

An Investigation of Factors Affecting Early Foreign Language Learning in the Netherlands

*¹SHARON UNSWORTH, ²LIV PERSSON,
³TINEKE PRINS and ^{4,5}KEES DE BOT

¹Department of Languages, Literature and Communication and Utrecht Institute of Linguistics, Utrecht University, Utrecht, The Netherlands, ²Utrecht Institute of Linguistics, Utrecht University, Utrecht, The Netherlands, ³Radboud University Nijmegen, Nijmegen, The Netherlands, ⁴Applied Linguistics, University of Groningen, Groningen, The Netherlands, ⁵Department of English, University of the Free State, South Africa
*E-mail: s.unsworth@uu.nl

1. EARLY ENGLISH IN THE NETHERLANDS

The most recent data available on language teaching in Europe indicate that the number of pupils in primary education that are being taught at least one foreign language is on the rise, with pupils being offered foreign languages at an increasingly younger age (Eurydice Network, Eurostat, and Education, Audiovisual and Culture Executive Agency, 2008). The teaching of English in Dutch primary schools has been mandatory in the (pen)ultimate year of schooling since 1986 (Thijs *et al.* 2011). However, schools also have the option of introducing a foreign language earlier in the curriculum. Currently, an estimated 850 primary schools—approximately 12 per cent of the total number of primary schools in the Netherlands—offer a foreign language from age 4 years, that is, the first year of primary school, and while there are a number of schools opting for French or German, the vast majority choose English (Europees Platform, n.d.).

In terms of approach, schools offering an early ‘English as a foreign language’ (EFL) programme differ from the regular programme in more respects than just starting age. The most noticeable differences include the amount of time devoted to providing children with English language input and teachers’ language proficiency. Whereas children in the regular programme have on average 45 min of instruction per week from age 10 years (Thijs *et al.* 2011), children in early EFL programmes may have between 15 and 220 min of English language exposure in the classroom per week from age 4 years onwards (Thijs *et al.* 2011; Persson 2012). In both cases, English is the language of instruction and typically, certain themes and topics, for example, the seasons, colours, animals, etc., are used as the starting point for the English classes, while the rest of the day, the language in the classroom is Dutch (Herder and de Bot 2005; Thijs *et al.* 2011).

This variability in the amount of time devoted to early EFL programmes is in part a reflection of the lack of regulation in this area, which also results in considerable variation in the proficiency level and qualifications of teachers participating in these programmes. Generally speaking, as is commonly the case in early language learning programmes across Europe (Nikolov and Mihaljevic Djigunovic 2011: 107), three patterns can be distinguished: schools will employ either (i) the class teacher, (ii) a (near-) native-speaker specialist teacher, or (iii) have the class and specialist teachers co-teach English (Thijs *et al.* 2011; Persson 2012; Europees Platform, n.d.). In the majority of Early English schools, classes are taught by a non-native-speaker (NNS) class teacher (Thijs *et al.* 2011) and these differ considerably in their mastery of English. In the Netherlands, an intermediate English proficiency—B1 within the CEFR framework¹—at the end of teacher training colleges is desired, and a B2 level is preferred, but teacher self-assessments indicate that this level is not always attained (Edelenbos *et al.* 2004; Thijs *et al.* 2011; Persson 2012).

Because early English education in the Netherlands is a bottom-up development, with limited governmental regulation, there is much variety in the amount of classroom exposure and teachers' English language proficiency. How this affects pupils' language development has not yet been subject to systematic investigation. This article aims to fill this gap by exploring the potential impact of these factors on children's receptive grammar and vocabulary skills at the end of their first and second year of instruction. The article is organized as follows. In the next section, we review some of the available literature on (very) early language learning programmes focusing specifically on the role of amount of input and teachers' language proficiency, as well as briefly considering other factors such as aptitude. Subsequently, data are presented from a study of 168 Dutch-speaking children learning English in the first 2 years of primary school. Finally, in the discussion section, we summarize the main findings and consider how they relate to previous studies; we furthermore outline possible implications for the implementation of early EFL programmes in the Netherlands. Note that our goal is not to evaluate whether early EFL programmes are more successful than programmes where EFL instruction starts later; irrespective of whether this is likely to be the case (see Muñoz 2008 for relevant discussion), we take the existence of such programmes as a given, and concern ourselves with the contribution of amount and type of input to early foreign language acquisition within EFL programmes in the Netherlands.

2. THE IMPACT OF AMOUNT OF INPUT AND TEACHERS' LANGUAGE PROFICIENCY IN (VERY) EARLY LANGUAGE LEARNING

In the Netherlands, as elaborated upon in the introduction, there is considerable variation in how early EFL is implemented, both in terms of input

quantity, that is, the amount of classroom time spent on English, and input quality, that is, whether teachers are native speakers (NSs) or NNSs of English and if non-native, how proficient they are in the foreign language. Establishing what constitutes a successful early EFL programme (in the Dutch context) entails establishing the extent to which these factors contribute to children's language outcomes.

Perhaps surprisingly, given the variation in how early language learning programmes are implemented across European countries (Edelenbos *et al.* 2006), rather little is known about how the amount of input and teacher's language proficiency may influence children's foreign language skills. While both length of exposure and intensity of exposure have recently been found to contribute to children's development of lexical and grammatical knowledge (Rohde 2010; Steinlein *et al.* 2010), in their recent review of early teaching and learning of foreign languages, Nikolov and Mihaljevic Djigunovic (2011: 97) write that 'no comparative study was found on how weekly exposure impacts on outcomes'.

In the Dutch context, Goorhuis-Brouwer and de Bot (2010) observe significant improvement in the English language skills of children with 3 hours of English after 1 year of English instruction; however, in the second year of instruction, no significant improvement was found. Many schools provide markedly less English input than this (Thijs *et al.* 2011). In a study involving a school with 1 hour of English classes per week, children were found to have only a very basic understanding of English vocabulary at the end of the first 2 years of primary school, and productive skills were virtually absent (Aarts and Ronde 2006). Thus, while time has been reported to be the most reliable predictor of children's success in early language learning programmes (Edelenbos and Johnstone 1996; Edelenbos *et al.* 2006), it remains unclear how much input is enough to lead to significant foreign language gains (de Bot and Philipsen 2007).

Investigations of teacher proficiency have largely been restricted to teachers' use of the first and second languages in the classroom (Inbar-Lourie 2010). Little attention has been paid to the (potential) relationship between children's developing foreign language skills and their teachers' own (foreign) language proficiency, quite possibly because, as Nikolov and Mihaljevic Djigunovic (2011: 107) point out, this is often seen as a delicate issue. In cases where NS teachers are compared with NNS teachers, differences in learner's foreign language skills were only found after several hundreds of hours of English education (Larson-Hall 2008; see also Árvá and Medgyes 2000 for one of the few studies on NS–NNS comparisons). In a study of input provided by teachers in bilingual preschool programmes, Weitz *et al.* (2010) observe that quality of language input, as measured in terms of didactic skills and strategies employed, had a greater impact on children's linguistic development than the amount of input (see also Bondi 2001; Jiménez Raya 2001; Driscoll *et al.* 2004; Enever 2011 for more on the desired profile of early foreign language teachers).

