provide a picture of the complexity of assessing language development among younger children, including challenges related to early identification, reliability and validity of tests, and transferability to different populations and ages. Our results also highlight the importance of considering cultural and linguistic factors in the development and administration of language tests and the need for a comprehensive approach to language assessment.

The results reported in the included studies are largely consistent with trends reported in other systematic reviews. Our review and other reviews emphasise the complexity of identifying universal tools that can be used for all children, the challenges of identifying appropriate tools for multilingual children, and the need to combine different tools for a more holistic understanding of the child's language development.

This scoping review does not provide a basis for making specific recommendations for practice. A full systematic review is needed to develop such a knowledge base. However, our work has raised important questions that this scoping review cannot answer. These include which tools are best suited for language screening in Norwegian child health centres, especially for multilingual children, and how language screening can best be adapted to their needs. Subject matter experts can use the findings in this scoping review to assess the measurement properties of a selection of tools in more detail.

The knowledge gaps identified in the review include a lack of evaluations for many tools, especially for the youngest age group (0-2 years), few Nordic studies, limited number of measurement properties evaluated for many tools, and a lack of suitable tools for multilingual children. These knowledge gaps provide direction for future research in this area.

Forord

Folkehelseinstituttet (FHI) fikk 23. mai 2023 i oppdrag av Helsedirektoratet å utføre en systematisk kartleggingsoversikt over forskningen på måleinstrumenter for å avdekke avvik fra normal språkutvikling hos barn 0-5 år. Måleinstrumentene må være egnet til bruk i helsestasjonen. Formålet med denne kartleggingsoversikten er å identifisere verktøy (måleinstrumenter, metoder) ment for å avdekke avvik fra normal språkutvikling hos barn 0-5 år, og å oppsummere hva som kjennetegner verktøyene og deres måleegenskaper. Kartleggingsoversikten skal brukes inn i arbeidet med å oppdatere anbefaling om undersøkelse av språk i nasjonal faglig retningslinje for helsestasjon, skolehelsetjenesten og helsestasjon for ungdom.

Finansiering

Helsedirektoratet finansierte oppsummeringen med midler fra nasjonalt kompetansemiljø for helsestasjon og skolehelsetjenesten. Forskergruppen tydeliggjorde problemstillingen og inklusjonskriteriene i dialog med oppdragsgiver, uten at oppdragsgiver hadde noen rolle i utarbeidelsen av oppsummeringen.

Bidragstere

Prosjektleder: Patricia Sofia Jacobsen Jardim

Interne prosjektmedarbeidere ved FHI: Tiril Cecilie Borge, Ingvild Kirkehei

Takk til eksterne fagfeller Malene Brekke, høyskolelektor/stipendiat, VID og Katrine Giæver, førsteamanuensis, OsloMet og intern fagfelle Kjetil Gundro Brurberg som har gjennomgått og gitt innspill til kartleggingsoversikten. En stor takk også til Signe Zenia Hulgaard-Bruvold, universitetslektor ved OsloMet; Monica Melby-Lervåg, professor ved Institutt for spesialpedagogikk, Universitetet i Oslo/direktør for forskningssenteret SpedAims og Hilde Lowell Gunnerud, førsteamanuensis i utdanningsvitenskap/spesialpedagogikk, Universitetet i Stavanger for innspill i prosessen med å identifisere relevante verktøy. En stor takk også til kollega Nikita Baiju for hjelp til dataauthenting.

Oppgitte interessekonflikter

Alle forfattere og fagfeller har fylt ut et skjema som kartlegger mulige interessekonflikter. Ingen oppgir interessekonflikter.

Folkehelseinstituttet tar det fulle ansvaret for innholdet i rapporten.

Kåre Birger Hagen
fagdirektør

Rigmor Berg
avdelingsdirektør

Patricia Jacobsen Jardim
Prosjektleder

Forkortelser

BESOS= Bilingual English Spanish Oral Screener
BESS = Bilston and Ettingshall Sure Start
BITS=Brigance Infant and Toddler Screen
BPVS= British Picture Vocabulary Scales
C-BiLLT =Computer-Based Instrument for Low Motor Language Testing
C-BiLLT-CAN =kanadiske-engelske versjonen av verktøyet
CDI =MacArthur Communicative Development Inventories
CELF =Clinical Evaluation of Language Fundamentals - Pre-School
CLT=Cross- linguistic Lexical Tasks
CN-Rep= The Children's Test of Nonword Repetition
COSMIN=Consensus-based Standards for the Selection of Health Measurement Instruments
DELV-NR Syntax subtest =Diagnostic Evaluation of Language Variation – Norm Referenced
DIF= Differential Item Functioning
E.R.I.K. - Early Referral Identification Kit
EL1=barn med engelsk som morsmål
ELIM =The Early Language Identification Measure
ELL=barn med engelsk som andrespråk
ENNI= Edmonton Narrative Norms Instrument
ERB =Early Repetition Battery
EVT= Expressive Vocabulary Test
FHI=Folkehelseinstituttet
FLI=Fox Language inventory
Fluharty-2=Fluharty Preschool Speech and Language Screening Test - Second Edition
FN=First Nations
FWT= First word test
GAPS=Grammar and Phonology Screening
GramBa= Grammatiktest för barn
IPSyn =Index of Productive Syntax
ITPA =Grammatic Closure fra Illinois Test of psycholinguistic Abilities
LCA =Latent component analysis
Lene= Leikki-ikäisen neurologisen
MCDI = MacArthur- Bates Communicative Development Inventories
MLUm =Mean length of utterance in morphemes
NELLI = Neurolingvistisk undersökningsmodell för språkstörda barn
NFN=Non-First Nations
NRDLS=New Reynell Developmental Language Scales

NRT=Nonword Repetition Test
NS=Normal språkutvikling
OA=OpenAlex
PLS-5=Preschool Language Scales 5th Edition
PPVT=Peabody Picture Vocabulary Test
PSrep Test =Preschool repetition test
QUILS:ES =Quick Interactive Language Screener: English-Spanish
QUILS=Quick Interactive Language Screener
RBUP= Regionsenter for barn og unges psykiske helse Øst og Sør
RDLS =Receptive Dutch Language Scale
ROC= Receiver operating characteristics
SATS= Screening Av Toåringers Språk
SLaM =Short language measure
SLC=Schlichting-tester for språkforståelse
SLPESI= The Speech and Language Pathology Early Screening Instrument
SPELT = Structured Photographic Expressive Language Test–Preschool
SSV=Spesifikke språkvansker
StatPed= statelig pedagogisk tjeneste
STEPS studien= Steps to the Healthy Development and Wellbeing of Children study
TENR=Test of Early Nonword Repetition
TPT=Toddler phonology test
TROG= Test for Reception of Grammar
TSD=Test for Sentence Development
USA= United states of America
VAS=visuell analog skala
WPPSI =Wechsler Preschool and Primary Scale of Intelligence

Innledning

Beskrivelse av problemet/tematikken

Språkutvikling hos barn er en kompleks og avgjørende del av deres kognitive og sosiale utvikling (1). Språk er en forutsetning for å kunne kommunisere og overføre meningsfull informasjon til hverandre (2) og innebefatter flere uttrykksformer slik som tegn, skrift eller tale (3).

Språk er ikke bare et verktøy for kommunikasjon, men også for sosial og emosjonell utvikling. Gjennom språket lærer barn å uttrykke følelser, forstå andres perspektiver og bygge sosiale relasjoner. Avvik fra normal språkutvikling kan derfor påvirke barnets evne til å delta aktivt i samspill med jevnaldrende og voksne. Forsinket språkutvikling kan ha en stor innvirkning på hvordan et barn opplever og interagerer med verden rundt seg (4). Problemer med språk og kommunikasjon i tidlig førskolealder kan påvirke både skoleårene til barnet samt voksenlivet (1;4;5). De fleste barn med språkforsinkelse tar igjen sine jevnaldrende innen 3 års alder (6).

Det er store individuelle variasjoner i barns språkutvikling og det finnes ikke noen eksakt alder for når et barn utvikler en konkret språklig ferdighet. Likevel følger barns kommunikasjon og språkutvikling visse utviklingsmønstre (1).

Kartlegging av barns språkutvikling

Barns språkutvikling kan kartlegges på ulike måter, avhengig av alderen til barnet og formålet med kartleggingen. Kartlegging av språk gjøres blant annet i barnehage og i helsestasjon. Bruk av språkkartleggingsverktøy for å avdekke avvik fra normal språkutvikling på norske helsestasjoner utføres ved 2 og 4 års konsultasjonene. Dette innebærer blant annet observasjon av barnet og samtaler med foreldre/omsorgspersoner. Kartleggingen er ment for tidlig identifikasjon av barn som avviker fra normal forventet språkutvikling, for å kunne gi den hjelpen og støtten barnet har behov for (5).

Tidlig kartlegging for å identifisere språkvansker er et omdiskutert tema. Oversiktsstudier har vist at kartleggingsverktøy for å identifisere språkvansker har ulike parametre for hva språkvansker er, inkludert barnets faktiske språkferdigheter, tidspunkt for kartleggingen og kliniske grenseverdier (5). I tillegg varierer verktøyene i format og hvem de skal administreres av. Eksempelvis baseres noen av kartleggingsverktøyene

på foreldrenes rapportering av barnets språkferdigheter imens andre blir utført av trente fagpersoner gjennom direkte vurdering og observasjon av barnet (5).

Språkkartlegging på norske helsestasjoner

Helsestasjonene møter så å si alle barn som bor i Norge (2) og helsesykepleiere står i en særegen posisjon til å kunne nå ut til alle barn og deres foreldre. Dette er sentralt for oppfølging av barn, og spesielt viktig for barn som ikke går i barnehage (7). I den nasjonale faglige retningslinjen anbefaler Helsedirektoratet at helsestasjonene gjennomfører systematisk observasjon av barnets kommunikasjon, språkforståelse og talespråk. Foreldrenes vurdering av barnets språk er også viktig, og benyttes inn i den endelige vurderingen av barnets språkutvikling. Dersom det er mistanke om forsinket språkutvikling, skal barnet henvises til videre språkutredning (2).

I Norge er det utviklet to språkkartleggingsverktøy til bruk i de fastsatte konsultasjonene i helsestasjonsprogrammet (2). SATS (Screening Av Toåringers Språk) er beregnet for toåringers og SPRÅK 4 for fireåringers. SATS fokuserer på barnets språkforståelse, og i tillegg skal samspill og oppmerksomhet vurderes (2). SPRÅK 4 er en systematisk observasjon med mål om å avdekke vesentlige trekk ved fireåringers språkferdigheter. I tillegg blir barnets samspill og oppmerksomhet vurdert (2).

Hvorfor det er viktig å utføre denne kartleggingsoversikten

Anbefalingen om språk i nasjonal faglig retningslinje har ikke blitt oppdatert siden 2006. Kapittelet skal oppdateres i 2024, og i den forbindelse er det nødvendig å få en oversikt over eksisterende kartleggingsverktøy ment for å avdekke avvik fra normal språkutvikling hos barn 0-5 år, og en oversikt over evalueringer av måleegenskapene til disse.

Mål og problemstilling

Hensikten med denne kartleggingsoversikten er å identifisere verktøy (måleinstrumenter, metoder) ment for å avdekke avvik fra normal språkutvikling hos barn 0-5 år, og å oppsummere hva som kjennetegner verktøyene og deres måleegenskaper. Verktøyene må være egnet til bruk i helsestasjonsprogrammets konsultasjoner. I denne rapporten har vi fokus på verktøy som kartlegger talespråk og språkforståelse.

Metode

Vi utførte en systematisk kartleggingsoversikt over studier som hadde undersøkt måleegenskapene til kartleggingsverktøy som kunne brukes ved helsestasjoner i Norge for vurdering av språkvansker hos barn 0-5 år.

Kartleggingsoversikten var et uavhengig arbeid av forskergruppen ved FHI. I samsvar med internasjonale metodeanbefalinger (8;9), og for å sikre at oppsummeringen tilfredsstilte oppdragsgivers kunnskapsbehov, konsulterte vi imidlertid med oppdragsgiver av kartleggingsoversikten i utarbeidelsen av prosjektplanen.

Beskrivelse av systematisk kartleggingsoversikt

En kartleggingsoversikt (engelsk: scoping review) er en kunnskapsoppsummering som kartlegger og narrativt beskriver eksisterende litteratur- eller forskningsgrunnlag på et bestemt temaområde. En slik oversikt er utarbeidet på en vitenskapelig, systematisk og transparent måte, og har ofte som mål å besvare et bredere forskningsspørsmål enn en full systematisk oversikt, samt *beskrive* den foreliggende forskningen på et spesifikt felt (8-10). Med en kartleggingsoversikt identifiserer, kartlegger og presenterer forskerne beskrivende informasjon om de inkluderte studiene/publikasjonene.

Kartleggingsoversikter inneholder vanligvis ikke analyser eller synteser av studienes forskningsresultater, ingen syntese av funn på tvers av studier eller vurdering av tillit til den samlede dokumentasjonen, slik en systematisk oversikt gjør. På den måten er en kartleggingsoversikt ikke egnet til å gi et klart svar på for eksempel spørsmål om effekt av et tiltak eller erfaringer med et tiltak, men gir i stedet et overblikk over forskningen som finnes. En kartleggingsoversikt er i så måte ikke egnet som beslutningsgrunnlag for politikktutforming (11). Presentasjonen av resultatene fra de inkluderte studiene er derfor vanligvis deskriptiv, som for kvalitative data kan innebære en organisering av overordnede kategorier og overordnet fremstilling av funn slik de blir presentert i studiene. Diskusjonen med drøfting av resultatene og vurdering av resultatenes betydning for praksis er følgelig noe begrenset i en kartleggingsoversikt, og vil være mindre omfattende og mer overordnet sammenlignet med en full systematisk oversikt (8;10).

Det finnes flere internasjonale metodebeskrivelser (8-10) samt en rapporteringsstandard (12) for kartleggingsoversikter. Vi fulgte disse anbefalingene og utarbeidet kartleggingsoversikten i følgende trinn:

1. Definere og tilpasse målene og spørsmålene
2. Utvikle og tilpasse inklusjon- og eksklusjonskriteriene med målene og spørsmålene
3. Beskrive den planlagte tilnærmingen til litteratursøk og valg av litteratur
4. Søke etter litteratur
5. Velge ut litteratur i henhold til inklusjons- og eksklusjonskriteriene
6. Vurdere studienes metodiske begrensninger
7. Hente ut data fra inkluderte studier/publikasjoner
8. Kartlegge kunnskapsgrunnlaget
9. Oppsummere kunnskapsgrunnlaget med hensyn til målene og spørsmålene

Prosjektplan

Kartleggingsoversikten er gjennomført i tråd med prosjektplanen, publisert her: <https://www.fhi.no/cristin-prosjekter/aktiv/verktøy-for-sprakkartlegging-av-barn-0-5-ar-prosjektplan-for-systematisk-kartleggingsoversikt/>. Vi avklarte prosjektplanen med Helsedirektoratet før vi utførte kartleggingsoversikten.

Vi har to avvik fra prosjektplanen. Begge skyldes manglende spesifisering i prosjektplanen. Det er:

1. Vi inkluderte masteroppgaver. Dette var ikke en publikasjonstype vi hadde spesifisert at vi skulle ta med, men vi vurderte disse publikasjonene som relevante.
2. Vi inkluderte studier fra engelskspråklige land som ikke var blant de vi spesifiserte i inklusjonskriteriene i prosjektplanen. Vi inkluderte dermed flere studier fra USA selv om dette ikke er et av landene spesifisert i inklusjonskriteriene. Rasjonale for denne avgjørelsen, som ble tatt i samråd med oppdragsgiver, var at vi mener vi ellers ville miste viktig informasjon om verktøy som er relevante for norsk helsestasjonssetting. Vi ekskluderte imidlertid studier fra USA som ikke omhandlet verktøy vi på forhånd identifiserte som relevante for norsk helsestasjonssetting (se beskrivelse av trinn 1 og 2 under).

Forskningsspørsmål

Hensikten med denne kartleggingsoversikten var å identifisere verktøy (måleinstrumenter, metoder) ment for å avdekke avvik fra normal språkutvikling hos barn 0-5 år, og å oppsummere hva som kjennetegner verktøyene og deres måleegenskaper. Vi hadde to forskningsspørsmål som skulle besvares i denne kunnskapsoppsummeringen:

1. Hva finnes av språkkartleggingsverktøy for å identifisere avvik fra normal språkutvikling hos barn 0-5 år, som egner seg å bruke i en norsk helsestasjonssetting?
2. Hva finnes av forskning som evaluerer måleegenskapene på kartleggingsverktøy identifisert i forskningsspørsmål 1?

Overordnet metodebeskrivelse

Metode for å identifisere kartleggingsverktøy (forskningsspørsmål 1)

For å besvare forskningsspørsmål 1 lette vi etter og identifiserte kartleggingsverktøy som var i bruk i Norden, samt i Canada, Sveits, Nederland, Australia og Storbritannia. Vi var mest interessert i kartleggingsverktøy som ble brukt i Norden, men for å finne flere potensielt nyttige verktøy valgte vi også å undersøke hva som fantes i noen utvalgte land utenfor Norden. Landene ble valgt i samråd med oppdragsgiver. Vi utførte denne kartleggingsoversikten trinnvis:

Trinn 1: Vi søkte etter informasjon om kartleggingsverktøy på nettsider til helsemyndigheter og andre relevante virksomheter. Vi prioriterte først å finne verktøy som ble brukt i de nordiske landene, for deretter å se på verktøy brukt i Canada, Sveits, Nederland, Australia og Storbritannia. Vi søkte i Google og Google Scholar og gjennomgikk nettsider og publikasjoner hos relevante institusjoner, som Statlig spesialpedagogisk tjeneste (Statped), Regionsenter for barn og unges psykiske helse Øst og Sør (RBUP) (Psyktest) og tilsvarende i andre nordiske land. Se vedlegg 1 for oversikt over kilder. Målet for dette steget var å danne en liste med kartleggingsverktøy som kunne være egnet for bruk i norske helsestasjoner.

Trinn 2: Vi sendte listen med kartleggingsverktøy til oppdragsgiver, fagfeller og et utvalg av fagpersoner, heretter omtalt som faggruppen. For å rekruttere fagpersoner kikket vi etter navn som gikk igjen på nettsider og i publikasjoner, samt personer med relevant kompetanse i sin ansattbeskrivelse. Vi kontaktet de identifiserte fagpersonene med spørsmål om de var villige til å gi innspill på listen over kartleggingsverktøy. De som ikke hadde anledning, henviste oss videre til andre aktuelle personer. Faggruppen bestod av: To representanter fra oppdragsgiver, de to eksterne fagfellene (Malene Brekke og Katrine Giæver), samt fagekspertene: Monica Melby-Lervåg, professor ved Institutt for spesialpedagogikk, UiO/direktør for forskningssenteret SpedAims; Hilde Lowell Gunnerud, førsteamanuensis i utdanningsvitenskap/spesialpedagogikk, UiS og Signe Zenia Hulgaard-Bruvold, universitetslektor ved OsloMet. Vi ba faggruppen om å komme med innspill på hvilke verktøy som manglet eller eventuelt burde fjernes fra listen. Hensikten med dette var å sikre at vi kun inkluderte verktøy som kunne brukes i en norsk helsestasjonssetting. Dette resulterte i en liste med verktøy som vi jobbet videre med i forskningsspørsmål 2.

Metode for å identifisere forskning som evaluerte kartleggingsverktøy (forskningsspørsmål 2)

Etter at forskningsspørsmål 1 var besvart, besvarte vi forskningsspørsmål 2 med følgende metoder (se detaljert metodebeskrivelse under):

Trinn 3: Vi utførte et databasesøk etter studier som undersøkte måleegenskapene til kartleggingsverktøy som ble identifisert i trinn 1 og 2. Studiene måtte ha hatt minst ett mål på validitet eller reliabilitet (se tabell 1).

Trinn 4: Vi valgte ut relevante studier i henhold til inklusjonskriteriene og oppsumerte verktøyene og deres måleegenskaper i henhold til kriteriene nevnt under «Ut-henting av data» (se tabell 1).

Inklusjonskriterier

Kartleggingsverktøyet måtte være egnet til å vurdere avvik fra normal språkutvikling hos barn 0-5 år på helsestasjon. Studier som vurderte måleegenskaper ved språkkartlegging, hvor språkkartlegging kun er en del av et større kartleggingsverktøy, ville også bli inkludert, hvis dette ble rapportert separat. Vår forståelse av måleegenskaper er basert på COSMIN (Consensus-based Standards for the Selection of Health Measurement Instruments) (13).

Vi hadde følgende inklusjonskriterier:

Populasjon	Barn 0-5 år som undersøkes for avvik fra forventet språkutvikling på helsestasjon eller lignende setting. Barna kan være både ettspråklige og flerspråklige.
Konsept	Kartleggingsverktøy/instrumenter for å oppdage avvik fra normal forventet språkutvikling.
Utfall	Måleegenskaper/psykometriske egenskaper ved kartleggingsverktøyet.
Studiedesign	Alle studier som har benyttet kvantitative metoder for å vurdere måleegenskaper ved et aktuelt kartleggingsverktøy.
Publikasjonsår	2000-2023
Land/Kontekst	Prioriterte land: Norge, Sverige, Danmark, Finland og Island. Andre land: Storbritannia, Australia, Sveits, Nederland og Canada. Studiene måtte omhandle verktøy egnet til bruk i en norsk helsestasjonssetting. For å vurdere egnethet baserte vi oss på innspill fra oppdragsgiver, fagfeller og andre fagekspertter. Det var viktig at kartleggingen med verktøyet kunne bli gjort innenfor den oppsatte timen for konsultasjon på helsestasjon og at verktøyet ikke var for omfattende å administrere. Det var også viktig at verktøyet kunne bli administrert av sykepleiere og at det ikke var nødvendig å ha en relasjon med barnet på forhånd.
Språk	Engelsk, norsk, dansk, svensk. Relevante studier publisert på andre språk enn de overnevnte vises i vedlegg 2.
Publikasjonstype	Tidsskriftsartikler, rapporter, masteroppgaver og PhD-avhandlinger.
Annet	Kartleggingsverktøyet må være oversatt og tilgjengelig på engelsk, norsk, svensk eller dansk. Studier hvor hovedandelen av populasjonen er 5 år eller yngre vil bli inkludert.

Eksklusjonskriterier

Vi ekskluderte følgende typer studier og publikasjoner:

- Studier som ikke er tilgjengelige i fulltekst
- Verktøy hvor foreldrene alene vurderer barnets språkutvikling
- Kartleggingsverktøy for bruk i barnehage, hvor det kreves en lengre observasjonstid av barnet eller relasjon til barnet
- Kartleggingsverktøy hvor det er spesifisert at det tar over en time å gjennomføre fordi de vil være for omfattende å bruke i dagens helsestasjonsprogram
- Studier som hverken måler validitet eller reliabilitet
- Verktøy hvor det er tydelig at de kun kan bli brukt for å stille diagnoser
- Verktøy hvor det er eksplisitt presisert at det krever en spesialkompetanse, eksempelvis i logopedi
- Verktøy kun til bruk blant barn med diagnoser eller funksjonsnedsettelse som påvirker språkutvikling, slik som hørselstap eller autisme

Litteratursøk

Søk i databaser

Bibliotekar (IK) utarbeidet en søkestrategi i samarbeid med prosjektgruppen og utførte søkene i følgende databaser:

- MEDLINE (Ovid)
- EMBASE (Ovid)
- ERIC (Ebsco)
- PsycINFO (Ovid)
- Epistemonikos
- Web of Science
- CINAHL (Ebsco)

Vi utførte søket i to deler. Første del inneholdt søkeord for språkkartlegging generelt, for eksempel *screening test and language disorders*. Søkene ble avgrenset med et søkefilter for psykometriske studier (for eksempel *validity, reliability*). En annen bibliotekar (Ingrid Harboe) fagfellevurderte disse søkestrategiene.

Andre del av søket inneholdt navnene på alle relevante kartleggingsverktøy som ble identifisert i forskningsspørsmål 1. Alle søkene ble avgrenset til å finne studier publisert f.o.m. år 2000. Se vedlegg 1 for fullstendig søkestrategi. Det siste søket ble avsluttet i oktober 2023.

Søk i andre kilder

I arbeidet med å finne relevante verktøy (forskningsspørsmål 1), utførte vi søk i Google, og Google Scholar og gjennomgikk nettsider og publikasjoner hos relevante institusjoner, som StatPed, RBUP (Psyktest). Dersom vi fant relevant forskning, tok vi dette med videre i arbeidet med forskningsspørsmål 2.

OpenAlex

TCB gjennomførte et søk i OpenAlex datasettet (14) som er tilgjengelig i EPPI Reviewer. Her benyttes allerede relevante studier som utgangspunkt for søk, fremfor en søkestreng. Vi benyttet blant annet studier hentet fra systematiske oversikter identifisert i et innledende søk. Av disse ble 69 studier (3;4;15-80) «matchet» i OpenAlex datasettet og disse benyttet vi som utgangspunkt for søk etter nye relevante publikasjoner. Søkeperioden var 1. januar 2000 – 15. august 2023. For mer detaljer om OpenAlex, se vedlegg 1.

Utvelging av litteratur

Vi (PSJJ/TCB/IK) utførte relevansvurderinger («screening») av titler og sammendrag fra litteratursøket opp mot inklusjonskriteriene. Vi piloterte inklusjonskriteriene på de 100 første referansene, for å sikre at prosjektmedarbeiderne hadde en felles forståelse av inklusjonskriteriene. Vi brukte maskinlæringsfunksjoner i det elektroniske verktøyet EPPI-Reviewer (81) for å hjelpe oss med å vurdere titler og sammendrag mer effektivt. Enkelt sagt betyr maskinlæring at vi tar i bruk algoritmer som gjør at programvaren er i stand til å lære og utvikle sin beslutningsstøtte basert på empiriske data som vi fører den med. Se strategi for bruk av maskinlæring i vedlegg 3.

Uenighet om vurderingene av titler/sammendrag løste vi ved diskusjon til felles enighet ble oppnådd, og vi innhentet relevante publikasjoner i fulltekst. To prosjektmedarbeidere (PJJ/TCB) gjorde uavhengige vurderinger av fulltekstene opp mot inklusjonskriteriene. Uenighet om vurderingene av fulltekster løste vi ved diskusjon eller ved å konferere med en tredje prosjektmedarbeider.

I flytdiagrammet (figur 1) presenterer vi antall referanser som ble inkludert og ekskludert på de ulike trinnene. Studier som vi ekskluderte etter fulltekstvurdering vises i vedlegg 4, med begrunnelse for eksklusjon

Uthenting av data

Én prosjektmedarbeider (IK, TCB, PSJJ) hentet ut data fra de inkluderte studiene, mens en annen (IK, TCB, PSJJ) kontrollerte dataene opp mot de aktuelle publikasjonene. Ved uenighet konsulterte vi en tredje prosjektmedarbeider for å bidra til enighet. Vi brukte et pilotert dataauthentingsskjema i Excel. Vi hentet følgende data fra de inkluderte studiene: informasjon om publikasjonen (sitering, type publikasjon), studiedesign og metode, type kartleggingsverktøy og kjennetegn (antall spørsmål, skåring o.l.), alder på barn det skulle brukes på (dvs. måletidspunkt), setting (hvor instrumentet ble brukt), kostnad, språk instrumentet fantes på, informasjon om opplæring for å kunne bruke instrumentet, og tidsbruk for instrumentet. For verktøy rettet mot flerspråklige barn hentet vi også spesifikk informasjon, som krav til fagkunnskap om flerspråklighet og

bruk av tolk osv. Vi hentet data tilsvarende domene i COSMIN-sjekklisten (13;82): reliabilitet (intern konsistens, reliabilitet, målefeil), validitet (innholdsvaliditet, begrepsvaliditet, kriterievaliditet) og responsivitet (se tabell 1).

Tabell 1: COSMIN oversikt over områder/domener, måleegenskaper og definisjoner

Område/ Domene	Måleegenskap	Definisjon
Reliabilitet	Intern konsistens	Måler hvor godt alle elementene eller spørsmålene i et instrument måler det samme konseptet. Høy intern konsistens indikerer at elementene er sammenhengende og måler det samme aspektet av språkutviklingen.
	Test-retest	Måler stabiliteten av resultatene over tid ved å administrere det samme instrumentet til de samme deltakerne i to separate tidspunkter. Høy test-retest pålitelighet indikerer at resultatene er stabile over tid.
	Inter-rater	Vurderer graden av enighet mellom forskjellige personer når de bruker instrumentet samtidig. Høy inter-rater reliabilitet antyder at flere personer kan bruke verktøyet og få lignende resultater.
	Intra-rater	Måler konsistensen av én enkelt persons resultater over tid. Dette gir innsikt i hvor pålitelig den samme personen kan være når de bruker instrumentet ved ulike anledninger.
	Målefeil	Tilfeldige og systematiske feil som ikke er knyttet til faktiske endringer i det som skal måles. Lav målefeil indikerer mer nøyaktige resultater og pålitelighet i instrumentet.
Validitet	Innholdsvaliditet	Vurderer i hvilken grad instrumentet dekker alle aspekter av det konseptet det er ment å måle. Høy innholdsvaliditet betyr at instrumentet er omfattende og relevant for målingen av avvik i språkutviklingen.
	Begrepsvaliditet	Begrepsvaliditet refererer til i hvilken grad et måleinstrument måler det konseptet eller fenomenet det er ment å måle. Det finnes flere typer begrepsvaliditet: Strukturell validitet: Er et mål på hvorvidt instrumentet reflekterer de ulike dimensjonene/aspektene ved det overordnede konseptet instrumentet er ment for å måle. Konvergent validitet: Måler graden av samsvar mellom resultatene av det aktuelle instrumentet og resultater fra andre målinger som antas å måle det samme konseptet (men ikke nødvendigvis en gullstandard). Kryss-kulturell validitet: Vurderer om instrumentet gir like resultater på tvers av ulike kulturelle grupper.
	Kriterievaliditet	Måler i hvilken grad resultatene fra instrumentet er korrelert med resultatene fra et «gullstandardinstrument» eller allerede fastsatte diagnoser. Blir også omtalt som «diagnostic accuracy». Begrepene sensitivitet og spesifisitet er knyttet til denne måleegenskapen. Sensitivitet er betegnelsen på om testen kan plukke ut de barna som har forsinket språkutvikling (tilfellene). Spesifisitet er i hvilken grad screeningstesten kan identifisere barn som har forventet språkutvikling. Samtidig validitet er et mål på hvor godt resultatene fra et instrument samsvarer med gullstandardmålinger/diagnostiseringer ved samme tid. Prediktiv validitet er et mål på i hvilken grad resultatene fra et instrument samsvarer med fremtidige skårer på gullstandardinstrumentet eller diagnoser.
Responsivitet		I hvilken grad et måleinstrument er i stand til å oppdage endringer eller respons i det som blir målt.

Kartlegging av kunnskapsgrunnlaget

Vi plasserte alle de inkluderte publikasjonene i en liste og sorterte dem i logiske kategorier. Vi sammenfattet informasjonen om studiene og deres resultater for instrumentene narrativt og vurderte resultatene fra de ulike studiene på tvers for hvert av verktøyene. Vi presenterte resultatene i tekst og tabeller. Hovedfremstillingen var i form av prosa med narrative presentasjoner av resultater og konklusjoner.

Vi har rapportert barnas alder på følgende måte: år;måneder (dvs. en gruppe barn med gjennomsnittlig alder på 5 år og 8 måneder vises som 5;8).

Iht. standarden for kartleggingsoversikter, utførte vi ingen datasynteser, dvs. vi utførte ikke en sammenstilling av individuelle studieresultater (f.eks. ingen metaanalyser) på tvers av studiene, eller vurderte tilliten til dokumentasjonen. Vi vurderte heller ikke studienes risiko for systematiske skjevheter.

Vi brukte sjekklisten «The PRISMA extension for scoping reviews» (12) for å sikre rapporteringskrav, se vedlegg 5.

Resultater

Identifisering av relevante kartleggingsverktøy (forskningsspørsmål 1)

Gjennom internettsøk identifiserte vi 46 verktøy som muligens kunne være relevante for bruk i norske helsestasjoner. Vi presenterte verktøyene for faggruppen, og ba om innspill. Etter å ha fått tilbakemeldinger fra faggruppen om verktøyenes relevans, hadde vi en liste med 32 verktøy, presentert i vedlegg 6.

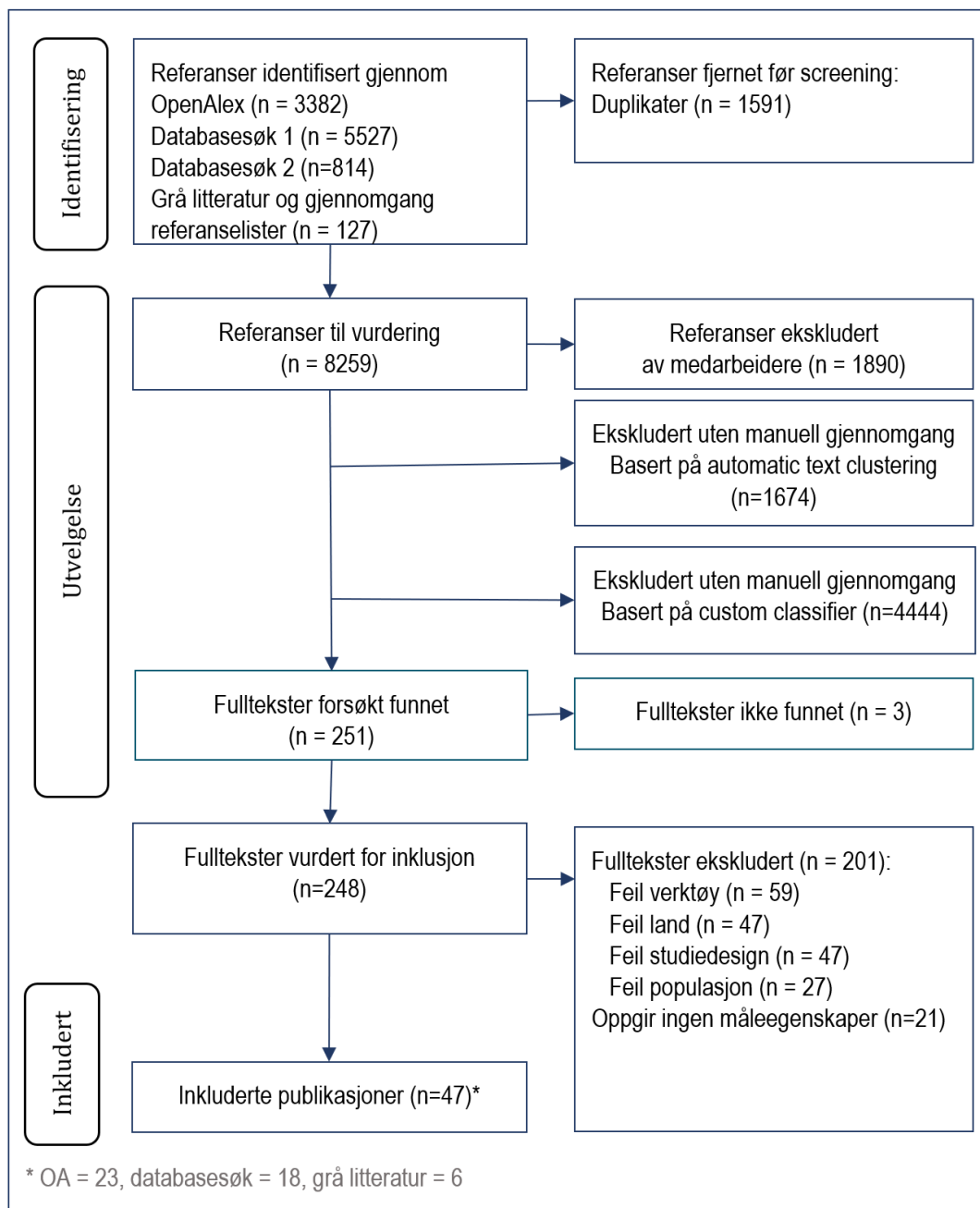
Disse verktøyene tok vi videre i arbeidet med forskningsspørsmål 2. Mens vi gjennomgikk søketreffene fra litteratursøket, fant vi evalueringer av ytterligere 14 kartleggingsverktøy fra ett av de relevante landene. I dialog med oppdragsgiver ble vi enige om at vi skulle inkludere alle verktøy som møtte inklusjonskriteriene, selv om de ikke allerede var identifisert i forbindelse med forskningsspørsmål 1. Disse verktøyene ble ikke relevansvurdert av faggruppen og vi søkte heller ikke spesifikt etter dem i databasesøkene. Derfor presenterer vi disse 14 kartleggingsverktøyene separat fra de 32 øvrige verktøyene.

Oppsummert identifiserte vi totalt 46 kartleggingsverktøy for å oppdage avvik fra normal forventet språkutvikling hos barn 0-5 år.

Identifisering av forskning som evaluerer kartleggingsverktøy (forskningsspørsmål 2)

Resultater av litteratursøket og utvelgelse av studier

Databasesøkene ga 9850 treff før fjerning av dubletter (figur 1). Etter fjerning av dubletter satt vi igjen med 8259 referanser som vi vurderte. Det inkluderer 127 referanser identifisert via referanselister og grå litteratur. Av disse ekskluderte vi 1890 publikasjoner som åpenbart ikke oppfylte inklusjonskriteriene våre, og vi ekskluderte totalt 6113 referanser uten manuell gjennomgang ved hjelp av to maskinlæringsfunksjoner. For detaljert informasjon om bruk av maskinlæring i vurderingsprosessen, se vedlegg 3. Vi fikk tak i og vurderte 248 publikasjoner i fulltekst, hvorav vi ekskluderte 201 (se vedlegg 4 for oversikt over alle ekskluderte referanser lest i fulltekst). Vi inkluderte 44 studier fordelt på 47 publikasjoner (51;57;60;74;83-125).



Figur 1: Flyttdiagram over utvelgelse av studier

Beskrivelse av de inkluderte studiene

Vi inkluderte 44 unike studier i denne oversikten, beskrevet i 47 publikasjoner (51;57;60;74;83-125). De 44 studiene beskriver 24 unike verktøy, hvorav ti av verktøyene er blant de som ble vurdert av faggruppen og 14 er blant de vi identifiserte senere i databasesøkene, som ikke er relevansvurdert av faggruppen. I resultatene under presenterer vi separat resultatene for de identifiserte verktøyene basert på hvorvidt de er relevansvurdert av faggruppen eller ikke.

Alle verktøyene i studiene er utført på en slik måte at det vil være sannsynlig at de kan

overføres til en helsestasjonssetting (se inklusjonskriterier) og administreres av en helsesykepleier. Elleve av studiene rapporterte at det var helsesykepleier som gjorde målingene og i de resterende studiene var verktøyet administrert av logoped, forskningsassistenter, sykepleiere, pedagoger eller studenter. I fem av publikasjonene var det uklart hvem som administrerte testen.

Kartleggingsverktøy vurdert av faggruppen (forskningsspørsmål 1)

Som beskrevet over, for forskningsspørsmål 1, etter relevansvurdering av faggruppen endte vi opp med en liste med 32 kartleggingsverktøy. Vi identifiserte imidlertid studier for kun ti av verktøyene som ble identifisert i forskningsspørsmål 1. Totalt 19 studier (21 publikasjoner) beskriver disse ti verktøyene:

- Brigance Infant and Toddler Screen (BITS)
- Early Language Identification Measure (ELIM)
- Expressive Vocabulary Test (EVT)
- New Reynell Developmental Language Scales (NRDLS)
- Peabody Picture Vocabulary Test (PPTV)
- Quick Interactive Language Screener (QUILS)
- SPRÅK 4
- Språkfyrer
- Språkscreening ved 2,5 år
- Språkscreening ved 3 år

I tabell 2 og 3 beskriver vi de ti verktøyene som ble identifisert fra forskningsspørsmål 1 og som vi fant forskning om. Først, i tabell 2 beskrives verktøyet, hvilke språkdome-ner som er målt samt hvor lang tid verktøyet tar å administrere. Flere av verktøyene anga ikke tidsestimat, men ut ifra informasjon om verktøyene er det sannsynlig at de vil kunne administreres i løpet av en konsultasjon på helsestasjon.

