Using parental questionnaires to investigate the heritage language proficiency of bilingual children

Kirsten Abbot-Smith[1], Patrycja Morawska-Patera[1], Magdalena Łuniewska[2] , Megan Spruce[1] and Ewa Haman[2]

[1]School of Psychology, University of Kent, UK

[2]Faculty of Psychology, University of Warsaw, Poland

Acknowledgments

Many thanks to Sarah Weinberg and Pearson Inc. for providing copies of their Polish translation of the CCC2, to Marek Patera and Nitika Shah for transcription, scoring and coding reliabilities, and to the parents and children of the Kent Child Development Unit and the primary schools in central Ashford, Kent, that participated. An earlier version of this study was submitted by Patrycja Morawska-Patera as her MSc thesis as part of an MSc by coursework in Developmental Psychology, University of Kent, UK. The cross-Linguistic Lexical Tasks CLT used in the study were designed as part of the networking programme COST Action IS0804 (Bi-SLI; http://bi-sli.org/). The CLT picture base and Polish adaptation of PABIQ questionnaire were developed as part of a grant from National Science Centre, Poland (no. 809/N-COST/2010/0; co-PIs: Ewa Haman & Zofia Wodniecka);

ABSTRACT

We asked whether parental questionnaires on the heritage language proficiency of bilingual children might elucidate how proficient bilingual children are in their heritage language. We tested 20 UK-based Polish-English bilingual children between 4;5 and 5;9 years on Polish and English versions of the Cross-Linguistic Lexical Tasks (CLTs). These comprise receptive and expressive picture tasks. Our bilingual group performed significantly worse on the Polish CLTs than on the English CLTs overall. They also performed significantly worse on the English CLTs than did an age- and gender-matched group of monolingual English-speaking children. Therefore our bilingual sample represent the type of bilinguals for whom education professionals have difficulty determining whether weak English is due to diminished English input versus an underlying Speech, Language or Communication Need.

Parents of the bilinguals completed a Polish adaptation of the Children's Communication Checklist 2. They also completed the Parents of Bilingual Children Questionnaire (PaBiQ), which includes Risk Factor measures ('No Risk Index' and children's 'Current Language Skills'). The PaBiQ also includes measures of the Amount and Length of Exposure to the majority language (English) prior to age four as well as the proportion of English in the current input.

For the bilingual sample the CCC2 General Communication Composite (GCC), which measures structural language, significantly predicted Polish CLT production,

uniquely accounting for 25% of the variance. The parent-rated PaBiQ ‘current Polish skills’ section predicted the Polish CLT comprehension. While the PaBiQ measure of Amount and Length of English Exposure was related to both Polish comprehension and production, it did not retain significance in a regression analysis. Therefore, parental questionnaires of the heritage language could provide a useful first step for education professionals when deciding whether to refer bilingual children for speech and language assessment. Large scale studies are needed to further develop these parental questionnaires.

Words = 297

Introduction

Bilingual children whose heritage language forms a large proportion of their input are on average delayed in their acquisition of the majority language (in UK: English) when they start primary school. This is well established for lexical and syntactic development (e.g. Bialystok, Luk, Peets & Yang, 2010; Hoff et al., 2012; Nicolls, Eadie & Reilly, 2011). While some have argued that the relative delay in bilingualism disappears when one uses Total Conceptual Vocabulary (e.g. Pearson, Fernandez & Oller, 1993; Junker & Stockman, 2002), others have found that typically-developing bilingual children can underperform monolingual peers even for TCV (see Core, Hoff, Rumiche & Senor, 2013; Hemsley, Holm & Dodd, 2010; Gross, Buac & Kaushanskaya, 2014; Thordardottir, Rothenberg, Rivard & Naves, 2006). One common reason for this delay is simply that most bilingual children have had reduced exposure to each of their languages; relationships between input frequency and lexical and syntactic development are well-established for monolingual language learners (e.g. Hurtado, Marchman & Fernald, 2008; Huttenlocher Vasilyeva, Cymerman & Levine, 2002). The degree to which a typically-developing child will underperform his or her monolingual peers when assessed in a given language depends to a large degree on the proportion of the input which the child hears in that particular language (Hoff et al., 2012; Pearson, Fernández, Lewedeg, & Oller, 1997; Thordardottir, 2011; Cattani et al., 2014).

Nonetheless, if both languages are assessed, typically-developing bilinguals can be clearly differentiated from bilingual children with Speech Language or Communication Needs (SLCN, Bishop et al., 2016) as significant language delay in bilingual children with SLCN is expected to be reflected in their both languages (not just in one for which they have had limited exposure). However, this does not help education professionals when they are deciding whether to refer a bilingual child for a speech and language assessment. Most education professionals do not speak the heritage language of bilingual children and are thus unable to directly determine whether the child is proficient in his or her heritage language (see also Boerma & Blom, 2017; Cattani et al. 2014).

A potential aid here might be the use of parental questionnaires about the child's use of the heritage language. One example is the Alberta Language and Development Questionnaire (ALDeQ, Paradis, Emmerzael & Duncan, 2010), which contains the following 'Risk Factor' scales: Early milestones, Current Language Skills and Family History. Paradis et al. (2010) found that there were significant differences with large effect sizes between typically-developing sequential bilinguals and sequential bilinguals with Developmental Language Disorder (DLD) for Early Milestones and Current Language Skills. No relationship was found between the latter and exposure to English (Paradis et al., 2010: 484).

The Parents of Bilingual Children Questionnaire (PaBiQ)

Tuller (2015) combined the ALDeQ with another questionnaire developed by Paradis (2011), the ALEQ, which asks parents about bilingual exposure and parental demographics. This combined and adapted questionnaire, the Parents of Bilingual Children Questionnaire (PaBiQ), provides researchers with the following Risk Factor measures; 1) 'Early Language Development' scale; 2) 'Family History' (of academic and language difficulties); 3) 'No Risk' index (which is the combination of (1) and (2)); and 4) Current Language Skills (for each language). In addition, the PaBiQ also provides measures to assess the relative dominance of the child's two languages in the input, both currently and retrospectively.

