


# Universal language development screening: comparative performance of two questionnaires

Philip Wilson <sup>1</sup>, Robert Rush,<sup>2</sup> Jenna Charlton,<sup>3</sup> Vicky Gilroy,<sup>4</sup> Cristina McKean,<sup>3</sup> James Law<sup>3</sup>

**To cite:** Wilson P, Rush R, Charlton J, *et al*. Universal language development screening: comparative performance of two questionnaires. *BMJ Paediatrics Open* 2022;**6**:e001324. doi:10.1136/bmjpo-2021-001324

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjpo-2021-001324>).

Received 15 October 2021  
Accepted 16 December 2021



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

<sup>1</sup>Centre for Rural Health, Institute of Applied Health Sciences, University of Aberdeen, Inverness, UK  
<sup>2</sup>Independent statistical consultant, Edinburgh, UK  
<sup>3</sup>School of Education, Communication and Language Sciences, Newcastle University, Newcastle upon Tyne, UK  
<sup>4</sup>Institute of Health Visiting, London, UK

**Correspondence to**  
Professor Philip Wilson; [p.wilson@abdn.ac.uk](mailto:p.wilson@abdn.ac.uk)

## ABSTRACT

**Background and objective** Low language ability in early childhood is a strong predictor of later psychopathology as well as reduced school readiness, lower educational attainment, employment problems and involvement with the criminal justice system. Assessment of early language development is universally offered in many countries, but there has been little evaluation of assessment tools. We planned to compare the screening performance of two commonly used language assessment instruments.

**Methods** A pragmatic diagnostic accuracy study was carried out in five areas of England comparing the performance of two screening tools (Ages and Stages Questionnaire (ASQ) and Sure Start Language Measure (SSLM)) against a reference test (Preschool Language Scale, 5th edition).

**Results** Results were available for 357 children aged 23–30 months. The ASQ Communication Scale using optimal cut-off values had a sensitivity of 0.55, a specificity of 0.95 and positive and negative predictive values of 0.53 and 0.95, respectively. The SSLM had corresponding values of 0.83, 0.81, 0.33 and 0.98, respectively. Both screening tools performed relatively poorly in families not using English exclusively in the home.

**Conclusion** The very widely used ASQ Communication Scale performs poorly as a language screening tool, missing over one-third of cases of low language ability. The SSLM performed better as a screening tool.

## INTRODUCTION

Early language development is a good indicator of cognitive capacity.<sup>1</sup> Satisfactory acquisition of language also requires an adequate social environment<sup>2</sup> and an ability to give attention to the communication of others. Delayed language acquisition in the third year of life is a strong predictor of later psychopathology<sup>3–5</sup> as well as reduced school readiness,<sup>6</sup> lower educational attainment,<sup>7 8</sup> employment problems<sup>3</sup> and involvement with the criminal justice system.<sup>9</sup>

Given our knowledge of the natural history of language disorders<sup>10–12</sup> and their potential mutability by social factors, it may seem surprising that language screening does not

## What is known about the subject?

- There is very little published research reporting the screening performance of the Ages and Stages Questionnaire (ASQ) Communication Scale compared with a reference test.
- No studies have reported on the screening performance of the Sure Start Language Measure (SSLM).
- No published research has reported good screening performance of questionnaires for speech and language development in primary care or in substantial population samples.

## What this study adds?

- The screening performance of the SSLM was superior to that of the ASQ Communication Scale.
- Neither instrument performed well among families in which English was not the only language spoken at home.
- Further work to refine the performance of language screening tools is required, particularly among families where English is not used exclusively at home.

yet meet internationally accepted screening criteria.<sup>13 14</sup> There is a lack of evidence on the benefit of early treatment<sup>15</sup> and on cost-effectiveness,<sup>16 17</sup> although there remains a widely held view that surveillance of language development is desirable and it has been adopted internationally.<sup>18 19</sup>

Parents are often best placed to report on their child's skills in the very early years and commonly want to discuss any concerns they may have about their child's development and language development, particularly around the age of 24 months, by which time parents compare their own child's development with that of others. Where national policies on developmental assessments exist, they generally include the use of structured instruments to assess language acquisition in the third year of life.<sup>18 20</sup> Many language screening measures exist, but few have been adequately evaluated



or compared with others; these are essential first steps in building the case for or against universal screening.<sup>14 21 22</sup>

Here we compare the screening performance of two commonly-used language screening tools: the Ages and Stages Questionnaire (ASQ)<sup>23–25</sup> and the Sure Start Language Measure<sup>26</sup> (SSLM). The ASQ is translated into many languages and is probably the most widely used developmental screening tool internationally, although there is only one published study of the concurrent validity of its communication domain against a reference language assessment in a low-risk population-based sample,<sup>27</sup> and the sample of children in the relevant age range is small. The SSLM has been developed for the English Sure Start programme<sup>26</sup> and has been used in its large-scale evaluation. It has nevertheless only undergone one small-scale assessment against the Reynell Developmental Language Scales and has shown good predictive validity.<sup>5</sup>

In the present paper, we examine the concurrent validity of the ASQ and SSLM used with children aged between 23 and 30 months against a reference test, the Preschool Language Scale V.5 (PLS-5),<sup>28</sup> in a substantial socially diverse sample in five areas of England.