Tabell 2: Beskrivelse av verktøyene identifisert i forskningsspørsmål 1 som det finnes forskning om (10 verktøy)

Navn på verktøy	Kort beskrivelse
Brigance Infant and Toddler Screen (BITS)	BRIGANCE Infant and Toddler Screens III er en individuell, kriterie- og normreferert screeningtest som er utviklet for å vurdere den generelle utviklingen av en rekke ferdigheter hos barn fra fødselen til slutten av første klasse (syv år og seks måneder). Screening-ingen omfatter: Finmotorikk, grovmotorikk, ekspressivt språk, reseptivt språk, selvhjelp og sosiale og emosjonelle ferdigheter. Dette kartleggingsinstrumentet er utviklet spesielt for å vurdere et bredt utvalg av barnets utvikling og ferdigheter, og gir persentilrangering, standardkårer og aldersekvivalente skårer. Hele testen tar ca. 10-15 minutter å gjennomføre.
The Early Language Identification Measure (ELIM)	ELIM (Early Language Identification Measure) ble utviklet for å skape et omfattende mål for å identifisere tidlige språkutviklingsproblemer hos barn. Måleinstrumentet består av fem seksjoner: (1) milepæler i utviklingen, (2) en sjekklister for ordforråd, (3) familiære og demografiske risikofaktorer, (4) fagpersoners observasjoner av barnet, og (5) foreldrenes bekymring. Alle kandidat spørsmålene i de fem seksjonene ble utarbeidet av forsker- teamet på grunnlag av relevant litteratur og ekspertkunnskap i samarbeid med en rekke fagpersoner og foreldre ved hjelp av pasient- og brukervedvirkningsprosesser. Det eneste unntaket var sjekklister for ordforråd. Denne listen med 50 ord, kjent som Sure

	Start Language Measure (SSLM), ble opprinnelig hentet fra Communicative Development Inventory (CDI).
Expressive Vocabulary Test (EVT)	Expressive Vocabulary Test er et individuelt administrert, normert mål på ekspressivt ordforråd og ordgjenfinning. Testen kan brukes på personer i alderen 2 år og 6 måneder til 90 år og eldre. Den måler tilegnelse av ekspressivt ordforråd og kan bidra med nyttig informasjon som en del av en språkevaluering. Den kan brukes som et screeningmål eller til å vurdere styrker og svakheter i det spesifikke semantiske domenet (dvs. ordkunnskap) og det generelle språkutviklingsområdet. Fordi den er samnormert med Peabody Picture Vocabulary Test, kan EVT-skårene sammenlignes direkte. EVT kan administreres digitalt og skåres ved hjelp av Q-globalt skårings- og rapporteringssystem. Testen tar 10-20 minutter å gjennomføre.
New Reynell Developmental Language Scales (NRDLS)	New Reynell Developmental Language Scales (NRDLS) er en standardisert test som ble publisert i 2011, og er en omfattende revisjon av den tidligere Reynell Developmental Language Scales, som tar hensyn til ny kunnskap om barns språktilegnelse og språkvansker. Testen dekker ulike aspekter ved språktilegnelse, inkludert ordforråd, setningsstruktur, verbmorfologi, inferens og vurdering av grammatikalitet hos barn i alderen 3-7 år. NRDLS gir en profil av barnets språkferdigheter og kan brukes til å identifisere språkforsinkelser eller språkvansker. Forståelse og produksjon testes på ulike skalaer, der vanskelighetsgraden øker etter hvert som barnet utvikler seg gjennom testen. Det gis veiledning om hvordan testen kan stoppes når barnet begynner å feile på enkelte oppgaver. Testen tar ca. 30-40 minutter å gjennomføre og er oversatt til norsk.
Peabody Picture Vocabulary Test, revised, 3rd and 4th edition (PPTV)	<p>Peabody Picture Vocabulary Test, Revised (utgitt 1981) måler reseptivt ordforråd og gir et raskt estimat av verbale evner, for personer i alderen 2½ til 40 år. Den engelskspråklige versjonen av PPVT-R består av 175 ordforrådsoppgaver med generelt økende vanskelighetsgrad. Barnet lytter til et ord som intervjueren sier, og velger deretter ett av fire bilder som best beskriver ordets betydning. PPVT-R består av 175 stimulusord og 175 tilhørende bilder. Hvert bilde inneholder fire svart-hvitt-tegninger, hvorav én best representerer betydningen av det tilsvarende stimulusordet. Det er også 5 treningsord og bilder. PPVT-R-oppgavene er nummerert i stigende vanskelighetsgrad.</p> <p>Peabody Picture Vocabulary Test, Third Edition (utgitt 1997) måler reseptivt ordforråd og er en screeningtest for verbale evner. Den er standardisert på over 2000 barn og over 700 voksne og er beregnet på personer mellom 2 år og ≥90 år. PPVT-III består av 204 stimulusord og tar i gjennomsnitt 11-12 minutter å gjennomføre. Undersøkeren sier en ledetekst, og testpersonen blir bedt om å peke på det bildet som passer best til det presenterte stimulusordet.</p> <p>Peabody Picture Vocabulary Test, Fourth Edition (PPVT-IV) (utgitt 2007) ble standardisert på en større populasjon (3540 personer) enn PPVT-III. PPVT-IV inneholder 228 stimulusord. Selv om noen av ordene er identiske med den forrige versjonens stimuli, er andre byttet ut og flere ord er inkludert. I tillegg presenteres bildevalgene i farger, noe som står i kontrast til de svart-hvite bildevalgene i den eldre PPVT-III. Andre endringer inkluderer utskifting av bildevalg for å utelukke utdatert teknologi som var representert i den tidligere versjonen.</p>
Quick Interactive Language Screener (QUILS)	Quick Interactive Language Screener (QUILS) er et verktøy for språkkartlegging i førskolealder. Det kartlegger barns ordforråd, syntaks og språktilegnelsesferdigheter. Verktøyet tar 15 minutter å administrere og er brukervennlig og egnet for ikke-spesialister. QUILS bruker visuelle og auditive stimuli, inkluderer trening og gir verbal oppmuntring. Den vurderer ordforståelse i åpen og lukket klasse, syntaktiske strukturer og metoder for språktilegnelse. Resultatene er aldersbaserte standardskårer og persentilrangeringer, med en grenseverdi på 25. persentil som indikerer mulige språkvansker. QUILS er beregnet for å identifisere risikobarn, ikke for å diagnostisere vansker.
Quick Interactive Language Screener: English-Spanish (QUILS:ES)	QUILS:ES er et nettbasert språkkartleggingsverktøy som er utviklet for engelsk-spanske tospråklige barn i alderen 3-5 år. I likhet med den engelske QUILS vurderer QUILS:ES ordforråd, syntaks og læringsprosesser. Verktøyet inneholder separate deler for engelsk og spansk, noe som gir et helhetlig bilde av tospråklige ferdigheter. Barnets beste totalskår og beste skår på hvert område fastsettes basert på de høyeste skårene i begge språkdelen. QUILS:ES er utviklet for å fange opp språklige variasjoner i grammatikk og ordbetydninger på begge språk, med engasjerende illustrasjoner som tar 15-20 minutter å

administrere. QUILES administreres og skåres automatisk, noe som gjør at det ikke er noen krav til spesiell kompetanse hos den som administrerer.

SPRÅK 4

Formålet med SPRÅK 4 er å kartlegge språkferdigheter hos fire år gamle barn som en del av 4 års kontrollen på helsestasjon. Kartleggingen består av en enkel bildemappe med bilder som er kjent for fireåringer, et kartleggings skjema med 33 spørsmål knyttet til bildene og en brukermanual. Språkområder som dekkes er ordforråd, grammatikk, fonologi og verbal hukommelse og administreres av helsesykepleiere. Oppgavene i SPRÅK 4 er utformet for å være håndterbare for helsesykepleier og tiltalende for barn. Verktøyet er spesielt tilpasset konteksten for den obligatoriske 4 årskontrollen og er basert på utviklingsmessige beskrivelser av språktilegnelse. Verktøyet forholder seg ikke eksplisitt til noen spesifikk teoretisk modell. Administrasjonstiden er på omtrent 10-15 minutter.

Språkfyran

Språkfyran, et standardisert språkscreeningsinstrumentet for 4-åringer i Sverige ble utviklet som en del av landets helseovervåkningsprogram for førskolebarn. Instrumentet har som mål å identifisere barn med risiko for språkvansker og henvise dem til logoped for videre utredning og støtte. Instrumentet ble utviklet for å avhjelpe begrensningene ved tidligere screeninginstrumenter som primært fokuserte på ekspressiv fonologi, noe som resulterte i overhenvising av barn med fonologiske vansker og underhenvising av barn med ordforråds-, grammatikk- og forståelsvansker. Det nye screeninginstrumentet inneholder oppgaver som vurderer ulike aspekter ved språkforståelse og -produksjon på fonologisk, morfosyntaktisk, leksikalsk-semantic og pragmatisk nivå. nonord-repetisjon, som har blitt foreslått som en markør for språkvansker, er også inkludert i instrumentet. Det konstruerte språkscreeningsinstrumentet besto av 37 oppgaver, inkludert oppgaver knyttet til ord- og setningsforståelse, leksikalsk organisering, leksikalsk tilgang, repetisjon av ord, repetisjon av nonord og generell vurdering av tale- og språkferdigheter. Instrumentet hadde et binært skåringssystem (skår 1 eller 0) med en maksimal skår på 37. Språkscreeningsinstrumentet tar mellom 5 og 15 minutter å administrere. Det administreres av helsesykepleiere på helsestasjoner i Sverige.

Språkscreening ved 2,5 år

Screeningen består av to metoder: et spørreskjema som besvares av foreldrene, og direkte observasjon av helsesykepleieren. Barnet må være 2,5 år +/- 2 måneder på tidspunktet for screeningen, og før barnet vurderes, oppsummerer helsesykepleieren barnets tidligere historie når det gjelder språk- og taleutvikling. Screening utføres av helsesykepleier, som observerer og kommuniserer med barnet etter et strukturert skjema ved hjelp av ni lekeobjekter: en kopp, et fat, en skje, en dukke, en hund, en ku, en ball, en stor bil og en liten bil. Hensikten er å vurdere barnets evne til å forstå utsagn med ett og flere ord, imitere, navngi og bruke utsagn med to ord. Under vurderingen bruker helsesykepleieren en protokoll og transkriberer ord og setninger ortografisk. Screening tar 15 minutter. Under kartleggingen blir foreldrene bedt om å svare på et spørreskjema om blant annet barnets ordforråd, verbal forståelse, bruk av kompenseringer og oralmotoriske ferdigheter, og om de er bekymret for barnets tale og språk.

Språkscreening ved 3 år

Rutinescreeningen ved 3 år involverer en todelt prosess som inkluderer et foreldrespørreskjema og en direkte observasjon utført av helsesykepleieren. Denne screeningen har som mål å vurdere barnets evne til å forstå enkle instruksjoner og uttrykke seg ved å danne setninger med minst tre ord. Under besøket blir barnet oppmuntret til å svare på fem spørsmål basert på seks bilder, som inkluderer et bilde av en bil, ball, kopp, lue, eple og blyant. Barnet kan vise at de forstår ved å peke på bildene eller bruke kroppsspråk. De fleste barn vil naturligvis svare verbalt og demonstrere dermed også sine verbale ferdigheter. I tillegg er det akseptabelt å ha kommentarer og dialog med foreldrene som en del av vurderingsprosessen. For de barna som ikke er villige til å snakke i det hele tatt, tilbys alternativer som dukkehusmøbler, små dukker eller andre leker for å oppmuntre dem til å snakke. Dette tiltaket har vist seg å være effektivt for å få selv sjenerte og nølende barn til å uttrykke seg. I tillegg til å evaluere barnets evne til å forstå og uttrykke seg verbalt, vurderer helsesykepleieren også deres konsentrasjonsevne. Barn som ikke kan svare på minst tre av de fem spørsmålene og/eller som ikke klarer å danne setninger med minst tre ord, ifølge vurderingen fra helsesykepleieren og/eller foreldrene, blir henvist til en logoped for videre oppfølging.

I tabell 3 beskriver vi de ti verktøyene identifisert i forskningsspørsmål 1 og de 19 studiene (21 publikasjoner) som beskriver verktøyene. Vi fant og inkluderte flest studier

om ulike versjoner av Peabody Picture Vocabulary Test (6 studier omtalt i 7 publikasjoner) og Språkscreening ved 2,5 år (n=3). Videre rapporterte to studier om Språkscreening ved 3 år og to studier om Quick Interactive Language Screener. For de resterende verktøyene var det kun én studie (n=6) som målte egenskapene ved verktøyet, men studien som omtale SPRÅK 4 ble omtalt i 2 publikasjoner.

Totalt antall deltakere kombinert for alle 19 studiene er i underkant av 10 000. For hvert kartleggingsverktøy er det store variasjoner på hvor mange barn verktøyene har blitt testet på, fra 165 på Expressive Vocabulary Test, til 4302 på Språkscreening ved 3 år. Aldersspennet på barna er mellom 0-7 år, men majoriteten er mellom 2-4 år gamle. Det er noe variasjon når det gjelder om studiene inkluderer flerspråklige barn eller ikke, og i flere av studiene er ikke dette beskrevet eller det er uklart. Det fremkommer tydelig i syv studier at de inkluderer flerspråklige barn i utvalget.

Åtte av studiene er publisert i tidsspennet 2001-2012 og elleve er publisert mellom 2013 og 2023. Majoriteten av studiene er fra USA (n=9), etterfulgt av Sverige (n=6) og Storbritannia (n=2). Videre indentifiserte vi fant én studie hver fra Canada og Norge. Vi indentifiserte ingen studier fra Danmark, Finland, Island, Australia, Nederland eller Sveits.

Tabell 3: Beskrivelse av verktøyene indentifisert i forskningsspørsmål 1 som det finnes forskning om (10 verktøy), og publikasjonene som beskriver verktøyene (n=21)

Verktøy (tool)	Studier	Antall barn	Alder barn (år; mnd)	Testet på flerspråklige?
Brigance Infant and Toddler Screen (BITS)	Glascoc (2002) (97)	408	0-2 år	Ja
Early Language Identification Measure (ELIM)	Law (2023) (105)	262	2;0-2;6 år	Ja*
Expressive Vocabulary Test (EVT)	Thomas-Tate (2006) (118)	165	Gjennomsnitt 4;3 år	Ikke beskrevet
New Reynell Developmental Language Scales (NRDLS)	Letts (2013) (106)	Pilot: 301 Standardisering: 1266	Pilot: 1;6 og 7;6 år Standardisering: 2;00-7;6 år	Ja*
Peabody Picture Vocabulary Test ** (PPTV)	Dollaghan (2009) (91) Finneran (2020) (93) Platt (2012) (111) Restrepo (2006) (113) † Spaulding (2013) (116) Webb (2008) (121) Wood (2019) (124) †	1658	2;6-6 år	Ja i Platt (2012)
Quick Interactive Language Screener (QUILS) ***	de Villiers (2021) (89) Pace (2022) (60)	551	3;0- 6;9 år	Ja i de Villiers (2021)
SPRÅK 4	Klem (2015) (100) † Klem (2016) (101) †	600. OBS: kun 206 ble testet med testene utenom språk 4	3;9- 5;0 år	Nei

Språkfyrn	Lavesson (2018) (104)	328	3;10 – 4;5 år	23 % av barna i studien var tospråklige
Språkscreening ved 2,5 år	Nayeb (2019) (57) Nayeb (2020) (109) Mattsson (2001) (51)	321	2;5-2;7 år	Ja* i Nayeb (2019) og Nayeb (2020)
Språkscreening ved 3 år	Westerlund (2001) (122) Westerlund (2007) (123)	4302	3 år	Ikke beskrevet

*Uklar andel, men verktøyet er testet på en populasjon hvor 50 % eller flere barn er flerspråklige, ** Verktøyet har flere ulike versjoner som blir vurdert, *** Verktøyet har også en spansk-engelsk versjon som blir vurdert. ¥ = publikasjoner med samme studieutvalg. † = publikasjoner med samme studieutvalg.

Igjen, 32 verktøy ble vurdert for relevans av faggruppen og vi søkte spesifikt etter disse verktøyene i databasesøket. Vi fant forskning om ti av verktøyene. For tre av verktøyene som ble relevansvurdert av faggruppen (The Renfrew Language Scales og Well-Comm (Early Years), CSBS DP), ble det imidlertid i etterkant vurdert at disse likevel ikke ville passe i en norsk helsestasjonssetting. Årsaken var at de enten viste seg å være for omfattende, fordi testen ble benyttet som et verktøy for videre utredning fremfor screening, eller fordi det var nødvendig at administrasjonen av verktøyet ble tatt opp-tak av og deretter transkribert som grunnlag for skåring av testen. Grunnen til at dette ikke ble fanget opp i forkant var at vi hadde lite tilgjengelig informasjon om verktøyene på tidspunktet da listen ble sendt til faggruppen.

For de resterende 19 verktøyene som faggruppen vurderte som relevante identifiserte vi ingen studier som beskrev minst ett mål på verktøyets validitet eller reliabilitet. Disse omtales derfor ikke videre i rapporten. Disse verktøyene er:

- Breuer-Weuffen-differensieringstest for førskolebarn
- British Picture Vocabulary Scale II,
- E.R.I.K. - Early Referral Identification Kit
- Edmonton Narrative Norms Instrument (ENNI)
- KAM-test
- Lene
- Looksee Checklist (tidligere Nipissing District Developmental Screen)
- Reynell Developmental Language Scales
- Rourke baby record
- SATS-2
- Schedule of Growing Skills
- Snøgubbstestet (Språk- og talscreening)
- Sprogvurdering 2
- Sprogvurdering 3-6
- Språk 5-6
- Språkundersøkning 18 måneder
- Stoke Speaks Out (Staged Pathway Toolkit)
- The Speech and Language Pathology Early Screening Instrument (SLPESI)
- Tribus

Verktøyenes måleegenskaper i henhold til COSMIN

I tabell 4 viser vi de ti verktøyene og hvilke måleegenskaper som blir målt i studiene vi inkluderte. Den vanligste rapporterte måleegenskapen for verktøyene var kriterievaliditet. Det var rapportert flest måleegenskaper om BITS (4 egenskaper). For NRDLS, PPVT, QUILS, Språkfyrn og SPRÅK 4 ble det rapportert tre måleegenskaper. For Språkscreening ved 3 år ble det rapportert to måleegenskaper. For de tre resterende verktøyene ble det kun rapportert én måleegenskap i henhold til COSMIN. Ingen verktøy er evaluert mht. responsivitet. Måleegenskaper evaluert per verktøy presenteres nedenfor, mens det presenteres per publikasjon i vedlegg 7 og 8. I vedlegg 8 er det mer detaljert beskrivelse av resultatene og det gjengis tallmateriale knyttet til reliabilitets- og validitetsmål.

Tabell 4: Oversikt over evaluerte måleegenskaper for verktøyene identifisert i forsknings spørsmål 1 som det finnes forskning om (10 verktøy)

Verktøy	Reliabilitet				Validitet				Responsivitet
	1	2	3	4	5	6	7	8	
BITS	x	x	x					x	
ELIM								x	
EVT							x		
NRDLS	x	x						x	
PPVT		x					x	x	
QUILS	x	x						x	
SPRÅK 4	x						x	x	
Språkfyrn	x		x					x	
Språkscreening ved 2,5 år								x	
Språkscreening ved 3 år			x					x	

1. Intern konsistens, 2. Test-retest (over tid), 3. Inter-rater (ulike personer på samme tid), 4. Intra-rater (samme person ved forskjellig tid), 5. Målefeil, 6. Innholdsvaliditet, 7. Begrepsvaliditet (strukturell, konvergent, kryss kulturell), 8. Kriterievaliditet.

Resultater om måleegenskaper ved hvert kartleggingsverktøy

Under presenterer vi resultater for hvert av de ti verktøyene slik de er rapportert i de inkluderte publikasjonene. Verktøyene presenteres i alfabetisk rekkefølge, og der det er flere publikasjoner som evaluerer samme verktøy presenteres publikasjonene i kronologisk rekkefølge. Vi presenterer resultatene overordnet for å bedre lesbarheten. For mer detaljerte resultater og tallmateriale se vedlegg 8.

Brigance Infant and Toddler Screen (BITS)

Glascie og medarbeidere (2002) beskrev utviklingen, standardiseringen og valideringen av Brigance Infant og Toddler Screen (BITS) (97). Den består av to deler; Brigance Infant Screen for barn i alderen 0 til 11 måneder og Brigance Toddler Screen for barn i alderen 12 til 23 måneder. Studien inkluderte 408 barn, hvorav 52 % var gutter, i alderen 2 uker til 24 måneder. BITS tester finmotorikk, reseptivt og ekspressivt språk,

grovmotorikk, selvhjelp og sosioemosjonell utvikling. Av de 408 barna ble 71 (17 %) også testet med andre språktester, inkludert Bayley Scales of Infant Development, Receptive-Expressive Emergent Language Test, Preschool Language Scale og Rosetti Infant Toddler Language Scale. Studien målte intern konsistens, test-retest og interrater reliabilitet, sensitivitet og spesifisitet. Forfatterne rapporterte at verktøyet kan administreres av en rekke fagpersoner ved hjelp av foreldreintervju eller/og direkte observasjon. Forfatterne konkluderte med at BITS er et verdifullt verktøy for helsepersonell som effektivt kan identifisere utviklingsproblemer hos små barn, men at kriterievaliditeten for det reseptive språkområdet alene ikke var tilfredsstillende.

Early Language Identification Measure (ELIM)

Law og medarbeidere (2023) utviklet og evaluerte kartleggingsverktøyet Early Language Identification Measure (ELIM) (105). ELIM brukes til å identifisere hvilke barn i alderen 24-36 måneder som har svakere språkferdigheter enn hva som er forventet i aldersgruppen. Verktøyet består av fem måleseksjoner som vurderer ulike aspekter med barns språkutvikling. Disse evaluerte forfatterne separat. De undersøkte også hvilken kombinasjon av de fem seksjonene som best kunne predikere barnas språkferdigheter. Verktøyet ble utviklet for å bli brukt av helsesykepleiere i lokalsamfunnet. 376 barn fra fem geografisk og demografisk ulike områder i England ble rekruttert. Flere av barna kom fra familier med lav sosioøkonomisk status og ved starten av studien hadde 25 % av barna avvik fra normal forventet språkutvikling. Studien testet ELIM opp mot et standardisert referansemål for språk, Preschool Language Scale-5 (PLS-5), ved 24-30-månedersundersøkelsen hos helsesykepleier i England. Måleegenskapene sensitivitet, spesifisitet, positiv prediktiv verdi og negativ prediktiv verdi ble regnet separat for ELIM sine fem seksjoner. Forfatterne rapporterte at kombinasjonen av den foreldrerapporterte sjekklisten for ordforråd og helsesykepleiers observasjon av barnets kommunikasjon og atferd var den optimale kombinasjonen for å forutsi språkferdigheter. Forfatterne konkluderer med at ELIM nøyaktig identifiserer barn med risiko for avvik fra normalforventet språkutvikling, og at verktøyet har potensial til å tidlig oppdage barn med svake språkferdigheter.

Expressive Vocabulary Test (EVT)

Thomas-Tate og medarbeidere (2006) ønsket å evaluere den kulturelle validiteten til Expressive Vocabulary Test (EVT) for å vurdere ordforrådsferdighetene til afroamerikanske førskole- og barnehagebarn (118). Studien omfattet totalt 165 barn med en gjennomsnittsalder på 4,23 år. Deltakerne kom fra mellomstore sentrale og perifere bysamfunn. EVT ble administrert til barna, sammen med andre standardiserte tester som Peabody Picture Vocabulary Test – Third edition (PPVT-III). Forfatterne rapporterte at barnas EVT-skårer ikke var nevneverdig lavere enn det standardiserte gjennomsnittet, var normalfordelte og positivt korrelert med skårene på PPVT. Forfatterne konkluderte med at EVT er et valid verktøy for å vurdere ordforrådsferdigheter hos denne spesifikke gruppen barn.

New Reynell Developmental Language Scales (NRDLS)

Letts og medarbeidere (2013) utviklet og beskrev New Reynell Developmental Language Scales (NRDLS) for vurdering av små barns språkforståelse og språkproduksjon (106). NRDLS ble testet ut på 301 barn og deretter standardisert på et utvalg på 1266 deltakere i aldersgruppen 2;00-7;06 år fra ulike regioner i England. Deltakerne ble vurdert til å ha en typisk utvikling og engelsk som førstespråk, men noen av barna hadde også erfaring med andre språk. Referansetester var Test for Reception of Grammar (TROG) og British Picture Vocabulary Scales (BPVS). Studien måler test-retest reliabilitet, validitet, sensitivitet og spesifisitet. Studien rapporterte at NRDLS har høy spesifisitet når det gjelder å identifisere barn med språkvansker, men lavere sensitivitet når det gjelder å identifisere barn som tidligere har blitt diagnostisert med språkvansker. Det rapportertes også om en lav test-retest-reliabilitetskoeffisient. Forfatterne konkluderte med at NRDLS er et verdifullt verktøy for å vurdere språktilegnelse og språkvansker hos små barn. De understreket likevel viktigheten av å bruke standardiserte tester, som NRDLS, i kombinasjon med andre kartleggingsverktøy og prosedyrer for å få en helhetlig vurdering av barnets språkferdigheter.

Peabody Picture Vocabulary Test (PPVT)

Hele seks studier (7 publikasjoner) evaluerte ulike versjoner av Peabody Picture Vocabulary Test. Vi presenterer hver studie under.

Restrepo og medarbeidere (2006) og ønsket å finne ut om Peabody Picture Vocabulary Test-Third Edition (PPVT-III) var egnet til å evaluere barns språkutvikling (113). Studien omfattet 210 barn i førskolealder fra en sørøstlig delstat i USA. Barnas alder varierte fra 4;0-5;2 år ($M = 4;6$, $SD = 0;5$). Utvalget bestod av 50 % gutter og 50 % jenter, 57,6 % afroamerikanske barn og 42,4 % europeisk-amerikanske barn. Ni prosent av barna i det totale utvalget hadde diagnostiserte særskilte behov. PPVT-III og Expressive Vocabulary Test (EVT), ble brukt til å vurdere barnas ordforråd. Studien gjennomførte en (Differential Item Functioning) DIF-analyse for å finne ut om det var noen oppgaver i PPVT-III som fungerte som kulturelt betingede oppgaver for barna i studien. Forfatterne rapporterte at DIF-analysen ikke identifiserte noen skjevheter, som indikerer at det ikke var noen tegn på kulturell skjevhet i PPVT-III. Studien konkluderte med at PPVT-III var hensiktsmessige mål for å evaluere ordforrådet hos afroamerikanske barn og barn med lavt utdanningsnivå hos mor. EVT-skårene kan imidlertid være mer representative for barnas reelle ordforrådsnivå i disse populasjonene.

Webb og medarbeidere (2008) ønsket å undersøke validiteten til PPVT-III for afroamerikanske barn sammenlignet med europeisk-amerikanske barn (121). Deltakerne bestod av samme studiepopulasjon som i Restrepo (2006), og var 210 barn (104 gutter og 106 jenter) som gikk på en offentlig førskole i to skoledistrikter i en sørøstlig delstat i USA. Barna var 4 og 5 år (gjennomsnitt 4 år og 6 måneder), hvor 57,6 % av barna var afroamerikanere og 42,4 % var europeisk-amerikanske barn. Studien utførte DIF-analyse for å undersøke om ulike testdeltakere, som afroamerikanske og europeisk-amerikanske barn, svarer ulikt på spesifikke testspørsmål, til tross for å ha de samme latente

evnene eller ferdighetene som testen måler. Forfatterne rapporterte om minimale forskjeller mellom afroamerikanske- og europeisk-amerikanske barn ved hjelp av DIF-analysen. Hovedforskjellen mellom barna i disse latente klassene kunne først og fremst tilskrives bruken av en bestemt svarstrategi. Forfatterne konkluderte med at de ved hjelp av analysene fant ut at det som skilte de to gruppene barn fra hverandre i henhold til ulike skårer, var hvordan de brukte en spesifikk svarstrategi, kalt N3C, snarere enn at oppgavene i testen var kulturelt skjeve.

Dollaghan og Campbell (2009) ønsket å undersøke om språkskåre ved 3 og 4 år predikerte språkskåre ved 6 år (91). Studiepopulasjonen bestod av 414 amerikanske barn som ble vurdert ved alderen 3, 4 og 6 år. Tre språktester ble brukt: Peabody Picture Vocabulary Test-Revised (PPVT-R), gjennomsnittlig lengde på morfemer og en nonord-repetisjonsoppgave. Studien rapporterer inter-rater reliabilitet, kriterievaliditet og prediktiv nøyaktighet. Forfatterne rapporterte at en lav skåre ved 3 år ikke økte risikoen for en lav skåre ved 4 eller 6 år for noen av målene, men en lav PPVT-R-skåre ved 4 år økte risikoen for en lav PPVT-R-skåre ved 6 år. Forfatterne rapporterer imidlertid at selv når korrelasjonene er så store, kan man ikke gå ut fra at tidligere skårer har en prediktiv verdi for det enkelte barn; til tross for den høye korrelasjonen mellom PPVT-skårer ved 4 og 6 år, hadde barn med lave skårer ved 4 år bare en marginalt økt sjanse for lav skår 2 år senere. Forfatterne konkluderer med at for disse generelt friske barna og disse språkmålene var den prediktive nøyaktigheten på individnivå generelt dårlig i alderen 3, 4 og 6 år.

Platt (2012) gjennomførte en studie med mål om å evaluere og undersøke eventuelle skjevheter i PPVT-III som et verktøy for å vurdere ordforrådsutviklingen hos flerspråklige barnehagebarn (111). Studien sammenlignet prestasjonene til barn med engelsk som morsmål (EL1) og barn med engelsk som andrespråk (ELL). Studiegruppen bestod av 129 barnehagebarn i alderen 46 til 70 måneder, der 59 % var ELL-barn og 41 % EL1-barn med forskjellig språkbakgrunn, inkludert koreansk, urdu, farsi og mandarin. Forfatteren rapporterte om betydelige forskjeller i de generelle standardpoengene mellom EL1- og ELL-barn, med bedre prestasjoner fra EL1-barna. Resultatene av en DIF-analyse antydte minimal kulturell skjevhet i testen mellom de to barnegruppene. Forfatteren konkluderte derfor med at forskjellene i testresultatene reflekterte faktiske forskjeller i det generelle reseptive ordforrådet, heller enn skjevheter i selve testen. Forfatteren konkluderte med at PPVT-III kan være egnet til å måle ordforrådsutvikling hos språklig mangfoldige barnehagebarn, men at det imidlertid er behov for videre forskning med et større utvalg og flere ulike språkgrupper for å undersøke testens mulige skjevheter og validitet.

Spaulding og medarbeidere (2013) ønsket å sammenligne den diagnostiske nytten av to versjoner av Peabody Picture Vocabulary Test (PPVT-III og PPVT-IV) for å identifisere spesifikke språkvansker hos barn (116). Studien inkluderte 40 førskolebarn med spesifikke språkvansker og 40 kontroller med typisk utvikling. Deltakerne var matchet med hensyn til alder, kjønn og sosioøkonomisk status, og representerte en rekke etnisiteter. Studien målte sensitivitet og spesifisitet. Forfatterne rapporterer en sterk sammenheng mellom prestasjonene på de to testene, men 35 % av barna presterte ulikt på de to versjonene. Forfatterne rapporterte at barn med spesifikke språkvansker presterte signifikant dårligere enn barn med typisk utvikling på begge testene, og en grenseverdi på 103 ble funnet å være optimal standardskår for begge testene som

kunne skille mellom barn med og uten spesifikke språkvansker for begge testversjonene. Forfatterne konkluderer med at selv om det er en sterk sammenheng mellom de to testversjonene, er det ikke sikkert at de er utskiftbare for mer enn 1/3 av barna. Forfatterne argumenterer også rundt misoppfatningen om at nyere testversjoner alltid er bedre til å identifisere språkvansker enn tidligere versjoner av samme instrument.

Wood og Schatschneider (2019) ønsket å undersøke nøyaktigheten og skjevhetene i Peabody Picture Vocabulary Test-Fourth Edition (PPVT-4) for spansk-engelske barn med to språk (124). Studien inkluderte 278 tospråklige-elever som gikk i barnehage (n=142) eller første klasse (n=136) og hadde engelsk som andrespråk, samtidig som minst ett familiemedlem snakket spansk hjemme. Studien analyserte faktorer som tverrspråklig overlapping, orddeling og ordfrekvens for å finne ut hvordan de påvirket nøyaktigheten av svarene på PPVT-4. Forfatterne rapporterte at de overnevnte faktorene var signifikante prediktorer for sannsynligheten for å svare riktig på testen. Forfatterne understreker viktigheten av å være oppmerksom på mulige skjevheter og utvise forsiktighet når man tolker ordforrådstester. Forfatterne konkluderte med at det er viktig å ta hensyn til skjevheter og de unike språklige erfaringene til flerspråklige barn når man vurderer ordforrådsferdighetene deres.

Finneran og medarbeidere (2020) undersøkte hvordan kulturell og språklig bakgrunn påvirker prestasjonene til afroamerikanske og latinamerikanske barn fra lavinntektsfamilier på Peabody Picture Vocabulary Test-Fourth Edition (PPVT-4) (93). Studien inkluderte 332 afroamerikanske og latinamerikanske barn fra lavinntektsfamilier i alderen 30 til 75 måneder. PPVT-4 ble administrert for å vurdere barnas ordforrådsfærdigheter. Studien rapporterte om klinisk signifikante forskjeller mellom afroamerikanske og latinamerikanske barn når det gjaldt nøyaktigheten på 14 PPVT-4-testoppgaver. Forfatterne konkluderte med at det er mulige kulturelle og språklige skjevheter i PPVT-4-oppgavene som ikke kan forklares av inntekt alene for unge latinamerikanske barn. De understreket behovet for alternative former for vurdering av språk og ordforråd som er mindre utsatt for medfødte skjevheter i forhold til kulturelt og språklig mangfoldige barn.

Quick Interactive Language Screener (QUILS)

Pace og medarbeidere (2022) ønsket å evaluere Quick Interactive Language Screener (QUILS) når det gjelder å identifisere språkvansker i to utvalg, et klinisk utvalg (studie 1) og et befolkningsutvalg (studie 2) (60). I studie 1 besto utvalget av totalt 63 barn i alderen 3;0- 6;9 år. Deltakerne var enspråklig engelsktalende. Studie 2 besto av 126 deltakere. Av disse ble 101 deltakere klassifisert som typisk utviklede, og 25 deltakere ble klassifisert til å ha språkvansker. Kjønnfordelingen var balansert. Utvalgets etnisitet bestod av afroamerikanske, hvite og barn med blandet etnisitet. Utvalget hadde middels og lav sosioøkonomisk status. QUILS screener og Structured Photographic Expressive Language Test (SPELT-3, SPELT P2), Diagnostic Evaluation of Language Variation – Norm Referenced (DELV-NR Syntax subtest), PPVT-IV og Expressive Vocabulary Test – 2nd Edition (EVT-2) ble brukt som instrumenter i studie 1, og QUILS og Preschool Language Scales 5th Edition (PLS-5) ble brukt som instrumenter i studie 2. Diskriminantanalyser ble gjennomført for å undersøke klassifiseringsnøyaktigheten til QUILS. Forfatterne rapporterte at QUILS hadde lav klassifiseringsnøyaktighet i studie 1

sammenlignet med studie 2, noe som tyder på at det kun bør brukes som screeninginstrument i sammenhenger som er tilnærmet lik populasjonsnivåene for diagnostiserte språkvansker. Forfatterne konkluderte med at det er viktig å bruke flere sammenfallende måleinstrumenter som en del av en omfattende evaluering for å kunne identifisere barn med språkvansker, og at det å reteste borderline-barn på et senere tidspunkt kan øke sannsynligheten for nøyaktig identifisering.

Quick Interactive Language Screener: English-Spanish (QUILS:ES)

de Villiers og kollegaer (2021) ønsket å evaluere et nyutviklet web-basert verktøy for tospråklige barn (89). Forfatterne fremhever problemer som ofte oppstår ved språkkartlegging av for flerspråklige barn og hvordan disse problemene kan bli løst med det nylig utviklede kartleggingsverktøyet QUILS:ES. Forfatterne påpeker at for å kartlegge språkferdighetene til flerspråklige barn må de bli vurdert på begge språkene de snakker, for å få et bilde av de faktiske språkferdighetene. QUILS:ES ble utviklet for å måle språkferdighetene til barn som både snakker engelsk og spansk. I denne studien var det 362 engelsk-spansk talende barn mellom 3-5 år, hvor halvparten var gutter. Omtrent 80 % av barna kom fra familier med lav sosioøkonomisk status og de resterende hadde middels sosioøkonomisk status. For å undersøke samtidig validitet ble PLS-5 brukt som sammenligningstest. PLS-5 måler ekspressivt og reseptivt språk. 44 barn ble testet med både den engelske og spanske versjonen av PLS-5. Videre ble 29 andre barn analysert med å sammenligne QUILS:ES med språkkartleggingsverktøyet BESOS. BESOS er et språkkartleggingsverktøy for barn mellom 4-7 år hvor subtestene både er på spansk og engelsk. Forfatterne rapporterer moderat til høy korrelasjon mot de andre språkverktøyene. Videre konkluderer forfatterne med at instrumentet har god intern reliabilitet, test-retest reliabilitet og validitet. De fremhever også at testen så langt ikke er testet på en klinisk populasjon med barn med språkproblemer.

Språk 4

To publikasjoner ser på måleegenskaper ved det norske screeningverktøyet SPRÅK 4, og benytter samme utvalg. Utvalget er et kombinert utvalg av fireåringer som hadde deltatt i to ulike studier av SPRÅK 4. Hele utvalget omfattet 600 barn i alderen 47-60 måneder med en gjennomsnittsalder på 52,3 måneder og 54 % gutter. Studien rekrutterte ikke spesifikt barn med språkvansker, men en liten andel (9,7 %) av utvalget ble identifisert av helsesykepleiere som barn med risiko for språkforsinkelse ved studiesstart. 206 av barna ble også testet med British Picture Vocabulary Scale-II (BPVS-II), deltesten Picture Naming fra Wechsler Preschool and Primary Scale of Intelligence-III (WPPSI-III), Test for Reception of Grammar-2 (TROG-2) og deltesten Grammatic Closure fra Illinois Test of psycholinguistic Abilities (ITPA), som til sammen utgjør det latente språkkriteriet.

Klem og medarbeidere (2015) undersøkte begrepsvaliditeten til SPRÅK 4 for språkferdigheter hos fireåringer (100). Forskerne brukte konfirmerende faktoranalyse (CFA) for å modellere den interne strukturen i SPRÅK 4 og fastslå dimensjonaliteten. Forfatterne konkluderte med at screeningverktøyet SPRÅK 4 har god begrepsvaliditet og kan

brukes til å identifisere barn med risiko for språkforsinkelse. De påpekte også at verktøyet kan brukes til å følge språkutviklingen over tid og til å evaluere effekten av språkintervensjoner. De understreket imidlertid at det er behov for ytterligere forskning på andre aspekter knyttet til screeningverktøyet før man kan trekke endelige konklusjoner om dets anvendelighet i praksis.

Klem og medarbeidere (2016) undersøkte deretter stabiliteten og utviklingen av språkferdigheter i førskolealder ved hjelp av SPRÅK 4 i samme utvalg som beskrevet over (101). Studien viste at SPRÅK 4 identifiserte en gruppe barn med svake prestasjoner som viste vedvarende språklige svakheter fra 4 til 6 år. Forfatterne rapporterte at resultater på SPRÅK 4 er en sterk samtidig prediktor for det latente språkkriteriet ved 4 års alder og viser en høyst signifikant indirekte sammenheng med språkprestasjoner to år senere. Forfatterne rapporterer at disse resultatene bekrefter de tidligere funnene til Klem og medarbeidere (100) om at resultater på SPRÅK 4 forklarer en betydelig del av det latente språkkriteriet ved 4 år. Forfatterne konkluderte med at tidlig identifisering av språkforsinkelser ved hjelp av SPRÅK 4 kan bidra til å identifisere barn som kan ha nytte av tidlig intervensjon og støtte.

Språkfyrn

Lavesson og medarbeidere (2018) utviklet og evaluerte Språkfyrn, et nytt språkscreeningsinstrument for svenske 4-åringer (104). Instrumentet ble konstruert basert på aktuell forskning og kunnskap om prediktorer og kliniske markører for språkvansker. Instrumentet inneholdt ulike oppgaver og ble brukt på en stor populasjon for å vurdere testegenskapene. Studien omfattet en populasjon på 352 barn i alderen 46-53 måneder som ble screenet av 11 helsesykepleiere i forbindelse med den ordinære 4-årskontrollen. Det endelige utvalget besto av 328 barn, hvorav 23 % var flerspråklige. Fem logopeder gjennomførte kartleggingen ved hjelp av et gullstandardisert språktestbatteri (kortversjonen av Test for Reception of Grammar (TROG-2), GramBa og en oppgave fra den nevrologvistiske kartleggingen NELLI). Forfatterne rapporterte at screeninginstrumentet viste tilfredsstillende reliabilitet og akseptabel sensitivitet og spesifisitet. Instrumentet hadde god validitet, og det teoretiske grunnlaget bidro til validiteten. Instrumentet var enkelt å administrere og skåre for helsesykepleiere, og inter-rater reliabiliteten var høy. Forfatterne konkluderte med at språkscreeningsinstrumentet var gyldig for å identifisere barn som trenger videre utredning av en logoped.

Språkscreening ved 2,5 år

Tre publikasjoner evaluerer språkscreening ved 2,5 års alder i Sverige.

Mattson og medarbeidere (2001) ønsket å evaluere språkscreeningsprogrammet for 2,5-åringer (51). Studien omfattet 105 2,5 år gamle barn rekruttert fra 24 helsestasjoner i Sverige. Screeninginstrumentene som ble brukt i studien, inkluderte et spørreskjema til foreldrene og vurderinger foretatt av helsesykepleiere og logopeder. Tjuefem av barna hadde et positivt screeningresultat, og hver av disse ble sammenlignet med 1-4 kontrollbarn med negativt screeningresultat, som ble testet samme dag på samme helsestasjon og valgt ut av helsesykepleier. Referansetestene ble utført samme dag som språkscreeningen hos 66 barn (63 %), innen 1 måned hos 95 barn (90 %) og innen 2

måneder hos alle barna. Screeningprogrammet ble evaluert ved hjelp av positive og negative prediktive verdier, sensitivitet og spesifisitet. Referansetesten benyttet var Reynell språktest. Forfatterne rapporterte at resultatene med hensyn til både sensitivitet og spesifisitet var lovende. Forfatterne konkluderer med at screeningprogrammet effektivt identifiserte barn med utviklingsmessige språkvansker ved 2,5 års alder.

