



UvA-DARE (Digital Academic Repository)

The construct of language proficiency in the study of bilingualism from a cognitive perspective

Hulstijn, J.H.

Published in:

Bilingualism : Language and Cognition

DOI:

[10.1017/S1366728911000678](https://doi.org/10.1017/S1366728911000678)

[Link to publication](#)

Citation for published version (APA):

Hulstijn, J. H. (2012). The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism : Language and Cognition*, 15(2), 422-433. DOI: 10.1017/S1366728911000678

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <http://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

OPINION

The construct of language proficiency in the study of bilingualism from a cognitive perspective*

JAN H. HULSTIJN
Amsterdam Center for Language and Communication,
University of Amsterdam

(Received: September 13, 2011; final revision received: October 22, 2011; accepted October 27, 2011)

This article aims at revitalizing the debate concerning the measurement of language proficiency (LP) in the study of bilingualism (Grosjean, 1998). A review is presented of the way in which LP was measured in a corpus of 140 empirical papers published in volumes 1–14 (1998–2011) of the journal Bilingualism: Language and Cognition. In 55% of these papers, in which the assessment of LP as an independent or moderating variable was a necessary or preferred requirement, LP was not measured with an objective LP test. Seldom were participants' LP scores used in explaining variance obtained in the dependent variable(s). After the discussion of some unresolved problems concerning cross-language comparisons of LP in bilinguals' languages, recommendations are offered for the measurement of LP. One of the recommendations is that, in studies investigating between-group contrasts, researchers carefully consider the assessment of participants' proficiency in the language(s) concerned, even in native-speaker comparison groups.

Keywords: bilingualism, language proficiency, language dominance, native speaker

Introduction

In the first issue of this journal, Grosjean (1998) argued that some of the difficulties encountered in the study of bilingualism from a cognitive perspective “could have been lessened, if not avoided, had close attention been paid to methodological and conceptual issues” (p. 131). The first of those issues is participant selection. Grosjean laments the lack of information concerning the history, stability, function, proficiency of the languages of bilingual participants in many studies. Criteria for being bilingual differ between studies and the variability within groups of participants is not adequately dealt with. In an article reviewing functional neuroimaging studies of bilinguals, Abutalebi, Cappa and Perani (2001, p. 188) conclude that language proficiency seems to be the most important factor, more important than age of acquisition, affecting the bilingual language system. Interpreting the results of a study on interlingual homograph interference in German–English bilinguals, Von Studnitz and Green (2002a, p. 20) round their article off with the following remark: “Our findings point to the necessity of distinguishing different levels of control if we are to understand bilingual performance in the everyday use of language”. Thus, there is no doubt

that language proficiency, along with the other variables mentioned by Grosjean (1998), plays an important role in our understanding of language processing in bilinguals (see also Grosjean, Li, Münte & Rodriguez-Formelis, 2003, p. 160; Kroll, Bobb & Wodniecka, 2006, p. 128; Dunn & Fox Tree, 2009).

This paper focuses on one aspect of the issue of participant selection, namely the definition and measurement of (levels of) language proficiency (LP) and its implications for (a) the definition and operationalization of language dominance (and the notion of ‘balanced bilingual’) and (b) the selection of native-speaker control groups. The paper is organized as follows. First, I survey all papers that have appeared in volumes 1–14 of this journal, *Bilingualism: Language and Cognition*, in order to ascertain to what extent researchers have taken heed to Grosjean’s advice with respect to the measurement of LP in participant selection. The following section discusses the ways in which LP was measured in some studies in the corpus. The next section addresses what is perhaps the most difficult issue in the assessment of LP in studies on bilingualism, concerning the question of whether it is possible to provide empirical evidence for the claim that a given bilingual is equally proficient (in a given language skill, e.g. listening) in both of his

* Work on this paper was funded by the Netherlands Organisation for Scientific Research (grant 360-70-230).

Address for correspondence:

Amsterdam Center for Language and Communication, University of Amsterdam, Spuistraat 134, 1012 VB Amsterdam, the Netherlands
j.h.hulstijn@uva.nl

or her languages. The next section proposes a distinction between basic and higher language cognition (BLC and HLC), which might be helpful in the conceptualization and measurement of LP in the cognitive study of bilingualism. The paper is rounded off with a list of recommendations and a final remark.

In my reading of the literature on bilingualism, more often than not, the notion of language proficiency, be it in a first language (L1) or second language (L2), is often taken for granted, and so are the notion of language dominance and the notion of native speaker. This paper aims at revisiting the discussion of these notions and revitalizing the debate concerning their relevance or irrelevance in the study of bilingualism, raising more questions than it can answer. The spirit of the article is not only to improve research methodology where possible but, above all, to stimulate reflection on some fundamental theoretical issues in which LP plays a crucial role. Thus, the study of bilingualism forces us to reflect on what it means to be “lingual” in the first place.