A number of studies have found significant differences between bilingual 5- to 8-year-olds with diagnosed DLD and typically-developing bilinguals the same age on either the PaBiQ 'No Risk' index (de Almeida et al., 2015; Boerma & Blom, 2017 for 5- and 6-year-olds) or sub-components thereof (Fleckstein, Prevost, Tuller, Sizaret & Zebib, 2016). de Almeida et al. (2017) found significant correlations with moderate effect sizes between the No Risk index and both NWR and Sentence Repetition (cf. dos Santos & Ferre, 2016). Unfortunately, de Almeida et al. (2007) conflated over the TD and DLD bilingual groups, which would artificially inflate the correlation (a "pooling" fallacy).

Indeed, there is, in fact, a fundamental problem with assessing the utility of parental questionnaires by examining results for children who already have a diagnosis of DLD (or other SLCN). This is the reverse 'halo' effect (e.g. Bryan & Wheeler, 1972); that is,

parents who have official confirmation that their child has SLCN might be more inclined to assess their child's language and communication skills as 'poor', also on retrospective measures of language development. In fact, the only study which has investigated the relationship between any of the PaBiQ Risk Factor measures and a direct assessment of language in typically-developing bilinguals (without conflating with bilinguals with DLD) is that of Hansen et al (2017a). This study found a moderate correlation between the PaBiQ 'Current Language Skills' scale (conflated over majority and heritage language) and a direct assessment of language in typically-developing 4-6-year-old Polish-Norwegian bilinguals.

The Children's Communication Checklist 2 (CCC2)

A frequently used questionnaire in the (monolingual) clinical literature is the Children's Communication Checklist 2 (Bishop, 2003), which allows the calculation of both a measure of structural language ability (General Communication Composite, GCC) and a measure of pragmatic language impairment (Social Interaction Deviance Composite, SIDC). Law, McBean and Rush (2011) found that the CCC2 GCC showed a moderate to strong significant correlation with the Core Language score of the Clinical Evaluation of Language Fundamentals (CELF, Semel, Wiig & Secord, 2006) in a sample of 167 monolingual, socially disadvantaged children (88% of whom had never been referred to SLTs). Unfortunately, the age range of Law et al.'s (2011) study was broad (5-12 years)

and correlations did not control for age.

The current problem

Thus, there is promising evidence that parental questionnaire measures which assess Risk Factors relating to the heritage language might have sufficient validity to enable education professionals to use these as a first pass when deciding whether to refer a child for speech and language assessment. To date, however, no study has examined the relationship between the CCC2 and a direct assessment of language in typically-developing bilingual children and indeed there are scarcely any studies which examine the relationship between these Risk Factor questionnaire measures and a direct assessment of language in typically-developing bilingual children.

What is the relationship between direct and parental measures of bilingual language development?

The current study therefore primarily investigates whether the parent-completed CCC2 (in translation) and / or the PaBiQ Risk Factor measures relate to the directly assessed vocabulary proficiency of Polish-English bilingual children in the first year of schooling in the UK. We also asked whether this relationship still holds when measures of input are taken into account. The decision to only assess vocabulary was made primarily because it is fairly straightforward (in comparison to syntax) to compare across languages. Moreover,

it is well-established that lexical development in one of a bilingual child's languages is highly predictive of syntactic development in that particular language (e.g. Marchman, Martínez-Sussmann & Dale, 2004; Conboy & Thal, 2006). All participants were assessed on noun and verb vocabulary in both English (majority language) and their heritage language, which was Polish. Both comprehension and production was assessed.

Do UK Polish-English bilingual reception-class children score significantly lower than their monolingual English peers on English vocabulary?

To verify that our sample of Polish-English bilinguals were likely to be from a population for which English-only screening or assessment would be problematic, we also carried out the English vocabulary measure with monolingual English-speaking age-matched peers.

In sum, our research questions were as follows:

1. Do UK Polish-English bilingual reception-class children score significantly lower than their monolingual English peers on English vocabulary?

2a) Do any of the Polish PaBiQ Risk Factor measures (No Risk Index or Current Polish Skills) or the CCC2-GCC relate to the Polish CLT vocabulary test?

2b) Does the relationship still hold when language input factors are taken into account?

Method

Participants

We tested 43 typically-developing 4- and 5- year-olds growing up in a medium-sized town in south-east England, none of whom had hearing difficulties or a diagnosed developmental disorder (according to parental report). Participants were recruited via three primary schools and also via word of mouth. The latter were tested in the Kent Child Development Unit (as arranging the testing in a school setting would be difficult for these children). For all the bilinguals, both parents were native Polish speakers who were born outside the UK. Two participants were removed from the dataset because their fathers were native English speakers. Around 75% of parents of the bilinguals reportedly only spoke Polish at home. For the rest, one or both parents spoke both Polish and English to their children. The age of first exposure to English ranged from birth to 38 months. All except three bilingual children were born in the south-east UK. The remaining three children were all born in Poland. All children were exposed to some English preschool childcare. The 'Proportion English input' (derived from the PaBiQ current input measures) ranged from 29% to 59% but for all except three children this was below 50% (median = 37.5%). Regarding SES, there were no relationships between the PaBiQ 'Years of Schooling' variables for either parent for any of the direct or questionnaire measures and thus parental education was not included in any analyses reported below. Details of participants who were included in the study are given in Table 1 below.

