
Article:
Schaefer, B., Bowyer-Crane, C., Herrmann, F. et al. (1 more author) (2015) Development of a tablet application for the screening of receptive vocabulary skills in multilingual children: A pilot study. Child Language Teaching and Therapy. pp. 1-13. ISSN 0265-6590

https://doi.org/10.1177/0265659015591634

Title: Development of a tablet application for the screening of receptive vocabulary skills in multilingual children – A pilot study

Dr Blanca Schaefer[1,3]

Dr Claudine Bowyer-Crane[2]

Dr Frank Herrmann[3]

Dr Silke Fricke[1]

[1]University of Sheffield, [2]University of York, [3]University of Chester

Corresponding author:

Dr Blanca Schaefer

University of Chester

Department of English

Parkgate Road

Chester

CH1 4BJ

Telephone: 07887395776

b.schaefer@chester.ac.uk

Published Article: Schaefer, B., Bowyer-Crane, C., Herrmann, F., & Fricke, S. (2015). Development of a tablet application for the screening of receptive vocabulary skills in multilingual children: A pilot study. *Child Language Teaching and Therapy*, first published online June 25, 2015. doi: 10.1177/0265659015591634

Abstract

For professionals working with multi-lingual children, detecting language deficits in a child's home language can present a challenge. This is largely due to the scarcity of standardised assessments in many children's home languages and missing normative data on multilingual language acquisition. A common approach is to translate existing English language vocabulary measures into other languages. However, this approach does not take into account the cultural and linguistic differences between languages. This pilot study explored whether English and home language receptive vocabulary skills can be objectively and reliably screened using a tablet application. Preliminary data on mono- and multilingual vocabulary skills was collected from 139 children aged 6-7 years. A tablet application was designed to assess children's receptive vocabulary in both English, and an additional eight languages using a four choice picture paradigm. Linguistically controlled and pre-recorded target items are presented orally via the tablet in each language and responses are made via the touchscreen and automatically scored. The English version of the test was administered to 67 mono- and 72 multilingual children, while 38 multilingual children also completed the test in their home language. Test criteria measures, including reliability and concurrent validity showed satisfactory results. These findings suggest that the tablet application could be a useful tool for professionals to screen receptive vocabulary skills in mono- and multilingual children. Limitations of the first version of the RVS and future steps are discussed.

Keywords: Multilingualism, vocabulary acquisition, tablet application, assessment, children, screening

INTRODUCTION

Across the UK and Europe, a high percentage of children are growing up bi- or multilingual, and speak a different language at home than the majority language spoken in their country of residence. In the following paper the term multilingual is used to acknowledge that many children may speak or are exposed to more than two languages. In England, 18.7% of primary school pupils are learning English as an additional language (DfE, 2014) and need support at an early stage in order to reach adequate levels of proficiency in the language of instruction (Tickell, 2011; Safford and Collins, 2012). In order to evaluate language proficiency, children need to be assessed in all of the languages they learn (Stow and Pert, 2015). However, standardised assessments are not readily available for all of the children's home languages. Although translated assessments are often used, these do not always consider cultural and linguistic differences between languages (e.g., Mueller Gathercole, 2013; Teoh et al., 2012). Moreover, the professionals administering language tests are not necessarily proficient in the children's home language. Thus it is difficult for speech and language therapists, teachers, and other educational/health professionals to identify multilingual children who have spoken language deficits as opposed to difficulties learning a new language. This distinction is crucial for identifying children's individual learning needs, for deciding which services will fund the required support and, as Stow and Dodd (2005) highlight, to ensure multilingual children receive equal support compared to their monolingual peers. This paper reports a pilot study investigating the development of a new receptive vocabulary screener designed to assess children's vocabulary knowledge in a range of languages in addition to English.

Considerations for language assessments

There is an ongoing discussion about how to best assess children's language abilities e.g. through standardised assessments, informal procedures or dynamic assessments; all of which contribute to the overall profile of a child's language skills. While standardised assessments provide norms with which to compare a child's performance to their age group, indicate specific deficits and track language development over time (e.g., Edwards et al., 2011), dynamic assessment allows a more in-depth exploration of language competence and cognitive strategies (e.g., Hasson and Dodd, 2014; Hasson et al., 2013). Informal explorations are more natural and might reveal more authentic language behaviour, including coping mechanisms children may use in every-day communication (e.g., King et al., 2014; Archibald et al., 2011). However, these assessments can be incredibly time-consuming. In contrast to these comprehensive assessments, screening tools allow practitioners to get a quick snapshot of children's language skills. While all assessment types have their benefits, screening tools are time-efficient and can help identify children who need more in-depth testing and may require language support. Like all assessments, screening tools should be grounded in a theoretical framework and contain carefully chosen and controlled test items (American Educational Research Association, 2014). An assessment with an unclear rationale or poorly controlled test items will yield unreliable results and may hinder the planning and allocation of intervention resources. The administration of screening tools must be transparent and easy enough to ensure that they can be used in different settings, with different children, and used by testers with different (professional) backgrounds (e.g. teachers, psychologists).