Survey of empirical studies published in 14 volumes of bilingualism

The article corpus

As the papers in this journal, in the 14 years of its existence, clearly illustrate, the study of bilingualism (and multilingualism) from a cognitive perspective attracts both linguists and psychologists. The cognitive study of bilingualism covers various bilingual POPULATIONS, differing in age (from one-year-old infants to seniors), type of bilingualism (early vs. late, balanced vs. unbalanced) or cognitive functioning (selection and control (facilitation, competition, inhibition) of languages in bilinguals). The TOPICS are concerned with, among others, language development (in bilingual children), the ways in which the languages in bilinguals affect each other in various linguistic domains in language reception and production (including code switching), the so-called critical-period issue, the linguistic and cognitive advantages and disadvantages of bilingualism, and language impairment (e.g., typically developing vs. cognitively impaired bilingual children, normal vs. brain-damaged bilingual adults). The LINGUISTIC DOMAINS, in grammar and lexis, vary from phonetics, phonology and morpho-syntax to semantics. The DATA range from corpora of extemporaneous speech (e.g., in studies of code switching and language contact) to participant performance (accuracy and/or speed) on researcher-designed and experimentally-controlled tasks, sometimes with additional measurement of eye movements, hand or head movements, or brain activity (ERP, fMRI). With the exclusion of single-case studies and corpus studies,

typical studies, reported in this journal, pertain to the comparison of groups of participants on some researcher-designed task or tasks (see below). In most of such comparison studies, the researchers apply participant-selection criteria. Typical criteria are age, language-acquisition history and environment (e.g., age of arrival in L2 country, length of residence in L2 environment, years of L2 instruction) and self-assessed proficiency in one or several domains of language use (listening, speaking, reading, and writing). Sometimes, selection criteria include performance on a researcher-administered language test or possession of a language certificate. Thus, proficiency in either or both languages of bilinguals, whether assessed by participants or measured in a more objective way, plays a crucial role in the cognitive study of bilingualism.

Under the assumption that the empirical studies published in *Bilingualism: Language and Cognition* form a representative sample of the literature of bilingualism (from a cognitive perspective) at large, I looked at all papers published in this journal from the first issue of volume 1 (1998) up till the last issue of volume 14 (2011). The journal distinguishes between (i) keynote articles followed by peer commentaries, (ii) guest-editor introductions to special issues, (iii) research articles, and (iv) research notes. Empirical studies, presenting new data or new analyses of existing data, fall into categories (iii) and (iv). Not all research articles and research notes are empirical, however. Some research articles provide a review of the literature and a new model or theory, while some research notes present critical comments of a paper published in an earlier issue or a reply from the authors of the article criticized. The papers, empirical in the sense just defined, number 224 (i.e. in categories (iii) and (iv)). Of those 224 empirical papers, 142 articles are concerned with various types of group comparisons. The most frequent types of comparisons are the following: (i) comparison of bilinguals dominant in either of two languages, (ii) comparison of bilinguals (languages A and B) with native speakers of either language, having no knowledge of the other language or with monolinguals of either language (knowing no other languages), (iii) comparison of L2 learners at different levels of L2 proficiency with native speakers of their L2, (iv) comparison of groups of bilinguals or groups of L2 learners with different first languages, (v) comparison of bilinguals of different ages, or (vi) combinations of these designs. Of the 142 comparison studies, 140 studies fall into what I call the LP category. An LP study is a study for which, in my personal judgment, the assessment of LP as an independent or moderating variable is a necessary or preferred requirement.¹

¹ Two studies measuring LP as the DEPENDENT variable, which was investigated in terms of NON-LANGUAGE-PROFICIENCY

The assessment of language proficiency

Of the 140 papers presenting group comparisons of the types just mentioned, 37 appeared in volumes 1–7 (1998–2004) while 103 appeared in volumes 8–14 (2005–2011). These 140 papers would have benefited – in my view – from LP assessment in one way or another. In only seven of the 37 papers in volumes 1–7, LP was assessed with an objective test (19%). Of the 103 LP papers in volumes 8–14, 56 (54%) reported the use of an objective LP measure – a significant increase. In 28 (20%) out of the 140 LP studies (13 of 37 studies in volumes 1–7 and 15 of 103 studies in volumes 8–14), LP was not assessed at all, i.e. neither with an objective nor with a subjective procedure.

Objective LP assessment used in participant selection

In the studies that used an objective test to assess candidate participants' LP, the procedure was in one of two forms. One procedure requested that candidate participants have an LP certificate issued by a testing institute of good reputation (e.g., in the case of English, Cambridge Certificate of Proficiency (UK), College English Test (China), IELTS (Australia and UK), Oxford Quick Placement Test (UK), Test for English Majors (China) and TOEFL (USA)). This screening procedure was adopted by Antón-Méndez (2011), Athanasopoulos (2006), Chen, Shu, Liu, Zhao and Li (2007), Cole and Pickering (2010), Dong, Gui and MacWhinney (2005), Felser, Sato and Bertenshaw (2009), Kormos and Sáfár (2008), Salamoura and Williams (2006, 2007), Silva and Clahsen (2008), Snape (2008), Trenkic (2008), and Williams (2006).