Nayeb og medarbeidere (2019) ønsket å finne ut om den rutinemessige språkscreeningsmetoden som brukes ved det svenske treårsbesøket, kunne utføres ved 2,5 år med opprettholdt sensitivitet, spesifisitet og prediktive verdier (57). De ønsket også å se om det å lempe på kriteriet for ordkombinasjoner fra tre til to ord ville forbedre testens nøyaktighet når det gjelder sensitivitet, spesifisitet og prediktive verdier i denne alderen. Bakgrunnen for dette: På 1990-tallet ble det utviklet to ulike screeningprogrammer, ett ved 2,5 år og ett ved 3 år, for å identifisere språk- og kommunikasjonsvansker i Sverige. Alle svenske helsestasjoner tok i bruk ett av disse to programmene. I 2016 ble det innført et nytt nasjonalt barnehelseprogram som innebar at en lege og en helsesykepleier skulle gjennomføre en felles vurdering ved tre år, noe som førte til at det ikke lenger var tid til rutinemessig språkscreening ved tre år. For å tilpasse seg dette flyttet alle sentrene språkscreeningen fra tre til 2,5 år. Fordi treårsscreeningen hadde en svært standardisert og enkel struktur, og for å spare kostnader forbundet med å iverksette nye rutiner, håpet man at formatet som ble brukt i treårsscreeningen, kunne tilpasses for barn som var 2,5 år gamle. Dette ville kreve et noe lempeligere kriterium for ordproduksjon og en klinisk vurdering som inkluderte både moderate og alvorlige funksjonsnedsettelse. I studien ble 105 svensktalende barn i alderen 2;5-2;7 år observert av totalt ti helsesykepleiere ved tre svenske helsestasjoner. Studien så på måleegenskapene positiv- og negativ predikert verdi, sensitivitet og spesifisitet. Forfatterne konkluderte med at en endring av screeningmetoden for 2,5-åringene var effektiv når det gjaldt å oppdage mulig språkforstyrrelse, noe som ga fordelen av tidligere oppdagelse og forberedelse for foreldrene. Det er imidlertid behov for ytterligere oppfølgingsstudier for å validere disse funnene.

Nayeb og medarbeidere (2020) ønsket å finne den mest nøyaktige metoden for å identifisere språkutviklingsforstyrrelser hos tospråklige barn i alderen 2,5 år (109). Studien sammenlignet fire screeningmodeller: screening på barnets morsmål, screening på svensk, screening på begge språk og en kombinasjon av screening på svensk og foreldrenes informasjon om barnets språkutvikling på morsmålet. Deltakerne i studien var tospråklige barn på 2,5 år og deres foreldre. Studien rekrutterte barn fra tre helsestasjoner i Gävle, Sverige. Deltakerne kom fra familier med lav sosioøkonomisk status. Screeningmodellene ble vurdert ved hjelp av direkte vurderinger av språkforståelse og språkproduksjon på begge språkene som barnet brukte. Studien så på måleegenskapene positiv- og negativ predikert verdi, sensitivitet og spesifisitet. Forfatteren rapporterte at kun screeningmodellen med direkte vurdering på begge språk oppnådde tilstrekkelig diagnostisk nøyaktighet, og konkluderer med at tospråklige barn bør screenes på begge sine språk. Dette er spesielt viktig for barn fra familier med lav sosioøkonomisk status som lever i komplekse språkmiljøer.

Språkscreening ved 3 år

Westerlund (2001) ønsket å evaluere den prediktive validiteten til en global vurdering av barns taleferdigheter foretatt av helsesykepleiere på helsestasjonen ved 3-årsalderen og vurdere nytten av denne vurderingen ved screening for språkvansker ved 4-årsalderen (122). Studien inkluderte 2243 barn ved 3-årsalderen som gjennomgikk en formalisert språkscreening og en global vurdering av barnas tale- og språkferdigheter, hvilket innebærer en vurdering av barnas taleferdigheter ved hjelp av en visuell analog skala (VAS). Ved 4-årsalderen gjennomgikk de samme barna en ny vurdering for å identifisere språkvansker. Den gjennomsnittlige VAS-verdien for den globale vurderingen av tale og språk ved 3-årsalder blant barna som besto 3-årsscreeningen, var signifikant forskjellig fra resultatet for barna som ikke besto screeningen. Forfatterne rapporterte at en global vurdering i forbindelse med 3-årsscreeningen ville være nyttig for selektiv screening av 4-åringene. Det ble målt moderat sensitivitet og spesifisitet for VAS terskelnivå på ≤ 50 mm. Forfatterne konkluderte med at den globale vurderingen er et nyttig verktøy for å redusere sykepleierens arbeidsmengde og identifisere barn som trenger videre utredning for språkvansker. Studien tyder på at kostnadene ved en generell 4-årsscreening kanskje ikke står i forhold til nytten, og at sykepleierens globale vurdering i tillegg til 3-årsscreeningen gir en relevant indikasjon på barnets evner i minst ett år.

Westerlund og Sundelin (2007) ønsket å vurdere hvor godt 3-årsscreeningen og, hvis tilgjengelig, medisinske journaler, predikerte utfall ved 4-årsscreeningen (123). Ved 3-årsscreeningen ble språkforståelse og språkproduksjon vurdert av over 60 helsesykepleiere. Deltakerne i studien var barn i 3- og 4-årsalderen, og totalt ble 2237 barn vurdert. Fireårsscreeningen besto av helsesykepleierens vurdering av fonologi og et spørreskjema til foreldrene. Studien målte inter-rater reliabilitet, sensitivitet, spesifisitet og positiv predikert verdi. Studien konkluderte med at screeningprosedyren er et verdifullt verktøy for å identifisere barn med risiko for alvorlige språkvansker.

Kartleggingsverktøy ikke vurdert av faggruppen (identifisert via litteratursøk)

Utover de forhåndsidentifiserte verktøyene, dvs. verktøyene identifisert i forsknings spørsmål 1, identifiserte vi via litteratursøk ytterligere 14 verktøy, evaluert mht. måleegenskaper i 25 studier beskrevet i 26 publikasjoner. Faggruppen vurderte ikke disse verktøyene for relevans. Imidlertid hadde vi tett kommunikasjon med oppdragsgiver om flere av verktøyene for nettopp å ta hensyn relevans. Disse 14 verktøyene er:

- Bayley Scales of Infant and Toddler Development-III (BSID-III)
- Bilston and Ettingshall Sure Start (BESS) Speech and Language Therapy Two Year Developmental Check
- Computer-Based Instrument for Low Motor Language Testing (C-BiLLT)
- Dutch well child language screening protocol
- First word test (FWT)
- Fluharty Preschool Speech and Language Screening Test - Second Edition (Fluharty-2)
- Fox Language inventory (FLI)

- Grammar and Phonology Screening (GAPS)
- Nonword Repetition Test (NRT)
- Preschool repetition test (PSrep Test)
- Short language measure (SLaM)
- Structured screening test
- Test for Sentence Development (TSD)
- Toddler phonology test (TPT)

I tabell 5 og 6 beskriver vi de 14 verktøyene som ble identifisert via litteratursøk. Først, i tabell 5 gis en overordnet beskrivelse av hvert verktøy, inkludert hvilke språkdome-ner som det dekker og hvor lang tid verktøyet tar å administrere.

Tabell 5: Beskrivelse av verktøyene identifisert via litteratursøk (14 verktøy)

Verktøy	Kort beskrivelse
Deltest i: Bay-ley Scales of Infant and Toddler Development-III (BSID)	Bayley Scales of Infant and Toddler Development-III (BSID-III) er et mye brukt kartleggings-verktøy for å evaluere sped- og småbarns utvikling. Det gir verdifull innsikt i barnets kogni-tive, motoriske og språklige utvikling. Skalaene kan brukes på barn i alderen 1 til 42 måne-der. BSID-III består av deltester som spesifikt vurderer reseptive og ekspressive språkfærdig-heter. Den reseptive språktesten vurderer barnets evne til å forstå talespråk, følge instruks-joner og reagere på verbale stimuli. Her presenterer testerer ulike oppgaver for barnet, for eksempel å identifisere kroppsdeler, følge enkle instruksjoner og peke på bilder som svar på talte ord. Deltesten ekspressivt språk evaluerer barnets evne til å kommunisere verbalt, in-kludert bruk av gester, ord og setninger. Her blir barnet bedt om å delta i aktiviteter som å navngi bilder, imitere talelyder og bruke ekspressivt språk for å beskrive scener eller hendel-ser.
Bilston and Ettingshall Sure Start (BESS) Speech and Language Therapy Two Year Developmental Check	Verktøyet BESS Speech and Language Therapy Two Year Developmental Check er en vurdering basert på foreldrerapport og direkte observasjon. Verktøyet består av 9 elementer for å vurdere tale- og språkfærdigheter, hvorav fem er basert på direkte observasjon og fire på foreldrerapport. Disse er delt inn i 3 delskalaer; Forståelse, Uttrykk og Tale. Direkte ob-servasjon: Barnet blir bedt om å følge enkle instruksjoner, delta i aktiviteter basert på lek og peke på og navngi ulike gjenstander og bilder ved hjelp av de lekene og bøkene som er til-gjengelige. Administratoren må observere barnets handlinger og svar og vurdere hvert av disse elementene enten i henhold til barnets evne til å fullføre oppgavens komponenter eller basert på foreldrerapport hvis barnet ikke responderer. Foreldrerapport: foreldrene gir ek-sempler på ord eller handlinger som demonstrerer barnets ferdigheter.
Computer-Based Instru-ment for Low motor Lan-guage Testing (C-BiLLT)	C-BiLLT, Computer-Based Instrument for Low Motor Language Testing, er et verktøy som er opprinnelig utviklet for å vurdere talespråkforståelse hos barn med alvorlige motoriske funk-sjonsnedsettelse, spesielt barn med cerebral parese. Testen administreres via en datama-skin og gir mulighet for ulike responsmetoder, for eksempel berøringsskjerm, inngangsbry-tere, rullestolhodestøtter, øyestyring og uavhengig øyestyring. Gir normdata for alderen 1;6 til 7;0 år, med mulighet for vurdering av eldre barn med mistanke om utviklingshemming. Testen består av 86 oppgaver som dekker substantiv, verb, dyr, objekter, personer og set-ninger med økende grammatisk kompleksitet. Visuell tilbakemelding gis for hver oppgave. Fleksibiliteten i tilgangsmetodene gjør C-BiLLT egnet for vurdering av barn med alvorlige motoriske utfordringer, og gir en omfattende evaluering av deres talespråkforståelse.
Dutch well child language screening pro-tocol	Den nederlandske språkscreeningprotokollen for 2-åring er en praksisbasert protokoll som er iverksatt i den nederlandske forebyggende helsetjenesten for barn. Formålet med proto-kollen er å bistå i identifisering av barn med mulige språkproblemer. Protokollen består av oppgaver og spørsmål som vurderer barnets oppnåelse av milepæler i språkutviklingen, for eksempel evnen til å bruke to-ordsetninger og forstå enkle, vanlige substantiver. Opplærte barneleger og barnesykepleiere administrerer screeningen, som innebærer å observere bar-nets atferd og stille standardiserte spørsmål til foreldrene. Protokollen har en klar grense-verdi og anvisninger for henvisning dersom ønsket atferd ikke blir observert under undersøkelsen.

First word test (FWT)	<p>First Words Test (FWT) er et verktøy utviklet for å evaluere barns tidlige ordforråd og språkferdigheter. Verktøyet er utviklet for bruk med barn i alderen 16 til 30 måneder, det vil si i den tidlige fasen av språkutviklingen. FWT består av to hovedkomponenter: Ordforståelse (vurderer barns evne til å forstå ord ved å presentere dem med visuelle bilder og spørre dem om å peke på riktig bilde som svarer til det hørte ordet; og ordproduksjon (måler barns evne til å uttale ord ved å be dem om å si ordet som svarer til det de ser på bildet. FWT kan brukes både av fagfolk og ufaglærte. FWT tar vanligvis et sted mellom 10 og 20 minutter å administrere. Tiden kan variere avhengig av barnets samarbeidsevne, oppmerksomhets-spenne, og antall ord som blir evaluert.</p>
Fluharty Pre-school Speech and Language Screening Test-Second Edition (FPSLST-2)	<p>Fluharty Preschool Speech and Language Screening Test-Second Edition (FPSLST-2) er et kartleggingsverktøy som er utviklet for å screene og evaluere tale- og språkferdigheter hos barn i førskolealder. Det er utviklet for å identifisere mulige tale- og språkforsinkelser eller -forstyrrelser hos barn mellom 3 år og 6 år og 11 måneder. FPSLST-2 består av seks deltester som hver er utviklet for å vurdere spesifikke aspekter ved barnets tale- og språkferdigheter: Artikulasjon, gjentakelse av setninger, følge direktiver, svare på spørsmål, beskrive handlinger og sekvensering av hendelser. I motsetning til andre tester har ikke FPSLST-2 en definert basalverdi eller et tak. Det betyr at alle deltestene gjennomføres for å gi et helhetlig bilde av barnets ferdigheter, og testeren vurderer prestasjonene på alle delprøvene. Det tar vanligvis mellom 15 og 20 minutter å gjennomføre alle de seks deltestene i FPSLST-2.</p>
Fox Language inventory (FLI)	<p>FLI er utviklet for screening av tale og språk ved 36 måneders alder. FLI er basert på det finske språket og er et mye brukt screeninginstrument i barnehager og i helsevesenet generelt. Det måler 8 språklige delferdigheter: benevning, språkforståelse, å følge instruksjoner, forståelighet av barnets tale, fortelling, morfologi, tall- og fargekunnskap og setningslengde. Screeningmaterialet består av 8 svart-hvitt strektegnede bilder, små leker og 8 fargetegnede bilder som viser hovedpersonen Fox i hverdagssituasjoner. FLI gir skårer på delområder og en totalskår for barnets språkkompetanse. De normative dataene for FLI kommer fra 100 friske, enspråklige finstaltende barn i alderen 3;0-3;6 år.</p>
Grammar and Phonology Screening (GAPS)	<p>GAPS ble utviklet for å gi en rask screeningstest for å vurdere om barn som er i ferd med å starte skolen har nødvendige grammatikk- og fonologiske ferdigheter for utdanning og sosial utvikling. GAPS-testen tar 10 minutter å gjennomføre og har to deltester: Fonologi (gjentakelse av ikke-ord) og Grammatikk (gjentakelse av setninger). Den er ment for bruk med barn i alderen 3 år 6 måneder til 6 år 6 måneder. Testen for gjentakelse av setninger inneholder 11 testsetninger. Instruksjonene i GAPS-manualen forklarer hvilke deler av setningene som må gjentas nøyaktig for en korrekt score. Forskeren informerer barnet om at de skal fortelle en historie til karakteren 'Bik'. Bik er et romvesen som bare forstår barn. Forskeren sier deretter setningene, som barnet gjentar nøyaktig til Bik. Et eksempel på setninger som finnes i GAPS-testen er: 'katten ønsket litt melk.' Testen for gjentakelse av ikke-ord inneholder åtte testelementer. Når 'historien' (gjentakelse av setninger) er fullført, informerer forskeren barnet om at de skal prøve å si noen av Bik's ord. Forskeren leser opp ikke-ordene for barnet å gjenta. Et eksempel på ikke-ord som finnes i GAPS-testen er 'tobilf'. Elementene må gjentas nøyaktig for en korrekt score. GAPS-testen tar ikke hensyn til barn som ikke kan uttale noen språkllyder; konsekvente artikulasjonsfeil markeres som feil. Persentilscore kan beregnes for hver deltest. Intern konsistens for grammatikkdelen er $\alpha = 0,858$ og for fonologidelen $\alpha = 0,729$.</p>
Nonword Repetition Test (NRT)	<p>Nonword Repetition Test (NRT) er en test for å sjekke språkferdigheter hos små barn. I testen gjentar barn nonord av forskjellig lengde og vanskelighetsgrad. Å gjenta nonord måler barnas korttidsminne for lyder, og resultatene kan endre seg når ordene blir lengre. Å gjøre slike oppgaver krever også evner innen å høre, tolke og si lyder, noe som varierer fra barn til barn. NRT-testen inneholder nonord av ulik lengde og er fonetisk enkle, med noen få vanskelige lyder. For å unngå påvirkning fra ordforråd, blir nonordene laget slik at de ikke tilsvarer ord på det språket de administreres på.</p>
	<p>Hver av de inkluderte publikasjonene bruker litt ulike versjoner av NRT, og ulikhetene presenteres for hver publikasjon.</p>

Preschool Repetition Test (PSRep)	Preschool Repetition Test (PSRep) er en av to tester som inngår i Early Repetition Battery, som er et britisk utviklet og standardisert testbatteri, basert på et nasjonalt stratifisert utvalg av barn, som vurderer fonologiske og morfosyntaktiske prosesseringsferdigheter hos barn i alderen 2 til 6 år. Her vurderes barnas gjenkjenning, gjenkalling og produksjon av virkelige ord og oppdiktete ord. Barna blir presentert for en dukke og bedt om å kopiere det dukken sier. Etter to øvingsoppgaver blir barnet presentert for 18 ord og 18 fonologisk matchende nonord. Det gis ett poeng for hvert ord som gjentas korrekt, med et maksimum på 36 poeng. Lengden på ord og nonord er likt fordelt mellom en-, to- og trestavelsesord. Testen tar 10-15 minutter å administrere
Short language measure (SLaM)	The Short language measure (SLaM) er et verktøy utviklet for å raskt evaluere et barns språklige ferdigheter i deres første skoleår og avgjøre om de trenger ytterligere språkvurdering. Det tar omtrent 10 minutter å gjennomføre og vurdere, og kan administreres av en logoped eller en utdanningsfaglig person. Vurderingen består av 40 testelementer delt inn i to språkopp-gaver. Barn får ett poeng for hvert element de svarer riktig på. Alle elementer blir administrert, og den endelige poengsummen blir beregnet ved å legge sammen poengene for alle elementene.
Structured screening test	Den strukturerte screeningtesten inneholdt 20 spørsmål som dekket både ekspressive og reseptive språkferdigheter, og var tidligere validert på lokalbefolkningen. Det ble gitt ett poeng for hvert spørsmål, og en poengsum på under 10 ble ansett som ikke bestått. Ellers lite informasjon om testen.
Test for Sentence Development (TSD)	Test for Sentence Development er en deltest i et nytt nederlandsk språkproduksjonsbatteri kalt Schlichting Test for Language Production. Den tar sikte på å vurdere den syntaktiske kunnskapen i språkproduksjonen hos små barn i alderen 1,8 til 6,3 år. Testen bruker funksjonell imitasjon som elisiteringsteknikk, der barna blir bedt om å imitere setninger med ulike syntaktiske strukturer. Testen kan gjennomføres på under 30 minutter, og skåringen tar ca. 10 minutter. Testen inneholder en rekke oppgaver som holder barna motiverte og engasjerte. Testen kan administreres av opplærte testadministratorer.
Toddler phonology test (TPT)	Toddler Phonology Test (TPT) er et kartleggingsverktøy som er utviklet for å evaluere den fonologiske tilegnelsen hos barn i alderen 2;0-2;11 år. Den gjør det mulig å sammenligne barnets tale med normative data og identifisere typiske utviklingsmessige og avvikende feilmønstre. Barna blir bedt om å navngi 31 bilder som er rettet mot 37 ord, for å vurdere engelske konsonanter og vokaler, stavelsesstrukturer og klynger i ord på opptil fire stavelser. TPT kan avdekke barn som gjør flere talefeil enn forventet, har mindre fonetisk inventar og/eller gjør atypiske feil, og dermed identifisere småbarn med risiko for fonologiske forstyrrelser. TPT tar ca. 10 minutter å gjennomføre og administreres av klinikere.

I tabell 6 beskriver vi de 14 verktøyene identifisert via litteratursøk og de 25 studiene som beskriver verktøyene. Verktøyene som vi inkluderte flest studier om var ulike former av Nonword Repetition Test (n=5), Computer-Based Instrument for Low motor Language Testing (C-BiLLT) (n=4 publisert i 5 publikasjoner) og Grammar and Phonology Screening (GAPS) (n=3).

Totalt antall deltagere kombinert for alle studiene er i overkant av 5800. For hver enkelt test er det stor variasjon når det gjelder hvor mange barn verktøyene har blitt testet på, fra 48 barn (Toddler phonology test (TPT)) til 1121 barn (Computer-Based Instrument for Low Motor Language Testing (C-BiLLT)). Aldersspennet til barna er fra 1-8 år, men majoriteten av barna er under 6 år, hvor en stor andel av studieutvalget er mellom 2-4 år gamle. Det er stor variasjon i hvorvidt studiene inkluderer flerspråklige barn eller ikke, og i flere av studiene er det ikke beskrevet. Det fremkommer tydelig i fem studier at verktøyet er testet på flerspråklige barn, og for tre av disse er formålet med studien å teste verktøyet på nettopp flerspråklige barn.

Fjorten av studiene er publisert i tidsspennet 2003-2013, mens 11 studier er publisert mellom 2014 og 2023. Majoriteten av studiene kom fra Storbritannia (n=9), etterfulgt av Nederland (n=4), USA (n=3), Canada (n=3) og Australia (n=2). Videre inkluderte vi

av skandinaviske studier to norske, én svensk og én finsk. Videre ble utvalg fra både Australia og Storbritannia benyttet én studie. Vi identifiserte ingen studier fra Danmark, Island eller Sveits.

Tabell 6: Beskrivelse av verktøyene identifisert via litteratursøk (n=14), og studiene som beskriver verktøyene (n=26)

Verktøy (tool)	Studier	Antall barn	Alder barn (år; mnd)	Testet på flerspråklige?
Bayley Scales of Infant and Toddler Development-III	Woods (2014) (125)	129	3-5 år	Nei
Bilston and Ettingshall Sure Start (BESS) Speech and Language Therapy Two Year Developmental Check	Adamson-Macedo (2009a (84), 2009b (83))	396	2-2;6 år	Nei (2)
Computer-Based Instrument for Low Motor Language Testing (C-BiLLT)	Bootsma (2023) (86) Fiske (2018) (94) ¥, Geytenbeek (2014) (96) Haddeland (2018) (99) ¥, Skipar (2019) (115)	1121	1;6-8;6 år	ca 20% (1), Nei (3)
Dutch well child language screening protocol	Visser-Bochane (2020) (115)	247	2 - 3 år	Nei
First word test (FWT)	Pickstone (2003) (110)	97	1;7-1;9 år	Nei
Fluharty Preschool Speech and Language Screening Test - Second Edition (Fluharty-2)	Anderson (2016) (85) Eisenberg (2019) (92)	491	3-6 år	Ja (1), Nei (1)
Fox Language inventory (FLI)	Korpilahti (2016) (102)	226	3 år	Nei
Grammar and Phonology Screening (GAPS)	Gardner (2006) (95) Nash (2011) (108) van der Lely (2011) (120)	305	3-8 år	Nei (3)
Nonword Repetition Test (NRT)	Deevy (2010) (90) Gray (2003) (98) Radeborg (2006) (112) Stokes (2009) (117) Thordardottir (2013) (119)	622	2- 6år	Ja (1), Nei (4)
Preschool repetition test (PSrep Test)	Chiat (2007) (87)	315	2-4 år	13 %
Short language measure (SLaM)	Matov (2020) (107)	126	5;1- 6;7 år	Nei
Structured screening test	Laing (2002) (103)	1598	2;6 til 3;0 år	Uklart/Nei
Test for Sentence Development (TSD)	Schlichting (2003) (114)	149	1,3 - 6;3 år	Ja
Toddler phonology test (TPT)	Claessen (2017) (88)	n=18 Australia, n=24 Storbritannia	Australia: 2;1- 3;0 Storbritannia: 2;1	Nei

Ja = uklar andel, men verktøyet er testet på en populasjon hvor 50 % eller flere barn er flerspråklige; Nei = enten ikke beskrevet, uklart eller spesifisert at studiepopulasjonen ikke består av noen flerspråklige barn. ¥ = Samme utvalg

Verktøyenes måleegenskaper i henhold til COSMIN

I tabell 7 viser vi verktøyene og hvilke måleegenskaper som blir målt. Den vanligste måleegenskapen rapportert i publikasjonene er kriterievaliditet. For seks av verktøyene fantes det kun rapportering på én måleegenskap i henhold til COSMIN. For tre av verktøyene ble det rapportert to måleegenskaper og for ett verktøy ble det rapporterte tre måleegenskaper. For tre verktøy ble det rapportert fire måleegenskaper og for ett verktøy – C-BiLLT – ble det rapportert syv måleegenskaper i henhold til COSMIN. Ingen verktøy er evaluert mht. responsivitet. Måleegenskaper evaluert per verktøy presenteres nedenfor, mens det presenteres per publikasjon i vedlegg 7 og 8. I vedlegg 8 er det er mer detaljert beskrivelse av resultatene og det gjengis tallmateriale knyttet til reliabilitets- og validitetsmål.

Tabell 7: Oversikt over evaluerte måleegenskaper for verktøy ikke relevansvurdert av faggruppen (14 verktøy)

Verktøy	Reliabilitet					Validitet			Responsivitet
	1	2	3	4	5	6	7	8	
BSID								x	
BESS	x	x	x						
C-BiLLT	x	x	x	x	x		x	x	
Dutch well child language screening protocol								x	
FWT			x						
FPSLST-2							x	x	
FLI								x	
GAPS	x		x				x	x	
NRT	x	x	x					x	
PSRep	x	x	x					x	
SLaM	x							x	
Structured screening test								x	
TSD							x		
TPT		x						x	

1. Intern konsistens, 2. Test-retest (over tid), 3. Inter-rater (ulike personer på samme tid), 4. Intra-rater (samme person ved forskjellig tid), 5. Målefeil, 6. Innholdsvaliditet, 7. Begrepsvaliditet (strukturell, konvergent, kryss kulturell), 8. Kriterievaliditet

Resultater om måleegenskaper ved hvert kartleggingsverktøy

Under presenterer vi resultater for hvert verktøy slik de er rapportert i de inkluderte studiene. Verktøyene presenteres i alfabetisk rekkefølge, og der det er flere studier som

evaluerer samme verktøy, presenteres de i kronologisk rekkefølge. Vi presenterer resultatene overordnet for å bedre lesbarheten. For mer detaljerte resultater og tallmateriale se vedlegg 6.

Bayley Scales of Infant and Toddler Development-III (BSID-III)

Woods og medarbeidere (2014) ønsket å undersøke den prediktive validiteten av tidlig språktesting og stabilitet i språkutviklingen, fra 3 til 5 år, i en kohort av premature barn (125). Språkvurdering ble utført ved 3 år ved hjelp av språkdomenet i BSID-III, og Wechsler Preschool and Primary Scale of Intelligence-III (WPPSI-II) ved 5 år. Studiepopulasjonen bestod av 72 gutter og 57 jenter, og flertallet av barna ble født mellom 28 og 29 fullgatte svangerskapsuker, mens 14 % ble født før 26. svangerskapsuke. Medianalderen ved vurderingene var 37 måneder for BSID-III og 60 måneder for WPPSI-II. Forfatterne rapporterte at til tross for høy spesifisitet for reseptivt språk, var sensitiviteten dårlig. For ekspressivt språk var den diagnostiske nøyaktigheten relativt lav og stabiliteten i språkvanskenivåene mellom 3 og 5 år viste bare middels samsvar. Forfatterne konkluderte med at det er vanskelig å predikere språkutviklingen hos premature barn ved hjelp av tidlig språktesting. Studien understreker behovet for kontinuerlige utviklingsvurderinger utover førskolealder for å forbedre nøyaktigheten i identifiseringen av språkvansker.

Bilston and Ettingshall Sure Start (BESS) Speech and Language Therapy Two Year Developmental Check

Adamson-Macedo og medarbeidere (2009a) ønsket å evaluere den interne reliabiliteten til Bilston and Ettingshall Sure Start (BESS) Speech and Language Therapy Two Year Developmental Check - et screeningverktøy utviklet for å avdekke tale- og språkvansker blant britiske førskolebarn (83). Studiepopulasjonen bestod av 152 foreldre med to år gamle barn som var tilknyttet BESS studien, et utvalg med en skjevhet mot en lavere sosioøkonomisk profil. BESS-verktøyet besto av ni spørsmål kategorisert i tre delskalaer: Forståelse, uttrykk og tale. Forfatterne rapporterte at BESS-verktøyet viste god intern reliabilitet og alle elementene av verktøyet korrelerte godt med totalskåren og var kompatible med verktøyet. Ved sammenligning med andre verktøy, var den intern reliabilitet lavere enn MCDI, mens den var nesten identisk med Reynell språktest.

Forfatterne konkluderte med at BESS-verktøyet er et sensitivt screeningverktøy for å identifisere tale- og språkforsinkelser hos to år gamle barn, spesielt i vanskeligstilte grupper.

Den samme forfattergruppen hadde i en annen publikasjon samme år som mål å vurdere de psykometriske egenskapene til BESS-verktøyet for to år gamle barn, med fokus på både intern og ekstern reliabilitet (84). Totalt 244 foreldre og deres toåringer deltok i to kohorter, hvorav 134 gutter og 110 jenter. Den interne konsistensen ble undersøkt ved hjelp av Cronbachs alfa, Kuder-Richardson-koeffisient, item-helhet-korrelasjoner og inter-rater korrelasjoner. Ekstern reliabilitet ble undersøkt ved hjelp av en test-retest-metode og Nevill Percentage of Agreement Test. Forfatterne konkluderte med at

BESS-verktøyet er pålitelig på tvers av ulike testere og effektivt identifiserer tale- og språkvansker hos to år gamle barn.

Computer-Based Instrument for Low motor Language Testing (C-BiLLT)

Fem publikasjoner evaluerte verktøyet Computer-Based instrument for Low motor Language Testing (C-BiLLT), som måler forståelsen av talespråk.

Geytenbeek og medarbeidere (2014) utviklet og evaluerte C-BiLLT blant barn med typisk utvikling og barn med cerebral parese (96). Studien inkluderte 806 barn med typisk utvikling, med en gjennomsnittsalder på 4;4 år og 87 ikke-talende barn med cerebral parese, men vi rapporterer kun resultatene for barn med typisk utvikling. Studien rapporterer om reliabilitet, intern konsistens og validitet. Konvergent validitet ble undersøkt ved å sammenligne testskårene i C-BiLLT med anerkjente instrumenter; Receptive Dutch Language Scale (RDLS) og Peabody Picture Vocabulary Test (PPVT-III). Forfatterne konkluderte med at C-BiLLT har god reliabilitet (målt ved ICC og SEM) og god validitet som mål på forståelse av nederlandsk talespråk (målt ved faktoranalyse og korrelasjonskoeffisienter).

Fiske (2018) og Haddeland (2018) benyttet samme utvalg i sine publikasjoner, som evaluerte C-BiLLT. Selv om C-BiLLT er opprinnelig utviklet for barn med store tale- og bevegelsesvansker, har den også potensiale for bruk hos alle barn. Studien inkluderte 187 typisk utviklede barn i alderen 1 til 7 år, hvor halvparten var under 6 år, fra barnehager og skoler i Oslo. Barna ble tilfeldig delt i to testrekkefølger og gjennomførte tester for språkforståelse (C-BiLLT, Reynell språktest, TROG-2, og ordgjenkjenning fra WPPSI-IV) og visuospatiale evner (terningsmønster fra WPPSI-IV og Ravens fargematrise). Testingen ble utført på barnas skoler og barnehager.

Fiske (2018) gjennomførte en valideringsstudie av den norske versjonen av C-BiLLT og dens egnethet til å evaluere språkforståelse hos barn (94). Korrelasjonsanalysen viste høyere korrelasjon mellom C-BiLLT-sumskårer og andre tester for språkforståelse enn mellom C-BiLLT og tester som måler ikke-verbal resonnering. Faktoranalysen resulterte i to faktorer, vokabularforståelse og grammatisk forståelse, som høyt korrelerte og forklarte variasjon i alle C-BiLLT-seksjoner. C-BiLLT viste seg sensitiv for å skille mellom barn i ulike aldersgrupper, spesielt blant de yngre barna, og for barn med lav språkforståelse. Imidlertid var det usikkerhet knyttet til C-BiLLTs evne til å differensiere blant eldre barn og høytfungerende barn. Forfatteren konkluderte med at den norske versjonen av C-BiLLT er egnet for å evaluere språkforståelse hos barn i det angitte aldersspennet, men oppfordrer til forsiktighet ved slutninger om eldre barn inntil ytterligere testing er gjennomført.

Haddeland (2018) hadde som formål å undersøke reliabiliteten til den norske oversettelsen av C-BiLLT gjennom vurdering av testens indre konsistens, stabilitet og inter-rater reliabilitet (99). Forfatteren konkluderte med at reliabiliteten til den norske oversettelsen av C-BiLLT er generelt god, med høy indre konsistens og tilfredsstillende inter-rater reliabilitet. Test-retest reliabiliteten er imidlertid noe lav på grunn av et kort tidsintervall, men resultater antyder forbedring ved økt tidsintervall.

Skipar (2019) ønsket å undersøke om C-BiLLT egnet seg til å vurdere barns språkforståelse i aldersgruppen 1;0 til 3;0 år. Studien inkluderte 48 typisk utviklede barn i

denne aldersgruppen, hvorav 27 var gutter og 21 var jenter. Referanseverktøyet var Reynell språktest. Forfatteren rapporterte om høy intern reliabilitet målt med Cronbach's alpha. Videre rapporterte forfatteren om god kriterievaliditet mellom C-BiLLT og Reynell språktest. Forfatteren konkluderte med at C-BiLLT er et godt egnet verktøy for å måle barnets språkforståelse i aldersgruppen 1;0 til 3;0 år.

Bootsma og medarbeidere (2023) ønsket å undersøke de psykometriske egenskapene og gjennomførbarheten til den kanadiske-engelske versjonen av verktøyet (C-BiLLT-CAN) hos barn med cerebral parese og barn med typisk utvikling (86). Studien undersøker begrepsvaliditet (konvergent og diskriminant validitet), intern konsistens og test-retest-reliabilitet for C-BiLLT-CAN. Antall deltakere varierer for de ulike analysene. Studiepopulasjonen med barn med typisk utvikling bestod av 80 barn (45 jenter) i alderen 1,5 til 8,5 år, hvorav 50 av barna var under 6 år. Forfatterne rapporterte at både validiteten og reliabiliteten var høy, inkludert konvergent validitet, diskriminerende validitet, intern konsistens, test-retest reliabilitet og målefeil. Forfatterne konkluderte med at C-BiLLT-CAN viste gode til utmerkede psykometriske egenskaper og validitet i et utvalg av barn med typisk utvikling, noe som indikerer at det er en adekvat test for å måle språkforståelse hos engelsktalende kanadiske barn.

Dutch well child language screening protocol

Visser-Bochane og medarbeidere (2020) evaluerte den nederlandske språkscreeningsprotokollen for 2-åringer og effekten av avvik fra protokollen fra fagpersoners side (74). Studien inkluderte totalt 124 toåringer, hvorav 61 barn ble identifisert ved hjelp av språkscreeningsprotokollen og 63 barn ikke ble identifisert. Barna ble rekruttert fra fire helsestasjoner i urbane og rurale områder i Nederland. Studien hadde et longitudinelt design, med oppfølgingsundersøkelser når barna fylte 3 år. Språkscreeningsprotokollen som ble brukt i studien, inkluderte omfattende språktester og en foreldrerapport, om nødvendig, for å vurdere barnas språkferdigheter. Screeningprotokollens samtidige og prediktive validitet ble vurdert ved å beregne sensitivitet og spesifisitet i forhold til en referansestandard (Schlichting-tester for språkforståelse (SLC), ordproduksjon og setningsproduksjon ved 2- og 3-års alder). Studien viste at protokollen for 2-årig språkscreening hadde god samtidig- og prediktiv validitet. Fagpersonalet viste god etterlevelse av protokollen, og avvik fra protokollen endret ikke protokollens sensitivitet og spesifisitet. Forfatterne konkluderte med at screeningprotokollen kan være en støtte for helsepersonell og bør iverksettes i større grad. De antydte også at det er mulig å forbedre spesifisiteten for å gjøre protokollen egnet for populasjonsbasert screening. Forfatterne konkluderte videre med at språkscreeningsprotokollen tilfredsstillende avdekket nåværende og senere språkproblemer.

First word test (FWT)

Pickstone (2003) undersøkte om FWT er et passende verktøy for en marginalisert urban populasjon og om verktøyet er egnet for å brukes av paraprofesjonelle (110). Paraprofesjonelle er i denne studien, som er utført i Storbritannia, sykepleiere kvalifisert til å arbeide med barns utvikling. Flere av sykepleierne var rekruttert i lokalsamfunnet til

å jobbe med «Sure Start» programmet. «Sure start» er et offentlig områdebasert program med mål om å gi barn i marginaliserte områder best mulig forutsetninger fra livets start. Programmet er initiert av den britiske regjeringen og retter seg mot barn 0-4 år og deres familier. Screening av språk inngår som en del av dette programmet. Studien inkluderte 97 barn i alderen 19-21 måneder med engelsk som førstespråk. Språkkartleggingen fant sted under et rutinebesøk i hjemmet. Det ble brukt et foreldrerapportert skjema i tillegg til FWT. Studien har oppgitt sensitivitets- og spesifisitetmålinger for det foreldrerapporterte skjemaet og FWT samlet; resultatene kan derfor ikke si noe om sensitiviteten til FWT alene. For inter-rater reliabilitet, rapporterte forfatteren at første forsøk oppnådde en moderat inter-rater reliabilitet (enighet) mellom testerne. I andre forsøk var inter-rater reliabiliteten høyere. Forfatteren påpeker at de dårligere testskårene i forsøk 1 kan skyldes at personalet ikke hadde erfaring med kartleggingsverktøyet og manglet erfaring med å vurdere tale på barn i den aldersgruppen.

Fluharty Preschool Speech and Language Screening Test-2nd Edition (FPSLST-2)

Anderson (2016) undersøkte effekten av Fluharty Preschool Speech and Language Screening Test - Second Edition (Fluharty-2) som et mål på tale- og språkferdigheter hos First Nations-barn i Rainy River District i Canada (85). Det har tidligere blitt stilt spørsmål ved om standardiserte tale- og språktester er effektive når det gjelder å måle tale- og språkferdigheter hos First Nations-barn. Effektiviteten av standardiserte tale- og språktester ble undersøkt ved å sammenligne testresultatene til to grupper: First Nations (FN) og Non-First Nations (NFN) barn over en treårsperiode (2009-2011). Studien omfattet 429 barn: 314 NFN-barn (73,2 %) og 115 FN-barn (26,8 %). Det var nesten like mange gutter som jenter: 214 gutter og 215 jenter. De fleste barna (69,9 %) var fem år gamle, 102 barn (23,8 %) var fire år gamle, og 27 barn (6,3 %) var seks år gamle. Det var generelle forskjeller i skåringsmønstre, der NFN-barn presterte signifikant bedre enn FN-barn. I studien ble det brukt DIF-analyse for å undersøke prestasjonsforskjeller mellom FN- og NFN-barn på spesifikke testoppgaver, samtidig som det ble kontrollert for den generelle testprestasjonen. Forfatterne rapporterte at seks fonologiske oppgaver viste små til moderate nivåer av uniform og ikke-uniform DIF, og én språkoppgave viste et moderat nivå av ikke-uniform DIF. Forfatterne konkluderte med at forskjellene i resultatene ikke var grunnet prestasjonsforskjeller mellom gruppene, men ble forklart med fraværet av visse lyder i Ojibwe-språket som snakkes i Rainy River-distriktet.

Eisenberg og medarbeidere (2016) ønsket å evaluere Fluharty-2 som verktøy for å screene språkferdigheter hos 3-åringer (92). Studien involverte 62 barn, som ble delt inn i to grupper: de som ikke besto testen og de som besto den. Disse barna ble rekruttert fra barnehager, samt barneleger og logopedier i en forstad til New Jersey, i USA. Fluharty-2 ble sammenlignet med Structured Photographic Expressive Language Test-Preschool, Second Edition (SPELT-2); mean length of utterance in morphemes (MLUm), finite verb morphology composite, og Index of Productive Syntax (IPSyn). Resultatene fra studien indikerte begrenset sammenfall mellom resultatene fra Fluharty-2 og de diagnostiske målingene. De barna som ikke besto Fluharty-2, presterte betydelig dårligere på hver av de diagnostiske testene sammenlignet med de som besto Fluharty-2, og skårene på Fluharty-2 var signifikant korrelert med skårene på de diagnostiske målene.

Studien konkluderte med at Fluharty-2 viste begrenset samsvar med de diagnostiske målene for bestått/ikke bestått for alle fire målinger. Forfatterne reiste også tvil om validiteten til to av de diagnostiske testene, SPELT-2 og MLU_m, for å diagnostisere språkproblemer hos 3-åringer. Videre konkluderte de med at Fluharty-2 mest sannsynlig ville forårsaket henvisning av både for få barn med risiko og for mange uten risiko til en videre vurdering, noe som gjør det til et ineffektivt verktøy for rutinescreening av språkferdigheter.