## METHODS

The study was conducted as a planned component of a larger programme of work funded by Public Health England, which was designed to optimise a universal programme for language assessment and early intervention in England—‘Identifying and Supporting Children’s Early Language Needs’.<sup>29</sup> The measure, the SSLM, which is used in the present paper, is an integral part of the Early Language Identification Measure (ELIM), created for the study in question. The screening data were collected during a scheduled review, offered by community child nursing (health visitor) teams as a mandated part of England’s Healthy Child Programme to all children at 2.0–2.5 years of age.<sup>23</sup>

## Participants

Participants were children aged 23–30 months and their parents/carers living in five areas identified by Public Health England: Derbyshire, Middlesbrough, the London Borough of Newham, Wakefield, and Wiltshire. The five sites were selected as they were considered to have good information technology systems, and they reflected a broad spread of socioeconomic status with a good representation of more disadvantaged areas.

## Procedures

Parents in the five sites who were due to attend for their child’s health visitor review of 2.0–2.5 years were sent, via post, a study information leaflet and consent form alongside the Ages and Stages Questionnaire, Third Edition (ASQ-3), which forms part of the universal developmental assessment offer to parents of all children aged 2.0–2.5 years in England.<sup>23</sup> Parents were

asked to return their signed consent form to the health visitor during their review.

Across sites, the review took place in either the home or a clinical setting. The health visitor recorded the ASQ-3 Communication Scale score, administered the SSLM, and collected demographic data and information about language use in the home as part of the ELIM. All parents who had the ELIM completed were then provided with a date to attend a second assessment with a speech and language therapist (SLT). Within this appointment, the SLT completed the PLS-5 with the child. The SLT assessment was carried out within 2–4 weeks of the initial ELIM assessment. The SLTs were blind to the results of the ELIM. Not all parents attended for the SLT assessment, and as the study period spanned the early phase of the UK national COVID-19 lockdown, a small proportion of participants were unable to attend for this reason.

## Measures

The ASQ-3 is a standardised parent-completed questionnaire used to screen for developmental delays in children<sup>25</sup>; it can be completed by parents in 12–18 min. Different versions are available for various ages: in the current study, these were the versions of 24, 27 and 30 months. There are five domains: fine motor, gross motor, communication, problem-solving and personal-social. Each domain contains six questions that can be answered with a yes (10 points), sometimes (5 points) or not yet (0 points), as well as nine open-ended questions. Scores obtained from each domain are compared with established cut-off points at 1 and 2 SDs that are used to identify children at risk of developmental problems. If the score on any domain falls below the 2-SD cut-off (‘refer’), referral for further assessment is advised. If the score on any domain is within the 1-SD and 2-SD cut-off point (‘monitor’), learning activities and monitoring of the child’s development are advised. We used data from the six communication domain questions for the present study. The ASQ-3 is a standard element in the Healthy Child Programme and the results are routinely reported in the UK.

The SSLM<sup>26</sup> is a 50-item word list originally adapted from the MacArthur Bates Communicative Development Inventory for use in Sure Start programmes in England in the early 2000s that takes approximately 5 min to complete. The clinician administering the measure ticks a box next to each word if the parent reports that their child is able to say that word. A total word count out of 50 is then recorded. For children living in families where English is not used exclusively, equivalent words in the other languages are acceptable. There is no recognised screening threshold for the SSLM.

The PLS-5<sup>28</sup> is a structured assessment of receptive and expressive language with items that range from preverbal, interaction-based skills to emerging language. The assessment is detailed and requires

practitioners to observe the child; assessment can take half an hour or more to complete. As part of the application of this measure, the person carrying it out has to have the requisite professional accreditation. In most cases, this measure would be carried out by an SLT who is who fully trained to administer and interpret standardised language assessments. The cut-off for low language ability was set at below or equal to the 10th centile following recent literature reporting a population prevalence of language disorder of 9.92%.<sup>30</sup>

### Analysis

Characteristics of children with full data and those without the Preschool Language Scale (PLS) assessment were compared using independent t-tests, analysis of variance, Mann-Whitney or  $\chi^2$  tests.

The optimal cut-off scores for the SSLM and ASQ-3 Communication Scale questions were determined from a receiver operating characteristic (ROC) curve in relation to the pre-selected PLS-5 threshold. The ROC curve illustrates the trade-off between sensitivity and specificity for each screening score. Areas under the ROC curve were used as an assessment of screening performance. Optimal cut-off scores from the ROC curves (best trade-off between sensitivity and specificity, derived using Youden's Index<sup>31</sup>) and the published above-threshold/monitor/refer classifications for the ASQ-3 Communication Scale score were used to obtain screening performance in terms of sensitivity, specificity, and positive and negative predictive values.