Many studies using child participants administered the Peabody Picture Vocabulary Test (PPVT) to assess productive vocabulary (Dressler, Carlo, Snow, August & White, 2011; Goetz, 2003; Nicoladis, 2006; Reyes & Hernández, 2006; Uchikoshi, 2006; Yan & Nicoladis, 2009). Some studies using infants matched infants on Mean Length of Utterance (MLU) in oral production. PPVT and MLU in young children can be seen as reflecting a combination of language development and language proficiency. Reyes and Hernández (2006) and Hernandez and Meschyan (2006) administered the Boston Naming Test in both languages of their bilingual participants to establish language dominance. Several studies (Ping & Liow, 2011; Von Studnitz & Green, 2002a, b), examining processing of English, screened participants with the LLEX vocabulary test

INDEPENDENT FACTORS, such as bilingual history, age, or non-language cognitive capabilities, were coded as not falling in the LP category. If, however, a study investigated some form of language control (as the dependent variable) in terms of level of LP (e.g., intermediate versus advanced L2 learners), it was coded as falling in the LP category.

developed by Meara (1994), which can be described as a computer-administered written lexical-decision task, with word stimuli taken from different frequency classes. Athanasopoulos (2009) and Athanasopoulos, Damjanovic, Krajciova and Sasaki (2011) screened subjects on the basis of their scores on Nation's Vocabulary Level test. Luk, de Sa and Bialystok (2011) used the PPVT (3rd edition) in screening adult subjects (early and late bilinguals and monolingual controls).

Some other types of objective LP assessment to screen subjects with are translation test (Hyltenstam, Bylund, Abrahamsson & Park, 2009), panel judgement of the speech of participants as near native (Pika, Nicoladis & Merentette, 2006) or a cloze test, or C-test (see below).

Fluency tasks to establish language (non)dominance

The adult Spanish–English bilinguals participating in the study of Knupsky and Amrhein (2007) were selected on the basis of a category naming task and a reading-speed task, administered in both languages (derived from measures developed by Soares & Grosjean, 1984). In the category-naming task, subjects had to produce as many exemplars of a certain category (e.g., animals) as possible in 30 seconds, in both their languages. This type of task, in the form of naming as many words beginning with F, A and S as one can, administered in participants' L1 and L2, was also used by Ayçiçeği-Dinn and Caldwell-Harris (2009), who call it a language fluency task.

Examining processing of morphologically complex Finnish words in monolingual speakers of Finnish and Swedish–Finnish bilinguals in Finland, Lehtonen and Laine (2003) not only collected self-assessment data from their participants but they also administered a digit-reading task in Finnish, thus providing two types of evidence “that the Finnish skills of the bilinguals were at the same level as those of the monolinguals” (p. 219). The researchers refer to a study of Chincotta, Hyönä and Underwood (1997), who found that digit reading times in bilinguals were faster in their dominant language when effects of the digits' word length in the two languages was controlled.

In a study of attention control in L2 processing, Taube-Schiff and Segalowitz (2005) screened their potential subjects (English–French bilinguals) not only on self-assessed L1 and L2 proficiency but also on a speeded word classification task (animacy judgment), which subjects performed faster in their L1 than in their L2. (For a treatise of fluency issues in bilingualism, see Segalowitz, 2010.)

Comments on LP assessment in the article corpus

In this section, some evaluative comments are made on some LP-assessment procedures observed in the article corpus.

Examples of careful screening

A fine example of careful screening is the study conducted by Tremblay (2011). Adult English–French participants were selected and assigned to three levels of L2 proficiency by means of (i) language-history information, (ii) self-assessed L2 proficiency, and (iii) an L2 cloze test (validated in another study). Factor analysis and regression analysis, using these three types of information, were conducted to produce three distinct proficiency groups. Another example of a study that applied many criteria, is the study of Sundara, Polka and Baum (2006), investigating the acoustic phonetics of coronal stop production by adult simultaneous bilingual and monolingual speakers of Canadian English and Canadian French. The researchers applied five selection criteria: three language-history criteria, self assessment (confirmed by a fluent bilingual in an oral interview), and rating of speech samples by panels of native listeners on a seven-point scale ranging from no ability to native-like ability.

Self-assessment of language proficiency

Four studies (Lee & Williams, 2001; Salomoura & Williams, 2006, 2007; Williams, 2006) used the self-assessment questionnaire of Bachman and Palmer (1989). This questionnaire consists of 24 questions, divided in a 3×3 fashion involving three traits and three question types. The three traits are grammatical competence (morphology and syntax), pragmatic competence (vocabulary, cohesion, organization) and sociolinguistic competence (register, nativeness, nonliteral language). The three question types are ability (e.g., How much English grammar do you know?), difficulty (e.g., How many different kinds of grammar mistakes do you make in English?), and recognition (e.g., How often can you tell when someone makes a grammar mistake?). (For a complete list of all questions, see the appendix in Bachman & Palmer, 1989.) The researchers administered the questionnaire to 116 mixed-ability nonnative speakers of English. Age ranged between 17 and 67 years, while length-of-residence in the US ranged between a few months and over ten years. Reliabilities of the nine scales ranged between .86 and .91 with the exception of grammar ability (.54). Exploratory and subsequent confirmatory factor analyses showed good fit of the 3×3 LP model. While this finding is quite remarkable given the literature on LP models (Bachman & Palmer, 1996; Hulstijn, 2011; Purpura, 2008), and while the findings might support the use of this questionnaire as a placement test in a language program, it might be of limited use in the cognitive study of bilingualism (depending on the study's research questions). First, note that the grammar ability scale was not reliable (.54) in the tested sample. Second, note that the