There are of course various other factors which have been found to affect child second language learning and acquisition, including, for example, children's socio-economic background, teachers' didactic skills and methodological

knowledge, language aptitude, and out-of-school exposure. In a recent study, Abrahamsson and Hyltenstam (2008: 500) observe that language aptitude, broadly defined as a specific talent for learning languages (Dörnyei and Skehan 2003), plays a 'significant (albeit modest) role in child second language acquisition'. Aptitude is generally thought to consist of both memory and analytical components (Carroll 1981); in research on early second/foreign language learning, developing L2 proficiency has been related to both (Genesee and Hamayan 1980; Alexiou 2009; Paradis 2011), and in particular to phonological short-term memory (Service 1992; Harley and Hart 1997; Masoura and Gathercole 1999; French and O'Brien 2008). Another factor observed to contribute to young learners' foreign language proficiency is out-of-school exposure, particularly through television media (Blondin *et al.* 1998; d'Ydewalle and Van de Poel 1999; Kuppens 2010; Lindgren and Muñoz 2013).

To summarize, research on early language learning in general and on early EFL programmes in the Netherlands specifically has not yet directly addressed the question of whether and how teachers' language proficiency and amount of weekly exposure affect children's developing language skills. The available evidence suggests that while children's language skills do develop in these limited input situations, this development may be restricted in terms of rate of acquisition. It is furthermore unclear how amount of input and teachers' language proficiency interact with other factors such as out-of-school exposure and aptitude.

3. RESEARCH QUESTIONS

The present study seeks to contribute to this line of research by addressing the following questions. In the context of the early EFL programmes in the Netherlands, where amount of input is generally limited:

- 1 To what extent do children develop their receptive vocabulary and grammar skills during the first and second year of instruction?
- 2 To the extent that such development exists, is it related to differences in quantity and quality of English input at school, as measured in terms of number of minutes of weekly classroom exposure and teachers' language proficiency?
- 3 To what extent do other factors, such as out-of-school exposure and aptitude, as measured by phonological short-term working memory, predict children's English language outcomes in these early stages of development?

4. THE STUDY

4.1 Participants

Data were collected from 168 young foreign language learners (87 girls) enrolled at early English schools ($n=14$) as well as 26 age-matched control

children from three different schools without early EFL programmes. Eleven children were exposed to a language other than or in addition to Dutch at home (but not English). The level of maternal education varied from technical and vocational training to university education.

In order to obtain an appropriate cross-sample of early EFL learners in the Netherlands, primary schools from different provinces of the Netherlands were included in the study. The participating early EFL schools had at least 2 years of experience with teaching English at the start of the experiment. The teachers reported using a combination of sources for their lesson material, including EFL textbooks specifically designed for young learners, such as *My Name is Tom*, material taken from websites and shared by colleagues. Teachers regularly made use of such resources to develop their own materials specific to the themes currently being addressed in class in general, such as winter and animals. Data collection took place in the autumn, at the beginning of the school year, when the children had just started primary education (baseline), at the end of the same school year, after having had 1 year of English lessons (post-test I), and again 1 year later (post-test II). The average age at the baseline measurement for the early EFL children was 4;4 (SD 0;3) and for the control children this was 4;3 (SD 0;2).

4.2 Establishing input quantity and teachers' language proficiency

Information was gathered from teachers concerning their proficiency level in English and the amount of input provided to children at school, that is, the total number of minutes of English classes per week. The teachers' speaking skills in English were estimated on the 6-point scale from the Common European Framework of Reference (Council of Europe 2009)¹ for *communication* and *interaction*, the components most relevant to the input children are likely to receive in the classroom. At each of the three test moments, teachers were asked to indicate which Can-Do statements of the CEFR scales most accurately described their English proficiency. In addition, at the baseline, trained native-speaker judges assessed a subset of the teachers' proficiency skills on the basis of recorded informal interviews. The correlations between the self-assessment and native-speaker assessments were high ($r = .89$, $p < .001$ for communication and $r = .90$, $p < .001$ for interaction), suggesting that the self-assessments were indeed reliable. The average teacher English language proficiency was C1 (range: A1 to C2 and natives), and children in the present study had—averaged over the 2 years—approximately 86 min of early English instruction per week (range: 40–225 min). For most schools, these minutes were spread out over several or all 5 days in the school week. It should be noted that the average NNS teachers' English language proficiency is likely to be lower than C1; the average in our sample is skewed by the inclusion of the NS teachers.

A parental questionnaire was used to collect additional biographical information and to estimate the amount of out-of-school exposure to English, for example, via television and other media (completion rate: 67%). The average number of minutes of out-of-school exposure per week was 38 min (SD 67 min) at the baseline, 144 min (SD 153 min) at post-test I and 155 min (SD 154 min) at post-test II. As the standard deviations indicate, there is considerable individual variation with respect to this variable, especially at the two post-tests. This is largely due to increased television viewing as the children grew older, and potentially also to the fact that as children become more proficient in English, many of them (or their parents) are likely to seek out more input opportunities outside of school.

4.3 Tasks and procedure

Two standardized tasks were used to assess children's English language development: to measure receptive vocabulary skills, the Peabody Picture Vocabulary Test—Fourth edition (PPVT-4; Dunn and Dunn 2007), and to measure receptive grammar skills, the Test for Reception of Grammar Version 2 (TROG; Bishop 2003). We opted for receptive rather than productive tasks because previous research (Aarts and Ronde 2006) indicated that productive skills were likely to be limited during these first 2 years of English language instruction, at least for those children with comparatively little input, and hence production tasks were likely to result in floor effects. Both tasks are picture selection tasks where children have to select one picture out of an array of four which best corresponds to either a word (PPVT) or sentence (TROG) which they are presented with auditorily.^{2,3} For example, in the PPVT, a child sees a picture of a pack of crayons, a cat, a brush, and a sock, and is asked *Where is 'cat'?*, or in the TROG, the child sees a picture of a sheep standing, a cat sitting, a boy running, and a sheep running, and is asked to point to *The sheep is running*. Raw scores are used in the analysis.