Data analysis was conducted using SPSS V.24.

### Role of the funding source

The project was funded as 'PHE - Corporate - Assessment Tool and Resources to Support Action by Health Visitors and Early Years Practitioners to Identify and Support Children with Early Speech, Language and Communication Needs ECM\_6378'. The study sponsor (Public Health England) approved the study design but played no part in the collection, analysis and interpretation of data, in the writing of this article or in the decision to submit the paper for publication.

### Patient and public involvement

Although there was substantial public engagement with the overall ELIM programme, members of the public were not involved in the design, analysis or interpretation of the research presented here.

## RESULTS

### Participant flow

Across the five sites, parents/carers of 894 children consented to participate in the study. **Figure 1** describes the flow of participants and data. Paired ASQ-3 and SSLM data were available for 811 (91%) of the children for whom consent was obtained. A full matched dataset

was obtained for 357 of these children who attended for the PLS-5 assessment.

Characteristics of the children taking part in the initial assessment, the 454 with ASQ-3 and SSLM screening data but with no PLS data and of the 357 with a full dataset including PLS-5 Language classification, SSLM score and ASQ-3 Communication Scale category are presented in **table 1**. Children attending for the PLS assessment were broadly similar to those who did not attend, with almost identical mean SSLM and ASQ language scale scores, but those who had the PLS were slightly older and were more likely to be from exclusively English-speaking families, and their mean area deprivation scores were lower.

In most cases, the reasons for non-attendance for the SLT assessment are not known, although 39 participants were unable to attend in March 2020 due to COVID-19.

Seventy-seven per cent (290/376) of the children in this sample had PLS scores above the 10th centile threshold while 23% (86) had scores below the threshold.

The age-appropriate versions of the ASQ-3 allow automatic correction for age within the sample, but the SSLM score is not age corrected. Online supplemental file 1 presents the scores for the whole sample of 865 children with SSLM data; children gained on average four new words per month within the age range 23–30 months, but correction for age does not significantly change the screening performance of the instrument.

The screening performance of the SSLM and ASQ-3 communication scales in relation to the PLS-5 cut-off is illustrated in **figure 2**.

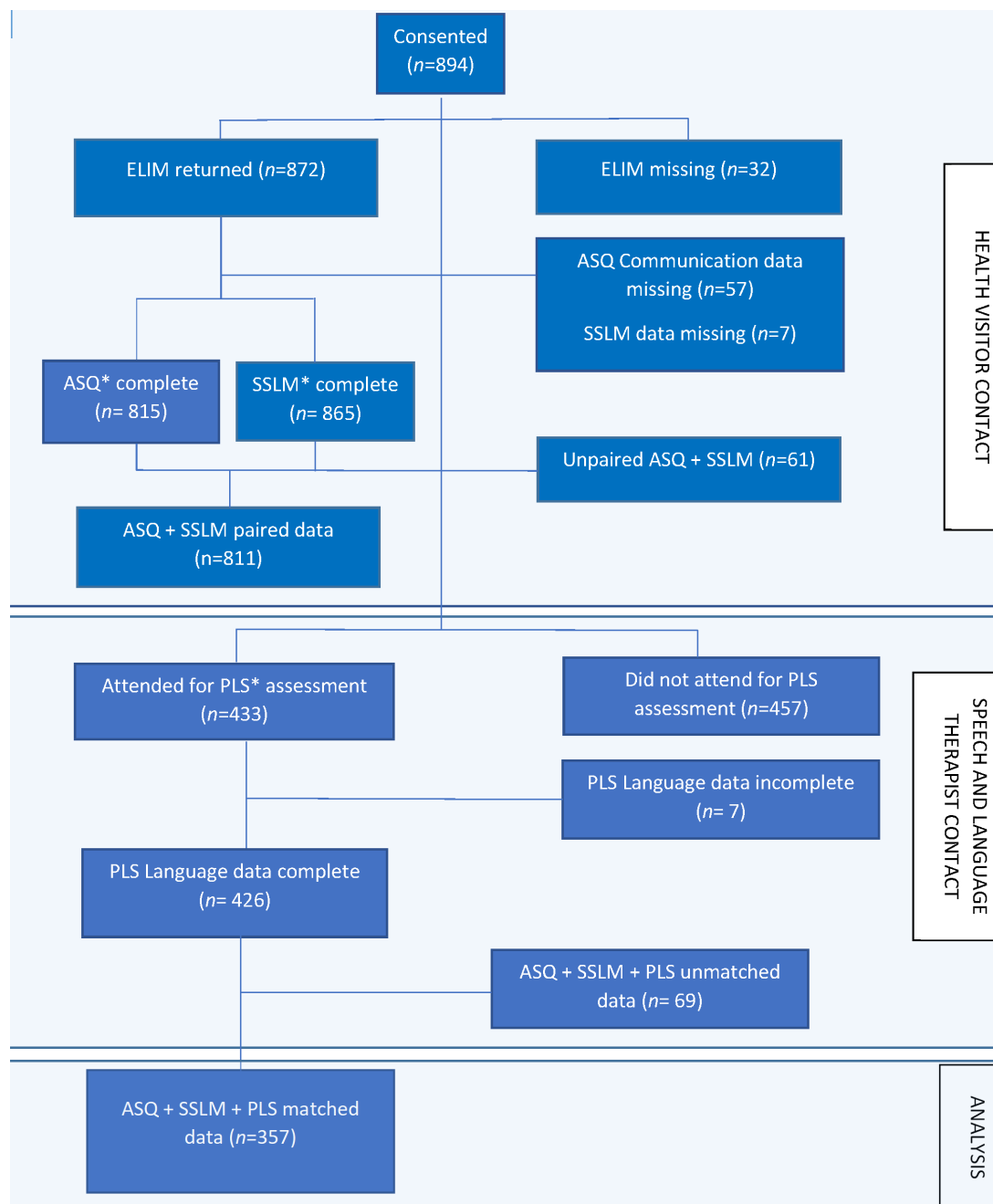
**Table 2** gives the performance of the ASQ-3 and SSLM language screening tools using both the optimal cut-off points derived from the ROC curves and, for the ASQ-3, the predefined cut-off points between the 'above-threshold' (normal), monitor and refer categories.