Importance of assessing vocabulary

Vocabulary forms the foundation for a range of language skills. A critical number of words in a child's lexicon is needed to enable grammar development and to trigger the differentiation

between different word forms (Edwards et al., 2011; Metsala and Walley, 1998; Edwards et al., 2004). A rich inventory and in depth knowledge of words, including words of different types (e.g. nouns, verbs, adjectives), helps children to understand word meanings when encountered in different contexts and grammatical structures. This enables them to comprehend language and form complex sentences in order to express their thoughts and ideas in spoken and written language. Vocabulary is also a very strong predictor of educational attainment. For example, studies across different languages have shown that reading comprehension is significantly influenced by vocabulary skills (Tong et al., 2012; Lervåg and Aukrust, 2010; Verhoeven and Van Leeuwe, 2008). Consequently, vocabulary skills are an important part of language acquisition making vocabulary assessment tools vital to identify children at risk for language difficulties as early as possible.

Assessing vocabulary skills in languages other than English

Norms

The challenges in language assessment are particularly evident when assessing language skills of children with a multilingual background (Williams et al., 2014; Pena and Hall, 2011; Van deVijver and Tanzer, 2004). For example, comparing a multilingual child's language performance on a test battery with standardised norms is not valid if the sample was only based on monolingual English-speaking children. Even if the normative sample comprises a mix of monolingual and multilingual children whose first language is English, the norms cannot be used to evaluate a child's language skills if their first language was not English. A repository of multilingual vocabulary acquisition data would help to understand which skills may be typical for different language combinations at different ages.

C*ultural appropriateness*

Assessments need to be culturally appropriate and avoid cross-cultural bias (Van de Vijver and Hambleton, 1996; Laing and Kamhi, 2003). Van de Vijver and Tanzer (2004) differentiate between three potential biases: construct, method, and item bias.

Firstly, construct bias occurs when the construct tested is not identical across different cultures (van der Vijver and Tanzer, 2004: 120). Life experience and communication styles within linguistic communities may influence how language is used (e.g., Simmons and Johnston, 2007), including word use. For example, the use of diminutives in different languages varies considerably. While English speakers do not use them frequently and only in relatively restricted contexts, speakers of Spanish, Russian or Greek show a more extensive and flexible use (e.g., King and Melzi, 2004).

Secondly, method bias includes three components, i.e. sample, instrument, and administration bias. Sample bias results when groups differ significantly in, for example, their cultural and educational background. Instrument bias occurs when assessment tools are used which are less familiar for a subgroup of participants. For example, for some children engaging with print material and looking collaboratively at pictures and books with adults might be a less familiar setting than for other children (e.g., van Steensel, 2006). Different cultural backgrounds and first languages can easily cause communication problems, including misunderstandings about how to complete a task, which can result in administration bias.

Thirdly, item bias can be caused by varying familiarity with stimulus items and their depiction. Word frequency will vary depending on socialisation practices. The frequency of words is influenced by how often the named object actually occurs in peoples' daily lives. For example, a teddy bear might be a frequent toy in Western cultures but might be rare in other cultures. Additionally there is the question of prototypes. Testing nouns in different languages requires

finding words that share a critical number of semantic features to be comparable. For example, a prototypical representation of a boat is difficult to find, because boats are used in a wide range of contexts (e.g. for fishing/leisure/travelling, on the sea/a lake/river) and in different styles, sizes, materials etc. The environment and people's experience play an important role as well. For example, the word *tree* may conjure up different prototypical pictures depending on the flora which surrounds a child. Therefore, it is a challenge to find a range of nouns which share sufficient semantic features to be tested across different languages. A further aspect which can cause item bias is the age of acquisition (AoA), i.e. the age at which a word is learnt (Kuperman et al., 2012). Factors such as those described above, including familiarity with stimulus items and word use, impact considerably on the age at which a word is acquired. In particular frequency and imageability have been identified as important predictors of early word learning (Ramey et al., 2013). Different environments, in particular the input from parents/carers, may lead to words being acquired at different ages (Goodman et al., 2008). Gender differences may play a role as well; boys may engage with some objects more than girls, resulting in higher familiarity and this finding has been reported in both adults (Laws, 2003) and children (Barbarotto et al., 2008).