Fox Language inventory (FLI)

Korpilahti og medarbeidere (2016) ønsket å evaluere effekten av tidlig språkscreening for å identifisere barn med mulig språkforsinkelse, ved å sammenligne to ulike screeninginstrumenter og to alderspunkter (102). Deltakerne i studien var barn fra STEPS-studien, en stor befolknings basert kohortstudie gjennomført i Finland. Utvalget i denne studien besto av 226 barn som tilfeldig var valgt ut fra den intensive oppfølgingsgruppen i STEPS-studien. STEPS er en populasjonsbasert studie som undersøker barns psykiske, sosiale og fysiske utvikling fra graviditet til tenåringsalder. Barna var 36 måneder gamle på tidspunktet for språkvurderingen, hvorav 120 var gutter og 106 var jenter. De to screeninginstrumentene som ble brukt i studien, var foreldrerapporterte MacArthur Communicative Development Inventories (CDI) ved 24 måneder og Fox Language Inventory (FLI) ved 36 måneder. Sistnevnte tar 10-15 minutter å gjennomføre og utføres av sykepleier. Utfall knyttet til «word finding» og språkforståelse på 36 måneder gamle barn ble sammenlignet med The Renfrew Word Finding Vocabulary Test og Reynell Developmental Language Scales III (language comprehension). Forfatterne konkluderte med at screening med FLI ved 36 måneder hadde noe høyere sensitivitet og spesifisitet enn CDI. Begge screeninginstrumentene var i stand til å identifisere barn som trengte spesiell støtte i språkutviklingen. Testskårene fra FLI korrelerte med sammenligningstestene med unntak av mål for setningslengde. Forfatterne konkluderte med at FLI er et pålitelig verktøy for språkscreening.

Grammar and Phonology Screening (GAPS)

Tre studier evaluerte Grammar and Phonology Screening (GAPS).

Gardner og kollegaer (2006) presenterer den teoretiske bakgrunnen, pilotstudien, påliteligheten og standardiseringen av testen Grammar and Phonology Screening (GAPS), som er utviklet for å kartlegge språkvansker og spesifikke lesevansker hos barn (95). Deltakerne i studien inkluderte totalt 668 barn i alderen 3;4 til 6;8 år fra ulike regioner i Storbritannia. Utvalget var nøye utvalgt for å kontrollere for regional beliggenhet, sosioøkonomisk status og inkluderte barn fra etniske minoriteter med engelsk som førstespråk. GAPS-testen besto av 11 testsetninger og åtte nonord for direkte imitasjon. Referansetestene var British Picture Vocabulary Scale - Revised (BPVS), to deltester fra Clinical Evaluation of Language Fundamentals - Pre-School (CELF) og The Children's Test of Nonword Repetition (CN-Rep). Forfatterne rapporterte at testen hadde god/svært god intern konsistens. Funnene viste også at bare halvparten av de barna i normalpopulasjonen som har en eller annen form for språkvanske, har både

fonologiske og grammatiske vansker. Testerens yrke forklarte bare en liten del av variasjonen i dataene. Korrelasjonsanalyser ble utført for å vurdere samsvar mellom to GAPS-deltester og referansetestene. GAPS-testen viste de sterkeste korrelasjonene med standardiserte språktester som tar opp lignende grammatikk- eller fonologikomponenter, og forble signifikante selv etter at det var tatt hensyn til alder. Forfatterne rapporterer at GAPS er rask å administrere, og med hell kan administreres av både profesjonelle og ikke-profesjonelle. De konkluderer med at testen identifiserer barn med risiko for språk- og lese- og skrivevansker i like stor grad som barn som er i fare for å utvikle språk- og lese- og skrivevansker.

Nash og kollegaer (2011) undersøkte om GAPS kunne identifisere barn med språk vansker i et tilfeldig utvalg med barn (108). 106 britiske barn i alderen 3-6 år fullførte GAPS testen samt sammenligningsverktøyene Early Repetition Battery (ERB) og språkskalene i CELF-Preschool 2 (clinical evaluation of language fundamentals-preschool 2). I henhold til konvergent validitet hadde begge GAPS sine to subkategorier (grammatikk og fonologi) positiv og signifikant korrelasjon med majoriteten av kategoriene i sammenligningstestene. Forfatterne rapporterte at GAPS identifiserte 2 av 10 barn med språkproblemer identifisert av CELF-Preschool 2. Dette gav lav sensitivitet. Spesifisiteten rapporterte de derimot til å være høy. GAPS hadde lignende måleegenskaper når verktøyet ble sammenlignet med foreldrebekymringer. Forfatterne konkluderer med at GAPS kan bli brukt av et bredt utvalg fagpersonell, men har er ikke et sensitivt screeningsverktøy for forskere.

van der Lely og medarbeidere (2011) gjennomførte en studie med mål om å evaluere GAPS-testens presisjon i å identifisere barn med språkvansker (120). Studien omfattet 51 deltakere, delt inn i to aldersgrupper: 3;6-6;6 år. Den ene gruppen bestod av barn uten språkforstyrrelser (n = 30), den andre av barn med språkforstyrrelser (n = 11). I tillegg inkluderte studien en tredje gruppe med barn i alderen 6;9-8;11 år og språkforstyrrelser (n = 10), som lå over testens aldersnormer. Forfatterne rapporter betydelig lavere poengsum på språktestbatteriet for barn med språkforstyrrelser sammenlignet med de med typisk utvikling. Ved den 5. persentilen oppnådde den 100 % sensitivitet og spesifisitet, nøyaktig skille mellom barn med og uten språkvansker. Selv ved en cut-off på 10 % opprettholdt den høy nøyaktighet og identifiserte 98 % av barna korrekt. Videre analyser ved hjelp av ROC (receiver operating characteristics) bekreftet GAPS-testens nøyaktighet. ROC-analysen inkluderte begge deltestene og viste at GAPS-testen var svært nøyaktig i å klassifisere barn med og uten fonologiske vansker. Forfatterne konkluderte med at resultatene indikerte at GAPS-testen har høy korrelasjon med andre språktester og er et svært nøyaktig screeningverktøy for å identifisere barn med nedsatte grammatiske og/eller fonologiske evner.

Nonword Repetition Test

Fem studier evaluerte ulike versjoner av nonword repetition tester (NRT).

Gray (2003) hadde som mål å undersøke den diagnostiske nøyaktigheten og test-retest-reliabiliteten til nonordrepetisjons- og sifferspennviddeoppgaver administrert til førskolebarn med spesifikke språkvansker (SSV) (98). I studien deltok 22 barn med SSV og 22 barn med normal språkutvikling (NS). Hvert NS-barn ble valgt ut for å matche et SSV-barn med hensyn til kjønn og alder (± 3 måneder). Alle barna var mellom

4,0 år (år, måneder) og 5,0 år (år, måneder) og snakket en standard dialekt av engelsk som hovedspråk ifølge foreldre- og lærerrapporter. Hver gruppe besto av 5 jenter og 17 gutter. Referansetesten var SPELT-III. Halvparten av øktene var barn med SSV og halvparten med NS. Forfatterne konkluderte med at skårene for repetisjon av nonord ga utmerket sensitivitet og spesifisitet for å skille mellom grupper. Imidlertid ble det funnet uakseptable nivåer av test-retest-reliabilitet for NS-gruppen på nonordrepetisjonsoppgaven. Forfatterne konkluderer videre med at disse foreløpige funnene tyder på at nonordrepetisjon er et lovende diagnostisk mål for SSV hos førskolebarn, med ytterligere forbedringer for å forbedre test-retest-reliabiliteten.

Radeborg og medarbeidere (2006) utviklet og evaluerte en svensk NRT blant førskolebarn (112). Studien inkluderte 200 barnehagebarn i alderen 4-6 år, hvor 57,5% var under 6 år. Testen ble konstruert for å vurdere språkutvikling og språkferdigheter, etter som fonologisk korttidsminne har vist seg å korrelere med disse ferdighetene. Testen går ut på å presentere ukjente nonord av ulik lengde for barna og vurdere deres evne til å gjenta dem korrekt. Nonordene ble konstruert i henhold til svenske fonotaktiske regler, og det ble kun brukt kombinasjoner av fonemer som er mulige på svensk, slik at nonordene skulle være enkle for barna å uttale. Ingen av nonordene tilsvarte et svensk ord, og likhet med virkelige ord ble unngått. Testen tar rundt 10-12 minutter å administrere. Forskerne presenterte 30 nonord av ulik lengde for barna og evaluerte deres evne til å gjenta dem korrekt. Basert på resultatene ble en forkortet versjon av testen bestående av 24 nonord vurdert som egnet for denne aldersgruppen. Forskerne fant at lengre nonord var vanskeligere for barna å gjenta, og observerte signifikante effekter av alder og dialektkompatibilitet mellom testerens og barnets. De viste også til samsvar mellom uavhengige skåringer som ble utført av de to testerne hver for seg. Videre ble det målt intern konsistens på 30 item-testen og 24 item-testen. Siden det ikke så ut til å være noen fordeler ved å bruke 30-item testen, anbefale de å bruke 24-item testen, som omfatter to-, tre-, fire- og femstavelses nonord. Det viste seg at manglende kompatibilitet mellom testerens aksent og den aksenten barnet er mest vant til, kan ha en negativ effekt på barnets prestasjoner, spesielt når det gjelder yngre barn. Avslutningsvis konkluderte studien med at den svenske NRT er et verktøy for å identifisere mulige språkvansker hos førskolebarn, noe som muliggjør tidlig intervensjon og støtte.

Stokes og Klee (2009) ønsket å undersøke nøyaktigheten av en nyutviklet test, Test of Early Nonword Repetition (TENR), for å diagnostisere språkvansker hos 2 år gamle barn (117). TENR inneholder 12 en-, to- og trestavelses nonord (4 av hver type) bestående av konsonanter og vokaler i tidlig utvikling. Ordene ble presentert i en bestemt rekkefølge fra én til tre stavelser. Barnet ble bedt om å imitere ikke-ordene som forskslederen sa, og fikk deretter lov til å trille en ball ned en rutsjebane som belønning. Barnet fikk beskjed om å «si det jeg sier, og så skyve ballen ned». Oppgaven ble introdusert ved at barnet ble bedt om å «si bamse», og deretter begynte oppgaven. Hver riktig konsonant eller vokal ble belønnet med et poeng, og den totale prosentandelen riktig ble beregnet. Studien inkluderte 232 britisk-engelsktalende barn i alderen 27 (± 3) måneder. Deltakerne ble vurdert med CDI som referansetest og to versjoner av TENR (1-3 stavelser og 1-4 stavelser). Studien målte sensitivitet og spesifisitet. Forfatterne konkluderte med at nytten av en NWR-test for å indikere språkforsinkelse må undersøkes nærmere hos både yngre og eldre barn, og at foreløpige resultater fra TENR 1-4-stavelsestesten ga gode resultater for yngre barn.

Deevy og medarbeidere (2010) evaluerte nøyaktigheten til NRT når det gjelder å identifisere spesifikke språkvansker, og vurderte om det er mulig å bruke den i universell screening av fire- og femåringer (90). Studien inkluderte totalt 92 amerikanske barn, hvorav 29 barn med spesifikke språkvansker og 47 barn uten språkvansker. Gjennomsnittsalderen på barna var 5,0 år (46 gutter og 30 jenter), majoriteten var hvite og alle var enspråklig engelsktalende. Referansetesten som ble brukt, var Structured Photographic Expressive Language Test, II (SPELT-II). Studien rapporterte om tilstrekkelig god interrater-reliabilitet, sensitivitet, spesifisitet og negativ- og positiv prediktiv verdi. Forfatterne konkluderte med at den diagnostiske nøyaktigheten til NRT for fire- og femåringer er lik den som er observert i studier med eldre barn. NRT alene er imidlertid ikke nok som språkscreeningsverktøy til å identifisere barn i den generelle befolkningen som kan være i risikozonen for språkvansker.

Thordardottir og Brandeker (2013) ønsket å undersøke effekten av tospråklig eksponering og språkvansker på skårene for nonordrepetsjon (NRT) og setningsimitasjon (119). De to studiene ble gjennomført med 5 år gamle fransk-engelske tospråklige barn. Målet med den første studien var å undersøke effekten av ulike grader av tospråklig eksponering på simultant tospråklige barns evne til å utføre NRT- og setningsimitasjonsoppgaver på hvert sitt språk. Studien omfattet 84 5-åringer med typisk utvikling (alder 52-69 måneder), hvorav 16 var enspråklige engelsktalende, 19 enspråklige fransktalende og 49 tospråklige. Resultatene viste at NRT-skårene ikke var sterkt påvirket av tidligere eksponering, og tospråklige barn med varierende eksponering oppnådde høye NRT-skårer uavhengig av lengden på nonordene. Den andre studien vurderte den diagnostiske nøyaktigheten av NRT- og setningsimitasjonsoppgaver hos enspråklige og tospråklige barn med og uten primære språkvansker. Studien omfattet fire grupper med 14 barn i hver gruppe: (1) tospråklige barn med primære språkvansker, (2) tospråklige barn med typisk språkutvikling, (3) enspråklige barn med primære språkvansker og (4) enspråklige barn med typisk språkutvikling. Alle gruppene var like gamle, med en gjennomsnittsalder på nesten 5 år. Forfatterne rapporterte at NRT oppnådde høy sensitivitet og identifiserte korrekt 85 % av barna som tidligere var identifisert som primære språkvansker, uavhengig av om barna var tospråklige eller enspråklige, og med samme grenseverdi for begge grupper. Spesifisiteten var høy for enspråklige barn, men falt til bare middels for tospråklige barn. Forfatterne konkluderer med at testen er effektiv til å identifisere språkvansker hos både enspråklige og mange tospråklige barn. Videre konkluderer forfatterne med at NRT-oppgaver kan fungere som et diagnostisk verktøy for å identifisere primære språkvansker hos tospråklige barn, og de understreker betydningen av språkeksponering for prestasjonene.

Preschool Repetition Test

Chiat og Roy (2007) undersøkte måleegenskapene til Preschool Repetition test (PSRep) (87). De gjennomførte to studier hvor de i studie 1 så på prestasjonsspennet hos normalt utviklede barn og variabler som påvirket prestasjonen, og i studie 2 sammenlignet skårene med barn som var klinisk henvist grunnet bekymring for barnets språkutvikling. Studien inkluderte 315 britiske barn med normal språkutvikling mellom 2;0- 4 år, og 168 klinisk henviste barn. Sistnevnte var mellom 2;8- 4 år og majoriteten var gutter. Tretten prosent av det normalspråklige utvalget besto av flerspråklige barn og for det kliniske utvalget var omtrent 1 av 4

flerspråkelige. I studie 1 ble reliabilitet målt og forfatterne rapporterer at testen er robust og skårer høyt på inter-rater reliabilitet, intern konsistens og test-retest reliabilitet. Forfatterne så også på samtidig validitet hvor PSRep testen viste delvis korrelasjon med sammenligningstesten British Picture Vocualurary Scale (BPVS). Forfatterne rapporterte at testprestasjonen for normalspråkelige barn ikke ble påvirket av kjønn og sosioøkonomisk status. I studie 2 ble måleegenskapene for inter-rater agreement rapportert. Denne rapporter forfatterne til å være høy og innenfor akseptable nivåer (0,98).

Short language measure (SLaM)

Matov og medarbeidere (2019) utviklet og validerte Short Language Measure (SLaM), et kortfattet språkvurderingsverktøy for barn i skolealder (107). Deltakerne ble rekruttert fra Foundation Year (det første året med formell utdanning i Victoria) fra tre barne-skoler i Melbourne, Australia (aldersgruppe = 5;1 måned til 6:7). I fase 1 ble 56 barn vurdert med 160 testoppgaver for retningsfølging og setningsgjenkalling, og de beste oppgavene ble valgt ut til SLaM. I fase 2 ble 126 barn vurdert med SLaM og et referansemål, Clinical Evaluation of Language Fundamentals-Fourth Edition (CELF-4), for å fastslå SLaMs nøyaktighet, validitet og reliabilitet. Totalt 40 testelementer ble valgt ut til SLaM i fase 1. Forfatterne konkluderte med at SLaM har utmerkede psykometriske egenskaper og kan brukes til å identifisere barn som trenger videre utredning.

Structured screening test

Laing og medarbeidere (2002) evaluerte to metoder for å identifisere tale- og språkproblemer hos førskolebarn, en strukturert screeningtest og en gullstandardtest (Reynell språktest) (103). Studien omfattet 37 helsesykepleiere i indre London, England, som ble tilfeldig fordelt til å bruke enten en strukturert screeningtest eller en foreldrestyrt metode. Totalt 582 foreldre takket ja til å delta, av 623 kvalifiserte barn i alderen 30-36 måneder. Totalt 458 barn fullførte alle de obligatoriske vurderingene, 282 for den strukturerte screeningtesten og 176 for den foreldrestyrte metoden. Studien målte sensitivitet og spesifisitet. Forfatterne rapporterte at begge metodene ikke fanget opp en betydelig andel av barn med alvorlige språkvansker og førte til overhenvisning til diagnostiske vurderinger. Forfatterne konkluderte med at screeningen sannsynligvis ikke er en effektiv metode for å håndtere tale- og språkproblemer hos førskolebarn i denne populasjonen. De antydte at det kan være mer effektivt å basere seg på foreldrenes observasjoner og helsepersonells kliniske vurdering. De påpeker imidlertid at det er uklart om disse resultatene kan overføres til andre populasjoner.

Test for Sentence Development (TSD)

Schlichting og Spelberg (2003) ønsket å vurdere syntaktisk utvikling hos nederlandske barn ved hjelp av Test for Sentence Development (TSD) (114). Studien utforsket de lingvistiske, psykometriske og kliniske aspektene ved TSD. Studien inkluderte et standardiseringsutvalg på 1049 barn, likt fordelt på aldersgrupper og kjønn. Aldersspennet for deltakerne var fra 1;3 til 6;3. Utvalget var representativt med hensyn til alder, kjønn, region, befolkningstetthet og sosioøkonomisk bakgrunn. Bare barn med foreldre

som begge hadde nederlandsk som morsmål, ble inkludert. TSD ble utviklet for å måle den syntaktiske kunnskapen i små barns språkproduksjon. Den besto av en rekke oppgaver av varierende kompleksitet og lengde. Studien fokuserte på konstruktvaliditet, som refererer til i hvilken grad en test måler det konstruktet den har til hensikt å måle. Når det gjelder reliabilitet, vurderte studien intern konsistens og test-retest-reliabilitet. Forfatterne konkluderte med at TSD viste tilfredsstillende validitet og reliabilitet, noe som gjør det til et pålitelig og gyldig kartleggingsverktøy for kliniske og forskningsmessige formål.

Toddler phonology test (TPT)

Claessen og medarbeidere (2017) evaluerte test-retest-reliabiliteten og den prediktive validiteten til Toddler Phonology Test (TPT) i to ulike studier (88). Studie 1 vurderte 18 australske barn mellom 25 og 35 måneder, og sammenlignet prestasjonene deres med normative data og vurderte dem på nytt en måned senere for å undersøke test-retest-reliabiliteten. Studie 2 vurderte den prediktive validiteten til TPTs kvantitative og kvalitative normative data ved å vurdere 24 britiske barn i alderen 25-35 måneder på nytt mellom fire og ni måneder etter den første vurderingen og ved å undersøke antall og type feil som ble begått. Positive korrelasjoner indikerte konsistente resultater på tvers av testene. Den andre studien, som fulgte opp 24 barn fire til ni måneder senere, viste at både kvantitative og kvalitative mål på TPT bidrar til å identifisere småbarn med risiko for fonologiske forstyrrelser. Forfatterne konkluderer med at resultatene antyder at TPT kan brukes til å vurdere småbarns fonologiske tilegnelse på en pålitelig og valid måte, noe som gjør det mulig å identifisere barn som er i risikozonen for å utvikle språklydsforstyrrelser på et tidlig tidspunkt.

Diskusjon

Hovedfunn

Under gjengir vi hovedfunnene i denne systematiske kartleggingsoversikten.

- Vi identifiserte totalt 46 verktøy for å kartlegge avvik fra normal forventet språkutvikling hos barn 0-5 år, som vurderes som passende for bruk i en norsk helsestasjonsetting.
- For 24 av disse verktøyene identifiserte vi forskning som evaluerte måleegenskaper til verktøyene.
- Av de verktøyene som anbefales brukt ved norske helsestasjoner i henhold til den nåværende retningslinjen, identifiserte vi kun to publikasjoner, hvor begge evaluerte SPRÅK 4.
- Ingen verktøy var evaluert mht. alle måleegenskapene iht. COSMIN.
- Verktøyet som var mest evaluert – evaluert mht. syv måleegenskaper – var C-BiLLT. Fire verktøy var evaluert mht. fire måleegenskaper: BITS, GAPS, NRT, PSRep. De resterende 19 verktøyene var evaluert på tre eller færre måleegenskaper hvorav ni verktøy var evaluert kun på én måleegenskap.
- Den vanligste måleegenskapen som ble vurdert var diagnostisk nøyaktighet; hvor godt verktøyet klarte å skille mellom barn med og uten språkvansker, enten samtidig eller longitudinelt.
- Flere studier så på kulturell validitet ved evaluering av barns språkferdigheter. To studier som evaluerte PPVT, rapporterte om kulturell skjevhet i vurdering av barna når de brukte PPVT-IV, mens tre studier rapporterte om ingen kulturell skjevhet når de brukte PPVT-III.
- For tospråklige barn rapporterte to studier at screening på begge språkene barna bruker, var den mest nøyaktige metoden for å identifisere barn med språkvansker.
- Flere studier presiserer at bruk av flere verktøy, for eksempel foreldrerapportert og profesjonelt administrert verktøy, forbedrer nøyaktigheten i vurderingen av språkutvikling.

Er kunnskapsgrunnlaget dekkende, pålitelig og anvendelig?

I denne kartleggingsoversikten gjorde vi et grundig arbeid i å a) identifisere verktøy for kartlegging av avvik fra normal språkutvikling hos barn 0-5 år som kan overføres til en

norsk helsestasjonssetting og b) identifisere publikasjoner som evaluerer disse verktøyene når det gjelder deres måleegenskaper. I tillegg til innhenting av innspill fra faggruppen utførte vi blant annet flere systematiske søk i tradisjonelle databaser, et søk i OpenAlex og søk etter grå litteratur. Vi antar derfor at kunnskapsgrunnlaget er dekkende for verktøyene identifisert i forskningsspørsmål 1. For de ytterligere verktøyene som ikke ble vurdert for relevans av faggruppen er det mulig at kunnskapsgrunnlaget er mindre dekkende, pga. vi gjennomførte ikke et eksplisitt søk etter studier på disse verktøyene. Det kan derfor hende det finnes relevante studier om disse verktøyene som vi ikke har identifisert.

Videre var både oppdragsgiver, fagfeller og fagekspertene (bortsett fra én) norske. Dette gjør at vi er ganske sikre på at vi identifiserte verktøy som er relevant for en norsk setting. Derimot involverte vi ikke fagekspertene fra de andre landene vi inkluderer studier fra, hvilket kan bety at det finnes ytterligere verktøy som brukes i disse landene som vi ikke identifiserte i forskningsspørsmål 1. Vi opplevde også hindringer med språket i søket etter relevante verktøy på ikke-engelske eller skandinaviske språk. Dette kan ha gjort at vi ikke identifiserte relevante verktøy på nettsider skrevet på islandsk, finsk og nederlandsk.

Det kan se ut som det mangler studier om måleegenskaper til en rekke av verktøyene som brukes i de relevante landene og som var vurdert som relevante av faggruppen. Vi identifiserte kun studier om måleegenskaper til 24 av de 48 (50 %) verktøyene vi identifiserte som relevante for en norsk helsestasjonssetting. For 24 av de 48 verktøyene fant vi altså ingen studier publisert siden 2000 om verktøyets måleegenskaper. I den grad disse verktøyene benyttes i praksis er det uten kunnskapsgrunnlag for at de måler det de skal.

Alt i alt var populasjonene i de inkluderte studiene ganske dekkende relativt til målgruppen for språkkartlegging på helsestasjon. De inkluderte studiene dekket barn fra ulike sosioøkonomiske bakgrunner, aldre og språkferdigheter. Derimot identifiserte vi ikke verktøy egnet for flerspråklige barn i den grad som var ønskelig. Flere av studiene mangler informasjon om hvorvidt det var flerspråklige barn i utvalget. Flerspråklige barn er en viktig gruppe å kartlegge språket til og også en gruppe som ikke nødvendigvis kan bli testet på identisk måte som ettspråklige barn. Flerspråklige barn utgjør en betydelig andel av barna som skal språkkartlegges ved norske helsestasjoner. Tall fra SSB viser at 20 % av alle barn i barnehage i 2022 var minoritetsspråklige barn, og det har vært en stabil økning siden 2015, hvor tallet var 15,3 % (126). SSB definerer minoritetsspråklige barn i barnehage som «barn med annet morsmål enn norsk, samisk, svensk, dansk og engelsk», så andelen flerspråklige vil være høyere enn 20 %.

Iht. metodeveiledninger for kartleggingsoversikter vurderte vi hverken risiko for systematiske skjevheter eller vår tillit til resultatene. Vi kan dermed ikke si noe om den metodiske kvaliteten til hver enkelt studie eller hvorvidt vi kan stole på resultatene som er rapportert. Vi presenterer resultatene slik de rapporteres i publikasjonene, og vi har ikke grunnlag for å trekke konklusjoner på tvers av funnen fra studiene eller komme med anbefalinger.

Styrker og svakheter ved kartleggingsoversikten

Vi utførte et omfattende litteratursøk etter kartleggingsverktøy som kan brukes i en norsk helsestasjonssetting og studier som vurderte måleegenskapene til disse. Dette inkluderte søk i OpenAlex (se vedlegg 1) og databasesøk. Det siste databasesøket ble gjennomført i oktober 2023. Dermed er det svært lite sannsynlig at det er publisert betydelig ny forskning mellom tidspunktet for siste søk og tidspunktet for publiseringen av denne oversikten. Vi antar også at det er lite sannsynlig at vi ville ha identifisert mange nye studier ved å gjøre et ytterligere litteratursøk etter verktøyene ikke relevansvurdert av faggruppen, gitt at 94 prosent av de inkluderte studiene enten ble identifisert via OpenAlex søket eller i første del av det tradisjonelle søket hvor verktøynavn ikke ble spesifisert.

Feltet språkkartleggingsverktøy for barn 0-5 år er et bredt og uoversiktlig felt, hvor det til tider har vært vanskelig å orientere seg. For noen av verktøyene var det lite informasjon om selve verktøyet og vanskelig å tyde hvilken setting verktøyet passet inn i. Det var derfor en styrke å involvere faggruppen for å vurdere relevans. Det kan likevel være verktøy som noen mener ikke burde vært inkludert i denne oversikten eller det kan være at vi har ekskludert verktøy som noen lesere av denne rapporten mener burde vært med. Dette skyldes at det ikke alltid er klare grenser for hvor verktøyene kan brukes og hvem som kan administrere dem.

En styrke med denne kartleggingen er at vi identifiserte flere verktøy fra nyere studier. Imidlertid er det verdt å påpeke at flere av de inkluderte verktøyene er evaluert i studier fra starten av 2000-tallet. Dette kan være problematisk, gitt at enkelte av disse verktøyene ikke nødvendigvis reflekterer eller er oppdatert i tråd med utviklingen av gjenstander og objekter som barnet forventes å identifisere. Et eksempel på hvordan et verktøy kan være utdatert er bruk av bilde av et kassefjernsyn istedenfor flatskjerm, såpestykker i stedet for dispensersåpe o.l. Slike gjenstander kan antas å være mindre kjent for barn født i senere tid, ettersom de er blitt mindre vanlige i moderne hjem. Videre foreligger det begrenset informasjon angående digitaliseringen av disse verktøyene og hvordan denne digitaliseringen eventuelt kunne forenkle deres bruk for helsepersonell.

I denne oversikten inkluderte vi ikke foreldrerapporterte verktøy, fordi disse faller utenfor omfanget av kartleggingsoversikten. Det er likevel viktig å få frem at disse muligens gir en annen innsikt i barnas språkforståelse enn hva observasjon av en helsesykepleier som ikke kjenner barnet gjør. Vi så heller ikke på kombinasjonen av verktøy som gis av helsesykepleier og foreldrerapporterte verktøy, som er praksis i dag på flere helsestasjoner.

Til slutt er det også viktig å få frem at vi utførte en kartleggingsoversikt, hvor vi ikke sammenfattet resultater på tvers av studier, slik det gjøres i en systematisk oversikt. Hensikten med en kartleggingsoversikt er å få en oversikt over hvilken forskning som finnes på et felt, og ikke nødvendigvis virkningen av tiltak eller verktøy. For å gi leseren

mer innsikt i studiene valgte vi å gjengi forfatterens konklusjon. Det er her viktig å påpeke at dette muligens ikke gir et like nyansert bilde av resultatene i hver studie som om vi hadde gjort en sammenfatning selv.

Overensstemmelse med andre litteraturoversikter

Flere kunnskapsoppsummeringer har undersøkt ulike aspekter ved språkkartleggingsverktøy, og det er tydelig at det er både utfordringer og fordeler knyttet til bruken av slike verktøy.

Vi identifiserte flere systematiske oversikter som omhandlet språkkartlegging for barn i alderen 5 år eller yngre. Vi observerer at til tross for at vi utførte en kartleggingsoversikt og ikke en systematisk oversikt, ser det ut som tendensene vi fant også går igjen i de systematiske oversiktene.

Feltner og medarbeidere (127) fant i likhet med vår rapport at sensitivitet- og spesifisitetsmål varierte mye. Forfatterne fant at det var ikke ett enkelt verktøy som konsekvent hadde bedre diagnostisk nøyaktighet enn de andre. Til forskjell fra vår oversikt inkluderte Feltner og medarbeidere også foreldrerapporterte verktøy. Foreldrerapporterte screeningverktøy for ekspressive språkferdigheter viste rimelig nøyaktighet, men den generelle konsistensen av screeningverktøyene varierte fra studie til studie.

Berkman og medarbeidere (128) derimot fant i sin systematiske oversikt at foreldrevurderte verktøy utkonkurrerte profesjonelt administrerte verktøy når det gjelder nøyaktighet, og studien pekte på variasjon i nøyaktigheten av profesjonelt administrerte verktøy, og behovet for mer forskning for å evaluere deres effektivitet på tvers av populasjoner og aldre. Dette fremhever begrensingen vi har i denne oversikten med at foreldrerapporterte verktøy ikke har vært en del av inklusjonskriteriene.

Forfatterne til flere av de inkluderte studiene i denne oversikten har påpekt at det kan være vanskelig å finne et kartleggingsverktøy som kan anbefales for alle barn, og vi ser at kartleggingsverktøyene fokuserer på ulike språkdomener, kulturelle aspekter og egner seg til ulike aldre. Dette samsvarer med hva Baker og medarbeidere (129) fant i sin systematiske oversikt. Baker konkluderer med at forskjellige verktøy har relative fordeler, men ingen alene oppfyller alle nødvendige kriterier for reliabilitet og validitet. Dette peker på kompleksiteten ved å utvikle et universelt verktøy som er egnet for alle aldersgrupper og dekker alle aspekter av språkutvikling.

Flere av de inkluderte studiene i vår kartleggingsoversikt omhandler ulike nonword repetition tests (NRT). I Schwob og medarbeidere (130) sin systematiske oversikt og metaanalyse undersøkte de den diagnostiske verdien av NRT for å oppdage språkutviklingsforstyrrelser hos både enspråklige og tospråklige. 35 studier ble inkludert i metaanalysen og resultatene viser at NRT diskriminerte tilfredsstillende mellom barn med og uten utviklingsforstyrrelser i begge språkgruppene, samtidig som sensitivitet og

spesifisitet generelt økte når testen ble kombinert med andre oppgaver (f.eks. foreldre-rapporterte verktøy).

Det å kombinere bruk av flere verktøy eller vurderingsmetoder er noe som det også konkluderes med i flere av våre inkluderte studier (105;106). Å kombinere bruken av flere verktøy er noe som også anbefales av Sansavini og medarbeidere (66), som i sin systematiske kartleggingsoversikt understreket betydningen av å kombinere ulike tilnærminger for å få en mer helhetlig forståelse av barnets språkutvikling. Det å kombinere bruk av flere verktøy, for eksempel både observasjon og foreldrerapport er en ting, men dersom tidsbegrensninger forhindrer dette, er det uenigheter i litteraturen om hvorvidt verktøy administrert av for eksempel en helsesykepleier eller et foreldre-rapportert verktøy burde benyttes. Stokes (131) sammenlignet evnen til sykepleiere og foreldre til å oppdage tale- og språkvansker. Foreldrespørreskjemaet viste seg å være sammenlignbart med profesjonell screening, og dette støttes av Madden (132), som vurderer nøyaktigheten av ASQ-3, et foreldrerapportert skjema, og indikerer at kvalitative spørsmål i foreldreskjemaet er en betydelig indikator for mulige tale- og språkproblemer. Matov og medarbeidere (107) argumenterer imidlertid at kun rapporter fra foreldre og lærere ikke er tilstrekkelig nøyaktige for å identifisere barn som trenger videre henvisning.

Flere av de inkluderte studiene (85;89;93;111;113;118;124) peker nettopp på utfordringer knyttet til å kartlegge språkutviklingen til flerspråklige barn sammenliknet med enspråklige barn. Dette temaet er noe som andre nordiske studier også har undersøkt, fra perspektivet til de som administrerer verktøyene. Disse studiene rapporterte ikke måleegenskaper, men ga likevel viktig informasjon om verktøyene.

Sivertsen og medarbeidere (133) ønsket å undersøke om New Reynell Developmental Language Scales (NRDLS) og Multilingual Toolkit gir fagpersoner tilstrekkelige ressurser for å kartlegge språk hos flerspråklige barn. Forfatteren benyttet semistrukturerte intervjuer med tre informanter, to logopeder og en PP-rådgiver. Forfatterne konkluderer med at verktøyene ble oppfattet som gode ressurser for å kartlegge språk hos flerspråklige barn, men at samarbeid med foreldre for å innhente informasjon om barnets morsmål og familiens kultur er avgjørende når barnet er flerspråklig. Videre undersøkte Nayeb og medarbeidere (134) helsesykepleieres oppfatninger av og praksis med språkkartlegging av tospråklige barn i alderen 30-36 måneder i Sverige. 863 sykepleiere som utførte språkkartlegging av tospråklige barn i Sverige minst en gang i måneden svarte på et spørreskjema, tilsvarende 89 % av de inviterte. Forfatterne rapporterte at helsesykepleierne rapporterte større usikkerhet og flere vanskeligheter med tolkning av resultater for tospråklige enn for enspråklige barn. Halvparten av sykepleierne forenklet kartleggingsprosessen for tospråklige barn, og 74 % utsette henvisninger til tale- og språktjenester, basert på deres oppfatninger av barnas svenske språkferdigheter. De fleste sykepleierne (82 %) trodde at språkutviklingen var tregere hos tospråklige barn, og dette var den sterkeste prediktoren for forenklete kartleggingspraksiser. Forfatterne konkluderte med at helsesykepleiere har behov for lett tilgjengelig infor-

masjon og klare retningslinjer om språkutviklingen til tospråklige barn for å sikre at tospråklige og enspråklige barn mottar likeverdige språkkartleggingstjenester.

Vi inkluderte to publikasjoner som evaluerte SPRÅK 4 (100;101), kartleggingsverktøyet som benyttes på norske helsestasjoner i dag. Ingen av disse publikasjonene omtaler bruk av SPRÅK 4 med flerspråklige barn. En evaluering fra 2008 påpekte begrensninger i dets bruk med flerspråklige barn. Tiltak som oversettelse av skjemaet ble derfor iverksatt for å forbedre kartleggingen av barn med mangelfulle norskkunnskaper. Resultatene av denne evalueringen (135) indikerte forbedringer, hvor helsesykepleiere rapporterte om økt sikkerhet, bedre oversikt over barn med mangelfullt morsmål og økt kompetanse på flerspråklige kartlegginger, men de påpekte også behovet for kvalitetssikring av oversettelsene og spesielle utfordringer knyttet til oversettelser. Til tross for identifiserte svakheter, rapporterte helsesykepleierne at det oversatte SPRÅK 4 skjemaet i SPRÅK 4 var nyttig i 90 % av tilfellene.

Resultatenes betydning for praksis

Iht. standarden for kartleggingsoversikter har vi ikke vurdert studienes risiko for systematiske skjevheter, syntetisert resultatene fra de inkluderte studiene eller vurdert tilliten til dokumentasjonen av syntetiserte resultater. Følgelig er vi tilbakeholdne med å vurdere resultatenes betydning for praksis. Det er flere spørsmål denne systematiske kartleggingsoversikten ikke kan svare på. Basert på uthentet data og hva de inkluderte studienes forfattere skriver, kombinert med resultater fra andre litteraturoversikter og primærstudier, ønsker vi å legge frem noen spørsmål som med fordel burde besvares før anbefalingen i nasjonal faglig retningslinje oppdateres. Spørsmålene er som følger:

1. Hvilke(t) språkkartleggingsverktøy er best egnet for språkkartlegging av barn 0-5 år på helsestasjonen?
2. Hvilke(t) språkkartleggingsverktøy er best egnet for å avdekke avvik fra normal språkutvikling hos flerspråklige barn?
3. Hvordan kan språkkartleggingen tilpasses for å bedre møte behovene til flerspråklige barn?
4. Burde samme verktøy benyttes for enspråklige og flerspråklige barn?
5. Når er de beste tidsperiodene å screene barn for forsinket språkutvikling?
6. Hvordan sikre at språkkartleggingsverktøy benyttes av helsesykepleiere i kartlegging av språkutvikling?
7. Hvordan kan helsesykepleiere styrke sin kompetanse i å identifisere forsinket språkutvikling og språkvansker hos barn, spesielt med tanke på flerspråklige barn?
8. Kan digitaliserte verktøy være et godt alternativ for språkkartlegging der det ikke stilles like høye krav til kompetanse på språkutvikling hos helsesykepleiere?
9. Hva er helsesykepleierens perspektiv på bruk av språkkartleggingsverktøyene som de benytter i deres arbeid?
10. Hvordan kan man optimalt utnytte foreldres/foresattes rolle i å bidra til språkkartleggingen, spesielt når barnet er flerspråklig?

Kunnskapshull

I forbindelse med forskningsspørsmål 1, hvor 34 verktøy ble identifisert som relevante for en helsestasjonssetting, identifiserte vi for 24 verktøy ingen studier som evaluerte deres måleegenskaper. Det er her viktig å få frem at vi søkte etter litteratur fra år 2000, så det kan være mulig at det finnes evalueringer fra tidligere år. For de verktøyene hvor vi identifiserte studier, var det gjennomgående få evalueringer gjort på barn 0-2 år. Videre var det få nordiske studier som vurderte relevante verktøy. Av de undersøkte verktøyene var det også begrenset hvor mange måleegenskaper som var evaluert, noe som indikerer et mulig gap i kunnskapen om evalueringer av måleegenskapene til verktøyene. Avslutningsvis er et viktig funn i denne kartleggingsoversikten at det ser ut til å mangle verktøy som er tilpasset flerspråklige barn, noe som fremhever behovet for mer forskning og utvikling innenfor dette spesifikke området.

Konklusjon

Hensikten med denne kartleggingsoversikten var å finne verktøy for å oppdage avvik i normal språkutvikling hos barn i alderen 0-5 år. Vi ønsket også å oppsummere kjennetegnene til disse verktøyene og deres måleegenskaper. Vi fant 24 ulike verktøy i 47 publikasjoner som anses relevante for norske helsestasjoner. Elleve av studiene (13 publikasjoner) var fra Norden. Hvilke måleegenskaper som ble evaluert varierte betydelig, og noen verktøy ble grundigere evaluert enn andre. For ni av de 24 verktøyene ble kun én måleegenskap evaluert. Diagnostisk nøyaktighet, målt med sensitivitet og spesifisitet, var den mest evaluerte måleegenskapen.

Kunnskapshullene identifisert i oversikten inkluderer manglende evalueringer av flere verktøy, både i form av antall studier gjennomført og antall måleegenskaper evaluert. Videre var det få evalueringer for den yngste aldersgruppen (0-2 år), få nordiske studier og mangel på egnede verktøy for flerspråklige barn. Disse kunnskapshullene gir retning for fremtidig forskning på området.

Denne kartleggingsoversikten gir ikke grunnlag for å fremme konkrete anbefalinger for praksis. Kartleggingsoversikten tydeliggjør imidlertid noen spørsmål som forskningen foreløpig ikke gir klare svar på, inkludert hvilke verktøy som er best egnet for språkkartlegging i en norsk helsestasjonssetting. Hvilke verktøy som egner seg for flerspråklige barn og hvordan språkkartleggingen kan best tilpasses deres behov ser ut til å være særlig aktuelle spørsmål.