Online supplemental file 2 presents ROC curves for children of families speaking only English at home and for children of families speaking languages other than English (including bilingual families). Screening performance for both ASQ-3 and SSLM is poorer among families not using English exclusively at home (area under the curve (AUC) 0.740 and 0.764, respectively) than among families speaking English exclusively (AUC 0.895 and 0.912, respectively).

## DISCUSSION

### Main findings

Using the lower predefined threshold for the ASQ communication (ASQ-3) domain, encompassing both the monitor and refer classifications, we missed the 35% of children with low language ability. The higher ASQ-3 refer threshold misses over half of such children. There appears to be little scope for improving the ASQ-3 screening performance through changing the



**Figure 1** Participant flow through the study. Figure created by coauthor JC. \*ASQ communication subscale, SSLM 50-word list (Q9 of ELIM) and PLS V.5 total Language standard score. ASQ, Ages and Stages Questionnaire; ELIM, Early Language Identification Measure; PLS, Preschool Language Scale; SSLM, Sure Start Language Measure.

threshold score category. The SSLM performs substantially better as a screening instrument, missing 17% of children with significant language pathology but with slightly poorer specificity than the ASQ-3 at this level of sensitivity. The SSLM threshold of 17.5/50 words appears to operate effectively across the age range of 23–30 months. The performance of both the ASQ-3 and SSLM is poorer among children living in families in which English is not the only language spoken at home than it is among children living in families in which English is spoken exclusively.

### Strengths

The findings are based in a real-world evaluation of an enhanced approach to a universal developmental review focused on language. The proportion (23%) of children attending for the PLS language assessment whose language attainment fell below the published population tenth centile is higher than anticipated but may reflect the characteristics of the five sites which were relatively socially disadvantaged.<sup>2</sup> This proportion is nevertheless very similar to other studies of 2-year-old children's language in community-ascertained samples.

**Table 1** Characteristics of children attending for initial health visiting team assessment for whom age-appropriate ASQ-3 and SSLM data are available, and for those who had subsequent assessment by a speech and language therapist

|  | All participating children<br>N (%) | Children with no PLS<br>N (%) | Children with full data available for the main analysis<br>N (%) |
|--|-------------------------------------|-------------------------------|--|
| Female                                       | 377 (47)                            | 213 (47)                      | 164 (46)   |
| Male   | 419 (52)                            | 230 (51)                      | 189 (53)   |
| Gender not known                             | 15 (2)                              | 11 (2.4)                      | 4 (1)  |
| First-born                                   | 345 (42)                            | 197 (43)                      | 148 (42)   |
| Age (months), mean (SD)***                   | 25.5 (1.7)                          | 25.2 (1.8)                    | 25.8 (1.5)   |
| English only spoken at home***               | 495 (61)                            | 247 (54)                      | 248 (70)   |
| Another language only spoken at home         | 119 (15)                            | 85 (19)                       | 34 (9)   |
| English plus another language spoken at home | 83 (10)                             | 51 (11)                       | 32 (9)   |
| Language spoken at home not known            | 114 (14)                            | 71 (16)                       | 43 (12)  |
| IDACI decile, † mean (SD)***                 | 4.8 (2.5)                           | 4.4 (2.4)                     | 5.2 (2.6)  |
| SSLM word count, median (IQR)                | 27 (26)                             | 26 (22)                       | 29 (29)  |
| ASQ-3 score, mean (SD)                       | 48.3 (16.4)                         | 48.7 (15.9)                   | 47.8 (16.9)  |
| Respondent—mother                            | 763 (94)                            | 426 (94)                      | 337 (94)   |
| Total  | 811                                 | 454                           | 357  |