The use of new technologies to assess child language

Recent advances in technology have led to new ways of assessing children with language difficulties. There is now a range of digital tools and programmes which support language testing, but many of them still require desktops or laptops. However, the number of apps for tablets and smartphones which allow monitoring, storing and analysing of, for example, health related data is growing. Therefore, the question is how technical devices can be used most effectively in different environments and employed by different end users to collect

meaningful data on child language. User-friendliness is one key element to ensure test objectivity and consistent data collection. Considering sparse resources and time constraints in clinical and educational settings the tool must be easy to administer and results need to be transparent and effectively summarised for the user. One advantage of tablets might be that they are engaging and popular amongst children which increases their motivation and compliance. Tablets might help to overcome cultural differences regarding test settings and stimulus presentation (i.e. avoiding method bias) since the use of tablets or other touch screen devices has been introduced across the world and is a familiar tool for young children (Chiong and Shuler, 2010; Geist, 2014; Geist, 2012). Moreover, an increasing number of schools use tablets in teaching activities and, therefore, have this resource available.

Bringing all aspects together the development of a high quality language screener for use in nursery and school to test mono- and multilingual children is much needed. The current pilot study aimed to address this need for a screening tool which is soundly based on theoretical, linguistic foundations, and which considers cultural aspects. It was designed to be administered by different professional user groups, i.e. SLTs, teachers, teaching assistants (independent of their language background and skills), and to test different groups of mono- and multilingual children. Therefore, the main research question was whether receptive vocabulary skills across different languages could be objectively and reliably screened, considering cross-linguistic and cross-cultural aspects using a newly developed tablet application.

METHOD

Participants

A total of 139 children (67 mono- and 72 multilingual children) with weak language skills based on a composite score, including raw scores from the Early Repetition Battery (nonwords; Seeff-Gabriel et al., 2008) and the expressive vocabulary and sentence structure subtests from the Child Evaluation of Language Fundamentals (CELF) – Preschool UK, (Semel et al., 2006) took part in the study. Children were selected from ten schools, the majority of which were in low socioeconomic areas (based on the schools' *Indices of Deprivation*, see Neighbourhood Statistics, 2010). All children were participating in a randomised controlled trial (RCT) evaluating the effectiveness of an oral language intervention for mono- and multilingual children and were assessed at several time points (for more details see Schaefer et al., in preparation). The data presented in this paper were collected in the Autumn term when children entered Year 1. They were aged between 6;03-7;04 years at the time of testing.

The English version of the *Receptive Vocabulary Screener* application (RVS) was completed by all mono- and multilingual children. Thirty-eight multilingual children completed the RVS in both their home language (Czech=6, Polish=2, Punjabi=12, Urdu=18) and in English. The children's home languages were identified based on school records and parental questionnaires. Considering the composition of the group of multilingual children two subgroups could be differentiated, i.e. a small group of children speaking an Eastern European language (Czech or Polish, n=8) and a group of children speaking an Indo-Iranian language (Urdu or Punjabi, n=30).

Material

The Receptive Vocabulary Screener (RVS)

The RVS was developed as a tablet application designed to facilitate the screening of mono- and multilingual children's receptive vocabulary in both English and their home language(s),

but which does not require the administrators to be proficient in those language(s). The test runs as a four choice picture paradigm in which children hear a word and select the appropriate picture. Responses are made directly via the tablet's touchscreen and are automatically recorded and scored. Core information about the child can be entered on the start screen (including e.g. name (or subject identifier), date of birth, languages the child speaks). A character called Meemo guides the children through the task using their home language. Meemo introduces the test, provides instructions at the beginning, guides the children through the first two practice items and provides encouragement throughout the testing (i.e. appearing on the screen, saying e.g. *Well done*). The scored responses are automatically exported to an Excel spreadsheet. The results indicate whether the child identified an item correctly and in the case of incorrect responses, the item selected by the child is recorded which allows for error analysis.