Table 1: Participant characteristics

	Monolingual English-speaking	**Polish-English bilinguals**
Age in months (n.s. p = .59)	M = 60.81 (range = 52-67 months)	M = 59.95 (range = 53-69 months)
Number	21	20
Gender	10 boys, 11 girls	9 boys, 11 girls
Percentage tested in schools (versus in a University Developmental Lab)	71%	35%

Testing materials and measures

Cross-linguistic Lexical Tasks (CLTs)

The particular lexical test we used to directly assess our bilingual sample was the Cross-linguistic Lexical Tasks (CLTs), which have been developed within COST Action IS0804 with the aim of assessing bilinguals in our age range (Haman, Łuniewska & Pomiechowska, 2015; see also Haman et al., 2017, for CLT results for monolinguals aged 3;0-6;11 across 17 languages; Hansen, Simonsen, Łuniewska & Haman, 2017b, for Polish bilinguals). The English versus Polish items are selected based on language-specific characteristics of word form complexity and age of acquisition. Detailed description of CLTs design may be found in Haman, Łuniewska, & Pomiechowska (2015; see also Haman et al., 2017; Simonsen & Haman, 2017). CLTs consist of four sub-tests per language: Noun Comprehension, Verb Comprehension, Noun Production and Verb Production. Each CLT sub-task contains 32 items, presented on a touch-screen computer. For the comprehension tests, each item consists of a choice of four pictures for which the

test question was played automatically before the picture set appeared. For nouns the question was always 'Where is X?' ('Gdzie jest X?' in Polish). For verbs the English question was either 'Who is Xing? or 'Where is something Xing?'. For the production tasks the elicitation question for nouns was 'What is this?'. For verbs it was either 'What is s/he doing?', 'What are they doing?', 'What is happening here?' or 'What is happening to him?'. For the Polish elicited verb production, the questions were 'Co on/a robi?' ('What is she/he doing?'), 'Co się tu dzieje?' ('What is happening here?'), or 'Co ktoś robi?' ('What is someone doing?'). All picture stimuli are coloured drawings designed for the CLTs. Order of items within each sub-task is semi-randomized (with two very simple items always appearing at the beginning of the sub-task). The location of each of the four option pictures within comprehension trials was also semi-randomized with the exception that each target picture cannot occur in a given location more than three times consecutively. The frequency with which each pictures occurs in each of the four positions across each comprehension sub-task was also controlled.

Polish adaptation of the 'Parents of Bilingual Children Questionnaire' (PaBiQ)

We used Kuś, Otwinowska, Banasik and Kiebzak-Mandera's (2012) Polish adaptation of Tuller's 'PaBiQ' questionnaire in the same format as used by Hansen et al. (2017a) and Mieszkowska, et al. (2017). The questionnaire consists of 60 questions which are grouped in sections on Early Language Development, 2) Familial history of language and school

difficulties, 3) Current Language Skills, 4) Language Input prior to 4 years, 5) Current Language Input and 6) Parental Education, Occupation and Language Proficiency. The scoring for these measures is outlined in Table 3 in the results section and also in more detail below.

PaBiQ Risk Factor Measures

As in Tuller's (2015) version of the PaBiQ, sections (1) and (2) combine to form the 'No Risk Index' (see Table 3 for details on scoring). The second Risk Factor measure is that of Current Language Skills, which here has two subsections: 3a) 'Current Polish Skills' and 3b) 'Current English Skills'. Each is based on nine questions regarding child's vocabulary, pronunciation, grammar and discourse skills as estimated by parents, the scoring of which is shown in Table 3. Kuś et al.'s (2012) version of the PaBiQ also contains an additional section, namely Polish Output'. This is a composite score of how often the child chooses Polish (measured on a Likert scale from 'never' to 'always'), when addressing her parents, other caregivers and friends (see Table 3 for scoring).

PaBiQ Measures: Language Input prior to age four

The PaBiQ language input measures are divided into the section on (4) Language Input Prior to the Age of Four Years and (5) Current Language Input. The section on language input prior to the age of four years asks questions on the age of first contact (here: with

English) and frequency of early contact, which is, estimated by parents on a 5-point Likert scale (from 'never heard English' to 'heard exclusively English'). As can be seen on Table 3, these two measures are multiplied to create the scale 'Amount and Quality of Early Exposure'. Another scale in this section is Length (in months) that the child has had contact with English. In the current study, 'Amount and Quality of Early Exposure' and 'Length of contact' were highly inter-correlated ($r(17) = .90$, $p < .001$) even when controlling for age using a partial correlation. To minimise the number of variables, we summed Amount and Quality of Early Exposure' and 'Length of contact' to create one measure 'PaBiQ Amount and Length of Early Exposure'.

PaBiQ Current Language Input

Thirteen questions assess current exposure to the majority language (here: English) versus the heritage language (here: Polish). This section is divided into questions about input in the home and questions tapping input outside the home. In the former the parent is asked to rate on a 5-point Likert scale the frequency with which a particular person (e.g. parent, sibling, grandparent) addresses the child in English vs. Polish, whereby input from parents and siblings is given greater weight than input from other sources. A child also receives a higher score if a greater number of people speak that language to child. Higher scores are also received for a particular language (Polish vs. English) if the child engages in a greater number of activities (e.g. book reading, film watching) in that language. The

section on language input outside the home asks questions about the frequency with which the child hears Polish vs. English from friends, visits Poland etc. (again with the use of Likert scale). There are also questions about the number of hours the child spends in clubs and whether the language used is Polish vs. English (see Mieszkowska et al., 2017, for a more detailed description of the Current Input Measure). In the current study, we created the variable 'Proportion English in input' by dividing the total English current input (household and outside the home) by the sum of total current English and total current Polish input.

Polish translation of the CCC2

The CCC2 (Bishop, 2003) consists of 10 sub-scales, each with 7 items, 5 of which are formulated in a negative manner (e.g. 'Forgets words s/he knows – e.g. instead of rhinoceros may say 'you know, the animal with the horn on its nose…') and 2 of which are formulated positively and are then reverse scored (e.g. 'Uses abstract words that refer to general concepts rather than something you can see – e.g. 'knowledge', 'politics', 'courage'). Parents rated each item on a Likert Scale from 0 to 3, where 0 = 'less than once a week or never' and 3 = 'several times (more than twice) a day (or always)'. The process of conversion to scaled scores reverses the valence of the raw scores. Thus, a child with a low raw score for a particular sub-scale will obtain a higher scaled score for that sub-scale than will a child the same age who obtained a high raw score. The GCC is the sum of the