Referanser

1. Socialstyrelsen. Stödja och stimulera kommunikations- och språkutveckling bland barn 0 till 2 år: ett kunskapsstöd för barnhälsovården. Stockholm: Socialstyrelsen; 2022.
2. Helsedirektoratet. Kapittel 4. Helsestasjon 0–5 år. I: Helsestasjon, skolehelsetjeneste og helsestasjon for ungdom: Nasjonal faglig retningslinje [nettdokument]. Oslo: Helsedirektoratet [oppdatert 30. juni 2023; lest 18. desember 2023]. Tilgjengelig fra: <https://www.helsedirektoratet.no/retningslinjer/helsestasjons-og-skolehelsetjenesten/helsestasjon-05-ar>
3. Bishop DVM, Snowling MJ, Thompson PA, Greenhalgh T, consortium C. CATALISE: A multinational and multidisciplinary Delphi Consensus Study. Identifying language impairments in children. PLoS One 2016;11(7):e0158753. DOI: 10.1371/journal.pone.0158753
4. Sim F, Thompson L, Marryat L, Ramparsad N, Wilson P. Predictive validity of preschool screening tools for language and behavioural difficulties: A PRISMA systematic review. PLoS One 2019;14(2):e0211409. DOI: 10.1371/journal.pone.0211409
5. So KKH, To CKS. Systematic review and meta-analysis of screening tools for language disorder. Front Pediatr 2022;10:801220. DOI: 10.3389/fped.2022.801220
6. Melby-Lervåg M. Språklig utvikling hos barn fra null til fem år. I: Lervåg A, red. Bergen: Fagbokforlaget; 2014. s. 167-88.
7. Ølstørn A. SATS og SPRÅK 4 - En deskriptiv tverrsnittstudie av systematisk språkkartlegging på helsestasjonene i Oslo kommune. Trondheim: NTNU; 2021.
8. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5(1):69. DOI: 10.1186/1748-5908-5-69
9. Peters M, Godfrey C, McInerney P, Munn Z, Tricco A, Khalil H. Chapter 11: Scoping reviews I: Aromataris E, Munn Z, red. JBI Manual for evidence synthesis. Adelaide South Australia: JBI; 2020. Tilgjengelig fra: <https://synthesismanual.jbi.global>
10. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. DOI: 10.1080/1364557032000119616
11. Pollock D, Davies EL, Peters MDJ, Tricco AC, Alexander L, McInerney P, et al. Undertaking a scoping review: A practical guide for nursing and midwifery students, clinicians, researchers, and academics. J Adv Nurs 2021;77(4):2102-13. DOI: <https://doi.org/10.1111/jan.14743>
12. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med 2018;169(7):467-73. DOI: 10.7326/m18-0850
13. Mokkink LB, Prinsen CA, Bouter LM, Vet HC, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and

- how to select an outcome measurement instrument. *Braz J Phys Ther* 2016;20(2):105-13. DOI: 10.1590/bjpt-rbf.2014.0143
14. Priem J, Piwowar H, Orr R. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. ArXiv 2022.
 15. Auza BA, Murata C, Peñaloza C. "Early detection of Spanish-speaking children with developmental language disorders: Concurrent validity of a short questionnaire and a screening test". *J Commun Disord* 2023;104:106339-. DOI: 10.1016/j.jcomdis.2023.106339
 16. Borkhoff CM, Atalla M, Bayoumi I, Birken CS, Maguire JL, Parkin PC. Predictive validity of the Infant Toddler Checklist in primary care at the 18-month visit and developmental diagnosis at 3-5 years: a prospective cohort study. *BMJ Paediatr Open* 2022;6(1):e001524-e. DOI: 10.1136/bmjpo-2022-001524
 17. Bruce B, Kornfalt R, Radeborg K, Hansson K, Nettelbladt U. Identifying children at risk for language impairment: screening of communication at 18 months. *Acta Paediatr* 2007;92(9):1090-5. DOI: 10.1111/j.1651-2227.2003.tb02583.x
 18. Bytešníková I, Smolík F. Brief children's dictionary questionnaire SDDS 16-42: Introducing a screening diagnostic tool for early detection of children with delay in their language skills' development. *Listy klinické logopedie* 2022;6(2):50-5. DOI: 10.36833/lkl.2022.009
 19. Cairney DG, Kazmi A, Delahunty L, Marryat L, Wood R. The predictive value of universal preschool developmental assessment in identifying children with later educational difficulties: A systematic review. *PLoS One* 2021;16(3):e0247299-NA. DOI: 10.1371/journal.pone.0247299
 20. Chaffee CA, Cunningham CE, Secord-Gilbert M, Elbard H, Richards J. Screening effectiveness of the Minnesota Child Development Inventory expressive and receptive language scales: Sensitivity, specificity, and predictive value. *Psychol Assess* 1990;2(1):80-5. DOI: 10.1037/1040-3590.2.1.80
 21. Clark B, Acton BV, Alton GY, Joffe A, R., Dinu I, Robertson CMT. Screening for language delay after life-saving therapies in term-born infants. *Cardiol Young* 2015;26(7):1343-51. DOI: 10.1017/s1047951115002577
 22. Conti-Ramsden G, Durkin K. Language development and assessment in the preschool period. *Neuropsychol Rev* 2012;22(4):384-401. DOI: 10.1007/s11065-012-9208-z
 23. de Araújo Barbosa AL, Lemos FF, Azoni CAS. Application of a vocabulary screening instrument for children between 3- and 7-years-old: a pilot study. *CoDAS* 2021;33(2):e20190154-NA. DOI: 10.1590/2317-1782/20202019154
 24. Dias DC, Rondon-Melo S, Molini-Avejonas DR. Sensitivity and specificity of a low-cost screening protocol for identifying children at risk for language disorders. *Clinics (Sao Paulo)* 2020;75(NA):e1426-NA. DOI: 10.6061/clinics/2020/e1426
 25. Doove BM, Feron FJM, van Os J, Drukker M. Preschool communication: Early identification of concerns about preschool language development and social participation. *Front Public Health* 2021;8(NA):546536-. DOI: 10.3389/fpubh.2020.546536
 26. Dworkin PH, Stancin T, Drotar D. Pediatric developmental screening: Understanding and selecting screening instruments: Commonwealth Fund [opdatert 26. februar 2008; lest 15. desember 2023]. Tilgjengelig fra: https://www.commonwealthfund.org/sites/default/files/documents/_media_files_publications_fund_manual_2008_feb_pediatric_developmental_screening_understanding_and_selecting_screening_instruments_pediatric_developmental_screening_pdf.pdf
 27. Eriksson M. Insufficient evidence for the validity of the Language Development Survey and the MacArthur-Bates Communicative Development Inventories as screening tools: A critical review. *Int J Lang Commun Disord* 2022;58(2):555-75. DOI: 10.1111/1460-6984.12800

28. Farabolini G, Ceravolo MG, Marini A. Towards a characterization of late talkers: The developmental profile of children with late language emergence through a web-based communicative-language assessment. *Int J Environ Res Public Health* 2023;20(2):1563-. DOI: 10.3390/ijerph20021563
29. Feldman HM. Evaluation and management of language and speech disorders in preschool children. *Pediatr Rev* 2005;26(4):131-42. DOI: 10.1542/pir.26.4.131
30. U. S. Preventive Services Task Force. Screening for speech and language delay in preschool children: Recommendation statement. *Pediatr* 2006;117(2):497-501. DOI: DOI: 10.1542/peds.2005-2766
31. Frisk V, Montgomery L, Boychyn E, Young R, vanRyn E, McLachlan D, 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/iy.0b013e3181bc4db6
32. Guiberson M, Rodriguez BL, Zajacova A. Accuracy of telehealth-administered measures to screen language in Spanish-speaking preschoolers. *Telemed J E Health* 2015;21(9):714-20. DOI: 10.1089/tmj.2014.0190
33. Holzinger D, Weber C, Barbaresi W, Beitel C, Fellingner J. Language screening in 3-year-olds: Development and validation of a feasible and effective instrument for pediatric primary care. *Front Pediatr* 2021;9. DOI: 10.3389/fped.2021.752141
34. Holzinger D, Weber C, Fellingner J. Validity and feasibility of a predictive language screening tool in 2-year-old children in primary pediatric care. *Front Pediatr* 2022;10:865457. DOI: 10.3389/fped.2022.865457
35. Holzinger D, Weber C, Jezek M. Identifying language disorder within a migration context: Development and performance of a pre-school screening tool for children with German as a second language. *Front Pediatr* 2022;10:814415. DOI: 10.3389/fped.2022.814415
36. Hutchins H, Robinson L, Charania S, Ghandour R, Hirsh-Pasek K, Zubler J. Psychometric assessment of pilot language and communication items on the 2018 and 2019 National Survey of Children's Health. *Acad Pediatr* 2021;22(7):1133-41. DOI: 10.1016/j.acap.2021.12.024
37. Hviding K. Kartleggingsverktøy og instrumenter for tidlig avdekking av utviklings-, atferds- og psykososiale vansker hos barn 0-6 år. Oslo: Nasjonalt kunnskapssenter for helsetjenesten; 2005. (Rapport fra Kunnskapssenteret Nr 10-2005). Tilgjengelig fra: <https://www.fhi.no/publ/eldre/kartleggingsverktoy-og-instrumenter-for-tidlig-avdekking-av-utviklings--atf/>
38. Jackson E, Levine D, de Villiers J, Iglesias A, Hirsh-Pasek K, Michnick GR. Assessing the language of 2 year-olds: From theory to practice. *Infancy* 2023;28(5):930-57. DOI: 10.1111/inf.12554
39. Jullien S. Screening for language and speech delay in children under five years. *BMC Pediatr* 2021;21(1):362. DOI: 10.1186/s12887-021-02817-7
40. Just J. Assessing the predictive value of the UK-CDI for early identification of developmental language delay [thesis]. Lincoln: University of Lincoln; 2018. Tilgjengelig fra: <https://eprints.lincoln.ac.uk/id/eprint/35704/1/Phd%20Janine%20Just%20170518%20-%20final%20151118.pdf>
41. Kaiser AP, Chow JC, Cunningham JE. A case for early language and behavior screening: Implications for policy and child development. *Policy Insights Behav Brain Sci* 2022;9(1):120-8. DOI: 10.1177/23727322211068886
42. Kas B, Jakab Z, Lórik J. Development and norming of the Hungarian CDI-III: A screening tool for language delay. *Int J Lang Commun Disord* 2022;57(2):252-73. DOI: 10.1111/1460-6984.12686

43. Klee T, Pearce KF, Carson DK. Improving the positive predictive value of screening for developmental language disorder. *J Speech Lang Hear Res* 2000;43(4):821-33. DOI: 10.1044/jslhr.4304.821
44. Klem MJ, Hagtvet B, E. Tidlig språkkartlegging, til barnets beste? *Norsk tidsskrift for logopedi* 2018;64(4):12-24.
45. Larson AL. Language screening for infants and toddlers: A literature review of four commercially available tools. *Commun Disord Q* 2016;38(1):3-12. DOI: 10.1177/1525740115627420
46. Law J, Boyle J, Harris F, Harkness A, Nye C. The feasibility of universal screening for primary speech and language delay: findings from a systematic review of the literature. *Dev Med Child Neurol* 2000;42(3):190-200. DOI: 10.1017/s0012162200000335
47. Lee NR, Chung SH, Song MK, Kong YH, Joo CU, Kim SJ. A comparative analysis of clinical screening test and language specific test in language delay children. *Chonnam Med J* 2020;56(1):44-9.
48. Longerbeam MS, Freeman LM. Language screening: How far have we come? *Curr Dev Disord Rep* 2020;7(3):116-23. DOI: 10.1007/s40474-020-00198-7
49. Lyngseth EJ. Erfaringer med bruk av TRAS-observasjoner i barnehagen. *Norsk pedagogisk tidsskrift* 2008;92(5):352-62. DOI: 10.18261/ISSN1504-2987-2008-05-02
50. Macy M. The evidence behind developmental screening instruments. *Infants Young Child* 2012;25(1):19-61. DOI: 10.1097/IYC.0b013e31823d37dd
51. Mattsson CM, Marild S, Pehrsson NG. Evaluation of a language-screening programme for 2.5-year-olds at Child Health Centres in Sweden. *Acta Paediatr* 2001;90(3):339-44.
52. McCauley RJ, Swisher L. Psychometric review of language and articulation tests for preschool children. *J Speech Hear Disord* 1984;49(1):34-42. DOI: 10.1044/jshd.4901.34
53. Miniscalco C, Westerlund M, Lohmander A. Language skills at age 6 years in Swedish children screened for language delay at 2½ years of age. *Acta Paediatr* 2005;94(12):1798-806. DOI: 10.1080/08035250500244242
54. Morgan L, Delehanty A, Dillon JC, Schatschneider C, Wetherby AM. Measures of early social communication and vocabulary production to predict language outcomes at two and three years in late-talking toddlers. *Early Child Res Q* 2020;51(NA):366-78. DOI: 10.1016/j.ecresq.2019.12.005
55. Mossabeh R, Wade KC, Finnegan K, Sivieri EM, Abbasi S. Language development survey provides a useful screening tool for language delay in preterm infants. *Clin Pediatr* 2012;51(7):638-44. DOI: 10.1177/0009922812439244
56. Mølland E, Haraldstad K, Abildsnes E, Håland ÅT, Köpp UMS, Fegran L, et al. Use of the Strengths and Difficulties Questionnaire in child and school health services among children aged 4 and 6 years in Southern Norway: clinical considerations. *BMC Pediatr* 2023;23(1):30-NA. DOI: 10.1186/s12887-023-03837-1
57. 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):1642-8. DOI: 10.1111/apa.14790
58. Nelson HD, Nygren P, Walker M, Panoscha R. U.S. Preventive Services Task Force Evidence Syntheses, formerly Systematic Evidence Reviews. I: Screening for speech and language delay in preschool children. Rockville (MD): Agency for Healthcare Research and Quality (US); 2006.
59. Nurse KM, Janus M, Birken CS, Keown-Stoneman CDG, Omand JA, Maguire JL, et al. Predictive validity of the Infant Toddler Checklist in Primary Care at the 18-month visit and school readiness at 4 to 6 years. *Acad Pediatr* 2022;23(2):322-8. DOI: 10.1016/j.acap.2022.09.004

60. Pace A, Curran M, Van Horne AO, de Villiers J, Iglesias A, Golinkoff RM, et al. Classification accuracy of the Quick Interactive Language Screener for preschool children with and without developmental language disorder. *J Commun Disord* 2022;100(NA):106276-. DOI: 10.1016/j.jcomdis.2022.106276
61. Pickstone C, Hannon P, Fox L. Surveying and screening preschool language development in community-focused intervention programmes: a review of instruments. *Child Care Health Dev* 2002;28(3):251-64. DOI: 10.1046/j.1365-2214.2002.00270.x
62. Prelock PA, Hutchins TL. Approaches to the screening and identification of communication disorders. I: Clinical guide to assessment and treatment of communication disorders Best practices in child and adolescent behavioral health care. Cham: Springer; 2018. s. 23-30.
63. Rainelli S, Bulotsky-Shearer RJ, Fernandez VA, Greenfield DB, L'opez M. Validity of the first two subtests of the preschool language assessment scale as a language screener for Spanish-speaking preschool children. *Early Child Res Q* 2017;38:10-22. DOI: 10.1016/j.ecresq.2016.08.001
64. Rescorla L. The Language Development Survey. *J Speech Hear Disord* 1989;54(4):587-99. DOI: 10.1044/jshd.5404.587
65. Ritoša A, Danielsson H, Sjöman M, Almquist L, Granlund M. Assessing School Engagement – Adaptation and Validation of “Engagement Versus Disaffection With Learning: Teacher Report” in the Swedish Educational Context. *Front Educ* 2020;5. DOI: 10.3389/educ.2020.521972
66. Sansavini A, Favilla ME, Guasti MT, Marini A, Millepiedi S, Di Martino MV, et al. Developmental language disorder: Early predictors, age for the diagnosis, and diagnostic tools. A scoping review. *Brain Sci* 2021;11(5):654-NA. DOI: 10.3390/brainsci11050654
67. Schonhaut L, Maturana A, Cepeda O, Seron P. Predictive validity of developmental screening questionnaires for identifying children with later cognitive or educational difficulties: A systematic review. *Front Pediatr* 2021;9:698549. DOI: 10.3389/fped.2021.698549
68. Schonhaut L, Pérez M, Armijo I, Maturana A. Comparison between Ages & Stages Questionnaire and Bayley Scales, to predict cognitive delay in school age. *Early Hum Dev* 2019;141:104933. DOI: 10.1016/j.earlhumdev.2019.104933
69. Schwartz AH, Murphy MW. Cues for screening language disorders in preschool children. *Pediatr* 1975;55(5):717-22. DOI: 10.1542/peds.55.5.717
70. Smolík F, Bytešníková I. Validity of the SDDS: A 40-item vocabulary screening tool for 18- to 42-month olds in Czech. *J Commun Disord* 2021;93(NA):106146-NA. DOI: 10.1016/j.jcomdis.2021.106146
71. So KKH, To CKS. Systematic review and meta-analysis of screening tools for language disorder. *Front Pediatr* 2022;10(NA):801220-NA. DOI: 10.3389/fped.2022.801220
72. van Agt HME, van der Stege HA, de Ridder-Sluis JG, de Koning HJ. Detecting language problems: accuracy of five language screening instruments in preschool children. *Dev Med Child Neurol* 2007;49(2):117-22. DOI: 10.1111/j.1469-8749.2007.00117.x
73. Vehkavuori S-M, Stolt S. Screening language skills at 2;0. *Infant Behav Dev* 2018;50(NA):174-9.
74. Visser-Bochane MI, Luinge M, Dieleman L, van der Schans CP, Reijneveld SA. The Dutch well child language screening protocol for 2-year-old children was valid for detecting current and later language problems. *Acta Paediatr* 2020;110(2):556-62. DOI: 10.1111/apa.15447
75. Wallace I, Berkman ND, Watson LR, Coyne-Beasley T, Wood CT, Cullen K, et al. Screening for speech and language delay in children 5 years old and younger: A

- systematic review. *Pediatr* 2015;136(2):e448-62. DOI: 10.1542/peds.2014-3889
76. Weiler B, Schuele CM, Feldman JI, Krimm H. A multiyear population-based study of kindergarten language screening failure rates using the Rice Wexler Test of Early Grammatical Impairment. *Lang Speech Hear Serv Sch* 2018;49(2):248-59. DOI: 10.1044/2017_lshss-17-0071
 77. Westerlund M, Berglund E, Eriksson M. Can severely language delayed 3-year-olds be identified at 18 months? Evaluation of a screening version of the MacArthur-Bates Communicative Development Inventories. *J Speech Lang Hear Res* 2006;49(2):237-47. DOI: 10.1044/1092-4388(2006/020)
 78. Westerlund M, Sundelin C. Screening for developmental language disability in 3-year-old children. Experiences from a field study in a Swedish municipality. *Child Care Health Dev* 2000;26(2):91-110. DOI: 10.1046/j.1365-2214.2000.00171.x
 79. Wright RK, Washington KN, Crowe K, Jenkins A, Leon M, Kokotek L, et al. Current methods of evaluating the language abilities of multilingual preschoolers: A scoping review using the International Classification of Functioning, Disability and Health—Children and Youth Version. *Lang Speech Hear Serv Sch* 2019;50(3):434-51. DOI: 10.1044/2019_lshss-18-0128
 80. Zhang X, Qin F, Chen Z, Gao L, Qiu G, Lu S. Fast screening for children's developmental language disorders via comprehensive speech ability evaluation—using a novel deep learning framework. *Ann Transl Med* 2020;8(11):707-. DOI: 10.21037/atm-19-3097
 81. Thomas JGS, Brunton J, Ghouze Z, O'Driscoll P, Bond M. EPPI-Reviewer: advanced software for systematic reviews, maps and evidence synthesis. EPPI-Centre, UCL Social Research Institute, University College London; 2020.
 82. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60(1):34-42. DOI: <https://doi.org/10.1016/j.jclinepi.2006.03.012>
 83. Adamson-Macedo EN, Patel R, Sallah D. An independent psychometric evaluation of a speech and language tool for two-year-old children from a Sure Start trailblazer site in the West Midlands. *Child Lang Teach Ther* 2009;25(2):191-214. DOI: <https://doi.org/10.1177/0265659009102977>
 84. Adamson-Macedo EN, Patel R, Sallah DK. An independent assessment of the internal reliability of the 'Bilston and Ettingshall sure start (BESS) speech and language therapy two year developmental check. *Early Child Dev Care* 2009;179(5):571-86. DOI: 10.1080/03004430701368408
 85. Anderson A. In search of an effective method of measuring First Nations children's speech and language development. *Diss Abstr Int Sect B Sci Eng* 2016;76(12-B(E)).
 86. Bootsma JN, Campbell F, McCauley D, Hopmans S, Grahovac D, Cunningham BJ, et al. Psychometric properties of the English language version of the C-BiLLT evaluated in typically developing Canadian children. *J Pediatr Rehabil Med* 2023;16(1):71-81. DOI: 10.3233/PRM-210101
 87. Chiat S, Roy P. The Preschool Repetition Test: An evaluation of performance in typically developing and clinically referred children. *J Speech Lang Hear Res* 2007;50(2):429-43. DOI: [https://doi.org/10.1044/1092-4388\(2007/030\)](https://doi.org/10.1044/1092-4388(2007/030))
 88. Claessen M, Beattie T, Roberts R, Leitao S, Whitworth A, Dodd B. Is two too early? Assessing toddlers' phonology. *Speech Lang Hear* 2017;20(2):91-101. DOI: 10.1080/2050571X.2016.1222723
 89. de Villiers J, Iglesias A, Golinkoff RM, Hirsh-Pasek K, Wilson M, S., Nandakumar R. Assessing dual language learners of Spanish and English: Development of the

- QUILS: ES. Rev Logop Foniatr Audiol 2021;41(4):183-96. DOI: <https://doi.org/10.1016/j.rlfa.2020.11.001>
90. Deevy P, Weil LW, Leonard LB, Goffman L. Extending use of the NRT to preschool-age children with and without specific language impairment. *Lang Speech Hear Serv Sch* 2010;41(3):277-88.
 91. Dollaghan CA, Campbell TF. How well do poor language scores at ages 3 and 4 predict poor language scores at age 6? *Int J Speech Lang Pathol* 2009;11(5):358-65. DOI: <https://doi.org/10.1080/17549500903030824>
 92. Eisenberg S, Victorino KR, Murray S. Concurrent validity of the Fluharty Preschool Speech and Language Screening Test–Second Edition at age 3: Comparison with four diagnostic measures. *Lang Speech Hear Serv Sch* 2019;50(4):673-82.
 93. Finneran DA, Heilmann JJ, Moyle MJ, Chen S. An examination of cultural-linguistic influences on PPVT-4 performance in African American and Hispanic preschoolers from low-income communities. *Clin Linguist Phon* 2020;34(3):242-55. DOI: 10.1080/02699206.2019.1628811
 94. Fiske S. Validering av C-BiLLT. En valideringsstudie av en ny språkforståelsestest tilpasset barn med store tale- og bevegelsesvansker. Masteroppgave i spesialpedagogikk, Institutt for spesialpedagogikk, Det utdanningsvitenskapelige fakultet. Oslo: Universitetet i Oslo; 2018. Tilgjengelig fra: <https://www.duo.uio.no/bitstream/handle/10852/64510/Validering-av-C-BiLLT-Masteroppgave-va-r-2018-Sara-Fiske.pdf?sequence=5&isAllowed=y>
 95. Gardner H, Froud K, McClelland A, van der Lely Heather KJ. Development of the Grammar and Phonology Screening (GAPS) test to assess key markers of specific language and literacy difficulties in young children. *Int J Lang Commun Disord* 2006;41(5):513-40. DOI: <https://doi.org/10.1080/13682820500442644>
 96. Geytenbeek JJ, Mokkink LB, Knol DL, Vermeulen R, Oostrom KJ. Reliability and validity of the C-BiLLT: A new instrument to assess comprehension of spoken language in young children with cerebral palsy and complex communication needs. *Augment Altern Commun* 2014;30(3):252-66. DOI: 10.3109/07434618.2014.924992
 97. Glascoe FP. The Brigance Infant and Toddler Screen: Standardization and validation. *J Dev Behav Pediatr* 2002;23(3):145-50. DOI: <https://doi.org/10.1097/00004703-200206000-00003>
 98. Gray S. Diagnostic accuracy and test-retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *J Commun Disord* 2003;36(2):129-51. DOI: 10.1016/S0021-9924%2803%2900003-0
 99. Haddeland A. Språkforståelsestesten C-BiLLT: Vurdering av reliabilitet og sammenligning av ulike responsmodaliteter. Masteroppgave i spesialpedagogikk, Institutt for spesialpedagogikk, Det utdanningsvitenskapelige fakultet. Oslo: Universitetet i Oslo; 2018. Tilgjengelig fra: <https://www.duo.uio.no/bitstream/handle/10852/64427/Masteroppgave---Anne-Lise-Haddeland.pdf?sequence=1&isAllowed=y>
 100. Klem MJ, Gustafsson J-E, Hagtvat BE. The dimensionality of language ability in four-year-olds: Construct validation of a language screening tool. *Scand J Educ Res* 2015;59(2):195-213. DOI: <https://doi.org/10.1080/00313831.2014.904416>
 101. Klem MJ, Hagtvat BE, Hulme C, Gustafsson J-E. Screening for language delay: Growth trajectories of language ability in low- and high-performing children. *J Speech Lang Hear Res* 2016;59(5):1035-45. DOI: <https://doi.org/10.1044/2016.jslhr-l-15-0289>

102. Korpilahti P, Kaljonen A, Jansson-Verkasalo E. Population-based screening for language delay: Let's talk STEPS Study. *Psychol* 2016;07(02):205-14. DOI: <https://doi.org/10.4236/psych.2016.72023>
103. 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. DOI: <https://doi.org/10.1136/bmj.325.7373.1152>
104. Lavesson A, Lövdén M, Hansson K. Development of a language screening instrument for Swedish 4-year-olds. *Int J Lang Commun Disord* 2018. DOI: <https://doi.org/10.1111/1460-6984.12374>
105. Law J, Charlton J, Wilson P, Rush R, Gilroy V, McKean C. The development and productivity of a measure for identifying low language abilities in children aged 24-36 months. *BMC Pediatr* 2023;23(1):495. DOI: 10.1186/s12887-023-04079-x
106. Letts C, Edwards S, Schaefer B, Sinka I. The New Reynell Developmental Language Scales: Descriptive account and illustrative case study. *Child Lang Teach Ther* 2013;30(1):103-16. DOI: <https://doi.org/10.1177/0265659013492784>
107. Matov J, Mensah F, Cook F, Reilly S, Dowell RC. The development and validation of the Short Language Measure (SLaM): A brief measure of general language ability for children in their first year at school. *Int J Lang Commun Disord* 2020;55(3):345-58. DOI: <https://doi.org/10.1111/1460-6984.12522>
108. Nash H, Leavett R, Childs HG. Evaluating the GAPS test as a screener for language impairment in young children. *Int J Lang Commun Disord* 2011;46(6):675-85. DOI: <https://doi.org/10.1111/j.1460-6984.2011.00038.x>
109. Nayeb L, Lagerberg D, Sarkadi A, Salameh EK, Eriksson ME. Identifying language disorder in bilingual children aged 2.5 years requires screening in both languages. *Acta Paediatr* 2020;110(1):265-72. DOI: <https://doi.org/10.1111/apa.15343>
110. Pickstone C. A pilot study of paraprofessional screening of child language in community settings. *Child Lang Teach Ther* 2003;19(1):49-65.
111. Platt AE. Examining the appropriateness of the PPVT-III as a measure of vocabulary development in linguistically diverse kindergarten children. DP - 2012. *Diss Abstr Int Sect A Humanit Soc Sci* 2012;72:2248.
112. Radeborg K, Barthelom E, Sjöberg M, Sahlen B. A Swedish non-word repetition test for preschool children. *Scand J Psychol* 2006;47(3):187-92.
113. Restrepo MA, Schwanenflugel PJ, Blake J, Neuharth-Pritchett S, Cramer SE, Ruston HP. Performance on the PPVT-III and the EVT: applicability of the measures with African American and European American preschool children. *Lang Speech Hear Serv Sch* 2006;37(1):17-27.
114. Schlichting J, Spelberg HCL. A test for measuring syntactic development in young children. *Lang Test* 2003;20(3):241-6. DOI: <https://doi.org/10.1191/0265532203lt256oa>
115. Skipar I. Språkforståelsestesten C-BiLLT : En undersøkelse av reliabiliteten og validiteten til den norske versjonen av språkforståelsestesten Computer-Based instrument for Low Motor Language Testing (C-BiLLT) i aldersgruppen 1;0-3;0 år. Mastergradsavhandling, studieprogram MPED: Fakultet for humaniora, Universitetet i Sørøst-Norge; 2019. Tilgjengelig fra: <https://www.duo.uio.no/bitstream/handle/10852/64427/Masteroppgave---Anne-Lise-Haddeland.pdf?sequence=1&isAllowed=y>
116. Spaulding TJ, Hosmer S, Schechtman C. Investigating the interchangeability and diagnostic utility of the PPVT-III and PPVT-IV for children with and without SLI. *Int J Speech Lang Pathol* 2013;15(5):453-62. DOI: 10.3109/17549507.2012.762042

117. Stokes SF, Klee T. The diagnostic accuracy of a new test of early nonword repetition for differentiating late talking and typically developing children. *J Speech Lang Hear Res* 2009;52(4):872-82. DOI: 10.1044/1092-4388%282009/08-0030%29
118. Thomas-Tate S, Washington J, Craig H, Packard M. Performance of African American preschool and kindergarten students on the expressive vocabulary test. *Lang Speech Hear Serv Sch* 2006;37(2):143-9.
119. Thordardottir E, Brandeker M. The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *J Commun Disord* 2013;46(1):1-16. DOI: 10.1016/j.jcomdis.2012.08.002
120. van der Lely HKJ, Payne E, McClelland A. An investigation to validate the Grammar and Phonology Screening (GAPS) Test to identify children with specific language impairment. *PLoS One* 2011;6(7):e22432-e. DOI: <https://doi.org/10.1371/journal.pone.0022432>
121. Webb ML, Cohen AS, Schwanenflugel PJ. Latent class analysis of different item functioning on the Peabody Picture Vocabulary Test -- III. *Educ Psychol Meas* 2008;68(2):335-51. DOI: 10.1177/0013164407308474
122. Westerlund M. Relationship between a global rating of speech ability at the age of 3 yrs and a phonological screening 1 yr later. A prospective field study. *Scand J Caring Sci* 2001;15(3):222-7. DOI: <https://doi.org/10.1046/j.1471-6712.2001.00016.x>
123. Westerlund M, Sundelin C. Can severe language disability be identified in three-year-olds? Evaluation of a routine screening procedure. *Acta Paediatr* 2007;89(1):94-100. DOI: <https://doi.org/10.1111/j.1651-2227.2000.tb01195.x>
124. Wood C, Schatschneider C. Item bias: Predictors of accuracy on Peabody Picture Vocabulary Test-Fourth Edition items for Spanish-English-speaking children. *J Speech Lang Hear Res* 2019;62(5):1392-402. DOI: 10.1044/2018_JSLHR-L-18-0145
125. Woods P, Rieger I, Wocadlo C, Gordon A. Predicting the outcome of specific language impairment at five years of age through early developmental assessment in preterm infants. *Early Hum Dev* 2014;90(10):613-9. DOI: <https://doi.org/10.1016/j.earlhumdev.2014.07.010>
126. SSB. 12272: Minoritetspråklige barn i barnehager 1-5 år, etter statistikkvariabel, region og år. Oslo: Statistisk Sentralbyrå [lest 13.12.2023]. Tilgjengelig fra: <https://www.ssb.no/statbank/table/12272/>
127. Feltner C, Wallace IF, Nowell S, Orr CJ, Raffa B, Middleton JC, et al. Screening for speech and language delay and disorders in children age 5 years or younger: An evidence review for the U.S. Preventive Services Task Force - Draft report. North Carolina: RTI International–University of North Carolina at Chapel Hill Evidence-based Practice Center; 2023. Tilgjengelig fra: <https://www.uspreventiveservicestaskforce.org/uspstf/document/draft-evidence-review/speech-language-delay-disorders-children-age-5-years-younger-screening>
128. Berkman ND, Wallace I, Watson L, Coyne-Beasley T, Cullen K, Wood C, et al. Screening for speech and language delays and disorders in children age 5 years or younger. A systematic review for the U.S. Preventive Services Task Force. Rockville (MD): Agency for Healthcare Research and Quality (US); 2015. Tilgjengelig fra: <https://www.ncbi.nlm.nih.gov/books/NBK305674/>
129. Baker S, Harding S, Holme C, Lewis R, Seifert M, Wren Y. Review of early language screening suitable for children in Wales from birth to 5 years. GSR report number 55/2022. Cardiff: Welsh Government, Government Social Research, Analysis for Policy; 2022. Tilgjengelig fra: <https://www.gov.wales/sites/default/files/statistics-and-research/2022->

- [08/review-of-early-language-screening-suitable-for-children-in-wales-from-birth-to-5-years.pdf](#)
130. Schwob S, Eddé L, Jacquin L, Leboulanger M, Picard M, Oliveira PR, et al. Using nonword repetition to identify developmental language disorder in monolingual and bilingual children: A systematic review and meta-analysis. *J Speech Lang Hear Res* 2021;64(9):3578-93. DOI: doi:10.1044/2021_JSLHR-20-00552
 131. Stokes SF. Secondary prevention of paediatric language disability: a comparison of parents and nurses as screening agents. *Int J Lang Commun Disord* 1997;32(2s):139-58. DOI: <https://doi.org/10.1111/j.1460-6984.1997.tb01628.x>
 132. Madden CA. Exploring the use of parent report in child speech and language screening. Auburn, Alabama: Auburn University; 2018. Tilgjengelig fra: https://etd.auburn.edu/bitstream/handle/10415/6187/Madden_Thesis.pdf?sequence=2
 133. Sivertsen B, Norvik M, Vulchanova M. The New Reynell Developmental Language Scales. Er den nye Reynell språktest et egnet verktøy for kartlegging av flerspråklige barn? Masteroppgave i Logopedi. Det humanistiske fakultet, Institutt for språk og litteratur: NTNU; 2022. Tilgjengelig fra: <https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3015520>
 134. Nayeb L, Wallby T, Westerlund M, Salameh EK, Sarkadi A. Child healthcare nurses believe that bilingual children show slower language development, simplify screening procedures and delay referrals. *Acta Paediatr* 2015;104(2):198-205. DOI: 10.1111/apa.12834
 135. Integrerings- og mangfoldsdirektoratet (IMDi). Forsøk med språkkartlegging av førskolebarn på helsestasjoner med bruk av SPRÅK 4 - Kartlegging av flerspråklige barn. Spredning av erfaringer fra de tre første årene. Oslo: IMDi; 2010. Tilgjengelig fra: <https://www.imdi.no/contentassets/9be2f14e92c44b129761fe6e5c50fe62/rapport-2010.-forsok-med-sprakkartlegging>
 136. Hoff KG, Helbæk M. Økonomistyring 1: Grunnleggende bedriftsøkonomisk analyse: Universitetsforl.; 2016.
 137. Muller AE, Ames HMR, Jardim PSJ, Rose CJ. Machine learning in systematic reviews: Comparing automated text clustering with Lingo3G and human researcher categorization in a rapid review. *Research Synthesis Methods* 2022;13(2):229-41.
 138. Thomas J, Brunton J, Graziosi S. EPPI-Reviewer 4.0: software for research synthesis. EPPI-Centre Software London: Social Science Research Unit, Institute of Education 2010.
 139. Tidemann A, Elster AC. Maskinlæring [nettdokument]: Store Norske Leksikon [oppdatert 26. juli 2023; lest 25. august 2023]. Tilgjengelig fra: <https://snl.no/maskinl%C3%A6ring>

Vedlegg 1: Søkestrategi

Søk etter kartleggingsverktøy

For å finne relevante kartleggingsverktøy utførte vi ulike søk i Google og Google Scholar. Eksempel på søk: kartlegging|screening språk barn, sprog børn screening, screening språk barn site:se

Vi benyttet oss også av av snøballmetoden, hvor vi startet med noen kilder og fulgte lenker, referanser eller tips videre fra disse.

Vi så etter oversiktsartikler eller nettsider med oversikter over kartleggingsverktøy brukt i ulike land, og hentet informasjon fra følgende kilder:

Universitetet i Stavanger. Oversikt over materiell til kartlegging og diagnostisering av språk. <https://www.uis.no/nb/lesesenteret/oversikt-over-materiell-til-kartlegging-og-diagnostisering-av-sprak>

Universitetet i Stavanger, Språkløyper

<https://sprakloyper.uis.no/sites/default/files/2021-06/04%20-%20Nedlasting%20-%20Observasjons%20og%20kartleggingsverkt%C3%B8y%281%29.pdf>

Statped læringsressurser

<https://www.statped.no/sprakforstyrrelser/utviklingsmessige-sprakforstyrrelser/kartlegging-av-utviklingsmessige-sprakforstyrrelser-dld/>

EMU Danmarks læringsportal. Om sprogvurdering.

<https://emu.dk/dagtilbud/sprogvurdering>

Rikshandboken. Barnhälsovård för proffessionen. Språkundersökning och screening.

<https://www.rikshandboken-bhv.se/metoder--riktlinjer/sprakundersokning-och-screening/#section-21812>

Lene - leikki-ikäisen lapsen neurologisen kehityksen arviointimenetelmä lastenneuvoloille [Lene - en vurderingsmetode for nevrologisk utvikling hos barn i lekealderen for helsestasjoner for barn] – innledning

<https://www.laakarilehti.fi/tieteessa/alkuperaistutkimukset/lene-leikki-ikaisen-lapsen-neurologisen-kehityksen-arviointimenetelma-lastenneuvoloille/>

Review of Early Language Screening Suitable for Children in Wales from Birth to 5 Years. <https://www.gov.wales/sites/default/files/statistics-and-research/2022-08/review-of-early-language-screening-suitable-for-children-in-wales-from-birth-to-5-years.pdf>

Søk i OpenAlex

OpenAlex er et «open source»-datasett med mer enn 250 millioner vitenskapelige objekter som tidsskriftpublikasjoner, stortingsmeldinger og konferanseabstrakter (14). Datakildene til OpenAlex inkluderer Microsoft Academic Graph, CrossRef, ORCID, PubMed, CENTRAL, Unpaywall, ISSN International Centre, preprint-servere samt en rekke andre fagområder og institusjonelle arkiver. Datasettet er satt sammen av åtte typer vitenskapelige enheter (works, authors, sources, institutions, concepts, publishers, funders, locations) og sammenhengene mellom dem. I stedet for å søke i henhold til ord som finnes i en studies tittel eller sammendrag, MeSH-termer eller nøkkelord gitt av forfatteren, tidsskriftet eller databasen, bruker OpenAlex dyp læring for å linke disse objektene sammen, i tillegg til bibliometriske- og siteringslikheter. Et slikt søk resulterer i langt færre identifiserte studier som er irrelevante. Videre bruker OpenAlex «web-crawling» for å innhente nye objekter/artikler, og inkluderer grå litteratur som institusjonelle rapporter, stortingsmeldinger og evalueringer publisert elektronisk.

Databasesøk

Søk 1: Søk etter studier om språkkartlegging generelt

Dato for alle søk: 31. august 2023. Søket ble oppdatert 12. oktober 2023, ved å legge til følgende tekstord: infant*, toddler*, internal* consisten, measurement error*, observation* error* or responsivity.

Søketreff totalt: 7227 (august) + 167 i oktober

Søketreff etter EndNote dublettkontroll: 5527 + 112

Søketreff etter dublettkontroll i EPPI: 4941 + 567

Databaser søkt: MEDLINE, Embase, PsycINFO, Cinahl, ERIC, Cochrane Library, Web of Science

MEDLINE, Embase, PsycINFO (Ovid)

Søkt hver for seg og slått sammen for dublettkontroll.