Item selection occurred in a series of stages. As a first step 63 verbs and 74 nouns from the Kuperman and colleagues' list of English word frequencies (2012) with an age of acquisition (AoA) between 3;06 and 6;00 were chosen. Only concrete nouns and verbs that could be simply presented visually and which were thought to be culturally unbiased were selected. Test items and pictures were reviewed by different professionals working with children from diverse cultural and linguistic backgrounds (including teaching assistants, SLTs, academic colleagues with a special interest in child language acquisition) to ensure they were culturally appropriate. The following languages were included in the first trial version of the RVS: English, Welsh, Urdu, Punjabi, Polish, Slovak, Czech, Portuguese, and Mandarin Chinese. An online questionnaire for native speakers was set up to select the final test items. Volunteers who are native speakers of one of the chosen languages and fluent in English were asked to

comment on whether the words existed in their language and had only one direct translation to avoid ambiguity. Additionally, they were asked to write down the translation of the word in their home language (for languages with non-Latin scripts, i.e. Punjabi, Urdu, Mandarin Chinese, translators were asked to write down the word the way they would pronounce it). They were also asked to comment on when they thought the word was acquired in their home language, selecting from a choice of eight categories (younger than three years; 3;00-3;05; 3;06-3;11/4;00-4;05/4;06-4;11/5;00-5;05/5;06-5;11; older than six years). As a reference point they were provided with the AoA for the equivalent English word. In total 32 volunteers completed the noun questionnaire and 20 completed the verb questionnaire. In a next step, based on the volunteers' answers, the research team selected 22 nouns and 22 verbs for which a direct translation existed (one exception: *bending* does not have a direct translation in Mandarin Chinese) and for which the AoA in English did not generally differ more than one year in comparison to the AoA in the other languages. Fifteen adults, two for each language (one female, one male, exception: Welsh) were recruited to translate and record the test items, instructions, and motivational phrases (e.g. *well done*) which were needed for the app to run in different languages. All audio files were recorded in a sound-attenuating booth by the translators (female and male) and edited by an acoustic phonetician to ensure high sound quality and consistent "loudness" across all items and speakers.

Distractor items for both nouns and verbs were selected to allow a four choice picture paradigm, i.e. children are asked to choose the correct item from a choice of four. Distractors for the noun items were categorised as follows:

1. Categorical distractor, i.e. same semantic category as target (e.g. target: *book*; categorical distractor: *newspaper*).

2. Meronymic or functional distractor, i.e. part of target (e.g. target: *monkey*; meronymic distractor: *tail*) or related to use and function of target (e.g. target: *egg*; functional distractor: *pan*).
3. Random distractor, i.e. not related to target or other distractors.

All distractors were matched to fit the test items' English AoA, i.e. being acquired between one year before or after the test item. Two practice items are introduced to familiarise the child with the test design. For the noun subtest the test items have an AoA between 3.7-6.0. Examples of noun test items and their distractors including AoA are shown in Table 1.

Please insert Table 1 about here

Distractors for the verb items were chosen using semantic categorisations according to Levin (1993), considering semantic and syntactic properties of English verbs:

1. Distractor of same specific verb class (e.g. target: *running* from overall verb class *verbs of motion* and specific subcategory *run verbs*; distractor item: *rolling* from same specific subcategory).
2. Distractor of same overall verb class (e.g. target: *laughing* from overall verb class *verbs involving the body*; distractor item: *sleeping*, from same overall verb class).
3. Distractor of different verb class (e.g. target: *knocking* from overall verb class *verbs of contact by impact*; distractor item: *hiding* from overall verb class *verbs of concealment*).

Examples of verb test items and their distractors including AoA are shown in Table 2. The AoA for the verbs range between 3.5 and 5.6 years.

Please insert Table 2 about here

Coloured illustrations were produced by an artist to depict all noun and verb items and their distractors to ensure consistent format and style.

Procedure

Trained research assistants tested all children individually in their school settings. The first pilot data with the app presented here was collected for the noun items as part of the RCT project. This meant that the testing period as well as the age of the children were determined by the ongoing project and it was not possible to select participants of a younger age group or to administer the verb section of the app. All children completed the English version of the RVS while a subsample of multilingual children were tested a second time in their home language. Of the nine languages included in the screening tool we were able to test five as part of this pilot study.

In addition, all children completed two standardised language tests; the British Picture Vocabulary Scales (BPVS, 3rd edition, Dunn et al., 2012) and the CELF Expressive Vocabulary subtest (Semel et al., 2006). In the BPVS children are presented with a word orally and are asked to identify the corresponding picture from a choice of four. The BPVS is normed for children between the ages of 3 and 16. The CELF Expressive Vocabulary subtest asks children to name different pictures, eliciting nouns and verbs and the test is normed for children aged 3;00 to 6;11.

RESULTS

The testers reported that the RVS was easy to administer and that children independent of their language background were engaged and motivated while using the app. Table 3 provides an overview of the raw scores for both groups (mono- and multilingual children) on the RVS, CELF, and BPVS.