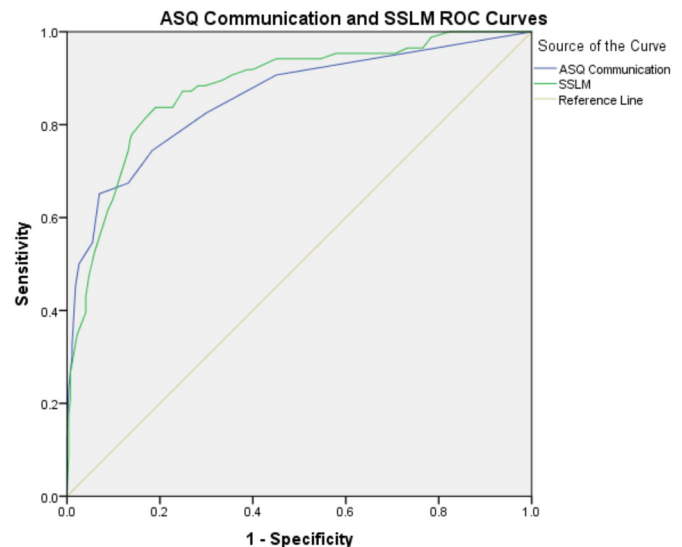
Characteristics of children with full data and those without the PLS assessment were compared using independent t-tests, analysis of variance, Mann-Whitney or  $\chi^2$  tests.

\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .

†IDACI taken from the family home postcode. The lower the IDACI decile, the more disadvantaged the post code.

ASQ-3, Ages and Stages Questionnaire, Third Edition; IDACI, Income Deprivation Affecting Children Index; PLS, Preschool Language Scale; SSLM, Sure Start Language Measure.

For example, the Early Language in Victoria Study<sup>32</sup> found 19.7% of children at 2 years of age fell below the 10th centile cut-point using the Communication Development Inventories,<sup>33</sup> suggesting samples in these studies may be more representative of the population than those used to norm the tests. The sample size was sufficient to draw sound conclusions, although only



**Figure 2** ROC curves illustrating screening performance of the ASQ-3 communication score and the age-adjusted SSLM word list against the Preschool Language Scale V.5 reference test. Figure created by coauthor RR. ASQ, Ages and Stages Questionnaire; ASQ-3, Ages and Stages Questionnaire, Third Edition; ROC, receiver operating characteristic; SSLM, Sure Start Language Measure.

around half the children undertook the optional gold-standard language assessment. Overall, the differences between those children attending the initial assessment only and those attending for both assessments were relatively minor: both the original sample and the sample with full data available have socioeconomic profiles very similar to the English population. Our dataset incorporating both assessments is therefore likely to be broadly generalisable to the population of children attending the universal review of 2.0–2.5 years in England.

### Weaknesses

Although uptake of the developmental review of 24–30 months in England is around 78%, no data were made available to the research team about the families failing to attend this review. It is likely that families with significant psychosocial difficulties would be over-represented among these non-attenders,<sup>34</sup> and this might have introduced some bias. Furthermore, as outlined previously, there was substantial attrition between the initial health visitor review and the research assessment carried out by the SLT. This level of attrition was probably unavoidable, given that the second research assessment was not part of routine care and involved an extra attendance by families. Finally, figure 1 illustrates that around 10% of the initial health visitor team reviews were associated with incomplete or inaccurate data recording—this is likely to reflect the pragmatic difficulties of evaluating development of a universal child health service innovation across five geographical areas with differing organisational configurations.



**Table 2** Screening performance of the SSLM score and the ASQ-3 communication score against the 10th population centile Preschool Language Scale V.5 cut-off