global scales of this questionnaire – and the scales in many other questionnaires used in studies of bilingualism – (i) do not differentiate between the four languages skills (listening, speaking, reading and writing), (ii) do not differentiate, within the grammar dimension, between control of phonetics, prosody, phonology, morphology and syntax, (iii) do not elicit information concerning fluency of language control, and (iv) in the answer options of the ability questions, only refer to “most native speakers” as a reference group and not to different types of native speakers in terms of educational or professional profile. Furthermore, it is doubtful whether non-experts are able to give a valid assessment of proficiency dimensions such as pronunciation, fluency of speech, size of productive vocabulary, and grammar control. Thus, global questionnaires, with their lack of detail, might be of limited use for many cognitive studies of bilingualism.

Language-program enrolment

Some studies used enrolment in different years, or levels of a language program to distinguish between proficiency groups (e.g., Ayoun, 1999; Sopata, 2005; Toribio, 2001). For example, Ayoun (1999) compared verb movement in French in three groups of English learners of French, who were in the 2nd, 3rd or 4th year of a French program. The results might have been different, however, had the participants been grouped on the basis of one or several LP tests, because it cannot be ruled out that, for example, the best student in the 2nd year was more proficient in French than the poorest student in the 3rd year. A similar non-optimal design, using enrolment as the between-group criterion rather than an objective proficiency test, was adopted in a study that the present author was involved in (Hulstijn & Marchena, 1989).

Cloze test and C-test

Several studies used a cloze test, i.e., a paper-and-pencil procedure whereby the test taker is given a text in which some words have been deleted and replaced by gaps and is instructed to fill acceptable words in the gaps (e.g., Kim, Montrul & Yoon, 2010; Montrul, 2004; Montrul & Ionin, 2010; Slabakova, 2005; Yan, 2007). In the C-test version, every second word of the text is gapped and the first letters (50% of the letters) are given. The candidate has to complete the incomplete words. (See Hulstijn (2010) for examples and discussion of cloze and C-test.) Hopp (2009) used a timed web-based C-test (Grotjahn, 2002) to screen subjects. More often than not, papers give insufficient information with respect to matters such as the type of text(s), the word-deletion procedure, number of deleted words, test instructions, and

time allowed. More often than not, information is missing with respect to performance of a native-speaker control group, or the test's internal consistency (e.g., Cronbach's alpha). In contrast, Montrul (2004) and Montrul and Ionin, (2010) used both a reliable cloze test and a vocabulary test of Spanish (taken from an exam issued by the ministry of education in Spain).

Dealing with the danger of circularity

In 63 studies in the corpus, many of which examined the development, fossilization or attrition of interlanguage grammars, an LP test was used as a means to establish different LP groups. Many studies of this type give the impression that LP is a clear construct that can be used independently in the grouping and comparison of interlanguage grammars. However, especially in the case of studies with a linguistic focus, eliciting the presence or absence of certain phonotactic, phonological, morphosyntactic, morpholexical, semantic or pragmatic features in the interlanguage of L2 learners as the dependent variable(s), it is not immediately obvious if and how LP should be measured independently from the feature under investigation. For instance, in a hypothetical study of the oral production of grammatical noun gender, how should LP be assessed without noun gender being part of the measure, especially when groups of bilinguals are compared with different language pairs, differing in the presence or absence of grammatical noun gender? This question illustrates the main argument of the present article, namely that LP assessment not "only" constitutes a methodological issue, but forces researchers to reflect on what LP actually is, the association of which with the linguistic feature under study they seek to establish.

The study of Guion, Harada and Clark (2004) gives a fine example of how the danger of circularity in LP assessment can be addressed. The researchers examined stress placement in bisyllabic nonwords of English (contrasting or not contrasting with Spanish stress placement) in native speakers of English and early and late Spanish–English bilinguals. The experimental task, producing the data for the dependent variable, required subjects to orally produce a concatenation of two aurally presented nonwords. Both bilingual groups were screened on their high proficiency in English by means of various language-history criteria and their performance on two LP tests: a listening-vocabulary and a listening-grammar subtest of the Test of Adolescent and Adult Language (TOAL; Hammill, Brown, Larsen & Weiderholt, 1994).² In addition, as a check of participants' familiarity with stress patterns of real English words, they had to read aloud 18 words with regular stress and 18 words with

irregular words. Mean correct scores by groups and word type were 96% or higher. The researchers succeeded in screening the bilingual subjects (i) with LP measures relevant to the purpose of the study (because knowing stress patterns in some common existing words is a prerequisite for performing in a challenging task eliciting stress patterns in pseudowords), and (ii) without falling into the trap of adopting a circular design (because knowing stress patterns in some common existing words is not the same as predicting stress patterns in pseudowords).