scaled scores for the following sub-scales A (Speech e.g. simplifies words by leaving out some sounds), B (Syntax, e.g. produces long and complicated sentences), C (Semantics, e.g. is vague in choice of words, making it unclear what s/he is talking about, e.g. saying 'that thing' rather than 'kettle'), D (Coherence, e.g. talks clearly about what s/he plans to do in the future (e.g. what s/he will do tomorrow)), E (Inappropriate Initiation, e.g. talks repetitively about things that no-one is interested in), F (Stereotyped Language, e.g. you can have an enjoyable, interesting conversation with him/her'), G (Context, e.g. gets confused when a word is used with a different meaning from usual) and H (Non-verbal communication, e.g. ignores conversational overtures from others). The SIDC is calculated in terms of the difference between sum of scales E, H, I (Social relations, e.g. appears anxious in the company of other children) and J (Interests, e.g. reacts positively when a new and unfamiliar activity is suggested) and the sum of scales A-D. The CCC2 was translated into Polish by Pearsons in consultation with a Speech and Language therapist and is closely based on the English original, whereby alternative words were given as examples for three items on sub-scale A (Speech). This has not yet been standardized for Polish.

Procedure

The parents of Polish-English bilingual children completed both questionnaires. Each child was tested individually either in a quiet area of the child's school (i.e. a separate classroom) or in a university developmental unit. The English version of the CLT vocabulary tests was administered by the fourth author (a native, monolingual English speaker) and the Polish version by the second author (a native Polish speaker). Half the bilinguals were administered the Polish version of the task first and within each language group the order of each sub-test was counterbalanced across participants.

Data coding and introductory analyses

Production was audio-recorded, transcribed and coded by the second author according the CLT procedure (Haman et al., 2015). Specifically, the following were included as 'correct' responses: mispronounciation (e.g. 'teethbrush' for 'toothbrush'), unexpected or incorrect inflection (e.g. 'chains' for 'chain'), derivations within syntactic category (e.g. 'froggie' for 'frog) and regional variants or synonyms (e.g. 'cuddling' for 'hugging'). A fuller description of the error analysis is provided in Kapalkova and Slancova (2017). Code-switching accounted for 2% of CLT responses and was counted as an error. Seven percent of the data was also coded by the third author (a native Polish speaker), blind to original codes, with excellent inter-rater reliability (k = .957). There were no effects for gender across any of our questionnaire or directly assessed variables for either Language Group (all $p > .2$). Therefore, we conflate across gender in all analyses.

Results

RQ1: Do UK Polish-English bilingual reception-class children score significantly lower than their monolingual English peers on English vocabulary?

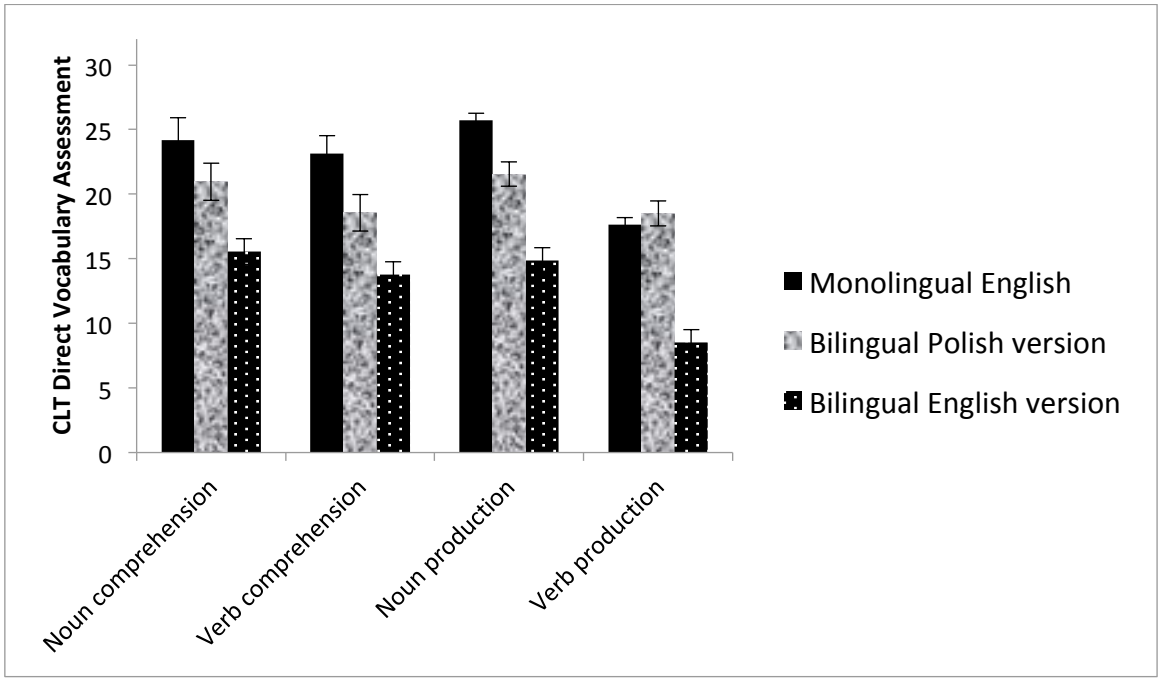
Figure 1 presents the CLT scores by Language Tested (Polish versus English), Language Group (monolinguals versus bilinguals), Modality (Comprehension versus Production test) and Syntactic Category (test of Noun vs. Verb knowledge). To investigate whether the bilinguals underperformed the monolinguals on the English-language version of the CLT we carried out linear mixed effects model with Group, Modality and Syntactic Category as fully crossed fixed effects. In all models, all predictor variables were centred. Random slopes were included for within-participant variables and the random slope structure was simplified until the model converged (Barr, Levy, Scheepers & Tily, 2013). To compute *p*-values we compared models using likelihood-ratio tests. The final model had random slopes for participants for both Modality and for Syntactic Category, but these were not crossed. There was a main effect of Syntactic Category (b = 4.32, SE = 1.12, χ^2= 49.09, p < .001) but this needs to be interpreted in the light of an interaction between this and

Modality (b = -5.80, SE = 0.79, χ^2= 42.371, p < .001). As can be seen in Figure 1, while there was no significant difference between Noun Comprehension vs. Production, Verb Production was significantly lower overall than was Verb Comprehension. However, our research question focussed on effects and interactions relating to the Group factor. We indeed found a main effect of Group (b = 9.51, SE = 1.12, χ^2= 43.35, p < .001), whereby the bilinguals scored significantly lower overall than did the monolingual on the English CLT overall. Importantly, this main effect of Group did not interact with any other factor (all p > .1).