|   | Overall |               | English-only group |                | Not English-only group |               |
|---|---------|---------------|--------------------|----------------|------------------------|---------------|
|   | Value   | 95% CI        | Value              | 95% CI         | Value                  | 95% CI        |
| ASQ-3 communication score (optimal cut points)                                  | ≤37.5*  |               | ≤47.5*             |                | ≤57.5*                 |               |
| Sensitivity   | 0.55    | 0.44 to 0.66  | 0.85               | 0.7 to 0.94    | 0.88                   | 0.72 to 0.97  |
| Specificity   | 0.95    | 0.91 to 0.97  | 0.84               | 0.78 to 0.88   | 0.55                   | 0.36 to 0.72  |
| Positive likelihood ratio   | 9.97    | 5.87 to 16.92 | 5.20               | 3.72 to 7.26   | 1.93                   | 1.3 to 2.87   |
| Negative likelihood ratio   | 0.48    | 0.38 to 0.61  | 0.18               | 0.09 to 0.38   | 0.22                   | 0.08 to 0.59  |
| Positive predictive value   | 0.53    | 0.39 to 0.65  | 0.37               | 0.29 to 0.45   | 0.18                   | 0.13 to 0.24  |
| Negative predictive value   | 0.95    | 0.94 to 0.96  | 0.98               | 0.96 to 0.99   | 0.98                   | 0.94 to 0.99  |
| Accuracy  | 0.91    | 0.87 to 0.93  | 0.84               | 0.79 to 0.88   | 0.58                   | 0.45 to 0.70  |
| AUC   | 0.86    | 0.81 to 0.91  | 0.90               | 0.83 to 0.96   | 0.74                   | 0.62 to 0.86  |
| ASQ-3 communication category using standard cut point: above threshold†/monitor |         |               |                    |                |                        |               |
| Sensitivity   | 0.65    | 0.54 to 0.76  | 0.75               | 0.59 to 0.87   | 0.42                   | 0.26 to 0.61  |
| Specificity   | 0.93    | 0.89 to 0.96  | 0.94               | 0.9 to 0.97    | 0.88                   | 0.72 to 0.97  |
| Positive likelihood ratio   | 9.41    | 5.93 to 14.91 | 13.00              | 7.3 to 23.16   | 3.50                   | 1.29 to 9.52  |
| Negative likelihood ratio   | 0.37    | 0.28 to 0.5   | 0.27               | 0.15 to 0.45   | 0.66                   | 0.48 to 0.9   |
| Positive predictive value   | 0.51    | 0.4 to 0.62   | 0.59               | 0.45 to 0.72   | 0.28                   | 0.12 to 0.51  |
| Negative predictive value   | 0.96    | 0.95 to 0.97  | 0.97               | 0.95 to 0.98   | 0.93                   | 0.91 to 0.95  |
| Accuracy  | 0.90    | 0.87 to 0.93  | 0.92               | 0.88 to 0.95   | 0.83                   | 0.72 to 0.91  |
| ASQ-3 communication category using standard cut point: monitor/refer            |         |               |                    |                |                        |               |
| Sensitivity   | 0.48    | 0.37 to 0.59  | 0.48               | 0.32 to 0.64   | 0.30                   | 0.16 to 0.49  |
| Specificity   | 0.98    | 0.95 to 0.99  | 0.99               | 0.97 to 1      | 0.94                   | 0.80 to 0.99  |
| Positive likelihood ratio   | 21.67   | 9.52 to 49.32 | 49.40              | 11.97 to 203.8 | 5.00                   | 1.19 to 21.09 |
| Negative likelihood ratio   | 0.54    | 0.44 to 0.66  | 0.53               | 0.39 to 0.71   | 0.74                   | 0.58 to 0.94  |
| Positive predictive value   | 0.71    | 0.51 to 0.85  | 0.85               | 0.57 to 0.96   | 0.36                   | 0.12 to 0.70  |
| Negative predictive value   | 0.94    | 0.93 to 0.95  | 0.94               | 0.93 to 0.96   | 0.92                   | 0.91 to 0.94  |
| Accuracy  | 0.93    | 0.90 to 0.95  | 0.94               | 0.9 to 0.97    | 0.88                   | 0.77 to 0.94  |
| SSLM  | ≤19.5*  |               | ≤16.5*             |                | ≤20.0*                 |               |
| Sensitivity   | 0.83    | 0.74 to 0.91  | 0.80               | 0.64 to 0.91   | 0.79                   | 0.61 to 0.91  |
| Specificity   | 0.81    | 0.76 to 0.85  | 0.87               | 0.82 to 0.91   | 0.85                   | 0.68 to 0.95  |
| Positive likelihood ratio   | 4.38    | 3.36 to 5.69  | 6.16               | 4.2 to 9.05    | 5.20                   | 2.28 to 11.88 |
| Negative likelihood ratio   | 0.21    | 0.13 to 0.33  | 0.23               | 0.12 to 0.43   | 0.25                   | 0.13 to 0.49  |
| Positive predictive value   | 0.33    | 0.27 to 0.39  | 0.41               | 0.32 to 0.50   | 0.37                   | 0.20 to 0.57  |
| Negative predictive value   | 0.98    | 0.96 to 0.99  | 0.98               | 0.96 to 0.99   | 0.97                   | 0.95 to 0.99  |
| Accuracy  | 0.81    | 0.77 to 0.85  | 0.86               | 0.81 to 0.90   | 0.84                   | 0.73 to 0.92  |
| AUC   | 0.88    | 0.84 to 0.92  | 0.91               | 0.87 to 0.96   | 0.76                   | 0.64 to 0.89  |
| Cut-point Youden's Index*   |         |               |                    |                |                        |               |

\*Youden's Index represents the optimal trade-off between sensitivity and specificity.<sup>31</sup>

†Above threshold represents normal language function.