To assess the memory component of aptitude, and more specifically phonological short-term memory, the phonological memory sub-tests (forward digit span and non-word repetition) of the *Comprehensive Test of Phonological Processing* (CTOPP; Wagner *et al.* 1999) were used. As noted by Paradis (2011: 223), testing phonological memory abilities in the children's non-native language will likely depress their scores, but given that the range of scores obtained is reasonable, this should not interfere with our analyses. To control for further individual variation in terms of non-verbal working memory and non-verbal intelligence, the hand movement part of the Kaufman Assessment Battery for Children (Kaufman and Kaufman 1983) and Raven's coloured progressive matrices (Raven *et al.* 1995) were also administered.

The tests were conducted as part of a larger test battery, and were administered individually. All items on the English language tests were pre-recorded

by NSs. The vocabulary test was always administered first. Scoring procedures followed the test manuals.

4.4 Data analysis

The data were analysed as follows. To assess the role of weekly classroom exposure ('minutes/week'), the EFL children were divided into three groups: the children in the first group had 1 hour or less ('60 min or less', $n = 57$) of weekly classroom exposure to English, the children in the second group had more than 1 hour but less than 2 ('60–120 min', $n = 65$), and the children in the third group had 2 hours or more of weekly classroom exposure to English ('120 + min', $n = 39$). This division was chosen partly in order to obtain relatively even groups, but more importantly to reflect the most common practices within schools offering early EFL in the Netherlands. To assess the role of teachers' language proficiency, EFL children were divided into four groups, again reflecting the most frequently occurring combinations. The children in the first group were taught by a NS teacher only ($n = 41$). The children in the second group were taught by a team consisting of an NS teacher alongside one or more NNS teachers who had an English proficiency level of CEFR-A and/or CEFR-B level ($n = 16$). The children in the third group were taught by an NNS teacher of English at CEFR-C level ($n = 43$), and finally, the fourth group consisted of children with an NNS teacher with an English proficiency level of CEFR-B ($n = 22$).⁴

For both variables, MANOVA analyses were conducted with time (baseline, post-test I, post-test II) as within-subjects variable and the relevant variable (minutes/week or teacher proficiency group) as between-subjects variable. Because the teacher proficiency groups differ in number of minutes of English per week, this latter variable was included in the analysis as a covariate. Likewise, teacher proficiency (on a 7-point scale from A1 through C2 to NS, averaged over teachers and over time) was also included as a covariate for the analysis with minutes/week as main factor. All univariate and multivariate outliers ($z > 3.3$) were removed, and to address problems with normality, the scores on the receptive grammar test were transformed (square root).

Finally, multiple regression analyses were used to determine which factors best predict children's outcomes at post-test I and post-test II for grammar and vocabulary separately.⁵ All independent variables which significantly correlated with the dependent variable were entered into a multiple regression analysis in the following steps: (i) scores on the same test at earlier test moment(s), (ii) invariable biographical factors (age, sex, maternal education), (iii) scores at same test moment on tests possibly contributing to individual variation but not directly measuring language abilities (phonological working memory, non-verbal working memory, and non-verbal intelligence), (iv) out-of-school exposure at same test moment, and (v) minutes/week and teacher proficiency at same test moment. In this analysis, we used dummy variables for teacher proficiency; 'NS teacher only' was used as the baseline

category, that is, the group with which all other groups are compared. Due to space limitations, we report only the details of the (significant) model which explains the most variance.

5. RESULTS

5.1 Weekly classroom exposure

Figures 1 and 2 present the mean raw scores for the three measurements for English vocabulary and grammar, respectively, when the children are divided according to weekly classroom exposure groups.

There was a main effect of time [vocab: $F(2,523) = 60.9$, $p < .001$, $n_p^2 = .19$; grammar: $F(2,523) = 56.5$, $p < .001$, $n_p^2 = .18$] and of group [vocab: $F(3,523) = 7.29$, $p < .001$, $n_p^2 = .04$; grammar: $F(3,523) = 20.1$, $p < .001$, $n_p^2 = .11$]; for vocabulary, there was no significant interaction between these two factors [$F(6,523) = 1.71$, $p = .12$, $n_p^2 = .02$], but a significant interaction was found for grammar [$F(6,523) = 4.87$, $p < .001$, $n_p^2 = .05$]. Post-hoc (Bonferroni) tests reveal that, averaged over time, students with 60 min or less score significantly lower than the other two groups ($p < .01$ for all pairwise comparisons). Furthermore, the vocabulary score of the control children was significantly lower than all the EFL groups ($p < .05$), except for the group with 60 min or less of English per week (MD (mean difference) = $-.62$, $p = 1.0$); on grammar, the control group scores lower than all the EFL groups ($p < .05$).

When teachers' language proficiency is included as a covariate and the above analysis repeated without control children, the overall results are the same. There is a main effect of time [vocab: $F(2,448) = 101.9$, $p < .001$, $n_p^2 = .32$; grammar: $F(2,448) = 88.3$, $p < .001$, $n_p^2 = .29$], group [vocab: $F(2,448) = 5.45$, $p = .005$, $n_p^2 = .02$; grammar: $F(2,448) = 12.0$, $p < .001$, $n_p^2 = .05$], with a significant interaction between the two for grammar [$F(4,448) = 5.33$, $p < .001$, $n_p^2 = .05$] but not for vocabulary [$F(4,448) = 1.48$, $p = .21$, $n_p^2 = .01$]; and there is also a significant effect of the covariate [vocab: $F(1,448) = 8.85$, $p = .003$, $n_p^2 = .02$; grammar: $F(1,448) = 6.99$, $p = .008$, $n_p^2 = .02$].

5.2 Teachers' language proficiency

Figures 3 and 4 present the mean raw scores for the three measurements for English vocabulary and grammar, respectively, when the children are divided according to teachers' language proficiency groups (EFL children only).

There was a main effect of time [vocab: $F(2,328) = 65.6$, $p < .001$, $n_p^2 = .29$; grammar: $F(2,328) = 54.1$, $p < .001$, $n_p^2 = .26$], and group [vocab: $F(3,328) = 5.08$, $p = .002$, $n_p^2 = .05$; grammar: $F(3,328) = 3.70$, $p = .012$, $n_p^2 = .03$]. For vocabulary there is a significant interaction between the two [$F(6,328) = 2.92$, $p = .009$, $n_p^2 = .05$] and for grammar, the interaction is approaching significance [$F(6,328) = 2.04$, $p = .061$, $n_p^2 = .04$]. This interaction is due to the children with an NNS English teacher at CEFR-B level developing more slowly than the