Use of LP assessment to explain variance in the dependent variable

A striking feature of almost all comparison studies using self assessment and/or objective testing of LP is that the information gathered with these procedures is only used for selecting and deselecting potential subjects or for assigning selected subjects to different proficiency groups (e.g., intermediate and advanced subjects). For example, in the former case, selected subjects performed above or below a certain criterion score to demonstrate evidence for near-native or non-native LP, respectively. The most common statistical analysis in the comparison studies is ANOVA. Thus, in the sample of studies surveyed, it is common practice to compare participant groups on whatever the dependent variable of the study is with ANOVA-type analyses. In other words, notwithstanding Grosjean's (1998, p. 135) advice to "make bilingual assessment measures covariate variables during the analysis of results or allow participants to be their own control when the study permits it", within-group individual differences are seldom taken into account in interpreting the results of the analyses. In most cases, the focus is on the interpretation of between-group differences. Thus, studies that found significant differences, for some dependent variable, between bilinguals and native controls or between intermediate and advanced L2 learners do not report how much variance in the dependent variable was actually accounted for by this significant difference. The exception to this observation is the study of Tokowicz, Michael and Kroll (2004). These researchers examined errors in English–Spanish and Spanish–English word translation in English–Spanish and Spanish–English bilinguals as a function of their study-abroad experience (SAE) and their working-memory capacity (WMC). Hierarchical linear regression was performed on the error data, not only with SAE and WMC as predictor variables, but also with self-assessed LP. Only some more recent studies report effect sizes. No studies were found that applied multilevel linear modelling (capable of handling hierarchical dimensions in participant groups), or structural equation modelling.

² In the corpus, the TOAL was also used by Chakraborty (2011).

Problems in comparing LP across languages

Perhaps the most difficult issue in the assessment of LP in studies on bilingualism concerns the question of whether it is possible to provide empirical evidence for the claim that a given bilingual is equally proficient (in a given language skill, e.g. listening) in both of his or her languages. In other words, is it possible to demonstrate that a given bilingual is less proficient (in a given skill) in one of his or her languages than in the other language? Perhaps more difficult to answer is the following question: Is it possible to produce evidence of tiny differences in a bilingual's proficiency (in a given skill) in the two languages? Addressing these conceptually and methodologically intriguing issues forms a daunting task. Unfortunately, this section raises more questions than it can answer. It should be seen as exploring some ideas, leaving aside for the moment what the scientific relevance of such questions might be (see Grosjean's (1989) criticism of the notion of balanced bilingualism and the concept of a bilingual as two monolinguals in one person).

Comparing apples and oranges

Let us assume for the moment that experts agree that test A is a valid and reliable test to assess control of a given component of LP in language A and that test B is an equally valid and reliable test to assess control of the same component in language B. Let us further assume that both tests consist of 100 independent items and that each response can be awarded with either 1 (correct) or 0 (incorrect). Can we say now that a score of 61 in each test reflects the same amount of control in both languages? Before answering this question, let us first consider the case of two tests measuring the same ability in the SAME language. For instance, with two versions of the listening subtest of the Test of English as a Foreign Language (TOEFL), one can empirically ascertain whether the two versions are equally difficult by administering them to the same individuals, representing a broad ability range. In this way, one can empirically ascertain test equivalence, both in terms of content (the ability assessed) and difficulty of the test for a given population. However, in the case of two tests measuring the same ability in different languages, there is no way of ascertaining whether a score of 61 on each test reflects equal ability in the languages tested. The main reason is that languages differ in the number and nature of elements in the phonotactic, phonological, morphological, syntactic, lexical and pragmatic domains. In other words, tests of knowledge or control in each of these domains in two languages measure different things (apples and oranges) although they may be referred to by the same label.

Knowledge of grammar

The most obvious example of comparing knowledge of apples and oranges is when one wants to construct a test of syntactic knowledge in language A and a test of syntactic knowledge in language B. For instance, whereas clitic pronouns may feature in language A but not in language B, ergative verbs may feature in B but not in A. How is one to establish equivalence of one test, tapping (among other things) knowledge of clitic pronouns in A, with another test, tapping (among other things) knowledge of ergativity in language B?³ Similarly, although perhaps to a less dramatic extent, a score of 61 in two 100-item tests assessing the perception of the vowels and consonant inventories in languages A and B may not be equated because the two inventories are likely to differ in size and complexity.