Please insert Table 3 about here

Monolingual children achieved on average higher scores on the BPVS and CELF than the multilingual children and group comparisons (Mann-Whitney-U Tests) revealed that this trend was significant on both of the standardised assessments, the BPVS (U = 1739.500, z = -2.84, p < 0.005, r = 0.24) and the CELF Expressive Vocabulary (U = 1671.500, z = -3.13, p < 0.002, r = 0.27). However, both groups showed considerable variability in their performance. There were no ceiling effects on the RVS, i.e. 50% or more of the children did not score 19 or 20 out of 20. However, mean performance was relatively high. The range of scores for the *RVS home language* was larger than for the *RVS English*, suggesting a greater degree of variability in test scores. In contrast to the standardised assessments, there was no significant difference between the mono- and multilingual children in their performance on the *RVS English* (Mann-Whitney-U Test, U = 2373.000, z = -0.17, p = 0.868, r = 0.01). Multilingual children performed better on the English version in comparison to their home language version (Wilcoxon Signed Ranks Test, z = -2.580, p < 0.010, r = 0.42). As regards the different subgroups of multilingual children, the children with an Eastern European home language scored significantly lower on the *RVS English* than the children with an Indo-Iranian home language (U = 59.000, z = -2.22, p < 0.026, r = 0.36). However, there was no significant group difference in the *RVS home language* (U = 118.000, z = -0.072, p = 0.942, r = 0.01).

A comparison of different types of distractors for the whole group (Friedman's ANOVA, $\chi^2(2)$ = 44.030, p < 0.001) showed that categorical or meronymic/functional distractors were significantly more often chosen than random distractors (Wilcoxon Signed Ranks Tests, z = -4.43, p < 0.001, r = 0.38 and z = -5.80, p < 0.001, r = 0.49 respectively), even after applying Bonferroni corrections. There was no significant difference comparing the frequency of

categorical and meronymic/functional distractors (Wilcoxon Signed Ranks Test, z = -1.28, p = 0.200, r = 0.11). The same pattern was confirmed for the group of mono- and multilingual children separately.

As a measure of internal consistency (scale reliability) standardised Cronbach's alpha was computed. For the *RVS English* (based on data from both mono- and multilingual children) the value was 0.60. The standardised Cronbach's alpha for the *RVS home language* (based on the data from the multilingual children) was 0.88. As a measure of concurrent validity nonparametric correlations (Spearman's rho) were computed between the RVS and the two standardised vocabulary measures BPVS and CELF Expressive Vocabulary. Significant and weak to moderate correlations were found for the monolingual children between the English version of the RVS and both standardised vocabulary measures (see Table 4). The correlation between the native language version of the RVS and the CELF was non-significant and weak while the correlation with the BPVS was moderate and statistically significant. Moreover, for the multilingual children the English version of the RVS showed significant and strong correlations with the two standardised vocabulary measures. Combining both language groups revealed moderate and highly significant correlations between the English version of the RVS, CELF, and BPVS.

Please insert Table 4 about here

DISCUSSION

The Receptive Vocabulary Screener (RVS) tablet application was developed to provide an efficient tool to screen multilingual children's receptive vocabulary in both English and their home language(s). Uniquely it allows the assessment of the same test items in different languages without requiring the administrators to be proficient in those language(s). It uses

a four choice picture paradigm in which target words (20 nouns and 20 verbs plus two practice items each) are pre-recorded and presented in the appropriate language(s). Items were carefully chosen with the help of native speakers to avoid item bias (Van de Vijver and Tanzer, 2004). Moreover, items and their distractors were linguistically controlled by systematically considering semantic-lexical criteria and developmental aspects (i.e. controlling for the English Age of Acquisition (AoA) of all test and distractor items). However, future revisions of the RVS need to include the extension of AoA values for the non-English test and distractor items. This will need the support from a larger group of translators and a more systematic establishment of translations and AoA values (e.g. calculating inter-rater reliability). Based on this extended data the test items/distractors may need amendment to improve test reliability and validity.

Test assistants reported back that the app was easy to use. One reason might be that responses are made via a touchscreen and are automatically recorded and scored. Those functions allow a simple, efficient and transparent assessment which considerably reduces instrument and administration bias (see Van de Vijver and Tanzer, 2004). All children were happy to engage with the tablet which supports the assumption that technical devices such as tablets increase children's motivation (Chiong and Shuler, 2010; Geist, 2014; Geist, 2012). Furthermore, all of them seemed equally familiar with the tablets, their use and the presented instruction to identify pictures on the screen. This underpins the view that touch screen devices are a familiar tool for young children nowadays, that children are able to use them independently or with little help and that therefore tablets may help balancing cultural differences and reducing the risk of method bias (Chiong and Shuler, 2010).