Since the bilingual children performed on average significantly poorer overall on the English CLT, we asked whether Polish might on average be their stronger language. To investigate this question we carried out a second linear mixed effects model, this time with the bilingual group only. Here we included Language Tested (Polish vs. English), fully crossed with both Modality and Syntactic Category. There was again a main effect for Syntactic Category (b = 3.40, SE = 0.68, χ^2 = 19.60, p < .01) and again an interaction between this and Modality, although here this was only of marginal significance (b = -2.60, SE = 1.34, χ^2 = 3.78, p = .052). In relation to our research question, there was indeed a main effect of Test Language (b = -6.73, SE = 0.67, χ^2 = 62.84, p < .001) whereby the bilinguals performed significantly worse on the English CLTs overall than they did on the Polish CLTs. There was also an interaction between Language Tested and Modality: Bilinguals performed particularly poorly on the English naming part (b = 3.25, SE = 1.34,

χ^2 = 5.68, p = .017). Neither the interaction between Language Tested and Syntactic Category (p = .3) nor the three-way interaction between Language Tested, Syntactic Category and Modality (p = .14) reached significance.

Fig 1: Mean correct performance in CLTs, by Group and Language Tested (error bars showing SE).



Primary research question: The relationship between direct and parental Risk Factor measures of bilingual language development

Our key research question investigated whether parent-completed Polish questionnaires (the GCC of the translation of the CCC2 and / or the PaBiQ Risk Factor measures) would relate to direct assessment of language ability of these children (i.e. Polish CLTs).

Descriptive statistics for questionnaire measures

Table 2 below shows the mean and range scaled scores obtained by the bilingual children on the Polish adaptation of the CCC2. We focus on the GCC because this assesses formal language according to Norbury, Nash, Baird and Bishop (2004). The GCC comprises the sum of scaled scores for sections A-H. Therefore a GCC score of 32 would be equivalent to child scoring 2 SD below the (monolingual) mean. The Polish GCC scale had high reliability (Cronbach's $\alpha = .844$)

Table 2: Descriptive statistics for Children's Communication Checklist 2, results obtained from the Polish adaptation. (Results for A-J are scaled scores, based on norms for monolingual English).

Scale	**Measures**	**Obtained mean**	**Lowest score**	**Highest score**
A	Speech	6.95	1	14
B	Syntax	7.55	2	14
C	Semantics	7.25	4	11
D	Coherence	7.65	4	12
E	Inappropriate initiation	8.9	6	14
F	Stereotyped Language	8.95	4	15
G	Use of context	9.35	5	14

H	Nonverbal communication	8.55	4	14
GCC	**Formal language** (sum of A-H above)	**65.55**	**40**	**97**
I	Social relations	8.4	2	14
J	Interests	9.05	5	18
SIDC	Measure of pragmatic language impairment	5.50	-9	16

Table 3 below shows the scale, obtained mean and obtained range for each of the PaBiQ scales, where the final column explains the scoring of each scale. The PaBiQ is not designed to be one coherent scale; rather, various sections measure different domains. We were primarily interested in the two Risk Factor scales: 'No Risk' and 'Current Polish Skills'. The 'No Risk' measures were all highly skewed as almost all children had no indications of either Family History or Early Developmental Delays. Current Polish Skills was, in contrast, normally distributed and had high reliability (Cronbach's $\alpha = .842$).

Table 3: Descriptive statistics for the Parents of Bilingual Children Questionnaire (PaBiQ).

Measure	**Mean**	**Lowest score**	**Highest score**	**Max. score poss.**	**Explanation of scoring**
Age at First Word	5.71	4	6	6	< 15 m: 6 points 16-24 m: 4 points > 24 m.: 0 points
Age at First Sentence	5.52	0	6	6	< 24 m: 6 points 25-30 m: 4 points > 30 m: 0 points
Parental Concern	1.43	0	2	2	YES: 0 points NO: 2 points
*Hearing Difficulties	1.9	0	2	2	4 questions: YES: 0 points

					NO: 0.5 points
1) Early Language Development	14.73	6.5	16	16	The above four scales combined.
2) Family History	9.8	9	10	10	Any score less than 10 indicates possible family history indicator
No Risk Index	24.53	16	26	26	Comprised of Early Language Development and Family History.
Current Language					
3a) Current Polish Skills	18.90	7	25	27	9 questions each rated on Likert scale of 0-3
3b) Current English Skills	15.05	7	23	27	9 questions each rated on Likert scale of 0-3
*3c) Polish Output	27.25	9	36	48	Comprised of two questions: 1) Language in which child addresses parents, siblings and friends (5-point scale, max score 8); 2) Language in which child addresses grandparents etc (5 point scale, max score 4).
4) Input prior to age of 4 years					
Age of the first contact	12.43	0	38	69	In months
Frequency of the first contact	2.67	2	3	4	Never = 0; Rare = 1; Sometimes = 2; Often = 3; Exclusively = 4.
Amount and quality of early exposure	94.30	20	144	192	Multiply 'Age of first contact' by 'Frequency of first contact'.
Length of contact	46.85	17	65	69	This is calculated by age of testing minus age of first contact.

5) Current input					
Proportion English in the input	.396	.29	.59	1.0	See text on p13

- These sections were added by Kuś et al. (2012) to Tuller's (2015) version of the PaBiQ.