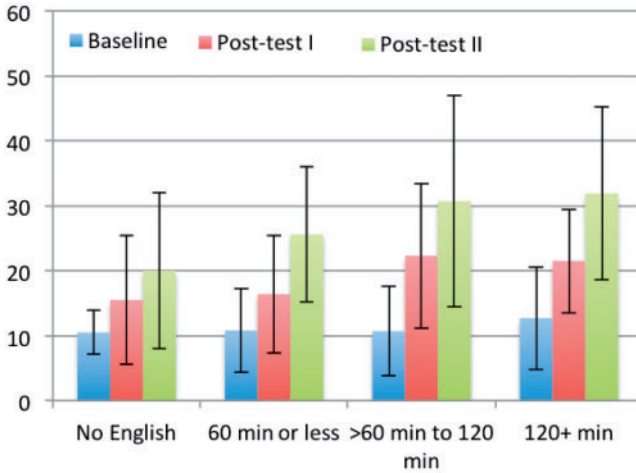


Figure 1: Average raw scores on receptive vocabulary in English (error bars ± 1 SD): number of minutes of English per week

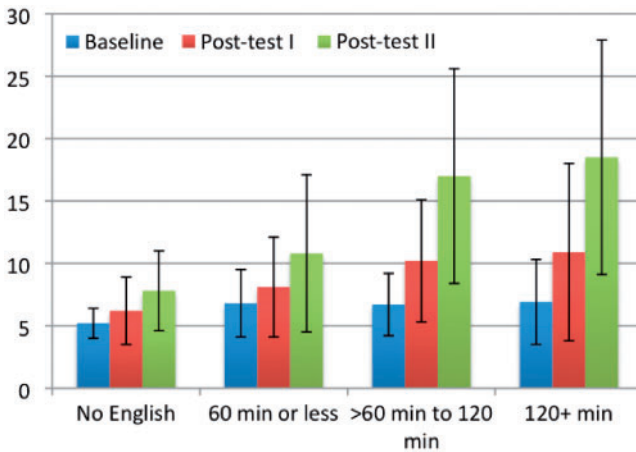


Figure 2: Average raw scores on receptive grammar in English (error bars ± 1 SD): number of minutes of English per week

other children: averaged over time, this group scored significantly lower than the other groups ($p < .05$ for all pairwise comparisons, except for with the group with an NS teacher only on grammar). Between the other three groups, there are no significant differences. The covariate, min of weekly exposure, is also significant [vocab: $F(1,328) = 5.98$, $p = .015$, $n_p^2 = .02$; grammar: $F(1,328) = 10.9$, $p = .001$, $n_p^2 = .03$].

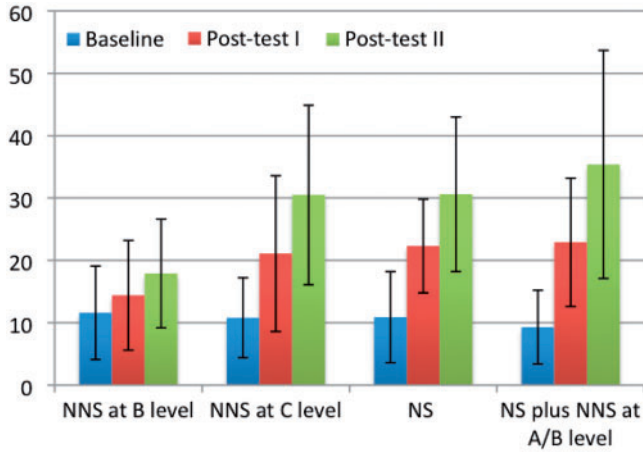


Figure 3: Average raw scores on receptive vocabulary in English (error bars ± 1 SD): teachers' language proficiency

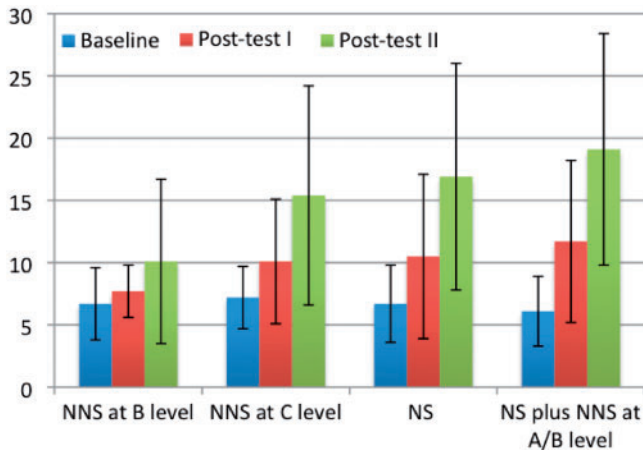


Figure 4: Average raw scores on receptive grammar in English (error bars ± 1 SD): teachers' language proficiency

5.3 Weekly classroom exposure, teachers' language proficiency and other factors

Table 1 presents an overview of the significant correlations between the independent and dependent variables.

The results of the regression analysis for vocabulary are presented in Tables 2 and 3 for post-test I and II, respectively. On post-test I, 16% of the variance is explained by teachers' English proficiency, that is, whether children

Table 1: Overview of the independent variables which significantly correlate with language outcomes (*r*-values)

| Factor | Vocabulary | | Grammar | |
|---|-------------|--------------|-------------|--------------|
| | Post-test I | Post-test II | Post-test I | Post-test II |
| Score on the same test at baseline | .16* | .21** | .25*** | .14* |
| Score on the same test at post-test I | — | .33*** | — | .51*** |
| Age | .22** | — | — | — |
| CTOPP (non-word repetition) | .16* | .26*** | .14* | .27*** |
| CTOPP (digit span) | — | — | .17* | .19** |
| Non-verbal working memory | — | .14* | — | .16* |
| Non-verbal intelligence | — | — | — | .28*** |
| Out-of-school exposure | — | .23** | .17* | .22** |
| Minutes of weekly classroom exposure | — | — | — | .21** |
| Teachers' language proficiency: dummy variable NNS-B | -.27** | -.36*** | -.20* | -.31*** |
| Teachers' language proficiency: dummy variable NS + NNS-A/B | — | .18* | — | .17* |

* $p < .05$; ** $p < .01$; and *** $p < .001$

Table 2: Results of regression analysis: vocabulary, post-test I

| Factor | Unstandardized coefficients | | Standardized coefficients |
|--|-----------------------------|----------------|---------------------------|
| | <i>B</i> | Standard error | β |
| Constant | 17.2 | 1.90 | |
| Score on the same test at baseline | .15 | .14 | .10 |
| Age | .68 | .35 | .19* |
| CTOPP (non-word repetition) | .62 | .36 | .17 |
| Teachers' language proficiency: dummy variable NNS-B | -7.27 | 2.48 | -.28** |

Adjusted $R^2 = .16$ [$F_{change}(1,96) = 8.59, p = .004$]; * $p = .05$, ** $p < .01$

were taught by an NNS teacher at CEFR B-level compared with an NS teacher, and at post-test II, 31% of the variance is explained by teachers' language proficiency, children's aptitude, and scores on the baseline.