PsycInfo 1806 to August Week 4 2023 (linje 1-16)

MEDLINE ALL 1946 to August 31, 2023 (linje 17-27)

Embase 1974 to 2023 August 31 (linje 28-40)

Søketreff totalt etter dublettkontroll: 4309 august + 93 oktober

#	Searches	Results
1	screening tests/ or screening/ or health screening/ or measurement/	586351
2	(neonatal birth 1 mo or preschool age 2 5 yrs).ag. or exp Early Childhood Development/ or exp Childhood Development/	252923
3	exp language disorders/ or language delay/ or exp language development/ or speech language pathology/ or language/ or vocabulary/	457667
4	1 and 2 and 3	619
5	("speech and hearing measures"/ or exp "communication and language measures"/ or Language tests/) and ((neonatal birth 1 mo or preschool age 2 5 yrs).ag. or exp Early Childhood Development/ or exp Childhood Development/)	305
6	4 or 5 [emneord]	911
7	(screening or ((assess* or detect* or identif* or diagnos*) adj3 (tool* or checklist* or check-list* or questionnaire* or method* or inventor* or instrument*))).ti,ab. or (assess* or detect* or identif* or diagnos*).ti.	7087391
8	(child or children or preschool* or pre-school* or infant* or toddler*).ti,ab.	4716825
9	((language or speech) adj5 (delay* or disorder* or difficult* or disability* or problem* or abilit*)).ti,ab. or (language or speech).ti.	296865
10	7 and 8 and 9 [tekstord]	11180
11	6 or 10	11797
12	(valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity).ti,ab.	9818428
13	exp psychometrics/	433481
14	12 or 13	9982431
15	11 and 14	4528
16	15 use psych	1624
17	(mass screening/ or early diagnosis/ or "diagnostic techniques and procedures"/ or diagnostic tests, routine/) and (child/ or child, preschool/ or child behavior/) and (language disorders/ or exp language development disorders/ or exp speech disorders/ or Communication	1210

	Disorders/ or *language/ or *verbal behavior/ or exp *speech/ or speech intelligibility/)	
18	(exp Language Disorders/di or exp language development disorders/di or exp speech disorders/di or Language tests/) and (child/ or child, preschool/ or child behavior/) [subheading diagnosis]	17047
19	17 or 18 [emneord]	17740
20	(screening or ((assess* or detect* or identif* or diagnos*) adj3 (tool* or checklist* or check-list* or questionnaire* or method* or inventor* or instrument*))).ti,ab,kf. or (assess* or detect* or identif* or diagnos*).ti.	7114906
21	(child or children or preschool* or pre-school* or infant* or toddler*).ti,ab,kf.	4811872
22	((language or speech) adj5 (delay* or disorder* or difficult* or disability* or problem* or abilit*)).ti,ab,kf. or (language or speech).ti.	298849
23	20 and 21 and 22 [tekstord]	11378
24	19 or 23	27185
25	exp "reproducibility of results"/ or exp "sensitivity and specificity"/ or Psychometrics/ or exp Diagnostic Errors/ or (valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity).ti,ab,kf.	10548443
26	24 and 25	8320
27	26 use medall	3810
28	mass screening/ or screening test/ or screening/ or developmental screening/	467595
29	child/ or preschool child/ or toddler/ or child development/	4474309
30	exp developmental language disorder/ or language disability/ or *language/ or *"speech and language"/ or *verbal communication/ or language ability/ or communication skill/ or language development/ or speech articulation/ or language delay/	221766
31	28 and 29 and 30	1019
32	developmental language disorder/di or language disability/di or language delay/di or Language test/ or exp speech test/ [Diagnosis]	21570
33	29 and 32	11713

34	31 or 33	12330
35	((screening or ((assess* or detect* or identif* or diagnos*) adj3 (tool* or checklist* or check-list* or questionnaire* or method* or inventor* or instrument*))).ti,ab,kf. or (assess* or detect* or identif* or diagnos*).ti.) and (child or children or preschool* or pre-school* or infant* or toddler*).ti,ab,kf. and (((language or speech) adj5 (delay* or disorder* or difficult* or disabilit* or problem* or abilit*))).ti,ab,kf. or (language or speech).ti.) [tekstord]	11378
36	34 or 35	21899
37	psychometry/ or "sensitivity and specificity"/ or exp validity/ or exp reliability/ or (valid* or reproducib* or reliab* or sensitiv* or specific* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity).ti,ab,kf.	10167379
38	36 and 37	6759
39	38 use oemezd	2921
40	limit 39 to embase [Limit not valid in Ovid MEDLINE®,APA PsycInfo; records were retained]	1281
41	(sprak 4 or sprak4 or spraak 4 or spraak4 or sats 2 or sats2 or spraakundersoekning or spraakscreening or Sprakundersoekning or sprakscreening or miniscalcos or spraakfyran or sprakfyran).ti,ab,kf.	7
42	16 or 27 or 40 or 41	6720
43	limit 42 to yr=>2000 -2010»	1828
44	limit 42 to yr=>2010 -Current»	4093
45	remove duplicates from 43	1631
46	remove duplicates from 44	3665
47	45 or 46	5059
48	remove duplicates from 47	5058

CINAHL (Ebsco)

790 august + 4 oktober

S21 S11 AND S14 AND S19

790

Limiters - Published Date: 20000101-20230931; Exclude MEDLINE records

S20	S11 AND S14 AND S19	1,383
S19	S15 OR S16 OR S17 OR S18	805,344
S18	TI (valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity) OR AB (valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity)	653,158
S17	(MH «Sensitivity and Specificity»)	94,477
S16	(MH «Psychometrics»)	33,702
S15	(MH «Reliability and Validity+»)	282,798
S14	S12 OR S13	920,980
S13	TI (child or children or preschool* or pre-school* or infant* or toddler*) OR AB (child or children or preschool* or pre-school* or infant* or toddler*)	527,691
S12	(MH «Child+»)	750,752
S11	S9 OR S10	7,747
S10	(MH “Speech and Language Assessment”)	4,021
S9	S3 AND S8	4,401
S8	S4 OR S5 OR S6 OR S7	53,827
S7	TI (language or speech) OR AB ((language or speech) N/5 (delay* or disorder* or difficult*))))	33,811
S6	(MH «Speech»)	6,201
S5	(MH “Language Disorders”) OR (MH “Speech Disorders”)	9,545
S4	(MH «Language»)	17,957
S3	S1 OR S2	578,120

S2	TI (screening or assess* or detect* or identif* or diagnos* or ((assess* or detect* or identif* or diagnos*) N/3 (tool* or check-list* or check-list* or questionnaire* or method* or inventor*))) OR AB (screening or ((assess* or detect* or identif* or diagnos*) N/3 (tool* or checklist* or check-list* or questionnaire* or method* or inventor*)))	560,394
S1	(MH «Health Screening»)	55,024

ERIC (Ebsco)

Søketreff: 527 august + 10 oktober

S16	S11 AND S14 Limiters - Date Published: 20000101-20230931	527
S15	S11 AND S14	587
S14	S12 OR S13	149,213
S13	TI (valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity) OR AB (valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity)	136,355
S12	DE "Validity" OR DE "Predictive Validity" OR DE "Test Validity" OR DE "Accuracy" OR DE "Evaluation" DE "Reliability" OR DE "Inter-rater Reliability" OR DE "Test Reliability" OR DE "Validated Programs"	48,557
S11	S3 OR S6 OR S10	1,711
S10	S7 AND S8 AND S9	954
S9	TI (language or speech) OR AB ((language or speech) N/5 (delay* or disorder* or difficult*))	75,087
S8	TI (child or children or preschool* or "pre-school" or infant* or toddler*) OR AB (child or children or preschool* or "pre-school" or infant* or toddler*)	321,848

S7	TI (screen* or assess* or detect* or diagnos*) OR AB (screening or ((assess* or detect* or identif* or diagnos*) N/3 (tool* or check-list* or check-list* or questionnaire* or method* or inventor*)))	72,584	EditS7
S6	S2 AND S4 AND S5	227	EditS6
S5	DE "Language Impairments" OR DE "Aphasia" OR DE "Dyslexia" OR DE "Speech Impairments" OR DE "Articulation Impairments" OR DE "Delayed Speech" OR DE "Stuttering" OR DE "Voice Disorders"	11,855	EditS5
S4	DE "Screening Tests" OR DE "Diagnostic Tests" OR DE "Preschool Tests"	13,550	EditS4
S3	S1 AND S2	664	EditS3
S2	(DE "Children" OR DE "Child Development") OR (DE "Preschool Children")	93,587	EditS2
S1	(DE "Language Tests") OR (DE "Speech Tests")	11,514	EditS1

Web of Science

Søketreff: 1474 august + 52 oktober

(TI=(screening or assessment or diagn* or detect*) or AB=(screening or ((assess* or detect* or identif* or diagnos*) Near/3 (tool* or checklist* or check-list* or questionnaire* or method* or inventor*)))) and TS=(child or children or preschool* or "pre-school" or infant* or toddler*) and (TI=(language or speech) or AB=(language or speech) Near/5 (test* or delay* or disorder* or difficult*)) and TS=(valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer varia* or internal* consisten* or measurement error* or observation* error* or responsivity) publication date 2000-2023

Cochrane Library

Søketreff: Trials 120 august + 0 oktober, systematic reviews 7 august + 8 oktober

- #1 MeSH descriptor: [Mass Screening] explode all trees
- #2 MeSH descriptor: [Early Diagnosis] explode all trees
- #3 MeSH descriptor: [Diagnostic Techniques and Procedures] this term only
- #4 MeSH descriptor: [Diagnostic Tests, Routine] explode all trees
- #5 #1 or #2 or #3 or #4
- #6 MeSH descriptor: [Child] explode all trees
- #7 MeSH descriptor: [Child Development] explode all trees
- #8 #6 or #7
- #9 MeSH descriptor: [Language Disorders] explode all trees
- #10 MeSH descriptor: [Language Development Disorders] explode all trees
- #11 MeSH descriptor: [Speech Disorders] explode all trees

- #12 MeSH descriptor: [Communication Disorders] explode all trees
- #13 MeSH descriptor: [Language] this term only
- #14 MeSH descriptor: [Verbal Behavior] explode all trees
- #15 MeSH descriptor: [Speech] explode all trees
- #16 MeSH descriptor: [Speech Intelligibility] explode all trees
- #17 #9 or #10 or #11 or #12 or #13 or #14 or #15 or #16
- #18 #5 and #8 and #17
- #19 (screening or ((assess* or detect* or identif* or diagnos*) near/3 (tool* or check-list* or check-list* or questionnaire* or method* or inventor* or instrument*)):ti,ab or (assess* or detect* or identif* or diagnos*):ti
- #20 (child or children or preschool* or (pre next school*)):ti,ab
- #21 ((language or speech) near/5 (delay* or disorder* or difficult* or disabilit* or problem* or abilit*)):ti,ab or (language or speech):ti
- #22 #19 and #20 and #21
- #23 MeSH descriptor: [Language Disorders] explode all trees and with qualifier(s): [diagnosis - DI]
- #24 MeSH descriptor: [Language Development Disorders] explode all trees and with qualifier(s): [diagnosis - DI]
- #25 MeSH descriptor: [Speech Disorders] explode all trees and with qualifier(s): [diagnosis - DI]
- #26 MeSH descriptor: [Language Tests] explode all trees
- #27 #23 or #24 or #25 or #26
- #28 #27 and #20
- #29 #22 or #28
- #30 MeSH descriptor: [Reproducibility of Results] explode all trees
- #31 MeSH descriptor: [Sensitivity and Specificity] explode all trees
- #32 MeSH descriptor: [Psychometrics] explode all trees
- #33 MeSH descriptor: [Diagnostic Errors] explode all trees
- #34 (valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia*):ti,ab
- #35 #30 or #31 or #32 or #33 or #34
- #36 #29 and #35 with Publication Year from 2000 to 2023, in Trials
- #37 #29 and #35 with Cochrane Library publication date Between Jan 2000 and Sep 2023
- #38 #36 or #37

Søk 2: Supplerende søk på verktøynavn

Dato for alle søk: 12.10.2023

Søketreff totalt: 1087

Søketreff etter EndNote dublettkontroll: 702

Søketreff etter EPPI dublettkontroll:

MEDLINE, Embase, PsycINFO (Ovid)

Embase 1974 to 2023 October 11

Ovid MEDLINE ALL 1946 to October 11, 2023

APA PsycInfo 1806 to October Week 1 2023

Søketreff: 645

#	Searches	Results
1	exp "reproducibility of results" / or exp "sensitivity and specificity" / or Psychometrics / or exp Diagnostic Errors / or (valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity).ti,ab,kf.	10548443
2	1 use medall	4434301
3	(valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity).ti,ab. or exp psychometrics /	9982431
4	3 use psyh	807781
5	psychometry / or "sensitivity and specificity" / or exp validity / or exp reliability / or (valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity).ti,ab,kf.	10167379
6	5 use oomezd	5311984
7	2 or 3 or 6	10582794
8	Breuer-Weuffen.ti,ab,kf,hw.	9
9	(Picture Vocabulary Scale* or british picture vocabulary test*).ti,ab,kf,hw.	166
10	CSBS DP Infant-Toddler Checklist.ti,ab,kf,hw.	22
11	"Communication and Symbolic Behavior Scales".ti,ab,kf,hw.	225
12	KAM next test.ti,ab,kf,hw.	0
13	Reynell* Development* Language.ti,ab,kf,hw.	344

14	(RDLS scale* or RDLS test*).ti,ab,kf,hw.	5
15	reynell* language development*.ti,ab,kf,hw.	19
16	(Sn?gubb* or sno?gubb*).ti,ab,kf,hw.	0
17	(Sprogvrudering or Spr?kunders?kning or spra?kunderso?kning or spr?kscreening or spra?kscreening or sprogscreening).ti,ab,kf,hw.	0
18	Brigance Preschool.ti,ab,kf,hw.	10
19	Early Referral Identification.ti,ab,kf,hw.	1
20	Edmonton* Narrative Norm*.ti,ab,kf,hw.	30
21	(Early Language Identification or (elim adj3 (measure or test or scale))).ti,ab,kf,hw.	5
22	Looksee.ti,ab,kf,hw.	0
23	Quick Interactive Language.ti,ab,kf,hw.	14
24	rourke baby record.ti,ab,kf,hw.	66
25	Schedule of Growing Skills.ti,ab,kf,hw.	51
26	Stoke Speaks Out.ti,ab,kf,hw.	0
27	Staged Pathway* Toolkit*.ti,ab,kf,hw.	0
28	Expressive Vocabulary Test*.ti,ab,kf,hw.	199
29	Peabody Picture Vocabulary Test*.ti,ab,kf,hw.	3180
30	Renfrew Language.ti,ab,kf,hw.	0
31	"Speech and Language Pathology Early Screening Instrument".ti,ab,kf,hw.	2
32	splesi.ti,ab,kf,hw.	0
33	(lene and (language or speech)).ti,ab,kf,hw.	11
34	wellcomm.ti,ab,kf,hw.	1
35	fluharty preschool.ti,ab,kf,hw.	27
36	(tribus and (sprak or spraak or language or speech or screening)).ti,ab,kf.	0
37	or/8-36	4270
38	7 and 37	1297
39	remove duplicates from 38 [verktøynavn]	1108
40	limit 39 to yr=>2000 -Current>	645

CINAHL (Ebsco)

Søketreff: 63

S9	S3 AND S8 Limiters - Published Date: 20000101-20231031; Exclude MEDLINE records	63	EditS9
S8	S4 OR S5 OR S6 OR S7	813,616	EditS8
S7	(MH «Reliability and Validity+»)	284,354	EditS7
S6	MH «Psychometrics»	34,051	EditS6
S5	MH «Sensitivity and Specificity»	94,937	EditS5
S4	TI (valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity) OR AB (valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity)	661,601	EditS4
S3	S1 OR S2	395	EditS3
S2	TI ((Snogubb* or snoegubb* or Sprogvurdering or Sprakundersokning or spraakundersoekning or sprakscreening or spraakscreening or sprogscreening)) OR AB ((Snogubb* or snoegubb* or Sprogvurdering or Sprakundersokning or spraakundersoekning or sprakscreening or spraakscreening or sprogscreening)) OR MW ((Snogubb* or snoegubb* or Sprogvurdering or Sprakundersokning or spraakundersoekning or sprakscreening or spraakscreening or sprogscreening))	0	EditS2
S1	TI (“Breuer-Weuffen” or “Picture Vocabulary Scale” or “british picture vocabulary test” or “CSBS DP Infant-Toddler Checklist” or “Communication and Symbolic Behavior Scales” or “KAM test” or “Reynell Development Language” or “RDLS scale” or “RDLS test” or “reynell language development” or “Brigance Preschool” or “Early Referral Identification” or “Edmonton Narrative Norms” or “Early Language Identification” or (elim N/3 (measure or test or scale)) or Looksee or “Quick Interactive Language” or “rourke baby record” or “Schedule of Growing Skills” or “Stoke Speaks Out” or “Staged Pathway Toolkit” or “Expressive Vocabulary Test” or “Peabody Picture Vocabulary Test” or “Renfrew Language” or “Speech and Language		

Pathology Early Screening Instrument" or ((splesi or lene) and (language or speech)) or wellcomm or "fluharty preschool" or (tribus and (sprak or spraak or language or speech or screening))) OR AB ("Breuer-Weuffen" or "Picture Vocabulary Scale" or "british picture vocabulary test" or "CSBS DP Infant-Toddler Checklist" or "Communication and Symbolic Behavior Scales" or "KAM test" or "Reynell Development Language" or "RDLS scale" or "RDLS test" or "reynell language development" or "Brigance Preschool" or "Early Referral Identification" or "Edmonton Narrative Norms" or "Early Language Identification" or (elim N/3 (measure or test or scale)) or Looksee or "Quick Interactive Language" or "rourke baby record" or "Schedule of Growing Skills" or "Stoke Speaks Out" or "Staged Pathway Toolkit" or "Expressive Vocabulary Test" or "Peabody Picture Vocabulary Test" or "Renfrew Language" or "Speech and Language Pathology Early Screening Instrument" or ((splesi or lene) and (language or speech)) or wellcomm or "fluharty preschool" or (tribus and (sprak or spraak or language or speech or screening))) OR MW ("Breuer-Weuffen" or "Picture Vocabulary Scale" or "british picture vocabulary test" or "CSBS DP Infant-Toddler Checklist" or "Communication and Symbolic Behavior Scales" or "KAM test" or "Reynell Development Language" or "RDLS scale" or "RDLS test" or "reynell language development" or "Brigance Preschool" or "Early Referral Identification" or "Edmonton Narrative Norms" or "Early Language Identification" or (elim N/3 (measure or test or scale)) or Looksee or "Quick Interactive Language" or "rourke baby record" or "Schedule of Growing Skills" or "Stoke Speaks Out" or "Staged Pathway Toolkit" or "Expressive Vocabulary Test" or "Peabody Picture Vocabulary Test" or "Renfrew Language" or "Speech and Language Pathology Early Screening Instrument" or ((splesi or lene) and (language or speech)) or wellcomm or "fluharty preschool" or (tribus and (sprak or spraak or language or speech or screening)))

ERIC (Ebsco

Søketreff: 68

S7	S3 AND S6 Limiters - Date Published: 20000101-20231131	68	EditS7
S6	S4 OR S5	691,536	EditS6
S5	DE "Validity" OR DE "Predictive Validity" OR DE "Test Validity" OR DE "Accuracy" OR DE "Evaluation" OR DE "Reliability" OR DE "Inter-rater Reliability" OR DE "Test Reliability" OR DE "Validated Programs"	61,506	EditS5

S4	TI (valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity) OR AB (valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer* varia* or internal* consisten* or measurement error* or observation* error* or responsivity)	661,601	EditS4
S3	S1 OR S2	Display	EditS3
S2	TI ((Snogubb* or snoegubb* or Sprogvurdering or Sprakundersokning or spraakundersoekning or sprakscreening or spraakscreening or sprogscreening)) OR AB ((Snogubb* or snoegubb* or Sprogvurdering or Sprakundersokning or spraakundersoekning or sprakscreening or spraakscreening or sprogscreening)) OR MW ((Snogubb* or snoegubb* or Sprogvurdering or Sprakundersokning or spraakundersoekning or sprakscreening or spraakscreening or sprogscreening))	Display	EditS2
S1	TI (“Breuer-Weuffen” or “Picture Vocabulary Scale” or “british picture vocabulary test” or “CSBS DP Infant-Toddler Checklist” or “Communication and Symbolic Behavior Scales” or “KAM test” or “Reynell Development Language” or “RDLS scale” or “RDLS test” or “reynell language development” or “Brigance Preschool” or “Early Referral Identification” or “Edmonton Narrative Norms” or “Early Language Identification” or (elim N/3 (measure or test or scale)) or Looksee or “Quick Interactive Language” or “rourke baby record” or “Schedule of Growing Skills” or “Stoke Speaks Out” or “Staged Pathway Toolkit” or “Expressive Vocabulary Test” or “Peabody Picture Vocabulary Test” or “Renfrew Language” or “Speech and Language Pathology Early Screening Instrument” or ((splesi or lene) and (language or speech)) or wellcomm or “fluharty preschool” or (tribus and (sprak or spraak or language or speech or screening))) OR AB (“Breuer-Weuffen” or “Picture Vocabulary Scale” or “british picture vocabulary test” or “CSBS DP Infant-Toddler Checklist” or “Communication and Symbolic Behavior Scales” or “KAM test” or “Reynell Development Language” or “RDLS scale” or “RDLS test” or “reynell language development” or “Brigance Preschool” or “Early Referral Identification” or “Edmonton Narrative Norms” or “Early Language Identification” or (elim N/3 (measure or test or scale)) or Looksee or “Quick Interactive Language” or “rourke baby record” or “Schedule of Growing Skills” or	Display	EditS1

“Stoke Speaks Out” or “Staged Pathway Toolkit” or “Expressive Vocabulary Test” or “Peabody Picture Vocabulary Test” or “Renfrew Language” or “Speech and Language Pathology Early Screening Instrument” or ((splesi or lene) and (language or speech)) or well-comm or “fluharty preschool” or (tribus and (sprak or spraak or language or speech or screening))) OR MW (“Breuer-Weuffen” or “Picture Vocabulary Scale” or “british picture vocabulary test” or “CSBS DP Infant-Toddler Checklist” or “Communication and Symbolic Behavior Scales” or “KAM test” or “Reynell Development Language” or “RDLS scale” or “RDLS test” or “reynell language development” or “Brigance Preschool” or “Early Referral Identification” or “Edmonton Narrative Norms” or “Early Language Identification” or (elim N/3 (measure or test or scale)) or Looksee or “Quick Interactive Language” or “rourke baby record” or “Schedule of Growing Skills” or “Stoke Speaks Out” or “Staged Pathway Toolkit” or “Expressive Vocabulary Test” or “Peabody Picture Vocabulary Test” or “Renfrew Language” or “Speech and Language Pathology Early Screening Instrument” or ((splesi or lene) and (language or speech)) or well-comm or “fluharty preschool” or (tribus and (sprak or spraak or language or speech or screening)))

Web of Science

Søketreff: 193

TS=(“Breuer-Weuffen” or “Picture Vocabulary Scale” or “british picture vocabulary test” or “CSBS DP Infant-Toddler Checklist” or “Communication and Symbolic Behavior Scales” or “KAM test” or “Reynell Development Language” or “RDLS scale” or “RDLS test” or “reynell language development” or “Brigance Preschool” or “Early Referral Identification” or “Edmonton Narrative Norms” or “Early Language Identification” or (elim N/3 (measure or test or scale)) or Looksee or “Quick Interactive Language” or “rourke baby record” or “Schedule of Growing Skills” or “Stoke Speaks Out” or “Staged Pathway Toolkit” or “Expressive Vocabulary Test” or “Peabody Picture Vocabulary Test” or “Renfrew Language” or “Speech and Language Pathology Early Screening Instrument” or ((splesi or lene) and (language or speech)) or well-comm or “fluharty preschool” or (tribus and (sprak or spraak or language or speech or screening)) or Snogubb* or snoegubb* or Sprogvurdering or Sprakundersokning or spraakundersokning or sprakscreening or spraacscreening or sprogscreening) and (valid* or reproducib* or reliab* or sensitiv* or specificit* or accurac* or predictive or psychometric* or false positive* or false negative* or fidelity or reference value* or likelihood ratio* or diagnostic* error* or observer varia* or internal* consisten* or measurement error* or observation* error* or responsivity)

Cochrane Library

Søketreff: 118

("Breuer-Weuffen" or "Picture Vocabulary Scale" or "british picture vocabulary test" or "CSBS DP Infant-Toddler Checklist" or "Communication and Symbolic Behavior Scales" or "KAM test" or "Reynell Development Language" or "RDLS scale" or "RDLS test" or "reynell language development" or "Brigance Preschool" or "Early Referral Identification" or "Edmonton Narrative Norms" or "Early Language Identification" or (elim Near/3 (measure or test or scale)) or Looksee or "Quick Interactive Language" or "rourke baby record" or "Schedule of Growing Skills" or "Stoke Speaks Out" or "Staged Pathway Toolkit" or "Expressive Vocabulary Test" or "Peabody Picture Vocabulary Test" or "Renfrew Language" or "Speech and Language Pathology Early Screening Instrument" or ((spleti or lene) and (language or speech)) or wellcomm or "fluharty preschool" or (tribus and (sprak or spraak or language or speech or screening)) or Snogubb* or snoegubb* or Sprogvrudering or Sprakundersokning or spraakundersokning or sprakscreening or spraakscreening or sprogscreening):ti,ab,kw with Cochrane Library publication date from Jan 2000 to Oct 2023, in Cochrane Reviews and Trials

Vedlegg 2: Mulig relevante studier publisert på andre språk enn engelsk, norsk, dansk eller svensk

Mulig relevante studier publisert på andre språk enn engelsk, norsk, dansk eller svensk.

Betz-Morhard K and Von Suchodoletz W. (2011). Language screening in infancy. Identification of developmental language disorders? [German]. *Padiatrische Praxis*, 77(4), pp.623-632.

Bouchard M G and Fitzpatrick E M; Olds J. (2009). Psychometric analysis of assessment tools used with francophone children. *Canadian Journal of Speech-Language Pathology & Audiology*, 33(3), pp.129-139. [French]

Brunner M, Pfeiffer B, Heinrich C and Proschel U. (2005). Evaluation of the new Heidelberg Preschool screening for auditory perception and language processing. [German]. *Folia Phoniatica et Logopaedica*, 57(1), pp.48-58.

Buschmann A and Sachse S. (2017). Early diagnosis of language delay. A diagnostic work-up and methods. *Fruh-forderung interdisziplinär*, 36(2), pp.82-92.

de Araújo Barbosa A L, Lemos F F, Azoni C A S. (2021). Application of a vocabulary screening instrument for children between 3- and 7-years-old: a pilot study. *CoDAS*, 33(2), pp.e20190154-NA.

Eikerling M and Lorusso M L. (2021). Da guckste in die Röhre?! Pilotierung computer-gestützter Screenings für bilinguale Kinder mit der Screening-Plattform MuLiMi. *Forschung Sprache*, 9(2), pp.29-37.

Euler H A, Holler-Zittlau I, Minnen S, Sick U, Dux W, Zaretsky Y, Neumann K. (2010). [Validity criteria of a short test to assess speech and language competence in 4-year-olds]. *HNO*, 58(11), pp.1116-23.

Gurgel L G, Plentz R D M; Joly M C R A; Reppold C T. (2010). Instrumentos de avaliação da compreensão de linguagem oral em crianças e adolescentes: uma revisão sistemática da literatura. *Revista Neuropsicologia Latinoamericana*, 2(1), pp.1-10.

Keilmann A, Moein G and Scholer H. (2012). [Does the SETK 3-5 detect clinically diagnosed language impairment?]. *HNO*, 60(1), pp.63-71.

Kiese-Himmel C and Rosenfeld J. (2012). [Evaluation of current assessment tools in early detection of developmental deviations in speech and language in the German preventive paediatric examination (Kindervorsorgeuntersuchung U8)]. *Gesundheitswesen*, 74(10), pp.661-72.

Kramer D, Kompauer I and Scholer H. (2013). [Normalisation of screening HASE for 4- to 41/2-year-old children]. *Gesundheitswesen*, 75(8-9), pp.492-5.

Leclercq A L and Veys E. (2014). Reflections on the selection of standardized tests in the diagnosis of specific language impairment. [French]. *ANAE - Approche Neuropsychologique des Apprentissages chez l'Enfant*, 26(131), pp.374-382.

MacLeod A A N, Sutton A, Sylvestre A, Thordardottir E and Trudeau N. (2014). Screening tool for speech sound development disorders: Theoretical bases and preliminary data. *Canadian Journal of Speech-Language Pathology & Audiology*, 38(1), pp.40-56.

-
- Melzer J, Rissling J K and Petermann F. (2015). Language diagnostics in preschool age: Analysis of current test procedures for children aged between 4 and 5 years. [German]. *Monatsschrift für Kinderheilkunde*, 163(1), pp.58-66.
-
- Metz D, Belhadj K E, Karpinski N and Petermann F. (2011). [The validity of the SET 5-10 (language level test for children aged between five and ten years): first analyses]. *Gesundheitswesen*, 73(10), pp.637-43.
-
- Munoz-Caicedo A, Zapata-Ossa H J and Perez-Tenorio L M. (2013). [Validating abbreviated developmental scale (ADS-1) criterion in the language-hearing domain]. *Revista de Salud Publica*, 15(3), pp.386-97.
-
- Neumann K, Holler-Zittlau I, van Minnen S, Sick U, Zaretsky Y and Euler H A. (2011). [Fool's gold standards in language screening. Sensitivity and specificity of the Hessian child language screening test (Kindersprachscreening, KiSS)]. *HNO*, 59(1), pp.97-109.
-
- Padovan N. (2014). Probir jezičnih sposobnosti u predškolskoj dobi: individualni ili grupni? Language screening for preschool children: Individual or group testing? *Logopedics*, 4(1), pp. 23-30.
-
- Ptok M and Dunkelmann A. (2011). [Phonological processing in preschool children. Test standardisation]. *HNO*, 59(3), pp.274-9.
-
- Ptok M, Lichte C, Bussing B, Schwemmler C and Teichler N. (2006). [Phoneme discrimination test material for preschool children: development and preliminary assessment]. *Laryngo- Rhino- Otologie*, 85(9), pp.640-8.
-
- Roy B and Alla F. (2002). ERTL4 and ERTLA6: Tools for identifying language and learning disorders for doctors' use. [French]. *ANAE - Approche Neuropsychologique des Apprentissages chez l'Enfant*, 14(1), pp.58-62.
-
- Sachse S and von Suchodoletz W. (2013). [Language development between the German preventive paediatric examinations at the age of 2 and 3 in children with and without language delay]. *Klinische Pädiatrie*, 225(4), pp.194-200.
-
- Sachse S, Anke B and Von Suchodoletz W. (2007). Early identification of language delay among two-year-olds - A comparison of methods. [German]. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 35(5), pp.323-331.
-
- Sachse S, Pecha A and von Suchodoletz W. (2007). Früherkennung von Sprachentwicklungsstörungen. *Monatsschrift Kinderheilkunde*, 155(2), pp.140-145.
-
- Steinbrink C, Schwanda S, Klatte M and Lachmann T. (2010). Do perceptual abilities at school entry predict reading and spelling abilities at the end of grades 1 and 2? Investigating the predictive validity of the "differenzierungsproben" 1 and 2 by Breuer and Weuffen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 42(4), pp.188-200.
-
- Tippelt S, Kuhn P, Grossheinrich N and Von Suchodoletz W. (2011). Diagnostic accuracy of language tests and parent rating for identifying language disorders. [German]. *Laryngo- Rhino- Otologie*, 90(7), pp.421-427.
-
- Touzain M and Billard C. (2019). Screening for oral language disorders in a new computerized battery for the first-line examination (modular computerized test battery (BMT-i)). [French]. *ANAE - Approche Neuropsychologique des Apprentissages chez l'Enfant*, 31(161), pp.494-504.
-

Vedlegg 3: Bruk av maskinlæring

Vi brukte maskinlæring for å hjelpe oss med å utføre flere av trinnene i denne kunnskapsoppsummeringen mer effektivt. Dette vedlegget beskriver *hvordan* vi brukte maskinlæring. I beskrivelsen benytter vi enkelte engelske begreper grunnet mangel på norske begrepsoversettelser innenfor maskinlæringsfeltet. I slutten av vedlegget er det en forklaring av begrepene vi har brukt for de ulike maskinlæringsfunksjonene som vi brukte i denne kunnskapsoppsummeringen.

Tabell 8: *Beskrivelse av maskinlæringsprosedyre i screeningprosessen*

Beskrivelse av fremgangsmåte	
Steg 1	Vi startet med pilotering av inklusjonskriteriene blant alle prosjektmedarbeidere på ca 100 referanser, for å sikre felles forståelse av seleksjonskriteriene.
Steg 2	For raskere å identifisere referanser som oppfylte inklusjonskriteriene i arbeidet med å vurdere titler og sammendrag benyttet vi « <i>priority screening</i> », og 810 referanser ble uavhengig vurdert av to medarbeidere
Steg 3	Deretter clustret vi de resterende uleste referansene (n=7566) med følgende parametre: Maximum hierarchy depth (1), Maximum cluster size (0,35), Minimum label length (1), minimum cluster size (0,00), single word label weight (0,5), og automatisk ekskluderte referanser i grupper som var åpenbart irrelevante ¹ .
Steg 4	Når 6087 referanser gjensto bygget og testet vi en custom classifier. Den ble bygget basert på 194 inkluderte og 385 ekskluderte referanser på tittel- og sammendragsnivå, og testet på 11 inkluderte og 39 ekskluderte på fulltekstnivå. Her oppnådde modellen 100% recall ved terskelverdien 50%, så alle uleste referanser som ble klassifisert til å ha under 50% sannsynlighet for å møte inklusjonskriteriene ble automatisk ekskludert (n=4439). Referansene

¹ Grupper med følgende clusternavn: Autism Spectrum Disorder, Cleft Palate, Cochlear Implants, Magnetic Resonance Imaging, Mental Retardation, Patients, Poor Readers, Stuttering

som ble klassifisert med 70-99% sannsynlighet vurderte to prosjektmedarbeidere uavhengig av hverandre, mens referansene som ble klassifisert med 50-69% sannsynlighet vurderte én prosjektmedarbeider.

Steg 5 Som et ledd i kvalitetssikringen gjorde vi også enkeltsøk blant de automatisk ekskluderte referansene på verktøynavn og identifiserte ingen relevante studier.

Begrepsforklaring

Algoritme kan forklares som en fullstendig presis og trinnvis beskrivelse av en prosedyre for operasjoner som er beregnet for å løse et problem (136).

Automatic text clustering er en prosess som analyserer fordelingen og konteksten av ord, deler av ord eller termer i tittelen og sammendrag. Funksjonen genererer grupper basert på fellestrekk i titler og sammendrag. Hver referanse kan tilordnes én eller flere grupper. Denne funksjonen kan brukes til å identifisere grupper med åpenbart relevante eller irrelevante referanser, mønster i datasett (137).

Classifiers er algoritmer som er trent til å gjenkjenne data med spesielle kjennetegn og som kategoriserer data etter om de innehar disse kjennetegnene eller ikke.

Custom classifier er en maskinlæringsfunksjon med en tilpasset binær modell for den gitte problemstillingen som kategoriserer referansene som relevant eller ikke relevant (x eller ikke x). Modellen trenes opp på treningsdata, inkluderte og ekskluderte referanser på tittel og sammendragsnivå, og «lærer» å anvende samme beslutningstaking på nye data. Custom classifiers er ofte brukt til å kategorisere gjenværende referanser etter sannsynlighet for å oppfylle/ikke oppfylle inklusjonskriteriene. Custom classifiermodellen i EPPI Reviewer presenterer referansene etter prosentvis sannsynlighet for å oppfylle inklusjonskriteriene (81;138).

Maskinlæring er en spesialisering innen kunstig intelligens hvor man bruker statistiske metoder for å la datamaskiner finne mønstre i store datamengder (139). Enkelt sagt betyr maskinlæring at vi tar i bruk algoritmer som gjør at datamaskinen er i stand til å lære fra og utvikle sin beslutningsstøtte basert på empiriske data som vi fører den med.

Priority screening er en rangeringsalgoritme i programvaren EPPI-Reviewer (81;138) som læres opp av forskernes avgjørelser om inklusjon og eksklusjon av referanser på tittel- og sammendragsnivå. Rangeringsalgoritmer er algoritmer som er trent til å gjenkjenne relevante data og til å presentere dataene etter relevans. Referanser som algoritmen anser som mer relevante basert på forskernes avgjørelser om inklusjon blir skjøvet frem i «køen». På denne måten får vi raskere overblikk over hvor mange referanser som muligens treffer inklusjonskriteriene enn om vi skulle lest referansene i tilfeldig rekkefølge.

Vedlegg 4: Ekskluderte studier lest i fulltekst

Tabell 9: Ekskluderte studier lest i fulltekst (n=206)

Referanse	Eksklusjonsårsak
Aikens N, West J, McKee K, Moiduddin E, Atkins-Burnett S and Xue Y. (2020). Screening approaches for determining the language of assessment for dual language learners: Evidence from Head Start and a universal preschool initiative. <i>Early Childhood Research Quarterly</i> , 51, pp.39-54.	Oppgir ikke måleegenskaper i henhold til COSMIN
Anaya J B, Pena E D and Bedore L M. (2018). Conceptual Scoring and Classification Accuracy of Vocabulary Testing in Bilingual Children. <i>Language and Speech & Hearing Services in the Schools</i> , 49(1), pp.85-97.	Feil populasjon
Antoniou F, Ralli A M, Mouzaki A, Diamanti V and Papaioannou S. (2022). Logometro®: The psychometric properties of a norm-referenced digital battery for language assessment of Greek-speaking 4–7 years old children. <i>Frontiers In Psychology</i> , 26(13), 900600.	Feil land
Arabpour M, Nakhshab M, Humphry S and Kazemi Y. (2023). Systematic Analysis of Language Transcripts' (SALT) transcribing method and Narrative Assessment Protocol (NAP) online coding method: are they interchangeable?. <i>Logopedics, Phoniatrics and Vocology</i> , 48(1), pp.35-43.	Feil land
Baker S, Harding S, Holme C, Lewis R, Seifert M and Wren Y. (2022). <i>Review of Early Language Screening Suitable for Children in Wales from Birth to 5 Years</i> . Welsh Government.	Feil studiedesign/metode
Barger B O, Rice C E, Wolf R and Roach A T. (2018). Better together: Developmental screening and monitoring best identify children who need early intervention. <i>Disability And Health Journal</i> , 11(3), pp.420-426.	Feil land
Beaulieu-Poulin C, Simard M-N, Babakissa H, Lefebvre F and Luu T M. (2016). Validity of the language development survey in infants born preterm. <i>Early Human Development</i> , 98, pp.11-16.	Feil verktøy
Berglund E. (2000). Language assessment of Swedish 3-year-olds using a protocol of natural speech: PLUS-3—productive language Use at 3. <i>Logopedics, Phoniatrics and Vocology</i> , 25(2), pp.51-62.	Feil verktøy
Berglund E and Eriksson M E. (2000). Communicative Development in Swedish Children 16-28 months old: The Swedish Early Communicative Development Inventory—Words and Sentences. <i>Scandinavian Journal Of Psychology</i> , 41(2), pp.133-144.	Feil verktøy

Berglund E and Eriksson M E. (2000). Reliability and content validity of a new instrument for assessment of communicative skills and language abilities in young Swedish children. <i>Logopedics Phoniatrics Vocology</i> , 25(4), pp.176-185.	Feil verktøy
Berkman N D, Wallace I, Watson L, Coyne-Beasley T, Cullen K, Wood C and Lohr K N. (2015). Screening for Speech and Language Delays and Disorders in Children Age 5 Years or Younger. Rockville (MD): Agency for Healthcare Research and Quality (US). Report No.: 13-05197-EF-1.	Feil studiede-sign/metode
Bleses D, Jensen P, Højen A and Dale P S. (2018). An educator-administered measure of language development in young children. <i>Infant Behavior & Development</i> , 52, pp.104-113.	Feil verktøy
Bloder T, Eikerling M and Lorusso M L. (2023). Evaluating the role of word-related parameters in the discriminative power of a novel nonword repetition task for bilingual children. <i>Clinical Linguistics & Phonetics</i> , pp. 1-18.	Feil popula-sjon
Bochane M V, van der Schans C P, Krijnen W P, Reijneveld S A and Luinge M. (2020). Validation of the Early Language Scale. <i>European Journal Of Pediatrics</i> , 180(1), pp.63-71.	Feil verktøy
Boerma T, Chiat S, Leseman P, Timmermeister M, Wijnen F and Blom E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. <i>Journal of Speech, Language and and Hearing Research</i> , 58(6), pp.1747-1760.	Feil popula-sjon
Bogue E L, DeThorne L S and Schaefer B A. (2014). A Psychometric Analysis of Childhood Vocabulary Tests. <i>Contemporary Issues In Communication Science And Disorders</i> , 41(Spring), pp.55-69.	Feil studiede-sign/metode
Bouchard M G, Fitzpatrick E M and Olds J. (2009). Psychometric analysis of assessment tools used with francophone children. <i>Canadian Journal of Speech-Language Pathology & Audiology</i> , 33(3), pp.129-139.	Feil studiede-sign/metode
Boyer J L, Flamant C, Boussicault G, Berlie I, Gascoin G, Branger B and Rozé J-C. (2014). Characterizing early detection of language difficulties in children born pre-term. <i>Early Human Development</i> , 90(6), pp. 281-6.	Feil land
Brandone A C, Golinkoff R M and Hirsh-Pasek K. (2008). Feasibility of computer-administered language assessment. <i>Perspectives on School-Based Issues</i> , 9(2), pp.57-65.	Feil land
Bruce B, Kornfält R, Radeborg K, Hansson K and Nettelblatt U. (2003). Identifying children at risk for language impairment: screening of communication at 18 months. <i>Acta Paediatrica</i> , 92(9), pp.1090-1095.	Oppgir ikke måleegen-skaper i henhold til COSMIN
Bucik N and Bucik V. (2003). Peabody Picture Vocabulary Test (PPVT-III): Psychometric properties and significance for application. <i>Psiholoska Obzorja / Horizons of Psychology</i> , 12(4), pp.91-108.	Feil land
Camilleri B and Botting N. (2013). Beyond static assessment of children's receptive vocabulary: the dynamic assessment of word learning (DAWL). <i>International Journal of Language & Communication Disorders</i> , 48(5), pp.565-81.	Feil popula-sjon
Campbell J M, Bell S K and Keith L K. (2001). Concurrent validity of the Peabody Picture Vocabulary Test-Third Edition as an intelligence and achievement screener for low SES African American children. <i>Assessment</i> , 8(1), pp.85-94.	Feil studiede-sign/metode