ASQ-3, Ages and Stages Questionnaire, Third Edition; AUC, area under the curve; SSLM, Sure Start Language Measure.

Relatively few children in the study (18%) were living in households where English was not spoken exclusively. The limited evidence we were able to gather suggests that both the ASQ-3 and SSLM performed more poorly in these groups. Further work is required with bilingual and non-English-speaking families, and in the meantime, we would recommend extreme

caution in the interpretation of language screening results from these children.

### Findings in context of the literature

Comparisons of developmental screening measures against a consistent reference test is rarely performed in low-risk populations<sup>20 23 27</sup> but is useful in selection

of effective instruments. Screening for poor language ability, although potentially valuable and very widely practised internationally,<sup>18</sup> does not yet meet accepted criteria for screening programmes,<sup>15</sup> and further work is required to demonstrate the effectiveness and cost-effectiveness of early interventions for screen-positive children. Our data suggest that the SSLM yields relatively robust results, and this could potentially be used for selecting participants for intervention in a trial of a comprehensive screening programme in the future.

## CONCLUSION

The SSLM appears to perform well as a language screening instrument for children aged 23–30 months, but further developmental work may be required to optimise its performance, particularly among families in which English is not used exclusively in the home.

**Acknowledgements** The authors would like to acknowledge other members of the “Early Identification” team: Professor Sue Roulstone and Caitlin Holme from the Bristol Speech and Language Therapy Research Unit, Rose Watson from Newcastle University, and Sheena Carr and Renvia Mason from Public Health England for their support throughout the project. Finally, we would like to thank the children and their families who attended the clinics and the health visitors and speech and language therapists who were involved in collecting data for the project.

**Contributors** PW drafted the manuscript. RR, JC and PW analysed the data. All authors contributed to the design of the study, edited the manuscript and approved the final submission. PW acts as guarantor for the paper.

**Funding** This work was funded by the Department of Education and managed by Public Health England and entitled Assessment Tool and Resources to Support Action by Health Visitors and Early Years Practitioners to Identify and Support Children with Early Speech, Language and Communication Needs ECM\_6378.

**Competing interests** The Sure Start Language Measure was originally developed for the Sure Start Programme in England by the late project chief investigator Professor James Law.

**Patient consent for publication** Not applicable.

**Ethics approval** This study involves human participants and obtained a favourable ethical opinion and approval from West Midlands/Black Country Research Ethics Committee (reference 19/WM/0114) in May 2019. Participants gave informed consent to participate in the study before taking part.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open access repository. The data from this project, together with other data from the study, will be available in the Newcastle University data repository. The protocol (initial accepted response to tender), the final report, the deidentified participant data and a data dictionary will be made available with publication from Newcastle University - My data (ncl.ac.uk). The reports are available currently. The data will be available by March 2022. Data access will be subject to a data access agreement but beyond that will be fully accessible. Those wishing to access the data will be given support from the investigator team if required.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is

properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## ORCID iD

Philip Wilson <http://orcid.org/0000-0002-4123-8248>

## REFERENCES

- 1 Pinker S. Colloquium paper: the cognitive niche: coevolution of intelligence, sociality, and language. *Proc Natl Acad Sci U S A* 2010;107 Suppl 2:8993–9.
- 2 Hart B. *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes Publishing, 1995.
- 3 Law J, Rush R, Schoon I, et al. Modeling developmental language difficulties from school entry into adulthood: literacy, mental health, and employment outcomes. *J Speech Lang Hear Res* 2009;52:1401–16.
- 4 Miniscalco C, Nygren G, Hagberg B, et al. Neuropsychiatric and neurodevelopmental outcome of children at age 6 and 7 years who screened positive for language problems at 30 months. *Dev Med Child Neurol* 2006;48:361–6.
- 5 Sim F, Haig C, O'Dowd J, et al. Development of a triage tool for neurodevelopmental risk in children aged 30 months. *Res Dev Disabil* 2015;45-46:69–82.
- 6 Duncan GJ, Dowsett CJ, Claessens A, et al. School readiness and later achievement. *Dev Psychol* 2007;43:1428–46.
- 7 Ripley K, Yuill N. Patterns of language impairment and behaviour in boys excluded from school. *Br J Educ Psychol* 2005;75:37–50.
- 8 Schoon I, Parsons S, Rush R, et al. Childhood language skills and adult literacy: a 29-year follow-up study. *Pediatrics* 2010;125:e459–66.
- 9 Snow PC, MB: P. Youth (In)justice: Oral language competence in early life and risk for engagement in antisocial behaviour in adolescence. *Trends Issues Crime Crim Justice* 2011;435.
- 10 Armstrong R, Scott JG, Whitehouse AJO, et al. Late talkers and later language outcomes: predicting the different language trajectories. *Int J Speech Lang Pathol* 2017;19:237–50.
- 11 McKean C, Mensah FK, Eadie P, et al. Levers for language growth: characteristics and predictors of language trajectories between 4 and 7 years. *PLoS One* 2015;10:e0134251.
- 12 McKean C, Reilly S, Bavin EL, et al. Language outcomes at 7 years: early predictors and co-occurring difficulties. *Pediatrics* 2017;139:e20161684.
- 13 Wilson JMG, Jungner G. *Principles and practice of screening for disease*. 34. Geneva: Public health papers, 1968.
- 14 Law J, Boyle J, Harris F, et al. Screening for speech and language delay: a systematic review of the literature. *Health Technol Assess* 1998;2:1–184.
- 15 Law J, Garrett Z, Nye C, et al. Speech and language therapy interventions for children with primary speech and language delay or disorder. *Cochrane Database Syst Rev* 2003;2015:CD004110.
- 16 Law J, Dockrell JE, Castelnuovo E, et al. Early years centres for pre-school children with primary language difficulties: what do they cost, and are they cost-effective? *Int J Lang Commun Disord* 2006;41:67–81.
- 17 Law J, Zeng B, Lindsay G, et al. Cost-effectiveness of interventions for children with speech, language and communication needs (SLCN): a review using the Drummond and Jefferson (1996) 'Referee's Checklist'. *Int J Lang Commun Disord* 2012;47:1–10.
- 18 Wilson P, Wood R, Lykke K, et al. 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. *Scand J Public Health* 2018;46:805–16.
- 19 van Agt HME, van der Stege HA, de Ridder-Sluiters H, et al. A cluster-randomized trial of screening for language delay in toddlers: effects on school performance and language development at age 8. *Pediatrics* 2007;120:1317–25.
- 20 Berkman ND, Wallace I, Watson L. *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 [Internet]. Evidence Syntheses, No. 120.* 2. Methods. Rockville (MD): Agency for Healthcare Research and Quality (US), 2015. <https://www.ncbi.nlm.nih.gov/books/NBK305677/>
- 21 Wallace IF, Berkman ND, Watson LR, et al. Screening for speech and language delay in children 5 years old and younger: a systematic review. *Pediatrics* 2015;136:e448–62.
- 22 Siu AL, US Preventive Services Task Force. Screening for speech and language delay and disorders in children aged 5 years or



- younger: US preventive services Task force recommendation statement. *Pediatrics* 2015;136:e474–81.
- 23 Bedford H, Walton S, Ahn J. *Measures of child development: a review*. London: London Policy Research Unit in the Health of Children, Young People and Families, 2013.
- 24 Marks KP, Madsen Sjö N, Wilson P. Comparative use of the ages and stages questionnaires in the USA and Scandinavia: a systematic review. *Dev Med Child Neurol* 2019;61:419–30.
- 25 Squires J, Bricker D, Twombly E. *ASQ-3 User's Guide*. Baltimore: Paul H. Brookes Publishing Co, 2009.
- 26 Harris F, Law J, Roy P. The development of a 50 word adaptation of the UK short form of the MCDI for use with two years olds in sure start programmes in England, 2004. Available: <https://myimpact.ncl.ac.uk/UploadFiles/273422/db164fd5-59ac-4701-862b-26905dbe78b3.pdf> [Accessed 11 Dec 2021].
- 27 Veldhuizen S, Clinton J, Rodriguez C, et al. Concurrent validity of the ages and stages questionnaires and Bayley developmental scales in a general population sample. *Acad Pediatr* 2015;15:231–7.
- 28 Zimmerman IL, Pond RE, Steiner VG. *Preschool Language Scale - Fifth Edition (PLS-5 UK)*. London: Pearson Assessment, 2014.
- 29 Law J, Charlton J, McKean C. *Identifying and supporting children's early language needs*. London: Public Health England, 2020. <https://myimpact.ncl.ac.uk/UploadFiles/270761/7d1567da-dbc1-460b-94ac-53f2e62f42f7.pdf>
- 30 Norbury CF, Gooch D, Wray C, et al. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *J Child Psychol Psychiatry* 2016;57:1247–57.
- 31 Youden WJ. Index for rating diagnostic tests. *Cancer* 1950;3:32–5.
- 32 Reilly S, Cook F, Bavin EL, et al. Cohort profile: the early language in Victoria study (ELVS). *Int J Epidemiol* 2018;47:11–20.
- 33 Reilly S, Wake M, Bavin EL, et al. Predicting language at 2 years of age: a prospective community study. *Pediatrics* 2007;120:e1441–9.
- 34 Ellis DA, McQueenie R, McConnachie A, et al. Demographic and practice factors predicting repeated non-attendance in primary care: a national retrospective cohort analysis. *Lancet Public Health* 2017;2:e551–9.