Relationship with production

Table 4 shows Pearson correlations with Polish CLT production. Whereas the English CLT measures all correlated with age (all $r < .47$), this was not the case for Polish CLT production. The marginally significant negative relationship between age and Polish Output indicates that the younger the child, the more likely s/he was to produce Polish, presumably because with increasing age the children had proportionally less exposure to Polish. There was a strong, negative relationship between the Amount and Length of Early Exposure to English and Polish CLT production, indicating that greater *relative* exposure to Polish led to higher Polish vocabulary scores. Both the PaBiQ 'Current Polish Skills' and CCC2 GCC related to Polish CLT production, with the latter showing a very high effect size.

Table 4 Relationships between parent ratings and Polish *production*

	PaBiQ Amount and Length of Early Exposure to English	PaBiQ Polish Output	PaBiQ 'No risk' index	PaBiQ Current Polish skills	CCC2 - GCC	Polish production

Age in months	-.05	-.42˅	.42˅	.31	.08	-.16
Proportion English in current input	.42˅	-61**	.3	-.15	-.02	-.34
PaBiQ Amount and Length of Early Exposure to English	-	-.40˅	-.35	-.47*	-.28	-.53*
PaBiQ Output_Polish	-	-	-.16	.19	-.05	.19
PaBiQ 'No risk' index	-	-	-	.7***	.43˅	.29
PaBiQ current Polish skills	-	-	-	-	.48*	.47*
CCC2 GCC	-	-	-	-	-	.69***

˅ = p < .1 * = p < .05 ** = p < .01 *** = p < .001

We then entered the significant variables from Table 4 into a hierarchical linear regression after first converting all variables to z scores. Following our primary research question, we wished to investigate the degree to which any of the Risk Factor measures (here: PaBiQ Current Polish Skills and CCC2-GCC) related to Polish production, once any significant language input variables had been controlled for. Table 5 below shows the results. The PaBiQ Amount and Length of Early Exposure to English was entered in the first step. This led to a model which was significant overall and accounted for 29 % of the variance. PaBiQ Current Polish Skills was entered in the second step. As the result, the model with these two factors was not significant and Current Polish Skills was not a significant factor within the model. Finally, in the third step we entered the CCC2-GCC. This model was

significant (F (3, 19) = 10.07, p = .006) and the R^2 change was .25; it also accounted for 52% of the variance in CLT Polish production. In this third model, only the GCC was significant; it uniquely accounted for 25% of the variance. The Amount and Length of Early Exposure to English was only of marginal significance, although it did uniquely account for 10% of the variance. The same pattern of results was found when comparing the full model with models in which one of these factors was removed. That is, when 'Current Polish Skills' was removed from the full model, this did not lead to a significant difference (F = 0.22, p = .88). Removing 'Amount and Length of Early Exposure to English' from the full model led to a marginally significant difference (F = 4.0, p = .063). Removing the CCC2-GCC from the full model did lead to a significant difference (F = 10.07, p = .006).

Table 5 Summary of multiple regression analysis for variables predicting Polish production (all variables converted to z scores)

Step	**Variable**	**b**	**SE**	**t**	**sr^2**
1	Constant			.00	
	PaBiQ Amount and Length of Early Exposure to English	-.53	.20	-2.68*	.29
2	PaBiQ Amount and Length of Early Exposure to English	-.40	.22	-1.80$^{\vee}$	.12
	Current Polish Skills	.28	.22	1.27	.06
3	PaBiQ Amount and Length of Early Exposure to English	-.36	.18	-2.00$^{\vee}$	.10
	Current Polish Skills	.03	.20	.147	.00
	CCC2 GCC	.57	.18	3.17**	.25

$^{\vee}$ = p < .1 * = p < .05 ** = p < .01 *** = p < .001

Relationship with comprehension

The same correlation matrix was run for Polish CLT comprehension, as shown in Table 6 below. There was a strong and significant relationship between this and the PaBiQ 'Current Polish skills' (r = .57, p = .009). There was also a marginal relationship with 'PaBiQ Amount and Length of Early Exposure to English' ($r(17)$ = -.40, p = .078). There were no relationships between the Polish CLT comprehension and either the GCC (p = .6) or any other PaBiQ variable (all p > .16).

Table 6 Relationships between parent ratings and Polish *comprehension*

	PaBiQ Amount and Length of Early Exposure to English	PaBiQ Polish Output	PaBiQ 'No risk' index	PaBiQ Current Polish skills	GCC	Polish comprehension
Age in months	-.05	-.42ᵛ	.42ᵛ	.31	.08	.26
Proportion English in input	.42ᵛ	-.61**	.30	-.15	-.02	-.27
PaBiQ Amount and Length of Early Exposure to English	-	-.40ᵛ	-.35	-.47*	-.28	-.40ᵛ
PaBiQ Output_Polish	-	-	-.16	.19	-.05	.33

PaBiQ 'No risk' index	-	-	-	.70**	.43ᵛ	.16
PaBiQ current Polish skills	-	-	-	-	.48*	.57**
CCC2 GCC	-	-	-	-	-	.13

ᵛ = p < .1 * = p < .05 ** = p < .01 *** = p < .001

We therefore entered the Current Polish Skills' and 'Amount and Length of Early Exposure to English' into a hierarchical linear regression after first converting all variables to z scores. The results are shown on Table 7. The PaBiQ Amount and Length of Early Exposure to English was entered in the first step. This led to a model that was marginally significant (p = .08) overall and accounted for 12% of the variance. PaBiQ Current Polish Skills was entered at the next step. This led to a significant model ($F(2,19) = 3.277$, p = .027 accounting for 27% of the variance overall and showing a R^2 change of .18. In this final model, only the PaBiQ parent-rated 'Current Polish skills' was a significant (positive) predictor, uniquely accounting for 18% of the variance. Again, the same pattern of results was found when comparing the full model with models in which one of these factors was removed. That is, removing 'Amount and Length of Early Exposure to English' from the full model did not lead to a significant difference (F = .608, p = .45). Removing 'Current Polish Skills' from the full model did lead to a significant difference (F = 4.75, p < .05).