The RVS was piloted on five languages, testing 139 children with the noun version of the tool. The choice of distractors seemed to be appropriate, showing a preference for the distractors closer to the target items than the random distractor. Given that this pattern was confirmed for both language groups, it can be assumed that there was no item bias. Hence, the item depiction seems appropriate and both groups of children seem to be familiar with the test items and distractors.

The participants had a mean age of 6;06 which was rather old given the age range the RVS was developed for. However, since the pilot study was part of a larger RCT-study it was not possible to assess younger children at the time of testing. Thus, overall test performance was rather high, producing lower variability than one could expect in a younger age group. There were no ceiling effects, however.

Statistically significant correlations between the RVS and the standardised vocabulary assessments were moderate when both groups were combined providing initial evidence that the RVS is a valid tool to assess receptive vocabulary skills in mono- and multilingual children. Overall higher correlations were found between the RVS and the BPVS which was expected considering that both tools assess receptive vocabulary with a four choice paradigm.

Standardised Cronbach's alpha, a measurement of internal consistency which is considered to reflect scale reliability, revealed a coefficient of 0.60 for the English version of the RVS. Given that reliabilities above 0.70 are generally desired if a test is to be used as a research tool, while reliabilities above 0.90 are sought-after for diagnostic and job selection purposes (e.g., Hammond, 2006), this reliability value seems relatively low. However, a minimum requirement of 0.55 is also often cited as appropriate for assessments administered in experimental group studies (e.g., Rost, 2007) and a value around 0.6 as acceptable for a newly developed measure (e.g., Nunnally, 1988). Since this was the first pilot study using the app,

including only noun items and testing children at the top of the target age range, it seems reasonable to assume that better reliability values could be achieved by testing three to six year old children in future studies followed by further improvements and developments of the test items based on a resulting larger sample. In addition, the value for the RVS home language was excellent, underpinning the suitability for multilingual children. It should also be mentioned that, although Cronbach's alpha is the most common estimate of reliability, there are other measures which can be computed. Since repeatability and stability across time are important aspects of screening items, test-retest- reliability should be calculated in future studies as well. Larger studies with more representative samples would also allow to run factorial components analyses to explore in-depth the relationship between test items.

It is important to note that due to time constraints the RVS was administered twice in one session with the multilingual children. In a follow up study the English and home language version should be administered in separate sessions and the items should be presented in randomised order. Moreover, the sample of multilingual children was biased since the group was small and had an unusual composition (East European and Indo-Iranian languages: Polish/Czech and Urdu/Punjabi). That children with an Indo-Iranian languages background (i.e. Urdu/Punjabi speaking) outperformed the Czech- and Polish speaking children on the English but not the home language version of the RVS may suggest that they form a specific subgroup of multilingual children. Issues including language exposure, language identity, and length of residence in the UK may play a role (MacLeod et al., 2013; Thordardottir, 2011). For example, parents may state on school records that their children speak Urdu, the official language of Pakistan, which also exists in written form. However, the family's home language may be another Indo-Iranian language such as Mirpuri or a mixture of Urdu, Punjabi and

Mirpuri (for a more in depth discussion see Stow and Dodd, 2003, 2005). This shows the need to differentiate between groups of multilingual children and to investigate children's language background in more detail to avoid sample bias. In following studies a larger range of languages should be tested and the composition of groups of children should be monitored.

IMPLICATIONS AND FUTURE STEPS

In sum, preliminary data suggests that the RVS could be a useful tool to screen receptive vocabulary skills in mono- and multilingual children. The informal positive feedback from the testers allows the assumption that the RVS may be a suitable tool for different professionals, including Speech and Language Therapists, teachers, and teaching assistants, and which can be used in clinical and educational settings. However, screening results must be treated with caution, because they only provide an indication of potential deficits. It is not presumed that this screening too which is only a starting point to assess lexical skills can replace a close collaboration with an interpreter within either of these settings. Future work will include an optimisation of the application (improving user friendliness, measure reaction time of responses, further checking linguistic properties of the test and distractor items), extending and adapting it for additional languages. The aim is to run further pilot studies with different age groups, with children from different socio-economic backgrounds and more groups of multilingual children, including the verb section of the app. Moreover, it is intended to present the app as a web-based application and link it to cloud computing for more extended data collection, which would facilitate building a repository of anonymised receptive vocabulary data.

Acknowledgements

We would like to thank Michael Crane for developing the first prototype of the tablet application, Louise Ainsworth for drawing the pictures for all test and distractor items, all volunteers and translators for helping us to choose and translate the test items, and all children who participated in the pilot study.

Conflict of Interest Statement

The Authors declare that there is no conflict of interest.