Knowledge of vocabulary

The situation is perhaps less dramatic in the case of productive vocabulary elicited by the same stimuli in two languages, as in the Peabody Picture Vocabulary Test. If a bilingual produced correct responses in each language to exactly the same 61 pictures, and no (correct) responses in each language to the remaining 39 items, one might say that this comes close to equivalence of lexical knowledge in two languages. But even then there are grounds for casting doubts on such a conclusion. For instance, it might well be that the words in languages A and B associated with the picture of, for instance, a shoe are translation equivalents of one another only in some contexts. Or it might be that the picture of a shoe is associated with only one high-frequency word in language A but to two high-frequency words in language B. If the test taker produces only one of these two words in language B, how do we know whether he or she also knows the other word? In sum, languages differ in the size and density of their word nets and it is thus difficult to assess knowledge of size and density in two languages in an equivalent way (see Li, Zhang & Wang, 2011, for a recent attempt).⁴

³ Pienemann, Keßler and Itani-Adams (2011) offer what might be seen as a partial solution to the cross-linguistic comparison of morpho-syntactic control of two languages, based on Pienemann's (2005) Processability Theory, which distinguishes four processing procedures, observed in developmental trajectories in L2 learners of typologically different languages. However, Pienemann et al. (2011) do not raise the question of whether LP is to be equated with the attainment of a developmental stage in an L2 learner's interlanguage and thus do not speak of what to do with observed individual differences in the control of a given processability procedure.

⁴ The question of which words someone knows should, of course be distinguished from the question of how many words someone knows. For bilinguals, it might well be the case that the size of the vocabularies in their languages is roughly the same but that the vocabularies may only partially overlap. Analysing the English vocabulary-test scores (Peabody Picture Vocabulary Test) of 1,738 children (3–10 years old) in Canada, Bialystok, Luk, Peets and Yang (2010), found that although

Fluency

Problems might also arise in the case of using fluency tasks in bilinguals' two languages to assess LP in them. If we ask bilinguals to read aloud a text in language A and one in language B (translation equivalents of each other) and then count the number of words read aloud in three minutes, a score of say 180 words per minute (or a score of 300 syllables per minute) in each language may not reflect equal proficiency because of between-language differences in phonological or morphological length and complexity of words or syllables or in syntactic differences. Even the pronunciation of a word-initial /p/ may differ in duration between languages. However, for the practical purposes of establishing a rough measure of language dominance in bilinguals, the examples just given of between-language differences may not invalidate the use of such tests; one might even correct the rough scores for overall between-language differences on the dimensions just mentioned (see Daller, Yldiz & de Jong, 2011).

Another way of dealing with between-language differences in the measurement of processing fluency might be to have bilinguals perform the tasks, in each language, under two conditions, a normal condition and a condition in which processing is inhibited by a nonverbal task factor. The ratio of the score in the impeded condition and the score in the normal condition might then be compared between languages. This procedure appears to by-pass a direct comparison of apples and oranges. The present author explored this idea with two of his students. In one study (Engelen, 2009), we administered dictation tasks in bilinguals' two languages, each task (i) without white noise and (ii) with white noise (at a signal-to-noise ratio of -3 dB). The participants (N = 29) were Dutch university students, split into a higher and a lower L2 English proficiency group by means of their scores on a receptive vocabulary test. The stimulus sentences consisted of syntactic prose, controlled for lexical frequency and sentence length; syntactic complexity was manipulated (five sentence types in each language). Thus, the study had a $2 \times 2 \times 2 \times 5$ design, with L2 proficiency as the between-subjects factor, and noise, language and syntactic complexity as the between-subject factors. The dependent variable was the number of correctly reproduced words. There was, as expected, an overall noise effect. Also, as expected, noise affected performance of high L2 proficiency participants less than

bilingual children (not speaking English at home) performed more poorly than monolingual children, this difference existed mainly in the domain of home-language vocabulary. Although this study was not concerned with between-language within-subjects comparisons, it suggests the likelihood of limited overlap of lexical knowledge in bilinguals. For a method using indexes of lexical richness in speech, elicited in bilinguals' two languages, to assess language dominance, see Treffers-Daller (2011).

performance of participants with low L2 proficiency in L2, but, unexpectedly, both groups were affected by noise more in L1 than in L2. The study suffered from some infelicities, however, one being that the voice of the L1 speaker who produced the L1 stimuli might have been more susceptible to noise than the L2 speaker's voice.

In another study (Gauvin and Hulstijn, 2010), 32 Dutch-English bilinguals (university students, split into a higher and a lower L2 proficiency group) read target sentences in both languages from a computer screen (in a blocked design), each sentence appearing both in normal and in poorly readable font. Sentence reading times formed the dependent variable. Degraded font affected reading times substantially, more so in L2 than in L1, as predicted for the participants, selected for their presumed L1 dominance. However, it was not found that participants with higher L2 proficiency were less affected by degraded font in L2 reading than participants with lower L2 proficiency. Thus, Engelen (2009) and Gauvin and Hulstijn (2010) could not entirely eliminate a comparison of apples and oranges because of potential interactions between the nonverbal inhibition factor and linguistic features of the languages. For instance, it cannot be ruled out, in principle, that white noise affects the perception of a tone language more than the perception of a non-tone language or that poorly readable font might have a stronger effect on the readability of a morphologically simple language than a morphologically complex language.