Table 7 Summary of multiple regression analysis for variables predicting Polish comprehension (all variables converted to z scores)

Step	Variable	b	SE	t	sr^2
1	Constant				
	PaBiQ Amount and Length of Early Exposure to English	-.403	.22	-1.87^{v}	.16
2	PaBiQ Amount and Length of Early Exposure to English	-.17	.22	-.78	.02
	Current Polish Skills	.49	.22	2.18*	.18

^{v} = p < .1 * = p < .05 ** = p < .01 *** = p < .001

Discussion

Our primary research question was whether parent ratings of 'Risk Factors' in their children's heritage language development would relate to direct measures of vocabulary development in the heritage language. This was indeed the case. This is an important question particularly for bilinguals who might be expected to underperform (in the majority language) peers who are monolingual in the majority language. In our particular sample the heritage language was Polish and the majority language was UK English. All children were primary school entry age. We first confirmed that our bilingual sample underperformed their monolingual English-speaking peers in the English version of the directly assessed vocabulary measure (CLT). The bilingual sample also performed significantly better overall in the Polish than on the English version of the CLT,

confirming that our sample is typical of those bilinguals whose English at school entry may cause concern for education professionals.

Regarding our primary research question, we used the Risk Factor measures from the Polish adaptation of the PaBiQ (Kuś et al., 2012), namely a) No Risk Index and b) Current Polish Skills, as well as a Polish translation of the CCC2-GCC. The No Risk Index was not related to either Polish CLT comprehension or production. The cause for this could be very little variance in the No Risk Index, as almost all parents reported no or scarcely any risk factors. The Current Polish Skills was strongly related to both Polish CLT comprehension and production and the CCC-GCC was very strongly related to Polish CLT production. However, both Polish CLT comprehension and production was also negatively related (marginally in the case of comprehension) to the PaBiQ measure of the Amount and Length of Exposure to English prior to the age of four years. Therefore, we investigated whether the relationship between indirect Risk Factor measures and the direct vocabulary measures still held if we controlled for the Amount and Length of Exposure prior to four, entering this in the first step of hierarchical regression analysis. There indeed remained a significant, strong, positive relationship between the CC2 GCC and the CLT Polish production, and between CLT Polish receptive vocabulary, there was a significant, positive relationship with the PaBiQ 'Current Polish skills'.

Thus, our findings provide some positive, preliminary findings that further investigation is warranted into determining the validity of parental questionnaires of the

heritage language of bilingual primary school children. That said, it is clear from Table 2, that the mean scaled scores of the CCC2 indicate that our sample of Polish parents gave ratings on average which are below monolingual norms (since a scaled score of 10 is always the normed mean for monolingual English-speaking children). There are many potential reasons for this (see Hansen et al., 2017a, for discussion) but it is clear that the norms for the English CCC2 cannot be straightforwardly applied to the Polish translation when used with bilingual children. Indeed, while the questions in the PaBiQ Current Heritage Language Skills are generic enough to be easily translated, some of the Polish CCC2 items require specialist knowledge of the target language in order to select good examples (e.g. words which a child might find difficult to pronounce). Moreover, the 'Current Polish Skills', and not the GCC, were related to Polish vocabulary comprehension, which is arguably more important since language comprehension has been found to more strongly predict long-term difficulties with language, social communication (e.g. Beitchman et al., 1994; Chiat & Roy, 2008). Thus, the current Polish adaptation of the CCC2 clearly needs a great deal more validation if it were ever to be used as a *clinical* tool. Nonetheless, the fact that both questionnaires predicted Polish vocabulary levels in some form suggests that if educational professionals can obtain either (or both) questionnaires in translation, they could be a useful first pass in order to rule out the need for referral for SLT assessment, at least for Polish-English bilingual reception class children whose parents rate their Polish highly enough.

To truly assess the utility for educational professionals of these types of parental questionnaires about a child's heritage language, we need large-scale studies to explore whether bilingual-specific cutting scores can be developed. The direct assessments used in these future studies should include measures of, not only the lexicon, but also of heritage language receptive and expressive syntax, speech sound production and pragmatics. Predictive validity should also be examined as the long-term overarching aim of this field is to help educational professional to refer for assessment those children who are likely to need long-term SLT and classroom support.

References

Barr D, Levy R, Scheepers C and Tily H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278.

Beitchman JH, Brownlie EB, Inglis A, Wild J, Mathews R, Schachter D, Kroll R, Martin S, Ferguson B and Lancee W (1994) Seven-year follow-up of speech/language-impaired and control children: Speech/language stability and outcome. *Journal of the American Academy of Child & Adolescent Psychiatry* 33: 1322-1330.

Bialystok E, Luk G, Peets KF and Yang S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*(4), 525-531.

Bishop D (2003) *The Children's Communication Checklist-2*. London: Psychological Corporation.

Bishop DV, Snowling MJ, Thompson PA and Greenhalgh, T (2016) CATALISE: A multinational and multidisciplinary delphi consensus study of problems with language development. Phase 2: terminology. *PeerJ Preprints*, 4, e2484v1.

Boerma T and Blom E. (2017). *Assessment of bilingual languages: what is testing both languages is not possible? Journal of Communication Disorders, 66*: 65-76.

Bryan T and Wheeler R. (1972). Perception of learning disabled children. *Journal of Learning Disabilities, 5*(8): 484-488.

Cattani A, Abbot-Smith K, Farag R, Krott A, Arreckx F and Floccia C (2014) How much exposure to English is necessary for a bilingual to perform like a monolingual peer in language tests? *International Journal of Language and Communication Disorders, 49*(6): 649-671.

Chiat S and Roy P (2008) Early phonological and sociocognitive skills as predictors of later language and social communication outcomes. *Journal of Child Psychology and Psychiatry* 49: 635–45.

Conboy B and Thal D (2006) Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development*. *77*:712–35.

Core C, Hoff E, Rumiche R and Señor, M. (2013). Total and conceptual vocabulary in Spanish–English bilinguals from 22 to 30 months: implications for assessment. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1637-1649.