For grammar, the regression analysis did not result in any significant model at post-test I. Table 4 presents the results for post-test II. A large proportion of

Table 3: Results of regression analysis: vocabulary, post-test II

| Factor | Unstandardized coefficients | | Standardized coefficients β |
|--|-----------------------------|----------------|--------------------------------------|
| | <i>B</i> | Standard error | |
| Constant | 5.12 | 5.82 | |
| Score on the same test at baseline | .39 | .18 | .20* |
| Score on the same test at post-test I | .22 | .13 | .16 |
| CTOPP (non-word repetition) | 1.17 | .49 | .21* |
| Non-verbal working memory | .54 | .47 | .10 |
| Out-of-school exposure | .009 | .006 | .14 |
| Teachers' language proficiency: dummy variable NNS-B | -10.1 | 3.30 | -.26** |
| Teachers' language proficiency: dummy variable NS + NNS-A/B | 3.15 | 3.85 | .08 |

Adjusted $R^2 = .31$ [$F_{change}(2,92) = 5.38, p = .006$]; * $p < .05$, ** $p < .01$

Table 4: Results of regression analysis: grammar, post-test II

| Factor | Unstandardized coefficients | | Standardized coefficients β |
|--|-----------------------------|----------------|--------------------------------------|
| | <i>B</i> | Standard error | |
| Constant | -.14 | .70 | |
| Score on same test at baseline | -.02 | .18 | -.01 |
| Score on same test at post-test I | .52 | .13 | .37*** |
| CTOPP (non-word repetition) | .06 | .04 | .15 |
| CTOPP (digit span) | -.01 | .06 | -.01 |
| Non-verbal working memory | .04 | .03 | .11 |
| Non-verbal intelligence | .04 | .02 | .16 |
| Out-of-school exposure | .00 | .00 | .07 |
| Minutes of weekly classroom exposure | .00 | .00 | .12 |
| Teachers' language proficiency: dummy variable NNS-B | -.50 | .25 | -.18* |
| Teachers' language proficiency: dummy variable NS + NNS-A/B | .22 | .28 | .07 |

Adjusted $R^2 = .42$ [$F_{change}(3,88) = 3.00, p = .035$]; * $p < .05$, *** $p < .001$

the variance (42%) is explained by two variables, mostly by children's scores on the same test at post-test I, but also by teachers' English proficiency, that is, whether children were taught by an NNS teacher at CEFR B-level compared with an NS teacher.

As noted above,² the PPVT (the vocabulary test used here) is often used for bilingual and second language learners even though it was developed and normed for monolingual English-speaking children. We can use these norms to transform the raw scores used in the analyses thus far into age equivalents. The control children have an age equivalent at the baseline measurement of 2 years and 3 months. At the post-test II, this becomes 2 years and 5 months. For children with 60 min or less English per week, the age equivalents are 2 years 2 months at baseline and 2 years 6 months at post-test II; the children in the other two EFL groups also begin at 2 years 2 months and end at 2 years 9 months. If we assume that the control children can be taken as an indicator of how English language proficiency develops in children who do not receive English instruction at school, then we can infer that children with more than 60 min per week acquire as much English in 2 years' time as young monolingual English-speaking children do in approximately 5 months.

5.4 Summary

To summarize, the children in early EFL programmes scored significantly higher than children who were not in such programmes and when analysed in groups, after controlling for the effect of teachers' language proficiency, those with more than 60 min of weekly classroom exposure scored on average significantly higher than those with 60 min or less. There was a significant improvement in scores after 1 and 2 years of instruction, and both these findings were observed for both receptive grammar and receptive vocabulary. When divided into teacher proficiency groups, and after controlling for minutes/week, children with an NNS teacher at CEFR-B level only were found to score significantly lower than the other groups and they developed significantly more slowly over time, at least for vocabulary. The results of the regression analyses suggest, however, that teachers' language proficiency rather than minutes/week is the best predictor of children's scores on both vocabulary and grammar, along with scores at earlier test moment(s), and for vocabulary at post-test II, phonological short-term memory.

6. DISCUSSION

In this article, we examined data from two receptive language tasks to investigate the development of 4-year-old children's receptive vocabulary and grammar skills in English as a function of their participation in early EFL programmes in the Netherlands.

6.1 Development over time

Our first research question asked to what extent children would significantly improve their receptive language skills during their first 2 years in an early EFL programme. The results showed that there was indeed a significant improvement between the baseline measurement upon children's entry to the programme and the two post-test measurements at the end of the first and second years of instruction, and that this improvement holds across all three quantity groups.

It is important to note that the scores for the control group also improve with time. This is not surprising, for the following reason: raw scores increase as children get older because the vocabulary of a child grows with age (Fenson *et al.* 2007). There are many words in English with cognates in Dutch, and because the Dutch vocabulary of the control children improves with age, this will likely help them with English. In a recent study comparing early language learning across seven different European countries, cognate linguistic distance was indeed found to be a significant predictor of receptive language skills (Lindgren and Muñoz 2013). The availability of cognates from Dutch applies to all children, irrespective of whether they participate in an EFL programme, and this means that a part of the improvement in the scores of the EFL children between the baseline and post-test I and II must be due to the fact that they are older, and consequently, their Dutch is better, and because Dutch and English are closely related languages, many words and constructions are the same.

After taking this age-related development into account, we nevertheless see that the EFL children still improve over time. More specifically, when we examined the age-equivalents for the standard scores on the vocabulary test, children with more than 60 min per week were found to acquire as much English in 2 years' time as young monolingual English-speaking children do in approximately 5 months. At first blush, that does not seem much, but it is considerable because EFL children hear and use *much* less English than monolingual English-speaking children immersed in an English-speaking environment all the time (Goorhuis-Brouwer and de Bot 2010).

In and of itself, the observation that there is a significant improvement after 2 years within an early EFL programme is promising in terms of the effectiveness of such programmes, but whether the magnitude of this improvement is sufficient depends on the programme's goals, and this can only be fully examined with data from older children who have had a longer period of instruction. As noted in the introduction, the goal of the present study was to document EFL development in the early stages of such programmes and to assess which factors contribute to this, not to assess the efficacy of early EFL programmes compared with a later start.

6.2 Amount of weekly classroom exposure

Our second research question asked whether differences between children reflect differences in amount and quality of English input at school, as

measured by minutes of weekly classroom exposure and teachers' language proficiency. In terms of weekly classroom exposure, we observed that all children participating in early EFL programmes scored significantly higher than the control children for grammar, but for vocabulary, there was no significant difference between the control children and the EFL children with 60 min or less of lessons per week. The observation that the EFL children almost always score higher than the controls is perhaps not so surprising, but it is an important result nonetheless because it shows that despite the sometimes rather limited amount of time devoted to English on a weekly basis (on average 52 min in the lowest input group), this additional exposure does appear to have an effect. After all, all three groups should really be considered as 'minimal input' groups when compared with immersion children. The finding that the children in the lowest input group and the control children do not significantly differ on vocabulary suggests that 60 min or less of English lessons per week may not provide sufficient input to go beyond that available outside of school and/or the helping hand which the children's closely related first language apparently provides is enough to acquire a basic vocabulary in English.