A proposal for assessing LP between languages in bilinguals

The ways to compare LP within bilinguals and between languages mentioned above all appear to have their problems, albeit some more than others. Perhaps a more feasible way to assess LP in bilinguals' languages, for the purpose of establishing language dominance in bilinguals, is to administer tests designed to tap roughly the same LP component in each language and compare bilinguals' performance to the performance of native-speaker (NS) reference groups in each language. It is crucial, however, that the NSs be selected with great care. First and foremost, never should a NS reference group consist of individuals with a higher intellectual profile (level of literacy, level of education, level of profession, intellectual level of leisure-time activities) than the profile of the bilingual target group. The effect of the linguistic norms of the standard language, in most societies, on NSs receiving education at the secondary and tertiary level should not be underestimated. Second, if the study's research question require LP measures in linguistic cognition that all NSs are assumed to possess (so-called BASIC LANGUAGE COGNITION, see next section), three scenarios offer themselves. Under the first scenario,

only NSs of lower intellectual profiles are selected to ensure that the LP tasks do not require the bilinguals to do what monolingual NSs are not able to do. Under the second, more ambitious scenario, NSs of a variety of intellectual profiles are selected. This increases the likelihood of variance in NS performance, which makes this scenario more challenging. Under the third scenario, which only applies when the bilingual target group has a high intellectual profile, the NS reference groups consist of individuals of the same, high intellectual profile. (Note that many cognitive studies of bilingualism in adults avail themselves of university students as participants.) In sum, several feasible solutions of the problems in LP assessment in bilinguals appear to exist, all based on comparison with NS reference groups, provided due attention is given to the LP component(s) involved (within or beyond basic language cognition) and the intellectual profile of the NS reference group(s).

As the reviewers of this paper point out, an additional method in the assessment of LP in bilinguals' languages would involve repeated administration of language tests, i.e. in longitudinal within-subject designs. This method is demanding but forms a very important complement to cross-sectional designs.

Basic and higher language cognition

As shown in the previous sections, some research treats the notion of LP as a rather unproblematic notion, just like age, length of residence, and amount of language instruction. The notion of LP should be taken seriously, however, because it is fundamental in the understanding of language acquisition and bilingualism. Hulstijn (2011) argues that, for the conceptualization and measurement of LP it may be useful to make a distinction between basic and higher language cognition (BLC and HLC). While BLC is restricted to the processing of ORAL language (listening and speaking) in utterances containing HIGH-FREQUENCY lexical, grammatical, phonotactic and prosodic elements, HLC is unrestricted in these respects. That is, HLC includes the processing of WRITTEN language (reading and writing, i.e. involving literacy skills) in sentences which may contain LOW-FREQUENCY lexical or grammatical elements. An informal way of distinguishing BLC from HLC is to regard BLC as the language knowledge shared by all adult native speakers (not affected by serious language-related mental disorders), whereas HLC exhibits individual differences in language control, potentially affected by attributes such as literacy, age, level of education, profession or leisure-time activities.⁵

⁵ Hulstijn (2011) gives a detailed definition of BLC and HLC, spells out the differences between these notions and the constructs of basic interpersonal communicative skills (BICS) and cognitive academic language proficiency (CALP), proposed by Cummins (19080a,

For the measurement of LP in studies of bilingualism, I recommend that researchers distinguish between the administration of tests assessing participants' ability to process simple oral utterances receptively or productively (i.e., in the BLC domain, not requiring literacy skills) and the administration of tests assessing participants' ability to process language in the HLC domain.

Levels of proficiency

To my knowledge, there is no linguistic, psycholinguistic or sociolinguistic theory on the basis of which one could define LEVELS of LP. Similarly, if I am not mistaken, there are no theories on the basis of which one could define levels of intelligence, levels of language-learning aptitude, levels of working-memory capacity, levels of language-learning motivation, or levels of socio-economic status (SES). Performance in tests developed for measuring these attributes in individuals is normally expressed as a score on one or several scales. Using scores, one could say that a person with an IQ of 107 is more intelligent than a person with an IQ score of 101, but not that the former person finds himself or herself at a higher level of intelligence than the latter. Similarly, someone with a backward digit span of 6 can be said to have a larger working memory span than someone with a score of 4, but not at a higher working-memory level. Although no theoretical basis exists for the distinction between levels of LP, in educational practice it is quite common, perhaps even unavoidable, to speak of LP levels, such as beginning, intermediate and advanced levels, or to adopt the level labels used by companies that produce language-proficiency tests. Under the auspices of the Council of Europe, a Common European Framework of Reference for Languages (CEFR) was developed, defining six global LP levels, best known as A1, A2, B1, B2, C1, and C2 (Council of Europe, 2001). In Hulstijn (2011), it is argued that the six levels of the CEFR fail to consistently distinguish between language development and language proficiency. The six levels cannot be regarded as forming a unidimensional ladder of language development because the higher levels (B2, C1 and C2) can only be attained by people with higher intellectual capacities.