De Almeida L, Ferre S, Morin, E, Prevost P, dos Santos C, Tuller L, Zebib R and Barthez M (2017). Identification of bilingual children with Specific Language Impairment in France. *Linguistic Approaches to Bilingualism,* 10.1080/10489223.2016.1192635

Fleckstein A, Prevost P, Tuller L, Sizaret E and Zebib R. (2016). How to identify SLI in bilingual children: a study on sentence repetition in French. *Language Acquisition*, Online First 0(0), 1-17.

Gross M, Buac M and Kaushanskaya M (2014). Conceptual scoring of receptive and expressive vocabulary measures in simultaneous and sequential bilingual children. *American Journal of Speech and Language Pathology, 23*(4): 574-586.

Haman E, Łuniewska M, Hansen P, Simonsen HG, Chiat S, Bjekić J, Blažienė A, Chyl K, Dabašinskienė I, Engel De Abreu P, Gagarina N, Gavarró A, Håkansson G, Harel E, Holm E, Kapalková S, Kunnari S, Levorato C, Lindgren J, Mieszkowska K, Montes Salarich L, Potgieter A, Ribu I, Ringblom N, Rinker T, Roch M, Slančová D, Southwood F, Tedeschi R, Müge Tuncer A, Ünal-Logacev Ö, Vuksanović J and Armon-Lotem S. (2017). Noun and verb knowledge in monolingual preschool children across 17 languages: Data from cross-linguistic lexical tasks (LITMUS-CLT). *Clinical Linguistics & Phonetics*, *31*(11-12), 818-843.

Haman E, Łuniewska M and Pomiechowska B (2015). Designing Cross-linguistic lexical tasks (CLTs) for bilingual preschool children. In S. Armon-Lotem, J., d. Jong, N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from Language Impairment*. pp196-240. Bristol: Multilingual Matters.

Hansen P, Łuniewska M, Simonsen HG, Haman E, Mieszkowska K, Kołak J and Wodniecka Z. (2017a). Picture-based vocabulary vs. parental questionnaires: a cross-linguistic study of bilingual assessment methods. *International Journal of Bilingualism,* Online First 0(0), 1-20.

Hansen P. Simonsen H, Łuniewska M and Haman E. (2017b). Validating the psycholinguistic aspects of LITMUS-CLT: evidence from Polish and Norwegian. *Clinical Linguistics and Phonetics*, *31*(11-12), 910-930.

Hemsley G, Holm A and Dodd B (2010) Patterns in diversity: Lexical learning in Samoan-English bilingual children. *International Journal of Speech-Language Pathology*, 12, 362-374.

Hoff E, Core C, Place S, Rumiche R, Senor M and Parra M. (2012) Dual language exposure and early bilingual development, *Journal of Child Language, 39*(1)*,* 1-27.

Hurtado N, Marchman VA and Fernald A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science, 11*(6): F31–F39.

Huttenlocher J, Vasilyeva M, Cymerman E and Levine S (2002) Language input and child syntax. *Cognitive Psychology, 45:* 337-374.

Junker DA and Stockman IJ (2002) Expressive vocabulary of German-English bilingual toddlers, *American Journal of Speech-Language Pathology,* **11**, 381-395.

Kapalkova S and Slancova D. (2017). The vocabulary profile of Slovak children with primary language impairment compared to typically developing Slovak children measured by LITMUS-CLT. *Clinical Linguistics and Phonetics, 31*(11/12): 893-909.

Kuś K, Otwinowska A, Banasik N and Kiebzak-Mandera D (2012) Kwestionariusz rozwoju językowego. Polish adaptation of PABIQ. Unpublished material. University of Warsaw.

Law J, McBean K and Rush R. (2011). Communication skills in a population of primary school-aged children in an area of pronounced social disadvantage. *International Journal of Language and Communication Disorder, 46*(6): 657-664.

Marchman VA, Martinez-Sussmann C and Dale P (2004) The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science*, 7: 212-224.

Mieszkowska K, Łuniewska M, Kołak J, Kacprzak A, Wodniecka Z and Haman E. (2017) Home Language will not take care of itself: vocabulary knowledge in trilingual children in the United Kingdom. *Frontiers in Psychology Volume 8, article 1358 doi:* 10.3389/fpsyg.2017.01358

Nicholls RJ, Eadie PA and Reilly S (2011). Monolingual versus multilingual acquisition of English morphology: what can we expect at age 3? *International Journal of Language and Communication Disorders*, 46: 449-463.

Norbury CF, Nash M, Baird G and Bishop DV (2004) Using a parental checklist to identify diagnostic groups in children with communication impairment: a validation of the Children's Communication Checklist—2. *International Journal of Language & Communication Disorders, 39*(3), 345-364.

Paradis J (2011). Individual differences in child English second language acquisition: Comparing child-Internal and child-external factors. *Linguistic Approaches to Bilingualism.* 1:3, 213-237

Paradis J, Emmerzael K. and Duncan TS (2010) Assessment of English language learners: using parent report on first language development. *Journal of Communication Disorders*, *43*, 474–497.

Pearson BZ, Fernandez SC, Lewedeg V and Oller DK (1997) The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics***,** 18*:* 41-58.

Pearson BZ, Fernandez SC and Oller DK (1993) Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, *43,* 93-120.

Semel E, Wiig EH and Secord WA. (2006). *Clinical Evaluation of Language Fundamentals, Fourth Edition:* Harcourt Assessment.

Simonsen H and Haman E. (2017). LITMUS-CLT: a new way to assess bilingual lexicons. *Clinical Linguistics and Phonetics*, *31*(11-12), 811-817.

Thordardottir, E, Rothenberg A, Rivard, M-E and Naves R (2006) Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders, 4*, 1-21.

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

Thordardottir E (2015). Proposed diagnostic procedures for use in bilingual and cross-linguistic contexts. In S. Armon-Lotem, J., d. Jong, N. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from Language Impairment*. pp331-354. Bristol: Multilingual Matters.

Tuller L (2015). Clinical Use of Parental Questionnaires in Multilingual Contexts. In S. Armon-Lotem, J., d. Jong, N. Meir (Eds.), *Assessing Multilingual Children: Disentangling Bilingualism from Language Impairment.* pp301–330. Bristol: Multilingual Matters.