With respect to amount of input, more fine-grained measures may be more revealing. Teachers may, for example, vary in terms of their relative use of the first and second languages (Inbar-Louie 2010), and in terms of how much input they provide and how rich and complex this is. At least the first of these two factors, as measured by total number of words produced by teachers, has been found to be positively related to early language learner's vocabulary growth (Bowers and Vasilyeva 2011). Similar effects have also been found to transpire over time; in a study on the lexical development of 4-year-old children learning Norwegian as a second language at primary school, Aukrust (2007) observes that while properties of teachers' input such as the number of different types and tokens per minute did not correlate with children's language abilities at the time, they were significant predictors of children's outcomes 2 years later. In addition, in the analysis presented here we did not take into account the intensity with which lessons took place; in other settings, intensity of language exposure, that is, the frequency with which lessons take place, has been found to be significantly related to children's language outcomes (Weitz *et al.* 2010). Furthermore, it is important to note that there may be a difference between allocated time and engaged time (Carroll 1963), that is, having 60 min of English lessons per week does not necessarily mean that the language-learning child attends to the input for this same amount of time.

6.3 Teachers' language proficiency

With respect to teachers' language proficiency, we found that after taking into account the variance explained by the variable minutes/week, the children with an NNS teacher at CEFR-B level only, scored significantly lower than children with an NNS teacher at the same level who co-teaches with an NS

teacher, children with an NS only, and children with an NNS teacher with a higher level of English language proficiency. Teachers' language proficiency, and more specifically the comparison between the children with an NNS B-level teacher only and those with an NS teacher only, was also found to be a significant predictor of children's scores at both post-test I and post-test II for vocabulary, and at post-test II for grammar. As mentioned above, previous studies have suggested that the level of lexical diversity and grammatical complexity provided by teachers may impact upon EFL children's language learning (Bowers and Vasilyeva 2011); assuming that this is indeed the case, the results of the regression analysis reflect the richer input offered by NS teachers compared with the limited input provided by lower proficiency (i.e., CEFR-B) NNS teachers.

While it is always difficult to interpret the non-significant differences between groups, what the present results seem to suggest—assuming that it is indeed robust—is that input from an NS teacher is not essential to reach a higher level of English language skills (because the children with a C-level NNS teacher scored similarly to the children with an NS only) but input from an NS teacher appears to compensate for input from a lower proficiency (B-level) NNS teacher. This (tentative) conclusion needs to be qualified as there is a confound in our sample between NS status and specialist teacher status. What this means is that we cannot tease apart the effect of NS input from the potentially crucial didactic insights of a specialist teacher trained to teach foreign languages to young learners. To the extent that the relevant groups exist, that is, children with NS teachers who are not specialists, this qualification could in principle be investigated empirically.

In this article, we have focused on teachers' oral *language* proficiency and its impact on children's language learning. It is of course the case that teacher proficiency involves much more than language proficiency alone. For example, factors such as the teacher's ability to use age-appropriate language, their use of L1 vs. L2 in the classroom, their use of feedback, focus on form/meaning, as well as their beliefs about learning and language use may all contribute to children's language learning (see Nikolov and Mihaljevic Djigunovic 2011 for review). While including these factors here goes beyond the scope of the present study, we fully acknowledge that it is quite possible that they may explain some of the (remaining) variance in our data. In the context of the larger project from which the data in the present study stem, lesson observation data have been collected at post-test II for a subset of the schools and teachers in order to capture some of the inter-teacher differences in terms of didactic and interaction skills (in a similar vein to Weitz *et al.* 2010) and to obtain more information concerning the input teachers provide to children, for example, in terms of number of different words and average length of utterance (following Bowers and Vasilyeva 2011). A complete analysis of these data is the subject of ongoing research. Related to this, we should acknowledge that variation in teaching/learning materials may also account for some of the variance in children's outcomes. It is important to note that due to the

dynamic and emerging nature of EFL education in Dutch primary school education, and the exploratory nature of the present study necessitated by this fact, it was not possible to keep these factors constant across groups.

6.4 Other factors

There was considerable individual variation in the EFL children's scores on the receptive vocabulary and grammar tests. In order to determine which factors best predict this variation, a series of multiple regression analyses were conducted. For vocabulary, the analysis explained 16% of the variance at post-test I and 31% at post-test II, and most of this was due to teachers' language proficiency, which also contributed to the 42% of variance explained at post-test II for grammar. The potential significance of this finding was discussed above.

For both vocabulary and grammar, children's scores on the same test at an earlier moment (the baseline for vocabulary and post-test I for grammar) were significant predictors of children's scores at post-test II. A similar finding is observed by French and O'Brien (2008) in a study of French-speaking children participating in a 5-month intensive English programme, that is, children's vocabulary scores at the start of the programme were significant predictors of the scores at its end. These authors suggest that as a result of ambient exposure to English, children may have already reached some critical level of vocabulary knowledge at the start of this programme. While it is possible that a similar explanation may hold for the children in the present study, out-of-school exposure at the baseline does not significantly correlate with children's baseline scores on vocabulary ($r = .06$, $p = .50$) or on grammar ($r = .12$, $p = .15$).⁶ It thus seems that in the present study, some other factor(s) must explain differences at the baseline.

In addition to previous test scores and teachers' language proficiency, phonological short-term working memory, as measured by the non-word repetition sub-test of the CTOPP, was also found to be a significant predictor of children's vocabulary scores at post-test I. This is in line with previous research showing that language aptitude (Genesee and Hamayan 1980; Alexiou 2009), and in particular, its memory component (Service 1992; Harley and Hart 1997; Masoura and Gathercole 1999; French and O'Brien 2008), is a significant predictor of early foreign language learners' proficiency.

In contrast to what has previously been observed (Lindgren and Muñoz 2013), however, out-of-school exposure—while significantly correlating with vocabulary scores at post-test I and grammar scores at both post-test I and II—was not found to be a significant predictor of children's outcomes. This may be because data for this variable were only available for a subset of the children; it is also possible that as children become older, and presumably spend more time, for example, on the internet or playing computer games, their exposure to English outside of school may increase, and hence the relationship between this variable and children's developing English language

skills may change over time, that is, at ages 4 and 5 years, the children are too young for this variable to be a significant predictor of their English language skills. Indeed, the children in the Lindgren and Muñoz study were older than those here.