REFERENCES

Archibald LM, Joanisse M and Edmunds A (2011) Specific language or working memory impairments: A small scale observational study. *Child Language Teaching and Therapy* 27(3): 294-312.

American Educational Research Association (2014) Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Barbarotto R, Laiacona M and Capitani E (2008) Does sex influence the age of acquisition of common names? A contrast of different semantic categories. *Cortex* 44(9): 1161.

Chiong C and Shuler C (2010) Learning: Is there an app for that? Investigations of young children's usage and learning with mobile devices and apps. New York: The Joan Ganz Cooney Center at Sesame Workshop.

Dunn LM, Dunn D, Styles B, et al. (2012) *The British Picture Vocabulary Scales.* London: GL Assessment.

Department of Education (DfE) (2014) Statistical First Release. Department for Education.

Edwards J, Beckman ME and Munson B (2004) The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research* 47(2): 421-436.

Edwards J, Munson B and Beckman ME (2011) Lexicon-phonology relationships and dynamics of early language development - A commentary on Stoel-Gammon's Relationships between lexical and phonological development in young children. *Journal of Child Language* 38(1): 35-40.

Edwards S, Letts C and Sinka I (2011) *The New Reynell Developmental Scales.* London: GL Assessment.

Geist E (2014) Using tablet computers with toddlers and young preschoolers. *YC Young Children* 69(1): 58-63.

Geist EA (2012) A qualitative examination of two year-olds interaction with tablet based interactive technology. *Journal of Instructional Psychology* 39(1): 26-35.

Goodman JC, Dale PS and Li P (2008) Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language* 35(3): 515-531.

Hammond SM. (2006) *Research Methods in Psychology.* London: Sage.

Hasson N, Camilleri B, Jones C, et al. (2013) Discriminating disorder from difference using dynamic assessment with bilingual children. *Child Language Teaching and Therapy* 29(1): 57-75.

Hasson N and Dodd B (2014) Planning intervention using dynamic assessments: A case study. *Child Language Teaching and Therapy* 30(3): 353-366.

King AM, Thomeczek M, Voreis G, et al. (2014) iPad® use in children and young adults with Autism Spectrum Disorder: An observational study. *Child Language Teaching and Therapy* 30(2): 159-173.

King K and Melzi G (2004) Intimacy, imitation and language learning: Spanish diminutives in mother-child conversation. *First Language* 24(2): 241-261.

Kuperman V, Stadthagen-Gonzalez H and Brysbaert M (2012) Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods* 44(4): 978-990.

Laing SP and Kamhi A (2003) Alternative assessment of language and literacy in culturally and linguistically diverse populations. *Language, Speech, and Hearing Services in Schools* 34(1): 44-55.

Laws KR (2003) Sex differences in lexical size across semantic categories. Personality and Individual Differences. *Cortex* 36(1): 23–32.

Lervåg A and Aukrust VG (2010) Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of Child Psychology and Psychiatry* 51(5): 612-620.

Levin B (1993) *English Verb Classes and Alternations - A Preliminary Investigation.* Chicago: The University of Chicago Press.

MacLeod AAN, Fabiano-Smith L, Boegner-Pag S, et al. (2013) Simultaneous bilingual language acquisition: The role of parental input on receptive vocabulary development. *Child Language Teaching and Therapy* 29(1): 131-142.

Metsala JL and Walley AC (1998) Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In: Metsala JL and Ehri LC (eds) *Word Recognition in Beginning Literacy.* Mahwah, New Jersey: Erlbaum, 89–120.

Mueller Gathercole VC (Ed.). (2013) *Solutions for the assessment of bilinguals*. Bristol: Multilingual Matters.

Nunnally JC (1988) *Psychometric Theory.* New Jersey, NY: McGraw-Hill, Englewood Cliffs.

Office for National Statistics. Neighbourhood Statistics (Lower Layer Super Output Areas - Indices of Deprivation and Classification). Available at http://www.neighbourhood.statistics.gov.uk/dissemination, accessed 10 May 2015

Pena ED and Hall TG (2011) Assessing preschool children dual language learners: Traveling a multiforked road. *Child Development Perspectives* 5(1): 28-32.

Ramey CH, Chrysikou EG and Reilly J (2013) Snapshots of children's changing biases during language development: Differential weighting of perceptual and linguistic factors

predicts noun age of acquisition. *Journal of Cognition and Development* 14(4): 573-592.

Rost DH (2007) *Interpretation und Bewertung pädagogisch-psychologischer Studien-Eine Einführung.* Weinheim: Beltz Verlag/UTB Wissenschaft.