Carscadden J, Corsiatto P, Ericson L, Illchuk R, Esopenko, Sterner E, Wells G D and Oddie S. (2010). A Pilot Study to Evaluate a New Early Screening Instrument for Speech and Language Delays Une étude pilote pour évaluer un nouvel instrument de dépistage des retards de la parole et du langage. <i>Canadian Journal of Speech-Language Pathology and Audiology</i> , 34(2), pp. 87-95	Feil verktøy
Castilla A P. (2010). Developmental measures of morphosyntactic acquisition in monolingual 3-, 4-, and 5-year-old Spanish-speaking children. <i>Dissertation Abstracts International: Section B: The Sciences and Engineering</i> , 71(4-B), pp.2362.	Feil land
Charles P, Belisle M, Tonita K and Smith J. (2015). Help Me Tell My Story: Development of an oral language measurement scale. <i>Journal of Applied Measurement</i> , 16(3), pp.278-297.	Feil verktøy
Chiat S and Polisenska K. (2016). A Framework for Crosslinguistic Nonword Repetition Tests: Effects of Bilingualism and Socioeconomic Status on Children's Performance. <i>Journal of Speech Language & Hearing Research</i> , 59(5), pp.1179-1189.	Oppgir ikke måleegenskaper i henhold til COSMIN
Christopoulos T T. (2022). A pilot implementation study of school-based universal screening for language impairment. <i>Dissertation Abstracts International: Section B: The Sciences and Engineering</i> , 83(6-B).	Oppgir ikke måleegenskaper i henhold til COSMIN
Clausen M C and Fox-Boyer A. (2022). Diagnostic validity, accuracy and inter-rater reliability of a phonological assessment for Danish-speaking children. <i>Journal Of Communication Disorders</i> , 95, pp.106168-106168.	Feil verktøy
Coyne-Martinson P A. (2000). Assessing the prereading language development of Kindergarten students: The reliability, validity, and sensitivity of Basic Language Assessment Story Tasks as dynamic indicators of basic early literacy skills. <i>Dissertation Abstracts International Section A: Humanities and Social Sciences</i> , 60(10-A), pp.3603.	Feil studiedesign/metode
Dahlén J, Drevenhorn E and Kalnak N. (2022). Assessment fidelity of a language screening instrument for 4-year-olds. <i>Logopedics Phoniatrics Vocology</i> , 48(4), pp.189-196,	Oppgir ikke måleegenskaper i henhold til COSMIN
Daub O, Cunningham B J, Bagatto M, Johnson A M and Smyth R. (2021). Adopting a Conceptual Validity Framework for Testing in Speech-Language Pathology. <i>American Journal Of Speech-Language Pathology</i> , 30(4), pp.1894-1908.	Feil studiedesign/metode
De Anda S, Cycyk L M, Moore H, Huerta L, Larson A L and King M. (2022). Psychometric Properties of the English-Spanish Vocabulary Inventory in Toddlers With and Without Early Language Delay. <i>Journal Of Speech Language And Hearing Research</i> , 65(2), pp. 672-691.	Feil land
de Koning H J, de Ridder-Sluiters J G, van Agt H M E, den Bergh C M M, van der Stege H A, Korfage I J, Polder J P and van der Maas P J. (2004). A cluster-randomised trial of screening for language disorders in toddlers. <i>Journal Of Medical Screening</i> , 11(3), pp.109-116.	Oppgir ikke måleegenskaper i henhold til COSMIN

Dempsey L F and Skarakis-Doyle E. (2001). The validity of the Joint Story Retell as a measure of young children's comprehension of familiar stories. <i>Journal of Speech-Language Pathology and Audiology</i> , 25(4), pp.201-211.	Feil verktøy
Denman D, Speyer R, Munro N, Pearce W M, Chen Y W and Cordier R. (2017). Psychometric Properties of Language Assessments for Children Aged 4–12 Years: A Systematic Review. <i>Frontiers In Psychology</i> , 8, 1515.	Feil studiede-sign/metode
Devescovi A and Caselli M C. (2007). Sentence repetition as a measure of early grammatical development in Italian. <i>International Journal Of Language & Communication Disorders</i> , 42(2), pp.187-208.	Feil land
Diepeveen F B, Dusseldorp E, Bol G, Oudesluys-Murphy A M and Verkerk P H. (2015). Failure to meet language milestones at two years of age is predictive of specific language impairment. <i>Acta Paediatrica</i> , 105(3), pp.304-310.	Feil studiede-sign/metode
Dockrell J. (2001). Assessing Language Skills in Preschool Children. <i>Child Psychology And Psychiatry Review</i> , 6(2), pp.74-85.	Feil studiede-sign/metode
Doove B M, Feron F J. M, van Os J and Drukker M. (2021). Preschool Communication: Early Identification of Concerns About Preschool Language Development and Social Participation. <i>Frontiers in public health</i> , 8(NA), pp.546536-546536.	Feil verktøy
Duggan C, Irvine A D, Hourihane J, Kiely M and Murray D M. (2023). ASQ-3 and BSID-III's concurrent validity and predictive ability of cognitive outcome at 5 years. <i>Pediatric Research</i> , 94, pp. 1465-1471.	Feil verktøy
Duncan L G, Gollek C and Potter D D. (2020). eLIPS: Development and Validation of an Observational Tool for Examining Early Language in Play Settings. <i>Frontiers In Psychology</i> , 11, 1813.	Feil verktøy
Eadie P, Ukoumunne O C, Skeat J, Prior M, Bavin E L, Bretherton L and Reilly S. (2010). Assessing early communication behaviours: structure and validity of the Communication and Symbolic Behaviour Scales—Developmental Profile (CSBS-DP) in 12-month-old infants. <i>International Journal Of Language & Communication Disorders</i> , 45(5), pp.572-585.	Feil verktøy
Edwards S, Letts C and Sinka I. (2011). The New Reynell Developmental Language Scales. <i>APA PsycTests Dataset</i> . doi.org/10.1037/t05236-000.	Feil studiede-sign/metode
Edwards S, Letts C and Sinka I. (2012). New Reynell Developmental Language Scales. <i>APA PsycTests Dataset</i> .	Feil studiede-sign/metode
Eriksson M. (2022). Insufficient evidence for the validity of the Language Development Survey and the MacArthur-Bates Communicative Development Inventories as screening tools: A critical review. <i>International journal of language & communication disorders</i> , 58(2), pp.555-575.	Feil verktøy
Everitt A, Hannaford P C and Conti-Ramsden G. (2013). Markers for persistent specific expressive language delay in 3-4-year-olds. <i>International Journal Of Language & Communication Disorders</i> , 48(5), pp.534-553.	Oppgir ikke måleegenskaper i henhold til COSMIN
Feltner C, Ina F W, Sallie N I, Colin J O, Brittany R, Middleton J C, Vaughan J and Kahwati C B L. (2023). <i>Screening for Speech and Language Delay and Disorders in Children Age 5 Years or Younger: An Evidence Review for the U.S. Preventive Ser-</i>	Feil studiede-sign/ metode

vices Task Force – Draft report. Available at: <https://www.uspreventiveservicestaskforce.org/uspstf/document/draft-evidence-review/speech-language-delay-disorders-children-age-5-years-younger-screening>

Floccia C, Sambrook T D, Delle L C, Kwok R, Goslin J, White L, Cattani A, Sullivan E, Abbot-Smith K, Krott A, Mills D, Rowland C, Gervain J and Plunkett K. (2018). Vocabulary of 2-year-olds learning English and an additional language: Norms and effects of linguistic distance: IV: Results for Studies 2 and 3: The UKBTAT model and its application to nontarget Additional Language learners. <i>Monographs of the Society for Research in Child Development</i> , 83(1), pp.61-67.	Feil verktøy
Fox C, Jones S, Gillam S L, Israelsen-Augenstein M, Schwartz S and Gillam R B. (2022). Automated Progress-Monitoring for Literate Language Use in Narrative Assessment (LLUNA). <i>Frontiers in Psychology</i> , 13, pp.10.	Feil populasjon
Friberg J. (2010). Considerations for test selection: How do validity and reliability impact diagnostic decisions?. <i>Child Language Teaching And Therapy</i> , 26(1), pp.77-92.	Feil verktøy
Frisk V, Montgomery L, Boychyn E, Young R, van Ryn E, McLachlan D and Neufeld J. (2009). Why Screening Canadian Preschoolers for Language Delays Is More Difficult than It Should Be. <i>Infants & Young Children</i> , 22(4), pp.290-308.	Oppgir ikke måleegenskaper i henhold til COSMIN
Fäldt A, Fabian H, Dahlberg A, Thunberg G, Durbeej N and Lucas. (2020). Infant-Toddler Checklist identifies 18-month-old children with communication difficulties in the Swedish child healthcare setting. <i>Acta Paediatrica</i> , 110(5), pp.1505-1512.	Feil verktøy
Gale R, Bird J, Wang Y, van Santen J, Prud'hommeaux E, Dolata J and Asgari M. (2021). Automated scoring of tablet-administered expressive language tests. <i>Frontiers in Psychology</i> , 12, 668401.	Feil populasjon
Gampe A, Kurthen I and Daum M M. (2018). BILEX: A new tool measuring bilingual children's lexicons and translational equivalents. <i>First Language</i> , 38(3), pp.263-283.	Feil verktøy
Geytenbeek J J, Mekkink L, Knol D, Vermeulen R and Oostrom K. (2012). Reliability and validity of the C-BiLLT: A new instrument to assess comprehension of spoken language in children with severe cerebral palsy. <i>Developmental Medicine and Child Neurology</i> , 6), pp.31.	Feil studiedesign/metode
Glad B and Kumlin K. (2007). Jämförelse mellan föräldrars, förskolepersonals och logopedstudenters bedömningar av små barns tal och språk. Linköpings Universitet, Hälsouniversitetet, Institutionen för Nervsystem och Rörelseorgan, Logopedprogrammet. LIU-INR/SLP-D--07/007—SE.	Feil studiedesign/metode
Glascoe F P. (2001). Are Overreferrals on Developmental Screening Tests Really a Problem?. <i>Archives Of Pediatrics & Adolescent Medicine</i> , 155(1), pp.54-54.	Feil studiedesign/metode
Greenslade K J, Plante E and Vance R. (2009). The Diagnostic Accuracy and Construct Validity of the Structured Photographic Expressive Language Test—Preschool: Second Edition. <i>Language Speech And Hearing Services In Schools</i> , 40(2), pp.150-160.	Feil land
Grether S. (2007). Detecting language problems: Accuracy of five language screening instruments in preschool children. <i>Developmental Medicine and Child Neurology</i> , 49(2), pp.84.	Feil studiedesign/metode

Guiberson M. (2020). Alternatives to traditional language sample measures with emergent bilingual preschoolers. <i>Topics in Language Disorders</i> , 40(2), pp.E1-E6.	Feil studiede- sign/metode
Gulbrandsen M A. (2021). <i>The Development and Pilot of a Norwegian Nonword Repetition Test Following the Framework of Language Impairment Testing in a Multilingual Setting</i> . Master Thesis. Bergen: University of Bergen, Faculty of Psychology.	Feil popula- sjon
Guo L Y, Schneider P and Harrison W. (2021). Clausal Density Between Ages 4 and 9 Years for the Edmonton Narrative Norms Instrument: Reference Data and Psychometric Properties. <i>Language and Speech & Hearing Services in the Schools</i> , 52(1), pp.354-368.	Feil popula- sjon
Guo L Y, Eisenberg S, Schneider P and Spencer L. (2019). Percent Grammatical Utterances Between 4 and 9 Years of Age for the Edmonton Narrative Norms Instrument: Reference Data and Psychometric Properties. <i>American Journal Of Speech-Language Pathology</i> , 28(4), pp.1448-1462.	Feil verktøy
Guo L Y, Eisenberg S, Schneider P and Spencer L. (2020). Finite Verb Morphology Composite Between Age 4 and Age 9 for the Edmonton Narrative Norms Instrument: Reference Data and Psychometric Properties. <i>Language Speech And Hearing Services In Schools</i> , 51(1), pp.128-143.	Feil verktøy
Guo L-Y and Schneider P. (2016). Differentiating school-aged children with and without language impairment using tense and grammaticality measures from a narrative task. <i>Journal of Speech, Language and and Hearing Research</i> , 59(2), pp.317-329.	Feil verktøy
Gutierrez-Clellen V F and Simon-Cereijido G. (2010). Using nonword repetition tasks for the identification of language impairment in Spanish-English-speaking children: Does the language of assessment matter?. <i>Learning Disabilities Research & Practice</i> , 25(1), pp.48-58.	Feil popula- sjon
Hadjadj O, Kehoe M and Delage H. (2022). Dynamic Assessment Identifies Morpho-syntactic Deficits in Mono- and Bilingual Children with Developmental Language Disorder. <i>Languages</i> , 7(4), pp.26.	Feil popula- sjon
Haitana T, Pitama S and Rucklidge J J. (2010). Cultural biases in the Peabody Picture Vocabulary Test III: Testing Tamariki in a New Zealand sample. <i>New Zealand Journal of Psychology</i> , 39(3), pp.24-34.	Feil popula- sjon
Haman E, Luniewska M and Hansen P. (2017). Noun and verb knowledge in monolingual preschool children across 17 languages: Data from Cross-linguistic Lexical Tasks (LITMUS-CLT). <i>Clinical Linguistics & Phonetics</i> , 31(11-12), pp.818-843.	Oppgir ikke måleegen- skaper i hen- hold til COSMIN
Hamilton A F, Plunkett K and Schafer G. (2000). Infant vocabulary development assessed with a British communicative development inventory. <i>Journal Of Child Language</i> , 27(3), pp.689-705.	Feil verktøy
Hannus S, Kauppila T, Pitkaniemi J and Launonen K. (2013). Use of language tests when identifying specific language impairment in primary health care. <i>Folia Phoniatica et Logopaedica:International Journal of Phoniatics and Speech Therapy and Communication Pathology</i> , 65(1), pp.40-46.	Feil popula- sjon
Hansen P, Luniewska M, Simonsen H G, Haman E, Mieszkowska K, Kolak J and Wodniecka Z. (2019). Picture-based vocabulary assessment versus parental questionnaires: A cross-linguistic study of bilingual assessment methods. <i>International Journal of Bilingualism</i> , 23(2), pp.437-456.	Oppgir ikke måleegen- skaper i hen- hold til COSMIN

Hansen P, Simonsen H G, Luniewska M and Haman E. (2017). Validating the psycholinguistic aspects of LITMUS-CLT: Evidence from Polish and Norwegian. <i>Clinical Linguistics & Phonetics</i> , 31(11-12), pp.910-930.	Oppgir ikke måleegenskaper i henhold til COSMIN
Hansson K and Nettelblatt U. (2002). Assessment of specific language impairment in Swedish. <i>Logopedics Phoniatrics Vocology</i> , 27(4), pp.146-154.	Feil populasjon
Hasson N, Dodd B and Botting N. (2012). Dynamic Assessment of Sentence Structure (DASS): Design and evaluation of a novel procedure for the assessment of syntax in children with language impairments. <i>International Journal of Language & Communication Disorders</i> , 47(3), pp.285-299.	Feil populasjon
Heilmann J J and Westerveld M E. (2013). Bilingual language sample analysis: Considerations and technological advances. <i>Journal of Clinical Practice in Speech-Language Pathology</i> , 15(2), pp.87-93.	Feil populasjon
Heilmann J, Tucci A, Plante E and Miller J F. (2020). Assessing Functional Language in School-Aged Children Using Language Sample Analysis. <i>Perspectives of the ASHA Special Interest Groups</i> , 5(3), pp.622-636.	Oppgir ikke måleegenskaper i henhold til COSMIN
Henderson D E, Restrepo M A and Aiken L S. (2018). Dynamic assessment of narratives among Navajo preschoolers. <i>Journal of Speech, Language and and Hearing Research</i> , 61(10), pp.2547-2560.	Feil land
Henderson D E. (2018). Dynamic assessment of narratives among Navajo Head Start children. <i>Dissertation Abstracts International: Section B: The Sciences and Engineering</i> , 78(9-BI).	Feil verktøy
Holzinger D, Weber C, Barbaresi W, Beitel C and Fellingner J. (2021). Language Screening in 3-Year-Olds: Development and Validation of a Feasible and Effective Instrument for Pediatric Primary Care. <i>Frontiers in Pediatrics</i> , 9, 752141.	Feil land
Holzinger D, Weber C and Diendorfer B. (2022). Development and Validation of a Language Screening for Implementation in Pre-School Settings. <i>Frontiers In Public Health</i> , 10, 866598.	Feil land
Holzinger D, Weber C and Fellingner J. (2022). Validity and feasibility of a predictive language screening tool in 2-year-old children in primary pediatric care. <i>Frontiers in pediatrics</i> , 10, 865457.	Feil land
Holzinger D, Weber C and Jezek M. (2022). Identifying Language Disorder Within a Migration Context: Development and Performance of a Pre-school Screening Tool for Children With German as a Second Language. <i>Frontiers in pediatrics</i> , 10, 814415.	Feil land
Hua Y. (2018). Test Review: Huer, M. B., & Miller, L. Tecel (2011), "Test of Early Communication and Emerging Language." Austin, TX: Pro-Ed. <i>Canadian Journal of School Psychology</i> , 33(2), pp.163-169.	Feil verktøy
Huang F L and Konold T R. (2014). A latent variable investigation of the Phonological Awareness Literacy Screening-Kindergarten assessment: Construct identification and multigroup comparisons between Spanish-speaking English-language learners (ELLs) and non-ELL students. <i>Language Testing</i> , 31(2), pp.205-221.	Feil land

Hviding K. (2005). Kartleggingsverktøy og instrumenter for tidlig avdekking av utviklings-, atferds- og psykososiale vansker hos barn 0-6 år. Oslo; Nasjonalt kunnskaps-senter for helsetjenesten.	Feil studiede-sign/metode
Irizarry-Perez D C, Pena E D and Bedore L M. (2021). Phonological predictors of nonword repetition performance in bilingual children. <i>Journal of Communication Disorders</i> , 94, pp.106156.	Oppgir ikke måleegen-skaper i henhold til COSMIN
Jackson E, Levine D, de Villiers, Iglesias A, Hirsh-Pasek K and Michnick G R. (2023). Assessing the language of 2 year-olds: From theory to practice. <i>Infancy : the official journal of the International Society on Infant Studies</i> , 28(5), pp.930-957.	Feil land
Jacobs E L. (2001). The effects of adding dynamic assessment components to a computerized preschool language screening test. <i>Communication Disorders Quarterly</i> , 22(4), pp.217-226.	Feil land
Janus M, Brinkman S A and Duku E K. (2011). Validity and psychometric properties of the Early Development Instrument in Canada, Australia, United States, and Jamaica. <i>Social Indicators Research</i> , 103(2), pp.283-297.	Feil verktøy
Jullien S. (2021). Screening for language and speech delay in children under five years. <i>BMC Pediatrics</i> , 21(1), pp.362.	Feil studiede-sign/ metode
Justice L M, Bowles R, Pence K and Gosse C. (2010). A scalable tool for assessing children's language abilities within a narrative context: The NAP (narrative assessment protocol). <i>Early Childhood Research Quarterly</i> , 25(2), pp.218-234.	Feil studiede-sign/metode
Swapna P K, Anand V and Babu S. (2017). Assessment of speech and language delay among 0-3 years old children using language evaluation scale. Trivandrum. <i>Journal Of Evolution Of Medical And Dental Sciences</i> .	Feil studiede-sign/metode
Kirk C and Vigeland L M. (2014). A Psychometric Review of Norm-Referenced Tests Used to Assess Phonological Error Patterns. <i>Language Speech And Hearing Services In Schools</i> , 45(4), pp.365-377.	Feil verktøy
Klee T, Pearce K F and Carson D K. (2000). Improving the positive predictive value of screening for developmental language disorder. <i>Journal of speech, language and and hearing research : JSLHR</i> , 43(4), pp.821-833.	Feil verktøy
Klem M J, Melby-Lervåg M and Hagtvet B E, Lyster S-A H, Gustafsson J-E and Hulme C. (2014). Sentence repetition is a measure of children's language skills rather than working memory limitations. <i>Developmental Science</i> , 18(1), pp.146-154.	Oppgir ikke måleegen-skaper i henhold til COSMIN
Knuijt S, Sondaar M and De Kleine M J K and Kollee L A. (2004). Validation of a Dutch language screening instrument for 5-year-old preterm infants. <i>Acta Paediatrica and International Journal of Paediatrics</i> , 93(10), pp.1372-1377.	Feil verktøy

Larson A L. (2016). Language Screening for Infants and Toddlers: A Literature Review of Four Commercially Available Tools. <i>Communication Disorders Quarterly</i> , 38(1), pp.3-12.	Feil studiede- sign/metode
Law J, Boyle J, Harris F, Harkness A and Nye C. (2007). The feasibility of universal screening for primary speech and language delay: findings from a systematic review of the literature. <i>Developmental Medicine & Child Neurology</i> , 42(3), pp.190-200.	Feil studiede- sign/metode
Lazewnik R, Creaghead N A, Smith A B, Prendeville J-A, Raisor-Becker L and Silberta N. (2019). Identifiers of language impairment for Spanish-English dual language learners. <i>Language, Speech and and Hearing Services in Schools</i> , 50(1), pp.126-137.	Feil verktøy
Lee N R, Chung S H, Song M K, Kong Y H, Joo C U and Kim S J. (2020). A Comparative Analysis of Clinical Screening Test and Language Specific Test in Language Delay Children. <i>Chonnam medical journal</i> , 56(1), pp.44-49.	Feil popula- sjon
Li'el N and Williams C K R. (2018). Identifying developmental language disorder in bilingual children from diverse linguistic backgrounds. <i>International Journal Of Speech-Language Pathology</i> , 21(6), pp.613-622.	Feil popula- sjon
Loakes D, Moses K, Simpson J and Wigglesworth G. (2012). Developing tests for the assessment of traditional language skill: A case study in an Indigenous Australian community. <i>Language Assessment Quarterly</i> , 9(4), pp.311-330.	Feil popula- sjon
Lohman D F, Korb K A and Lakin J M. (2008). Identifying academically gifted English-language learners using nonverbal tests: A comparison of the Raven, NNAT, and CogAT. <i>Gifted Child Quarterly</i> , 52(4), pp.275-296.	Feil land
Longerbeam M S and Freeman L M. (2020). Language Screening: How Far Have We Come?. <i>Current Developmental Disorders Reports</i> , 7(3), pp.116-123.	Feil studiede- sign/metode
Lous J and Lauritsen M-B G. (2018). Comparison of the Reynell Developmental Language Scale II and the Galker test of word-recognition-in-noise in Danish day-care children. <i>International Journal of Pediatric Otorhinolaryngology</i> , 109, pp.104-111.	Oppgir ikke måleegen- skaper i hen- hold til COSMIN
Lousada M, Valente A R and Mendes A. (2016). Validation of a Paediatric Speech and Language Screening (RALF). <i>Folia Phoniatrica Et Logopaedica</i> , 68(6), pp. 247-251.	Feil land
Luinge M. (2007). The language screening instrument SNEL. <i>Educational And Child Psychology</i> , 24(4), pp.20-30.	Feil verktøy
Lung F W, Shu B C, Chiang T L, Chen P F and Lin L L. (2009). Predictive validity of Bayley scale in language development of children at 6-36 months. <i>Pediatrics International</i> , 51(5), pp.666-9.	Feil land
Lüke C, Grimminger A, Rohlfing K J, Liszkowski U and Ritterfeld U. (2016). In Infants' Hands: Identification of Preverbal Infants at Risk for Primary Language Delay. <i>Child Development</i> , 88(2), pp.484-492.	Feil land
MacLeod A A N, Sutton A, Sylvestre A, Thordardottir E and Trudeau N. (2014). Screening tool for speech sound development disorders: Theoretical bases and preliminary data. <i>Canadian Journal of Speech-Language Pathology & Audiology</i> , 38(1), pp.40-56.	Feil studiede- sign/metode

Marcotte A M, Clemens N H, Parker C P and Whitcomb S A. (2016). Examining the Classification Accuracy of a Vocabulary Screening Measure With Preschool Children. <i>Assessment For Effective Intervention</i> , 41(4), pp.230-242.	Feil land
Marcotte A M, Parker C, Furey W and Hands J L. (2014). An Examination of the Validity of the Dynamic Indicators of Vocabulary Skills (DIVS). <i>Journal of Psychoeducational Assessment</i> , 32(2), pp.133-145.	Feil land
McGinty C. (2000). An investigation into aspects of the Mayo early language screening test. <i>Child Care Health And Development</i> , 26(2), pp.111-128.	Feil land
McKean C, Law J, Mensah F, Cini E, Eadie, Frazer K and Reilly S. (2016). Predicting Meaningful Differences in School-Entry Language Skills from Child and Family Factors Measured at 12 months of Age. <i>International Journal Of Early Childhood</i> , 48(3), pp.329-351.	Feil verktøy
McLeod S and Verdon S. (2014). A Review of 30 Speech Assessments in 19 Languages Other Than English. <i>American Journal Of Speech-Language Pathology</i> , 23(4), pp.708-723.	Feil studiede-sign/metode
Metz D, Belhadj K, Karpinski E N and Petermann F. (2011). [The validity of the SET 5-10 (language level test for children aged between five and ten years): first analyses]. <i>Gesundheitswesen</i> , 73(10), pp.637-43.	Feil studiede-sign/metode
Miller-Whitehead M. (2001). <i>A Review of the Literature on Early Childhood Cognitive Academic Language Assessments Suitable for Use in Ethnically Diverse School Systems</i> . Available at: https://eric.ed.gov/?id=ED459473 .	Feil studiede-sign/metode
Miniscalco C, Westerlund M and Lohmander A. (2007). Language skills at age 6 years in Swedish children screened for language delay at 2(1/2) years of age. <i>Acta paediatrica</i> , 94(12), pp.1798-1806.	Feil popula-sjon
Morgan L, Delehanty A, Dillon J C, Schatschneider C and Wetherby A M. (2020). Measures of Early Social Communication and Vocabulary Production to Predict Language Outcomes at Two and Three Years in Late-Talking Toddlers. <i>Early childhood research quarterly</i> , 51, pp.366-378.	Oppgir ikke måleegen-skaper i henhold til COSMIN
Nair M K C, Harikumar G S, Mini A O, Indulekha S, Letha S and Russell P S. (2013). Development and validation of Language Evaluation Scale Trivandrum for children aged 0–3 years — LEST (0–3). <i>Indian Pediatrics</i> , 50(5), pp.463-467.	Feil land
Nasution L K W and Trevino P. (2012). The validity and reliability of Modified Peabody Picture Vocabulary Test IV (PPVT IV) in 48-59-month-old children. <i>Indonesian Journal Of Physical Medicine and Rehabilitation</i> , pp.54-54.	Feil land
Nayeb L, Sarkadi A, Eriksson M, Salameh E K and Lagerberg D. (2020). Identifying language disorder in bilingual children is possible only if both languages are assessed. <i>European Journal of Public Health</i> , 30(Supplement_5), pp. 1316.	Feil studiede-sign/metode
Nelson H D, Nygren P, Walker M and Panoscha R. (2006). Screening for Speech and Language Delay in Preschool Children . Rockville (MD): Agency for Healthcare Research and Quality (US).	Feil studiede-sign/ metode
Nelson H D, Nygren P, Walker M and Panoscha R. (2006). Screening for speech and language delay in preschool children: Systematic evidence review for the US preventive services task force. <i>Pediatrics</i> , 117(2), pp.e298-e319.	Feil studiede-sign/metode

Nielsen M S. (2010). Vurdering af danske børns sprog – En undersøgelse af eksisterende værktøjer og udvikling af nyt værktøj til opfølgning på sprogvurdering af tre-årige. Ph.d.-afhandling. Syddansk Universitet, Det Humanistiske Fakultet.	Feil studiede-sign/metode
Orellana C I, Wada R and Gillam R B. (2019). The use of dynamic assessment for the diagnosis of language disorders in bilingual children: A meta-analysis. <i>American Journal of Speech-Language Pathology</i> , 28(3), pp.1298-1317.	Feil verktøy
Ortiz J A. (2021). Using Nonword Repetition to Identify Language Impairment in Bilingual Children: A Meta-Analysis of Diagnostic Accuracy. <i>American Journal of Speech-Language Pathology</i> , 30(5), pp.2275-2295.	Feil verktøy
Pankratz M, Plante E, Vance R and Insalaco D. (2007). The Diagnostic and Predictive Validity of The Renfrew Bus Story. <i>Language Speech And Hearing Services In Schools</i> , 38(4), pp.390-399.	Feil verktøy
Patterson J L, Rodriguez B L and Dale P S. (2020). Dynamic assessment language tasks and the prediction of performance on year-end language skills in preschool dual language learners. <i>American Journal of Speech-Language Pathology</i> , 29(3), pp.1226-1240.	Feil land
Pavelko S L. (2019). Diagnostic Accuracy of the Sampling Utterances and Grammatical Analysis Revised (SUGAR) Measures for Identifying Children With Language Impairment. <i>Language, Speech & Hearing Services in Schools</i> . Apr 2019; 50(2): 211-223. <i>Language and Speech & Hearing Services in Schools</i> , 50(3), pp.458-458.	Feil land
Peña E D, Spaulding T J and Plante E. (2006). The Composition of Normative Groups and Diagnostic Decision Making: Shooting Ourselves in the Foot. <i>American Journal Of Speech-Language Pathology</i> , 15(3), pp.247-254.	Feil verktøy
Perona K, Plante E and Vance R. (2005). Diagnostic Accuracy of the Structured Photographic Expressive Language Test: Third Edition (SPELT-3). <i>Language Speech And Hearing Services In Schools</i> , 36(2), pp.103-115.	Feil land
Petersen D B, Tonn P, Spencer T D and Foster M E. (2020). The classification accuracy of a dynamic assessment of inferential word learning for bilingual English/Spanish-speaking school-age children. <i>Language, Speech and and Hearing Services in Schools</i> , 51(1), pp.144-164.	Feil land
Pickstone C, Hannon P and Fox L. (2002). Surveying and screening preschool language development in community-focused intervention programmes: a review of instruments. <i>Child: care and health and development</i> , 28(3), pp.251-264.	Feil studiede-sign/metode
Pinheiro Ladc, Silva A P D and Hage S R V. (2020). Morphosyntactic Evaluation Protocol (MEP): validation of content. <i>Codas</i> , 32(6), pp.e20190148.	Feil land
Potratz J R, Gildersleeve-Neumann C and Redford M A. (2022). Measurement Properties of Mean Length of Utterance in School-Age Children. <i>Language and Speech & Hearing Services in the Schools</i> , 53(4), pp.1088-1100.	Feil studiede-sign/metode
Rainelli S, Bulotsky-Shearer R J, Fernandez V A, Greenfield D B and L'opez M. (2017). Validity of the first two subtests of the preschool language assessment scale as a language screener for Spanish-speaking preschool children. <i>Early Childhood Research Quarterly</i> , 38, pp.10-22.	Feil land
Rasheed M A, Kvestad I, Shaheen F, Memon U and Strand T A. (2023). The predictive validity of Bayley Scales of Infant and Toddler Development-III at 2 years for later general abilities: Findings from a rural, disadvantaged cohort in Pakistan. <i>PLOS Global Public Health</i> , 3(1), pp.e0001485-e0001485.	Feil land

Redmond S, Ash A, Christopoulos T T and Pfaff T. (2019). Diagnostic Accuracy of Sentence Recall and Past Tense Measures for Identifying Children's Language Impairments. <i>Journal Of Speech Language And Hearing Research</i> , 62(7), pp.2438-2454.	Feil verktøy
Reuterskiold-Wagner C, Sahlen B and Nyman A. (2005). Non-word repetition and non-word discrimination in Swedish preschool children. <i>Clinical Linguistics and Phonetics</i> , 19(8), pp.681-699.	Feil populasjon
Richter J and Valla L. (2013). Måleegenskaper ved den norske versjonen av Bayley Scales of Infant and Toddler Development (BSID, BSID-II, BSID-III). <i>PsykTestBarn</i> .	Feil studiedesign/metode
Risling J-K, Melzer J and Petermann F. (2016). Construction of a language level test for children between 3 and 5 years of age. <i>Zeitschrift fur Entwicklungspsychologie und Padagogische Psychologie</i> , 48(2), pp.69-79.	Feil land
Roman A S, Pisoni D B and Kronenberger W G. (2014). Assessment of Working Memory Capacity in Preschool Children Using the Missing Scan Task. <i>Infant and Child Development</i> , 23(6), pp.575-587.	Feil land
Roux S, Kammermeyer G and Stuck A. (2011). Procedure for assessing the demand for language fostering in the year before school 106orwegian. <i>Psychologie in Erziehung und Unterricht</i> , 58(1), pp.1-14.	Feil land
Roy B and Alla F. (2002). Two French-language tests for identifying language problems in children aged 4 yrs and aged 6 yrs. <i>A N A E Approche Neuropsychologique des Apprentissages chez l'Enfant</i> , 14(1; 66), pp.58-62.	Feil land
Ryan J J, Glass L A, Sullivan D K, Gibson C and Bartels J. (2009). PPVT-III alternate forms reliability and stability among inner-city primary school students. <i>Individual Differences Research</i> , 7(2), pp.70-75.	Feil populasjon
Ryu S H and Sim Y J. (2019). The validity and reliability of DDST II and bayley iii in children with language development delay. <i>Neurology Asia</i> , 24(4), pp.355-361.	Feil populasjon
Sachse S and von Suchodoletz W. (2008). Early identification of language assessment or parent report?. <i>Journal of Developmental and Behavioral Pediatrics</i> , 29(1), pp.34-41.	Feil studiedesign/metode
Sachse S and von Suchodoletz W. (2008). Early Identification of Language Delay by Direct Language Assessment or Parent Report?. <i>Journal Of Developmental And Behavioral Pediatrics</i> , 29(1), pp.34-41.	Feil studiedesign/metode
Samuelsson C, Scocco C and Nettelblatt U. (2003). Towards assessment of prosodic abilities in Swedish children with language impairment. <i>Logopedics Phoniatrics Vocology</i> , 28(4), pp.156-166.	Feil populasjon
Sanchez J D. (2011). Assessment of English language development: A validity study of a district initiative. <i>Journal of Applied Measurement</i> , 12(2), pp.106-123.	Feil land
Sansavini A, Favilla M E, Guasti M T, Marini A, Millepiedi S, Di Martino M V, Vecchi S, Battajon N, Bertolo L, Capirci O, Carretti B, Colatei M P, Frioni C, Marotta L, Massa S, Michelazzo L, Pecini C, Piazzalunga S, Pieretti M, Rinaldi P, Salvadorini R, Termine C, Zuccarini M, D'Amico S, De Cagno A G, Levorato M C, Rossetto T and Lorusso M L.(2021). Developmental Language Disorder: Early Predictors, Age for the Diagnosis, and Diagnostic Tools. A Scoping Review. <i>Brain sciences</i> , 11(5), pp.654.	Feil studiedesign/metode

Scheffler F, Vogel D, Astern R, Burgess J, Conneally T and Salerno K. (2007). Screening for communication and cognitive disorders in infants and toddlers. <i>Pediatric Nursing</i> , 33(6), pp.473-80.	Feil verktøy
Schneider P, Hayward D and Dube R V. (2006). Storytelling from pictures using the Edmonton Narrative Norms Instrument. <i>Journal of Speech-Language Pathology and Audiology</i> , 30(4), pp.224-238.	Feil popula-sjon
Seager E and Abbot-Smith K. (2016). Can early years professionals determine which preschoolers have comprehension delays? A comparison of two screening tools. <i>Child Language Teaching And Therapy</i> , 33(1), pp. 67-79.	Feil verktøy
Sharp K M and Gathercole V C. (2013). Can a novel word repetition task be a language-neutral assessment tool? Evidence from Welsh-English bilingual children. <i>Child Language Teaching & Therapy</i> , 29(1), pp.77-89.	Oppgir ikke måleegenskaper i henhold til COSMIN
Sim F, Haig C, O'Dowd J J, Thompson L, Law J, McConnachie A, Gillberg C and Wilson P. (2015). Development of a triage tool for neurodevelopmental risk in children aged 30 months. <i>Research In Developmental Disabilities</i> , 45-46, pp.69-82.	Feil verktøy
Sim F, Thompson L, Marryat L, Ramparsad N and Wilson Philip. (2019). Predictive validity of preschool screening tools for language and behavioural difficulties: A PRISMA systematic review. <i>PLOS ONE</i> , 14(2), pp.e0211409.	Feil studiede-sign/metode
Simos Panagiotis G, Sideridis G D, Protopapas A and Mouzaki Angeliki. (2011). Psychometric evaluation of a receptive vocabulary test for Greek elementary students. <i>Assessment for Effective Intervention</i> , 37(1), pp.34-49.	Feil popula-sjon
Siu A L. (2015). Screening for Speech and Language Delay and Disorders in Children Aged 5 Years or Younger: US Preventive Services Task Force Recommendation Statement. <i>Pediatrics</i> , 136(2), pp.e474-e481.	Feil studiede-sign/metode
Sivertsen B. (2022). <i>The New Reynell Developmental Language Scales. Er den nye Reynell språktest et egnet verktøy for kartlegging av flerspråklige barn?</i> Masteroppgave i Logopedi, NTNU.	Feil studiede-sign/ metode
Skarakis-Doyle E, Dempsey L and Lee C. (2008). Identifying language comprehension impairment in preschool children. <i>Language, Speech and and Hearing Services in Schools</i> , 39(1), pp.54-65.	Feil verktøy
Skarakis-Doyle E, Miller L T and Reichheld M. (2000). Construct validity as a foundation of evidence-based practice: The case of the Preschool Language Assessment Instrument. <i>Journal of Speech-Language Pathology and Audiology</i> , 24(4), pp.180-191.	Feil verktøy
Slott M, Vach W and Bleses D. (2008). Evaluation of methods used to assess language development of 3-4-year-old Danish children. <i>Logopedics Phoniatrics Vocology</i> , 33(4), pp.190-207.	Feil studiede-sign/metode
So K K H and To C K S. (2022). Systematic Review and Meta-Analysis of Screening Tools for Language Disorder. <i>Frontiers in pediatrics</i> , 10, pp.801220.	Feil studiede-sign/metode
Solna G and Cervenkova B. (2022). Validation study and introduction of the new TEPO sentence comprehension test for children aged 3-8 years. [Czech]. <i>Ceska a Slovenska Neurologie a Neurochirurgie</i> , 85(6), pp.477-483.	Feil land
Sonnander K. (2007). Early identification of children with developmental disabilities. <i>Acta Paediatrica</i> , 89, pp.17-23.	Feil studiede-sign/ metode

Spaulding T J, Plante E and Farinella K A. (2006). Eligibility Criteria for Language Impairment. <i>Language Speech And Hearing Services In Schools</i> , 37(1), pp.61-72.	Feil studiede- sign/metode
Spencer T D, Goldstein H, Kelley E S, Sherman A and McCune L. (2017). A curriculum-based measure of language comprehension for preschoolers: Reliability and validity of the Assessment of Story Comprehension. <i>Assessment for Effective Intervention</i> , 42(4), pp.209-223.	Feil verktøy
Spencer Trina D, Thompson M S, Petersen D B, Liu Y and Restrepo M A. (2023). <i>Reliability and Validity Evidence for the English and Spanish Preschool Narrative Language Measures-Listening</i> . : Grantee Submission, pp.. Available at: https://search.eric.ed.gov/search/doc?direct=true&db=eric&AN=ED629218&site=ehost-live .	Feil verktøy
Spencer-Smith M M, Spittle A J, Lee K J, Doyle L W and Anderson P J. (2015). Bayley-III cognitive and language scales in preterm children. <i>Pediatrics</i> , 135(5), pp.e1258-e1265.	Feil verktøy
Stolk Y, Kaplan I and Szwarc J. (2017). Review of the strengths and difficulties questionnaire translated into languages spoken by children and adolescents of refugee background. <i>International Journal of Methods in Psychiatric Research</i> , 26(4), pp.21.	Feil verktøy
Stolt S, Matomäki J, Lind A, Lapinleimu H, Haataja L and Lehtonen L. (2014). The prevalence and predictive value of weak language skills in children with very low birth weight – a longitudinal study. <i>Acta Paediatrica</i> , 103(6), pp.651-658.	Oppgir ikke måleegenskaper i henhold til COSMIN
Stott C, Merricks M J, Bolton P and Goodyer Ian. (2002). Screening for Speech and Language Disorders: the reliability, validity and accuracy of the General Language Screen. <i>International Journal Of Language & Communication Disorders</i> , 37(2), pp.133-151.	Feil verktøy
Thordardottir Elin, Kehayia E, Mazer B, Lessard N, Majnemer A, Sutton A, Trudeau N and Chilingaryan G. (2011). Sensitivity and specificity of French language and processing measures for the identification of primary language impairment at age 5. <i>Journal of Speech, Language and and Hearing Research</i> , 54(2), pp.580-597.	Feil land
Tresoldi M, Ambrogi F, Favero E, Colombo A, Barillari M R, Velardi P and Schindler A. (2015). Reliability, validity and normative data of a quick repetition test for Italian children. <i>International Journal Of Pediatric Otorhinolaryngology</i> , 79(6), pp.888-894.	Feil verktøy
Troost J, Brunner M and Proschel U. (2004). Validity of the Heidelberger Vorschulscreening (HVS). <i>Diagnostica</i> , 50(4), pp.193-201.	Feil land
Ukrainetz T A and Blomquist C. (2002). The criterion validity of four vocabulary tests compared to a language sample. <i>Child Language Teaching and Therapy</i> , 18(1), pp.59-78.	Feil studiede- sign/metode
Vach W, Bleses D and Jørgensen R N. (2010). Construction of a Danish CDI short form for language screening at the age of 36 months: Methodological considerations and results. <i>Clinical Linguistics & Phonetics</i> , 24(8), pp.602-621.	Feil verktøy
Valtonen R, Ahonen T, Lyytinen P and Tolvanen A. (2007). Screening for developmental risks at 4 years of age: predicting development two years later. <i>Nordic Psychology</i> , 59(2), pp.95-108.	Feil studiede- sign/metode
van Agt H M E, van der Stege H A, de Ridder-Sluis J G and de Koning H J. (2007). Detecting language problems: accuracy of five language screening instruments in	Feil verktøy

preschool children. <i>Developmental Medicine & Child Neurology</i> , 49(2), pp.117-122.	
Vehkavuori S-M and Stolt S. (2018). Screening language skills at 2;0. <i>Infant behavior & development</i> , 50(NA), pp.174-179.	Feil verktøy
Veldhuizen S, Clinton J, Rodriguez C, Wade T J and Cairney J. (2015). Concurrent Validity of the Ages and Stages Questionnaires and Bayley Developmental Scales in a General Population Sample. <i>Academic Pediatrics</i> , 15(2), pp.231-237.	Feil verktøy
Wackerle-Hollman A K, Rodriguez M I, Bradfield T A, Rodriguez M C and McConnell S R. (2015). Development of Early Measures of Comprehension: Innovation in Individual Growth and Development Indicators. <i>Grantee Submission</i> , 40(2), pp.81-95.	Feil land
Wallace I, Berkman N D, Watson L R, Coyne-Beasley T, Wood C T, Cullen K and Lohr K N. (2015). Screening for Speech and Language Delay in Children 5 Years Old and Younger: A Systematic Review. <i>Pediatrics</i> , 136(2), pp.e448-62.	Feil studiede-sign/metode
Warren R, Kenny M, Bennett T, Fitzpatrick-Lewis D, Ali M U, Sherifali D and Raina P. (2016). Screening for developmental delay among children aged 1-4 years: a systematic review. <i>CMAJ Open</i> , 4(1), pp.E20-E27.	Feil verktøy
Washington J A and Craig H K. (2004). A Language Screening Protocol for Use With Young African American Children in Urban Settings. <i>American Journal Of Speech-Language Pathology</i> , 13(4), pp.329-340.	Feil studiede-sign/metode
Webster R I, Majnemer A, Platt R W and Shevell M I. (2004). The predictive value of a preschool diagnosis of developmental language impairment. <i>Neurology</i> , 63(12), pp.2327-2331.	Feil popula-sjon
Weiler B, Schuele C M, Feldman J I and Krimm H. (2018). <i>A Multiyear Population-Based Study of Kindergarten Language Screening Failure Rates Using the Rice Wexler Test of Early Grammatical Impairment</i> . : Grantee Submission. Available at: https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED622627&site=ehost-live .	Feil land
Westergren T, Mølland E, Haraldstad K, Håland Å T, Köpp U M S, Fegran L and Abildsnes E. (2021). Implementation of the 109orwegian 'Starting right' child health service innovation: implementation adjustments, adoption, and acceptability. <i>BMC Health Services Research</i> , 21(1), pp..	Oppgir ikke måleegen-skaper i henhold til COSMIN
Westerlund M and Sundelin C. (2000). Screening for developmental language disability in 3-year-old children. Experiences from a field study in a Swedish municipality. <i>Child: care and health and development</i> , 26(2), pp.91-110.	Oppgir ikke måleegen-skaper i henhold til COSMIN
Westerveld M F and Vidler K. (2015). The use of the Renfrew Bus Story with 5-8-year-old Australian children. <i>International Journal of Speechlanguage Pathology</i> , 17(3), pp.304-13.	Feil popula-sjon
Wetherby A M, Allen L D, Cleary J, Kublin K and Goldstein Howard. (2002). Validity and Reliability of the Communication and Symbolic Behavior Scales Developmental Profile With Very Young Children. <i>Journal Of Speech Language And Hearing Research</i> , 45(6), pp.1202-1218.	Oppgir ikke måleegen-skaper i henhold til COSMIN