6.5 Limitations

The present study reports the results of two tasks only; while they each target different linguistic domains—with no real differences in results between the two—they both test the same modality. In addition, as a consequence of being standardized tasks, from a linguistic point of view, they are both rather general in nature. For a more complete assessment of children's developing English language skills, more specific tasks tapping into more detailed aspects of the children's linguistic knowledge and examining productive as well as receptive skills are required.⁷

To completely understand the contribution of input quantity and teacher proficiency, more data are needed from lower proficiency NNS teachers, and for each teacher proficiency group, children should be included with varying amounts of weekly exposure. Indeed, in the original set-up of this study, that is exactly what we had intended to do; it turned out, however, that in our sample, which appears to be a reasonably accurate reflection of the current teaching provision in early English programmes in the Netherlands, not all such groups in fact exist, that is, where more time is devoted to teaching English, teachers are more likely to be NSs and consequently, children following EFL programmes for more than 2 h with a low proficiency NNS teacher do not appear to exist.

Finally, as noted above, our investigation of input quality was restricted to just one aspect of this variable, namely teachers' English language proficiency. For a more complete investigation, more factors need to be incorporated.

7. CONCLUSION

In this article we investigated the impact of input quantity and quality in early foreign language learning in the Netherlands. Our results indicate that when grouped according to number of minutes devoted to English per week, more than 60 min of English classes per week appear to be necessary to lead to significant gains, at least on the tasks used here. With respect to quality of input, measured in terms of teachers' self-assessment of their spoken proficiency, the results suggest that children taught by intermediate level (CEFR B-level) non-native teachers only score significantly lower than higher proficiency NNS teachers, NS teachers, or NNS teachers of the same proficiency level but who teach alongside an NS. The present findings underscore the importance of teacher's own proficiency level, even in such a 'minimal input' context, while at the same time indicating that NS teachers are not necessary for development in the children's English language development

to occur. Language aptitude was also found to be a significant predictor for vocabulary; this result suggests that it may be worth paying more attention to individual child-related factors, such as aptitude, in future studies on early foreign language learning. Whether these findings hold for other language skills, for data collected using different tasks, or for the same children when tested after more than 2 years in an early English programme will be the subject of future research. By systematically investigating the variability inherent to early language learning classrooms, both in terms of the quantitative and qualitative aspects of the available linguistic input as well as other teacher- and child-related factors, we endeavour to formulate well-founded and empirically feasible recommendations for how best to implement early EFL programmes in the Netherlands and elsewhere.

ACKNOWLEDGEMENTS

This study was supported by funding from the Dutch Ministry of Education, Culture, and Science, the Europees Platform, and Utrecht Institute of Linguistics (Uil OTS). The authors are indebted to participating schools for allowing us to gather the data and to our student assistants for their sterling work in collecting the data. We would also like to thank Xiaoyan Xu for the statistical analyses, and three anonymous reviewers for their comments and suggestions.

NOTES

- 1 The Common European Framework of Reference for Languages (henceforth CEFR) helps to describe the level of proficiency on the basis of six broad levels. The levels range from basic (A1 and A2) through intermediate (B1 and B2) to advanced foreign language proficiency (C1 and C2) (Council of Europe 2009).
- 2 Both these tasks were originally designed and normed for monolingual populations; they are, however, regularly used in research with bilingual and second language populations, and the edition of the vocabulary task which we used includes more items in the lower sets specifically to allow for differentiation between low proficiency English learners.
- 3 The PPVT-4 (vocabulary) has two versions available, A and B; these were alternated within children between test moments. The TROG (grammar) has only one version available and this was used on all three occasions.
- 4 To ensure large enough groups, the differences between the sub-levels, for example, C1 and C2, were ignored for the purposes of this analysis. A number of children were also excluded because (i) they did not form a sizeable group, that is, 8 children with an NNS teacher at A or A/B level or (ii) because they were impossible to categorize, that is, 14 children who had an NNS teacher at C-level at post-test I and a NNS teacher at A or B level at post-test II.
- 5 Because a number of the possible predictor variables, that is, minutes/week and teacher proficiency, vary at the class level, we explored the option of using a multilevel regression model. It was, however, the case that this data set does not contain enough variance at the class level to justify such an

analysis. We fully acknowledge, however, that with a larger data set, such an analysis may be possible and would likely be preferable.

- 6 We cannot rule out that the lack of a significant correlation with out-of-school exposure may be due to the non-normal distribution of scores for

this variable, that is, many children did not have any exposure to English out of school.

- 7 As noted above, the data presented here are part of a larger project, and this project incorporates several such tasks.

REFERENCES

- Aarts, R.** and **S. Ronde.** 2006. 'Tweetalig onderwijs met vervroegd Engels in het basisonderwijs,' *Levende Talen Tijdschrift* 7: 3–14.
- Abrahamsson, N.** and **K. Hyltenstam.** 2008. 'The robustness of aptitude effects in near-native second language acquisition,' *Studies in Second Language Acquisition* 30: 481–509.
- Alexiou, T.** 2009. 'Young learners' cognitive skills and their role in foreign language vocabulary learning' in M. Nikolov (ed.): *Early Learning of Modern Foreign Languages. Processes and Outcomes.* Multilingual Matters, pp. 46–61.
- Árva, V.** and **P. Medgyes.** 2000. 'Native and non-native teachers in the classroom,' *System* 28: 355–372.
- Aukrust, V. G.** 2007. 'Young children acquiring second language vocabulary in preschool group-time: Does amount, diversity, and discourse complexity of teacher talk matter?,' *Journal of Research in Childhood Education* 22: 17–37
- Bishop, D.** 2003. *Test for reception of grammar (TROG-2)* Version 2. Pearson Assessment.
- Blondin, C., M. Candelier, P. Edelenbos, R. Johnstone, A. Kubanek-German, and T. Taeschner.** 1998. *Foreign Languages in Primary and Pre-School Education: Context and Outcomes. A Review of Recent Research Within the European Union.* CILT.
- Bondi, M.** 2001. 'Towards a language profile for the primary foreign language teacher' in M. Jiménez Raya (ed.): *Effective Foreign Language Teaching at the Primary Level.* Peter Lang, pp. 39–50.
- Bowers, E. P.** and **M. Vasilyeva.** 2011. 'The relation between teacher input and lexical growth of preschoolers,' *Applied Psycholinguistics* 32: 221–41.
- Carroll, J. B.** 1963. 'A model for school learning,' *Teacher College Record* 64: 723–33.
- Carroll, J. B.** 1981. 'Twenty-five years of research in foreign language aptitude' in K. C. Diller (ed.): *Individual Differences and Universals in Language Learning Aptitude.* Newbury House, pp. 83–118.
- Council of Europe.** 2009. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (10th print.). Cambridge University Press.
- de Bot, K.** and **K. Philipsen.** 2007. 'Early English: Waarom (zo)?' in K. Philipsen, E. Deelder, and M. Bodde-Alderlieste (eds): *Early English: A Good Start.* Europees Platform voor het Nederlandse Onderwijs, pp. 105–15.
- Dörnyei, Z.** and **P. Skehan.** 2003. 'Individual differences in second language learning' in C. Doughty and M. Long (eds): *The Handbook of Second Language Acquisition.* Blackwell, pp. 589–630.
- Driscoll, P., C. Martin, L. Graham-Matheson, H. Dismore, and R. Sykes.** 2004. 'A systematic review of the characteristics of effective foreign language teaching to pupils between the ages 7 and 11,' *Research Evidence in Education Library.* EPPI-Centre, Social Science Research Unit, Institute of Education.
- Dunn, L. M.** and **D. M. Dunn.** 2007. *Peabody Picture Vocabulary Test*, 4th edn. Pearson.
- d'Ydewalle, G.** and **M. Van de Poel.** 1999. 'Incidental foreign-language acquisition by children watching subtitled television programmes,' *Journal of Psycholinguistic Research* 28: 227–44.
- Edelenbos, P.** and **R. M. Johnstone** (eds). 1996. *Researching Languages in Primary School. Some European Perspectives.* CILT.
- Edelenbos, P., J. H. A. L. De Jong, and G. J. Westhoff.** 2004. *Vreemdetalenonderwijs in Nederland: Een situatieschets.* Nationaal Bureau Moderne Vreemde Talen.