Safford K and Collins F (2012) EAL and English. *NALDIC ITE Support Materials*. Available at: http://www.naldic.org.uk/Resources/NALDIC/Teaching%20and%20Learning/Documents/EALEnglish.pdf (accessed 25 February 2015).

Schaefer B, Fricke S, Bowyer-Crane C, et al. (in preparation) Effects of an oral language intervention programme for reception-aged children with English as Additional Language and monolingual children with language weaknesses.

Seeff-Gabriel B, Chiat S and Roy P (2008) Early repetition battery. Oxford: Harcourt Assessment.

Semel E, Wiig EH and Secord WA (2006) *Clinical Evalulation of Language Fundamentals - Preschool, second edition (CELF Preschool-2).* Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.

Simmons N and Johnston J (2007) Cross-cultural differences in beliefs and practices that affect the language spoken to children: Mothers with Indian and Western heritage. *International Journal of Language & Communication Disorders* 42(4): 445-465.

Stow C and Dodd B (2003) Providing an equitable service to bilingual children in the UK: A review. *International Journal of Language & Communication Disorders* 38(4): 351-377.

Stow C and Dodd B (2005) A survey of bilingual children referred for investigation of communication disorders: A comparison with monolingual children referred in one area in England. *Journal of Multilingual Communication Disorders* 3(1): 1-23.

Stow C and Pert S (2015) *SLT assessment and intervention: Best practice for children and young people in bilingual settings*. London: Royal College of Speech and Language Therapy.

Teoh WQ, Brebner C and McCormack P (2012) Assessing the language skills of children from culturally and linguistically diverse backgrounds: The expressive vocabulary performance of Singaporean English-Mandarin bilingual pre-schoolers. *International Journal of Speech-Language Pathology* 14(3): 281-291.

Thordardottir E (2011) The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism* 15(4): 426-445.

Tickell DC (2011) The Early Years: Foundations for life, health and learning - An Independent Report on the Early Years Foundation Stage to Her Majesty's Government. London: Report for the Department for Education.

Tong L, McBride - Chang C, Shu H, et al. (2012) Longitudinal predictors of spelling and reading comprehension in Chinese as an L1 and English as an L2 in Hong Kong Chinese children. *Journal of Educational Psychology* 104(2): 286-301.

Van de Vijver F and Hambleton RK (1996) Translating tests: Some practical guidelines. *European Psychologist* 1(2): 89-99.

Van deVijver F and Tanzer NK (2004) Bias and equivalence in cross-cultural assessment: An overview. *Revue Européenne de Psychologie Appliquée* 54: 119–135.

Van Steensel R (2006) Relations between socio-cultural factors, the home literacy environment and children's literacy development in the first years of primary education. *Journal of Research in Reading* 29(4): 367-382.

Verhoeven L and Van Leeuwe J (2008) Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology* 22(3): 407-423.

Williams ME, Sando L and Soles TG (2014) Cognitive tests in early childhood: Psychometric and cultural considerations. *Journal of Psychoeducational Assessment* 32(5): 455-476.

Table 1: Examples for noun test items.

Target item	AoA	Categorical distractor	AoA	Meronymic/ functional distractor	AoA	Random distractor	AoA
night	3.6	day	3.5	bed	2.9	pig	3.8
bridge	5.6	road	4.6	tunnel	5.9	dragon	5.6

Table 2: Examples for verb test items.

Test item	AoA	Same specific verb class	AoA	Same overall verb class	AoA	Different overall verb class	AoA
smile	3.5	frowning	4.5	bleeding	4.3	flying	3.1
knocking	4.6	hitting	4.8	scratching	5.6	hiding	4.5

Table 3: Descriptive statistics for RVS, BPVS, CELF for monolingual and multilingual children.

Children	**Measure**	**n**	**M**	**SD**	**Min-Max**
Monolingual	BPVS raw score	67	69.31	14.20	39-92
	CELF raw score	67	22.79	7.09	6-36
	RVS English	67	17.12	2.14	8-20
Multilingual	BPVS raw score	72	61.65	15.10	15-85
	CELF raw score	72	18.19	8.59	2-34
	RVS English	72	17.04	2.13	12-20
	RVS home language	38	15.13	4.63	5-20

Table 4: Correlations between the RVS, CELF and BPVS for monolingual and multilingual children and both groups combined.

Groups	n	App version	CELF Expressive Vocabulary	BPVS	RVS English
Monolingual	67	RVS English	0.324**	0.306*	
Multilingual	72	RVS English	0.508***	0.597***	
		RVS home language	0.214ns	0.391*	0.673***
Mono- and multilingual	139	RVS English	0.410***	0.456***	

$p < 0.05$*, $p < 0.01$**, $p < 0.001$***