Wilbourn M P, Kurtz L E and Kalia V. (2012). The Lexical Stroop Sort (LSS) picture-word task: A computerized task for assessing the relationship between language and executive functioning in school-aged children. <i>Behavior Research Methods</i> , 44(1), pp.270-286.	Feil land
Wilder A and Redmond S. (2022). The Reliability of Short Conversational Language Sample Measures in Children With and Without Developmental Language Disorder. <i>Journal Of Speech Language And Hearing Research</i> , 65(5), pp.1939-1955.	Feil land
Wilks T, Gerber R J and Erdie-Lalena C. (2010). Developmental Milestones: Cognitive Development. <i>Pediatrics In Review</i> , 31(9), pp.364-367.	Feil verktøy
Willinger Ulrike, Schmoeger M, Deckert M, ; Eisenwort B, Loader B, Hofmair A and Auff E. (2017). Screening for specific language impairment in preschool children: Evaluating a screening procedure including the Token Test. <i>Journal of Psycholinguistic Research</i> , 46(5), pp.1237-1247.	Feil verktøy
Wilson P, Rush R, Charlton J, Gilroy V, McKean C and Law J. (2022). Universal language development screening: comparative performance of two questionnaires. <i>BMJ Paediatrics Open</i> , 6(1), pp.e001324-e001324.	Feil verktøy
Wilson P, Wood R, Lykke K, Graungaard A H, Ertmann R K, Andersen M K K, Haavet O R, Lagerløv P, Abildsnes E, Dahli M P; Mäkelä M, Varinen A and Hietanen M. (2018). International variation in programmes for assessment of children's neurodevelopment in the community: Understanding disparate approaches to evaluation of motor, social, emotional, behavioural and cognitive function. <i>Scandinavian Journal Of Public Health</i> , 46(8), pp.805-816.	Feil studiede-sign/metode
Wilson S B and Lonigan C J. (2009). An evaluation of two emergent literacy screening tools for preschool children. <i>Annals Of Dyslexia</i> , 59(2), pp.115-131.	Feil verktøy
Wright K R, Washington K N, Crowe K, Jenkins A, Leon M, Kokotek L, Raisor-Becker L and Westby C. (2019). Current Methods of Evaluating the Language Abilities of Multilingual Preschoolers: A Scoping Review Using the International Classification of Functioning, Disability and Health-Children and Youth Version. <i>Language, Speech and and Hearing Services in Schools</i> , 50(3), pp.434-451.	Feil studiede-sign/metode
Xue Y, Bandel E, Vogel C and Boller K. (2022). Psychometric properties of parent- and staff-reported measures and observational measures of infant and toddler development in Early Head Start. <i>Early Childhood Research Quarterly</i> , 61, pp.132-144.	Feil verktøy
Yoon J A, An S W, Choi Y S, Seo J S, Yoon S J, Yeon K S and Shin Yong B (2022). Correlation of Language Assessment Batteries of Toddlers With Developmental Language Delay. <i>Annals Of Rehabilitation Medicine</i> , 46(5), pp.256-262.	Feil land
Zimmerman I L and Castilleja N F. (2005). The role of a language scale for infant and preschool assessment. <i>Mental Retardation and Developmental Disabilities Research Reviews</i> , 11(3), pp.238-246.	Feil verktøy

Vedlegg 5: PRISMA sjekklisten

Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
TITLE			
Title	1	Identify the report as a scoping review.	1-2
ABSTRACT			
Structured summary	2	Provide a structured summary that includes (as applicable): background, objectives, eligibility criteria, sources of evidence, charting methods, results, and conclusions that relate to the review questions and objectives.	5-11 (Norwegian and English)
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known. Explain why the review questions/objectives lend themselves to a scoping review approach.	17
Objectives	4	Provide an explicit statement of the questions and objectives being addressed with reference to their key elements (e.g., population or participants, concepts, and context) or other relevant key elements used to conceptualize the review questions and/or objectives.	19
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists; state if and where it can be accessed (e.g., a Web address); and if available, provide registration information, including the registration number.	19
Eligibility criteria	6	Specify characteristics of the sources of evidence used as eligibility criteria (e.g., years considered, language, and publication status), and provide a rationale.	21
Information sources*	7	Describe all information sources in the search (e.g., databases with dates of coverage and contact with authors to identify additional sources), as well as the date the most recent search was executed.	22-23

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
Search	8	Present the full electronic search strategy for at least 1 database, including any limits used, such that it could be repeated.	79-93
Selection of sources of evidence†	9	State the process for selecting sources of evidence (i.e., screening and eligibility) included in the scoping review.	23
Data charting process‡	10	Describe the methods of charting data from the included sources of evidence (e.g., calibrated forms or forms that have been tested by the team before their use, and whether data charting was done independently or in duplicate) and any processes for obtaining and confirming data from investigators.	23-24
Data items	11	List and define all variables for which data were sought and any assumptions and simplifications made.	24-25
Critical appraisal of individual sources of evidence§	12	If done, provide a rationale for conducting a critical appraisal of included sources of evidence; describe the methods used and how this information was used in any data synthesis (if appropriate).	N/A
Synthesis of results	13	Describe the methods of handling and summarizing the data that were charted.	25
RESULTS			
Selection of sources of evidence	14	Give numbers of sources of evidence screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally using a flow diagram.	26-27
Characteristics of sources of evidence	15	For each source of evidence, present characteristics for which data were charted and provide the citations.	27,31,45
Critical appraisal within sources of evidence	16	If done, present data on critical appraisal of included sources of evidence (see item 12).	N/A
Results of individual sources of evidence	17	For each included source of evidence, present the relevant data that were charted that relate to the review questions and objectives.	27-56
Synthesis of results	18	Summarize and/or present the charting results as they relate to the review questions and objectives.	27-56
DISCUSSION			
Summary of evidence	19	Summarize the main results (including an overview of concepts, themes, and types of evidence	57

SECTION	ITEM	PRISMA-ScR CHECKLIST ITEM	REPORTED ON PAGE #
		available), link to the review questions and objectives, and consider the relevance to key groups.	
Limitations	20	Discuss the limitations of the scoping review process.	58-59
Conclusions	21	Provide a general interpretation of the results with respect to the review questions and objectives, as well as potential implications and/or next steps.	62
FUNDING			
Funding	22	Describe sources of funding for the included sources of evidence, as well as sources of funding for the scoping review. Describe the role of the funders of the scoping review.	12

JBI = Joanna Briggs Institute; PRISMA-ScR = Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews.

* Where *sources of evidence* (see second footnote) are compiled from, such as bibliographic databases, social media platforms, and Web sites.

† A more inclusive/heterogeneous term used to account for the different types of evidence or data sources (e.g., quantitative and/or qualitative research, expert opinion, and policy documents) that may be eligible in a scoping review as opposed to only studies. This is not to be confused with *information sources* (see first footnote).

‡ The frameworks by Arksey and O'Malley (6) and Levac and colleagues (7) and the JBI guidance (4, 5) refer to the process of data extraction in a scoping review as data charting.

§ The process of systematically examining research evidence to assess its validity, results, and relevance before using it to inform a decision. This term is used for items 12 and 19 instead of "risk of bias" (which is more applicable to systematic reviews of interventions) to include and acknowledge the various sources of evidence that may be used in a scoping review (e.g., quantitative and/or qualitative research, expert opinion, and policy document).

From: Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann Intern Med.* 2018;169:467–473. doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850).

Vedlegg 6: Identifiserte kartleggingsverktøy (forskningsspørsmål 1)

Tabell 10: Identifiserte kartleggingsverktøy referansevurdert av faggruppen

Identifiserte kartleggingsverktøy (forskningsspørsmål 1)
Breuer-Weuffen-differensieringstest for førskolebarn
Brigance (Preschool Screen/ Early Childhood Screens III)
British Picture Vocabulary Scale II
CSBS DP™ Infant-Toddler Checklist
E.R.I.K. - Early Referral Identification Kit
Edmonton Narrative Norms Instrument (ENNI)
ELIM - Early Language Identification Measure
Expressive Vocabulary Test
KAM-test
Lene
Looksee Checklist (formerly NDDS) - Nipissing District Developmental Screen
New Reynell Developmental Language Scales
Peabody Picture Vocabulary Test
Quick Interactive Language Screener™ (QUILS™)
Renfrew Language Scales
Reynell Developmental Language Scales
Rourke baby record
SATS-2
Schedule of Growing Skills
Snøgubbstestet (Språk- och talscreening)
Speech and Language Pathology Early Screening Instrument (SLPESI)
Sprogvurdering 2
Sprogvurdering 3-6
SPRÅK 4
Språk 5-6
Språkfyrn
Språkscreening ved 2,5 års alder
Språkscreening ved 3 år
Språkundersøkning 18 mnd

Stoke Speaks Out (Staged Pathway Toolkit)

Tribus

WellComm (Early Years)

Vedlegg 7: Måleegenskaper evaluert per publikasjon

Tabell 11: COSMIN Måleegenskaper evaluert per publikasjon

Publikasjon	Reliabilitet					Validitet			Responsivitet
	1	2	3	4	5	6	7	8	
Adamson-Macedo (2009a)	x								
Adamson-Macedo (2009b)	x	x	x						
Anderson (2016)							x		
Bootsma (2023)	x	x			x		x		
Chiat (2007)	x	x	x					x	
Claessen (2017)		x						x	
de Villiers (2021)	x	x						x	
Deevy (2010)			x					x	
Dollaghan (2009)		x						x	
Eisenberg (2019)								x	
Finneran (2020)							x		
Fiske, (2018)	x							x	
Gardner (2006)	x		x					x	
Geytenbeek (2014)	x		x	x			x		
Glascoe (2002)	x	x	x					x	
Gray (2003)		x	x					x	
Haddeland (2018)	x		x						
Klem (2015)	x						x	x	
Klem (2016)								x	
Korpilahti (2016)								x	
Laing (2002)								x	
Lavesson (2018)	x		x					x	
Law (2023)								x	
Letts (2013)	x	x						x	
Matov (2020)	x							x	
Mattsson (2001)								x	
Nash (2011)							x		
Nayeb (2019)								x	
Nayeb (2020)								x	
Pace (2022)								x	
Pickstone (2003)			x						

Platt (2012)								x	
Radeborg (2006)	x								
Restrepo (2006)								x	
Schlichting (2003)								x	
Skipar, (2019)	x								x
Spaulding (2013)									x
Stokes (2009)									x
Thomas-Tate (2006)								x	
Thordardottir (2013)									x
van der Lely (2011)									x
Visser-Bochane (2020)									x
Webb (2008)								x	
Westerlund (2001)			x						x
Westerlund (2007)			x						x
Wood (2019)								x	
Woods (2014)									x

1. Intern konsistens, 2. Test-retest (over tid), 3. Inter-rater (ulike personer på samme tid), 4. Intra-rater (samme person ved forskjellig tid), 5. Målefeil, 6. Innholdsvaliditet
7. Begrepsvaliditet: (strukturell, konvergent, kryss kulturell), 8. Kriterievaliditet.

Vedlegg 8: Detaljerte resultater per publikasjon

Verktøy identifisert i forskningsspørsmål 1

Brigance Infant and Toddler Screen

Glascoe og medarbeidere (2002): Intern konsistens, evaluert med Guttman Lambda-koeffisienter, varierte fra 0,94 til 0,97 på tvers av spedbarns- og småbarns versjonen. Test-retest- og interrater-reliabiliteten ble evaluert på nytt for 8 % av utvalget (n = 36), og reliabilitetskoeffisientene varierte fra 0,98 til 0,99 for både spedbarns- og småbarns versjonen. Sensitiviteten for spedbarnscreeningen overordnet var 77 % og spesifisiteten 86 %. På småbarnscreeningen var den overordnede sensitiviteten 76 % og spesifisiteten 85 %, som forfatterne rapporterer er innenfor standardene for screeningtestens nøyaktighet. Korrelasjonen mellom BITS og kriteriemålene var ikke tilfredsstillende for ekspressive språkferdigheter, med $r=0,5$ for både sped- og småbarn. For reseptivt språk var korrelasjonen bedre; 0,89 og 0,70 for henholdsvis sped- og småbarn.

Early Language Identification Measure (ELIM)

Law og medarbeidere (2023): Måleegenskapene sensitivitet, spesifisitet, positiv sannsynlighetsratio, negativ sannsynlighetsratio, positiv prediktiv verdi og negativ prediktiv verdi ble regnet separat for ELIM sine fem seksjoner. Resultatene viste at for del 1: utviklingsmilepæler var sensitiviteten på 0.88 og spesifisitet på 0.55. Del 2 omhandlet ordlisten rapportert av foreldre og hadde sensitiviteten på 0.85 og spesifisitet på 0.84. Del 3 omhandlet populasjons risikofaktorer og hadde 0.7 i sensitivitet og 0.59 i spesifisitet. Del 4 var observasjon av fagpersoner og hadde sensitivitet på 0.91 og spesifisitet på 0.65. Del 5 omhandlet foreldre sine bekymringer og hadde sensitivitet på 0.73 og spesifisitet på 0.74. Videre rapporterte forfatterne at kombinasjonen av den foreldrerapportert sjekkliste for ordforråd og helsesykepleiers observasjon av barnets kommunikasjon og atferd var den optimale kombinasjonen for å forutsi språkferdigheter, med en sensitivitet på 0,98 og en spesifisitet på 0,63.

New Reynell Developmental Language Scales

Letts og medarbeidere (2013): Interne reliabilitetskoeffisienter ved hjelp av Kuder-Richardson-formelen ble beregnet på grunnlag av dataene som ble samlet inn fra alle barna. Disse var 0,95 for språkforståelse og 0,96 for språkproduksjon, noe som tyder på at NRDLs-skalaene har en høy grad av intern konsistens. Test-retest-reliabilitetskoeffisientene (40 barn) var 0,45 for språkforståelse og 0,62 for språkproduksjon, noe

som ikke er høyt. Samtidig validitet med BPVS og TROG ble vurdert ved å korrelere resultatene for 132 barn for prestasjoner på disse testene og på NRDLs. Alle korrelasjonene (0,63-0,78) var signifikante på 0,01-nivå (tosidig). De rapporterte korrelasjonene for samtidig validitet er høye nok til å antyde at de tre testene måler beslektede ferdigheter, men ikke så høye at de antyder at disse ferdighetene er nøyaktig de samme. Den diagnostiske nøyaktigheten til NRDLs varierte avhengig av seksjon, med høyere sensitivitet for produksjonsskårer sammenlignet med forståelsesskårer. Diskriminantanalysen viste at NRDLs skiller effektivt mellom barn med språkvansker og barn med typisk utvikling.

Peabody Picture Vocabulary Test

Spaulding og medarbeidere (2013): viste en sterk sammenheng mellom prestasjonene på de to testene, men 35 % av barna presterte ulikt på de to versjonene. Barn med spesifikke språkvansker (SSV) presterte signifikant dårligere enn TD-barn på begge testene. En grenseverdi på 103 ble funnet å være optimal standardskår for begge testene som kunne skille mellom barn med og uten SSV for begge testversjonene. Ved bruk av denne grenseverdien forble sensitiviteten konstant på 80 %, noe som indikerer hvor stor andel av barna med spesifikke språkvansker som ble korrekt identifisert som å ha lidelsen, mens spesifisiteten var 75 % på PPVT-III og 70 % på PPVT-IV, noe som indikerer hvor stor andel av barna uten SSV som ble korrekt identifisert som ikke å ha lidelsen.

Dollaghan og medarbeidere (2009): Inter-rater reliabiliteten varierte fra 91 % til 94 %. Alle PPVT-R-skårer ble verifisert av en annen uavhengig bedømmer. Når det gjelder diagnostisk nøyaktighet, var korrelasjonskoeffisienten for PPVT 0,73 (95% CI: 0,68, 0,77) fra 3 til 4 år, 0,77 (95%CI 0,73, 0,81) fra 4 til 6 år og 0,69 (95%CI 0,64, 0,74) fra 3 til 6 år. På individnivå var den prediktive nøyaktigheten generelt dårlig i alderen 3, 4 og 6 år. En lav skåre ved 3 år økte ikke risikoen for en lav skåre ved 4 eller 6 år for noen av målene, men en lav PPVT-R-skåre ved 4 år, som økte risikoen for en lav PPVT-R-skåre ved 6 år.

Webb og medarbeidere (2008): Differential item functioning-analyser ble gjennomført for å avdekke DIF, og standard Mantel-Haenszel-test ble brukt for å avdekke DIF-elementer. Forfatterne rapporterte om minimale forskjeller mellom AA- og EA-barn ved hjelp av DIF-analysen. Videre utførte de latent klasseanalyse, og inspeksjonen av svarmønstrene viste at hovedforskjellen mellom barna i disse latente klassene først og fremst kunne tilskrives bruken av en bestemt svarstrategi, N3C. Denne strategien brukes av barnet når det er flere objekter til stede, hvorav vanligvis bare ett er ukjent, for å tilordne navn til nye objekter i omgivelsene.

Platt (2012): Resultatene viste signifikante forskjeller i de generelle standardskårene mellom EL1- og ELL-barn, der EL1-barna presterte bedre. En Mantel-Haenszel DIF-analyse ga et detaljert bilde av hvordan de enkelte testoppgavene fungerte psykometrisk i de to språkgruppene. Denne analysen viste begrensede indikasjoner på DIF i de første

168 oppgavene i PPVT-III, noe som tyder på minimal testskjevhet i PPVT-III mellom de to gruppene av barn, noe som forfatteren rapporterer at indikerer at forskjellene i skårer gjenspeiler forskjeller i det generelle reseptive ordforrådet snarere enn skjevheter i testen.

SPRÅK 4:

Klem 2015: Resultatene viste at verktøyet har en dimensjonal struktur som støttes av én generell språkfaktor, som står for 65-70 % av variansen i deltestene. Når det gjelder konvergent validitet, fant forskerne at SPRÅK 4 -verktøyet var signifikant korrelert med den eksterne språkfaktoren ($r = 0,60$, $p < .001$), noe som indikerer at verktøyet måler det samme konstruktet som kriteriet.

Språkfyrn

Laversson og medarbeidere (2018): Screeninginstrumentet viste tilfredsstillende reliabilitet (Cronbachs alfa var 0,80) og akseptabel sensitivitet, spesifisitet. Instrumentet hadde en sensitivitet på 0,84 og en spesifisitet på 0,96. Instrumentet hadde god validitet, og det teoretiske grunnlaget bidro til validiteten. Instrumentet var enkelt å administrere og skåre for helsesykepleiere, og inter-rater reliabiliteten var høy (0,99). Instrumentet var enkelt å administrere og skåre for helsesykepleiere, og inter-rater reliabiliteten var høy (0,99)

Språkscreening ved 2,5 år i Sverige

Mattson og medarbeidere (2001): Resultatene viste at den positive prediktive verdien ble beregnet til 0,52 (95 % KI: 0,31-0,72, den negative prediktive verdien til 0,96 (95 % KI 0,89-0,99), sensitiviteten til 0,69 og spesifisiteten til 0,93.

Nayeb og medarbeidere (2020): Resultatene viste at bare screeningmodellen med direkte vurdering på begge språk oppnådde tilstrekkelig nøyaktighet. Denne modellen hadde en sensitivitet på 88 %, spesifisitet på 82 %, positiv prediktiv verdi på 67 % og negativ prediktiv verdi på 94 %.

Nyaeb og medarbeidere (2019): Resultatene for treords- og toordskriteriet var følgende: sensitivitet (100 % versus 91 %), spesifisitet (81 % versus 91 %), positiv prediktiv verdi (38 % versus 56 %) og negativ prediktiv verdi (100 % versus 99 %). Treordskriteriet identifiserte 29 barn med mulig språkforstyrrelse, inkludert 11 tilfeller som senere ble bekreftet, mens toordskriteriet identifiserte 18 mulige tilfeller, inkludert 10 bekreftede tilfeller.

Språkscreening ved 3 år

Westerlund og medarbeidere (2001): Resultatene viste at en global vurdering i forbindelse med 3-årsscreeningen ville være nyttig for selektiv screening av 4-åringene. Det spesifikke terskelnivået på VAS ble satt til ≤ 50 mm med sensitivitet (74 %) og spesifisitet (68 %). Skalaens inter-rater reliabilitet er tidligere beregnet til $r = 0,85$.

Westerlund 2007: Den prosentvise Inter-rater reliabiliteten for 3-årscreeningen varierte fra 0,75 til 1,0 for språkforståelse og fra 0,92 til 1,0 for språkproduksjon. Screeningprosedyren ved 3 år hadde en sensitivitet på 86 %, en spesifisitet på 95 % og en positiv prediktiv verdi på 15 % når det gjaldt å identifisere barn med språkavvik i henhold til aldersnormen.

Quick Interactive Language Screener (QUILS)

Pace og medarbeidere (2022): Resultatene viste at QUILS hadde høy spesifisitet i studie 1 (92,3 %) og middels spesifisitet i studie 2 (77 %), mens sensitiviteten varierte fra uakseptabel i studie 1 (0,65) til adekvat i studie 2 (0,76). Noe som tyder på at det kun bør brukes som screeninginstrument i sammenhenger som er tilnærmet lik populasjonsnivåene for lidelsen.

De Villiers og kollegaer (2021): Mot PLS-5 var korrelasjonen på de engelske verktøyene var moderat til høy ($r = 0,69$, $p < .001$) og korrelasjonen til den spanske versjonen av verktøyene var noe lavere, men fremdeles signifikant ($r = 0,449$, $p < .002$). Videre ble 29 andre barn analysert med å sammenligne QUILS:ES med det spansk-engelske språkkartleggingsverktøyet BESOS. Korrelasjonen var moderat. For å undersøke intern konsistens ble det gjennomført separate Rasch-analyser for den engelske og spanske versjonen. Den interne reliabiliteten var god med høye Cronbach alpha skårer for både den engelske versjonen (0,89) og spanske versjonen (0,85). Det ble også målt test-retest reliabilitet. Det ble gjennomført testen med to ukers mellomrom hvor test-retest-reliabiliteten var høy (0,89).

Verktøy identifisert i databasesøk

Bayley Scales of Infant and Toddler Development-III (Bayley-III)

Woods og medarbeidere (2014): Til tross for høy spesifisitet (0,93) for reseptivt språk, var sensitiviteten dårlig, da 79 % av barna med spesifikke språkvansker ved 5 år skåret innenfor normalområdet ved 3 års alder. For ekspressivt språk var den diagnostiske nøyaktigheten relativt lav (spesifisitet 0,77, sensitivitet 0,54), og stabiliteten i språkvanskenivåene mellom 3 og 5 år viste bare middels samsvar (Cohens κ 0,383)

BESS Speech and Language Therapy Two Year Developmental Check

Adamson-Macedo og medarbeidere (2009a): BESS-verktøyet viste god intern konsistens på de kvantitative elementene, med en Cronbach's alpha = 0,75. Videre viste BESS-verktøyet god intern konsistens på de kvalitative elementene, målt ved Kuder-Richardson-statistikken (KR20 = 0,66). Videre analyser indikerte også at alle elementene av verktøyet korrelerte godt med totalskåren og var kompatible med verktøyet. Ved sammenligning av de totale alfa-skårene på skalaen viste BESS-verktøyet svakere skåre enn MCDI, mens den var nesten identisk med Reynell verktøyet

Adamson-Macedo og medarbeidere (2009b): Studien viste at verktøyet hadde en god intern konsistens med en god alfa-koeffisient på totalskalaen for de kvantitative elementene ($r_s=0,75$). Alfa-koeffisientene for inter-rater reliabilitet for forskeren ($r_s=0,76$) og BESS-personalet ($r_s=0,75$) var nesten identiske, noe som tyder på at verktøyet er pålitelig på tvers av administratorer med ulik kompetanse. Test-retest-metoden, som ble brukt på 101 BESS-barn i løpet av to uker, viste en robust korrelasjonskoeffisient ($r_s=0,96$, $p<0,01$), noe som tyder på konsistente skårer som ikke påvirkes av ulike administrasjonstidspunkter. Ved hjelp av Nevills Percentage of Agreement Test vurderte studien test-retest-stabiliteten og avdekket en enighet på 93 % og mer for hele verktøyet, hvilket indikerer god stabilitet på tvers av de to tidspunktenes.

Computer-Based Instrument for Low motor Language Testing (C-BiLLT)

Geytenbeek og medarbeidere (2014): Studien viste at både intra-rater reliabiliteten og inter-rater reliabiliteten ($n = 37$) var utmerket for barn med typisk utvikling, med ICC på henholdsvis 0,97 (95 % KI = 0,94 - 0,99) og 0,96 (95 % KI = 0,94 - 0,97). Cronbachs alfa og Guttman's lambda2 beregnet med 75 ledd for barn med typisk utvikling var henholdsvis 0,88 og 0,93. Når det gjelder begrepsvaliditet, ble den sterkeste korrelasjonen funnet mellom C-BiLLT-skårene og RDLS-skårene hos barn med typisk utvikling, $r = 0,93$ ($n = 423$). En noe lavere korrelasjon enn forventet (men fortsatt høy) ble funnet mellom C-BiLLT-skårene og PPVT-III-skårene hos barn med typisk utvikling, $r = 0,88$ ($n = 117$).

Bootsma og medarbeidere (2023): Resultatene viste at C-BiLLT-CANs konvergente validitet var god til svært god (Spearman's $\rho > 0,78$), og den diskriminante validiteten var høyere enn antatt (Spearman's $\rho > 0,8$). Intern konsistens (Cronbachs alfa = 0,96), test-retest-reliabilitet (ICC $> 0,9$) og målefeil (SEM < 5 %) var utmerket.

Fiske (2018): Korrelasjonsanalysen viste høyere korrelasjon mellom C-BiLLT-sumskårer og andre tester for språkforståelse enn mellom C-BiLLT og tester som måler ikke-verbal resonnering. C-BiLLT kan dermed betraktes som et rent mål for barns språkforståelse. Faktoranalysen resulterte i to faktorer, vokabularforståelse og grammatisk forståelse, som høyt korrelerte og forklarte variasjon i alle C-BiLLT-seksjoner. Dette støtter at C-BiLLT representerer en enhetlig språklig evne. C-BiLLT viste seg sensitiv for å skille mellom barn i ulike aldersgrupper, spesielt blant de yngre, og for barn med lav språkforståelse. Imidlertid var det usikkerhet knyttet til C-BiLLTs evne til å differensiere blant eldre barn og høytfungerende barn, og dette kan kreve ytterligere testing på et mer variert utvalg.

Haddeland (2018): Studien støtter tidligere funn om at bruk av ulike responsmodaliteter ikke signifikant påvirker testresultatene, noe som indikerer muligheten for tilpasninger i testadministrasjon for barn med tale- og bevegelsesvansker.

Skipar 2019: Analysen ble gjennomført både for C-BiLLTs dimensjoner og for alle spørsmålene, og resultatene viste en høy Cronbach's Alpha på henholdsvis 0,89 og 0,98.

Validiteten ble undersøkt ved å se på sammenhengen mellom C-BiLLT og Reynell språktest. Resultatene viste en signifikant korrelasjon på $r=0,89$ mellom C-BiLLT og Reynell.

Dutch well child language screening protocol

Visser-Bochane og medarbeidere (2020): Studien viste at protokollen for 2-årig språkscreening hadde god samtidig (sensitivitet 0,72-0,79, spesifisitet 0,86) og prediktiv validitet ((sensitivitet 0,80-0,82, spesifisitet 0,74-0,79). Fagpersonalet viste god etterlevelse av protokollen, og avvik fra protokollen endret ikke protokollens sensitivitet og spesifisitet.

First word test (FWT)

Pickstone 2003: Studien har oppgitt sensitivitets- og spesifisitetsmålinger for det foreldrerapporterte skjemaet og FWT samlet, så resultatene kan derfor ikke si noe om sensitiviteten til FWT alene. For inter-rater reliabilitet, rapporterte forfatterne at første forsøk oppnådde en moderat inter-rater reliabilitet (enighet) mellom testerne med en Kappaskår på 0,55 (95 % KI 0,45-0,65). I andre forsøk var inter-rater reliabiliteten høyere med en Kappaskår mellom 0,78 og 1,0.

Fluharty Preschool Speech and Language Screening Test - Second Edition

Eisenberg og medarbeidere (2019): Resultatene fra studien indikerte begrenset sammenfall mellom resultatene fra Fluharty-2 og de diagnostiske målingene. De barna som ikke besto Fluharty-2, presterte betydelig dårligere på hver av de diagnostiske testene sammenlignet med de som besto Fluharty-2, og skårene på Fluharty-2 var signifikant korrelert med skårene på de diagnostiske målene. Den eneste betydelige sammenhengen ble funnet i forhold til IPSyn når det gjaldt bestått/ikke bestått. Likevel var selv denne sammenhengen moderat for bestått (80 %) og middels for ikke bestått (68 %).

Fox Language Inventory (FLI)

Korpilathi og medarbeidere (2016): Resultatene fra språktestene viste at 8.8 % av barna hadde forsinket språkutvikling sammenlignet med 10.9 % for FLI alene. Sensitivitet lå på 85,7 % og spesifisitet på 73.7 % (KI:40.2-100.0). Dette var mer nøyaktige måleegenskaper enn de for det foreldre rapporterte verktøyet som ble testet på barn som var 24 måneder.

Grammar and Phonology Screening (GAPS)

Van der Lely (2011): Resultatene indikerte betydelig lavere poengsum på språktestbatteriet for barn med språkforstyrrelser (SSV) sammenlignet med de med typisk utvikling. GAPS-testen viste imidlertid svært gode resultater når det gjaldt barn innenfor testnormene. Ved den 5. persentilen oppnådde den 100 % sensitivitet og spesifisitet, nøyaktig skille mellom barn med og uten språkvansker. Selv ved en cut-off på 10 % opprettholdt den høy nøyaktighet og identifiserte 98 % av barna korrekt.

Selv om den 5. persentilen viste seg å være den mest nøyaktige grenseverdien, var den 10. persentilen fortsatt egnet til å anbefale videre utredning, til tross for en noe høyere

andel falske positive. Dette kunne føre til klassifisering av et barn som potensielt med fonologiske svakheter. Videre analyser ved hjelp av ROC (receiver operating characteristics) bekreftet GAPS-testens nøyaktighet. ROC-analysen inkluderte begge deltestene og viste at GAPS-testen var svært nøyaktig i å klassifisere barn med og uten fonologiske vansker. Arealet under kurven (AUC) var perfekt (AUC = 1,0) ved 5. persentilen og bare litt mindre nøyaktig ved 10. og 15. persentilen. Når deltestene ble undersøkt individuelt, viste både grammatikk- og fonologi-deltestene høy nøyaktighet ved 10. persentilen (henholdsvis AUC = 0,955 og AUC = 0,938).

Nash (2011): I henhold til konvergent validitet hadde begge GAPS sine to subkategorier (grammatikk og fonologi) positiv og signifikant korrelasjon med majoriteten av kategoriene i sammenligningstestene. Videre identifiserte GAPS 2 av 10 barn med språkproblemer som CELF-Preschool 2 hadde identifisert noe som gav en lav sensitivitet på 20 %, spesifisiteten derimot ble målt til å være 96 %. GAPS hadde lignende måleegenskaper når verktøyet ble sammenlignet med foreldrebekymring

Gardner og kollegaer: Resultatene viste at testen hadde god/svært god intern konsistens, med en Cronbachs alfa på 0,858 for setningsrepetisjonskomponenten og 0,729 for nonordrepetisjonskomponenten. Funnene viste også at grammatiske og fonologiske vansker for mange barn ikke går hånd i hånd, og at bare halvparten av de barna i normalpopulasjonen som har en eller annen form for språkvanske, har både fonologiske og grammatiske vansker. Testerens yrke forklarte bare en liten del av variasjonen i dataene. Korrelasjonsanalyser ble utført for å vurdere samsvar mellom to GAPS-deltester og referansetestene. GAPS-grammatikkprøven (setningsrepetisjon) korrelerte sterkt ($p < 0,001$) med alle de fire testene. GAPS' fonologitest (ikke-ordrepetisjon) viste lavere, men signifikante korrelasjoner med BPVS, CELF Sentence og Word Structure. GAPS' fonologitest hadde en sterk korrelasjon med CN-Rep. GAPS-testen viste de sterkeste korrelasjonene med standardiserte språktester som tar opp lignende grammatikk- eller fonologikomponenter, og forble signifikante selv etter at det var tatt hensyn til alder.

Non word repetition:

Gray og medarbeidere (2003): Lydbånd eller WAV-filer fra 40 % av nonordrepetisjonsøktene ble valgt ut for uavhengig skåring av en trent lytter. Halvparten av øktene var barn med SSV og halvparten med NL. Den gjennomsnittlige punkt-til-punkt-overensstemmelsen for riktig/feil-skåring var 94 % (spredning: 70-100 %). Den første administrasjonen av nonordrepetisjonsoppgaven, der det totale antallet nonord som ble repetert korrekt, ble brukt, resulterte i utmerket sensitivitet og spesifisitet. 95 % av barna med SSV ble identifisert som SSV, og 100 % av barna med NL ble identifisert som NL. Sensitiviteten og spesifisiteten gikk noe ned ved den andre administreringen, men var fortsatt høy. Spesifisiteten ble ytterligere redusert ved tredje administrering.

Når det gjelder test-retest-reliabilitet, var det en signifikant korrelasjon mellom skårene for nonordrepetisjon ved hvert administrasjonstidspunkt for SSV-gruppen, med sterkeste korrelasjon fra tidspunkt 1 til 2 (0,72). Verken tid 1-2 (0,40) eller tid 2-3 (0,22)

var signifikant korrelert for NL-gruppen, men tid 1-3 (0,49) nådde signifikans på $P < 0,05$ -nivå. Det er imidlertid verdt å merke seg at barna repeterte den samme listen med nonord på tidspunkt 1 og 2, og en annen liste med ord på tidspunkt 3, og at listene var balansert på tvers av gruppene.

Radeborg og medarbeidere (2006): Forskerne fant at lengre nonord var vanskeligere for barna å gjenta, og observerte signifikante effekter av alder og dialektkompatibilitet mellom tester og barnet. Uavhengige skåringer ble utført av de to testerne hver for seg, og det ble oppnådd 96 % samsvar mellom de to. Den interne konsistensen for 30-itemtesten og den forkortede 24-itemtesten, beregnet ved Cronbachs alfa, var på henholdsvis 0,74 og 0,73. Siden det ikke så ut til å være noen fordeler ved å bruke 30-itemtesten, anbefaler forfatterne å bruke 24-itemtesten, som omfatter to-, tre-, fire- og femstavelses nonord. Det viste seg at manglende kompatibilitet mellom testerens aksent og den aksenten barnet er mest vant til, kan ha en negativ effekt på barnets prestasjoner, spesielt når det gjelder yngre barn. Studien konkluderte med at den svenske nonord-repetisjonstesten er et verktøy for å identifisere potensielle språkvansker hos førskolebarn, noe som muliggjør tidlig intervensjon og støtte.

Deevy (2010): Interrater-reliabiliteten var 94 %. Analyse av diagnostisk nøyaktighet, basert på en grenseverdi på 66 % på NRT som ville gi den største inndelingen i grupper, resulterte i 86 % sensitivitet og 91 % spesifisitet, med 95 % konfidensintervaller på henholdsvis 73 %-99 % og 83 %-99 %, noe som anses som tilstrekkelig og godt for sensitivitet og spesifisitet. Den positive og negative prediktive verdien var henholdsvis 86 % og 91 %.

Stokes (2009): For TENR 1-3-stavelsestesten var sensitiviteten og spesifisiteten henholdsvis 75 % (95 % KI=51 %-90 %) og 90 % (95 % KI=85 %-94 %). For TENR 1-4-stavelsestesten var sensitiviteten og spesifisiteten henholdsvis 88 % (95 % KI = 53 %-99 %) og 94 % (95 % KI = 87 %-98 %), dvs. at TENR 1-4-stavelsestesten viste bedre diagnostisk nøyaktighet.

Thordardottir og medarbeidere (2013): Studien omfattet fire grupper med 14 barn i hver gruppe: (1) tospråklige barn med primære språkvansker (PLI) (BI-PLI), (2) tospråklige barn med typisk språkutvikling (BI-TD), (3) enspråklige barn med PLI (M-PLI) og (4) enspråklige barn med typisk språkutvikling (M-TD). Alle gruppene var like gamle, med en gjennomsnittsalder på nesten 5 år. NWR oppnådde høy sensitivitet og identifiserte korrekt 85 % av barna som tidligere var identifisert som PLI, uavhengig av om barna var tospråklige eller enspråklige, og med samme grenseverdi for begge grupper. Spesifisiteten var høy for enspråklige barn (100 %), men falt til bare middels for tospråklige barn (79 %).

Short language measure (SLaM)

Matov og medarbeidere (2019): Resultatene viste at SLaM hadde en nøyaktighet på 94 %, (sensitivitet på 94 %, spesifisitet på 93 %). Målet viste også en samtidig validitet på 0,89 og en reliabilitet (intern konsistens) på 0,93.

Structured screening test

Laing og medarbeidere (2002): Den strukturerte screeningtesten hadde en sensitivitet på 66 % og en spesifisitet på 89 % for barn med alvorlige språkvansker, og en sensitivitet på 54 % og en spesifisitet på 90 % for barn med behov for behandling. Den foreldrestyrte metoden hadde en sensitivitet på 56 % og spesifisitet på 85 % for barn med alvorlige språkvansker, og en sensitivitet på 58 % og spesifisitet på 90 % for barn med behov for terapi.

Test for Sentence Development

Schlichting og kollegaer (2003): Korrelasjonene med språktestene varierer fra 0,63 til 0,72, mens korrelasjonene med de ikke-språklige testene varierer fra 0,42 - for en test som skal måle ikke-verbal intelligens - til 0,56. Når det gjelder reliabilitet, har studien vurdert intern konsistens og test-retest-reliabilitet. Den interne konsistensen ble målt ved hjelp av lambda-2-koeffisienten, med en gjennomsnittsverdi på 0,85. Test-retest-reliabiliteten ble funnet å være .75 med seks måneders intervall mellom testene.

Toddler phonology test (TPT)

Claessen og medarbeidere (2017): Resultatene fra den første studien viste ingen endring i Percent Consonants Correct (PCC) og Percent Vowels Correct (PVC) målene i løpet av en måned, men en liten signifikant nedgang i atypiske feil. Positive korrelasjoner indikerte konsistente resultater på tvers av testene. Den andre studien, som fulgte opp 24 barn fire til ni måneder senere, viste at både kvantitative og kvalitative mål på TPT bidrar til å identifisere småbarn med risiko for fonologiske forstyrrelser. Atypiske feil, særlig når det gjelder stavelsesform, viste seg å være en mer pålitelig markør i de longitudinelle casestudiene.

Utgitt av Folkehelseinstituttet
Januar 2024
Postboks 222 Skøyen
NO-0213 Oslo
Telefon: 21 07 70 00
Rapporten kan lastes ned gratis fra
Folkehelseinstituttets nettsider
www.fhi.no