- Edelenbos, P., R. Johnstone,** and **A. Kubanek.** 2006. *Languages for the Children of Europe. Published Research, Good Practice and Main Principles.* Final Report of the EAC 89/04, lot 1 study. European Commission.
- Evener, J.** (ed.) 2011. *ELLiE Early Language Learning in Europe.* British Council.
- Europees Platform.** 'Overview of number of early english schools', available at <http://www.europeesplatform.nl/sf.mcgi?3632&cat=767>. Accessed 1 March 2013.
- Eurydice Network, Eurostat, and Education, Audiovisual and Culture Executive Agency.** 2008. *Key Data on Teaching Languages at School in Europe.* Education Audiovisual and Culture Executive Agency.
- Fenson, L., V. A. Marchman, D. Thal, P. S. Dale,** and **E. Bates.** 2007. *MacArthur-Bates Communicative Development Inventories (CDIs)*, 2nd edn. Brookes.
- French, L. F.** and **I. O'Brien.** 2008. 'Phonological memory and children's second language grammar learning,' *Applied Psycholinguistics* 29: 463–87.
- Genesee, F.** and **E. Hamayan.** 1980. 'Individual differences in second language learning,' *Applied Psycholinguistics* 1: 95–110.
- Goorhuis-Brouwer, S.** and **K. de Bot.** 2010. 'Impact of early english language teaching on L1 and L2 development in children in Dutch schools,' *International Journal of Bilingualism* 14: 289–302.
- Harley, B.** and **D. Hart.** 1997. 'Language aptitude and second language proficiency in classroom learners of difference starting ages,' *Studies in Second Language Acquisition* 19: 379–400.
- Herder, A.** and **K. de Bot.** 2005. *Vroeg Vreemdetalenonderwijs in Internationaal Perspectief. Literatuurstudie in Opdracht van het Europees Platform.* Expertisecentrum taal, onderwijs en communicatie, University of Groningen.
- Inbar-Louie, O.** 2010. 'English only? The linguistic choices of teachers of young EFL learners,' *International Journal of Bilingualism* 14: 351–67.
- Jiménez Raya, M.** 2001. 'Towards a teaching skills profile of the primary foreign language teacher' in M. Jiménez Raya (ed.): *Effective Foreign Language Teaching at the Primary Level.* Peter Language, pp. 25–38.
- Kaufman, A. S.** and **N. L. Kaufman.** 1983. *Kaufman Assessment Battery for Children.* Pearson.
- Kuppens, A. H.** 2010. 'Incidental foreign language acquisition from media exposure,' *Learning, Media and Technology* 35: 65–85.
- Larson-Hall, J.** 2008. 'Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation,' *Second Language Research* 24: 35–63.
- Lindgren, E.** and **C. Muñoz.** 2013. 'The influence of exposure, parents and linguistic distance on young European learners' foreign language comprehension,' *International Journal of Multilingualism* 10: 105–29.
- Masoura, E.** and **S. Gathercole.** 1999. 'Phonological short-term memory and foreign language learning,' *International Journal of Psychology* 34: 383–88.
- Muñoz, C.** 2008. 'Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning,' *Applied Linguistics* 29: 578–96.
- Nikolov, M.** and **J. Mihaljevic Djigunovic.** 2011. 'All shades of every color: an overview of early teaching and learning of foreign languages,' *Annual Review of Applied Linguistics* 31: 95–119.
- Paradis, J.** 2011. 'Individual differences in child english second language acquisition: comparing child-internal and child-external factors,' *Linguistic Approaches to Bilingualism* 1: 213–37.
- Persson, L.** 2012. 'Engels voor kleuters: Een eerste verkenning naar de vormgeving van vroeg vreemde-talenonderwijs in Nederland,' *Levende Talen Tijdschrift* 13: 25–32.
- Raven, J. C., J. H. Court,** and **J. Raven.** 1995. *Coloured Progressive Matrices Manual.* Lewis.
- Rohde, A.** 2010. 'Receptive L2 lexical knowledge in bilingual preschool children' in K. Kersten, A. Rohde, C. Schelleter, and A. K. Steinlein (eds): *Bilingual Preschools.* Vol. 1. Wissenschaftlicher, pp. 45–68.
- Service, E.** 1992. 'Phonology, working memory, and foreign-language learning,' *The Quarterly Journal of Experimental Psychology A* 45: 21–50.
- Steinlein, A. K., G. Håkansson, A. Housen,** and **C. Schelleter.** 2010. 'Receptive L2 grammar knowledge development in bilingual preschools' in K. Kersten, A. Rohde, C. Schelleter, and A. Steinlein (eds): *Bilingual Preschools.* Wissenschaftlicher, pp. 60–100.
- Thijs, A., D. Tuin,** and **B. Trimbos.** 2011. *Engels in het basisonderwijs.* SLO (Nationaal expertisecentrum leerplanontwikkeling).

Wagner, R., J. Torgesen, and **C. Rashotte.** 1999. *Comprehensive Test of Phonological Processing*. Pro-Ed.

Weitz, M., S. Pahl, A. Flyman Mattsson, A. Buyl, and **E. Kalbe.** 2010. 'The Input Quality Observation Scheme (IQOS): the

nature of L2 input and its influence on L2 development in bilingual preschools,' in K. Kersten, A. Rohde, C. Schelletter, and A. Steinlen (eds): *Bilingual Preschools*. Wissenschaftlicher, pp. 5–44.