Recommendations

This short section lists five recommendations for empirical research on bilingualism, particularly for group-comparison studies as mentioned in the section "The article corpus". I wish to stress with great emphasis,

1980b), and formulates several hypotheses, making the distinction between BLC and HLC empirical.

however, that these recommendations are to be seen only as proposals that may or may not apply, depending on the research question of the study. As every researcher and instructor in graduate courses knows, it does not make sense to talk about the methodology of a study without taking its purpose, research questions and theoretical embedding into account. It is only on the basis of the study's theoretical embedding, its purpose and research question that the way in which LP is (or is not) measured can be based (see Hulstijn, 2010, for details). Bearing this crucial caveat in mind, I propose the following recommendations for studies comparing performance of groups of human subjects, as mentioned in the subsection "The article corpus".

1. Motivate the absence or presence of an objective measurement of LP to select candidate participants.
2. In case an objective LP test is used, motivate the type of test chosen, given the age, literacy and educational level of participants. Describe the test's target group (age, literacy and educational level), skills measured, task(s), and materials in sufficient detail, and report its validity (if known) and its psychometric characteristics (e.g., internal consistency) for the target group.
3. For the assessment of language dominance, (i) use tests of oral reception and production rather than tests of reading or writing, (ii) exclude linguistic elements that not all adult NSs may be familiar with, and (iii) administer the test to NSs too, in order to verify whether NSs of lower intellectual, educational or professional profiles perform at ceiling. If assessment of LP only takes place in the form of a questionnaire (to be filled out by the participants themselves or their parents or teachers), make its items refer to communicative skills in the realm of BLC only (reception and production of oral language involving only high-frequency elements), or in BLC and HLC (unrestricted oral and written language use) separately.
4. Consider whether it is appropriate to analyse the data (in addition to, or instead of using ANOVA) with multilevel linear mixed modelling or similar techniques that allow the researcher to determine to what extent the LP-assessment data (within and across participant groups) account for the amount of variance observed in the dependent variable(s).⁶
5. In the discussion section of papers, discuss not only the between-group findings but also the variance obtained within groups, and consider the potential

meaning of the findings for our understanding of notions such as "language proficiency", "native speaker", "language dominance", and "(un)balanced bilinguals".

Concluding remarks

The paradox of science is that, in order to understand the phenomena that we are interested in, we have to do "un-justice" to them, as it were. We classify them into categories, disregarding differences between them. In the literature on bilingualism, people are categorized as native speakers, as monolingual and bilingual native speakers, as intermediate or advanced non-native speakers, or as balanced or unbalanced bilinguals. These categories are not only convenient but perhaps even necessary because science is necessarily reductionist in nature. This paper has attempted to make a case for the individual LP differences that exist WITHIN these conventional categories. The idea is that we may increase our understanding of phenomena of bilingualism by taking these individual LP differences into account in our measurement and in our statistical analyses. The risk is, of course, that our understanding will decrease rather than increase because we might lose sight of the categories (groups) and hence of the between-groups contrasts that we sought in the first place. But that risk is inherent in science, making scientific work so exciting for researchers. By paying special attention to the measurement of LP, we are forced to reflect on what the notion of LP is, what it means to be "lingual", in the first place, and what role LP may play in the phenomena that we seek to explain.

References

- Abutalebi, J., Cappa, S. F., & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition*, 4, 179–190.
- Antón-Méndez, I. (2011). Whose? L2-English speakers' possessive pronoun gender errors. *Bilingualism: Language and Cognition*, 14, 318–331.
- Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition*, 9, 89–96.
- Athanasopoulos, P. (2009). Cognitive representation of colour in bilinguals: The case of Greek blues. *Bilingualism: Language and Cognition*, 12, 83–95.
- Athanasopoulos, P., Damjanovic, L., Krajciova, A., & Sasaki, M. (2011). Representation of colour concepts in bilingual cognition: The case of Japanese blues. *Bilingualism: Language and Cognition*, 14, 9–17.
- Ayçiçeği-Dinn, A., & Caldwell-Harris, C. L. (2009). Emotion-memory effects in bilingual speakers: A levels-of-processing approach. *Bilingualism: Language and Cognition*, 12, 291–303.
- Ayoun, D. (1999). Verb movement in French L2 acquisition. *Bilingualism: Language and Cognition*, 2, 103–125.

⁶ On multilevel linear mixed modelling, see, for example, Field (2009, chapter 19).

- Bachman, L. F., & Palmer, A. S. (1989). The construct validation of self-ratings of communicative language ability. *Language Testing*, 6, 14–29.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13, 525–531.
- Chakraborty, R. (2011). Influence of L2 proficiency on speech movement variability: Production of prosodic contrasts by Bengali–English speakers. *Bilingualism: Language and Cognition*, 14, 489–505.
- Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, 10, 161–174.
- Chincotta, D., Hyönä, J., & Underwood, G. (1997). Eye fixations, speech rate and bilingual digit span: Numeral readings indexes fluency not word length. *Acta Psychologica*, 97, 253–275.
- Cole, R. L., & Pickering, S. J. (2010). Phonological and visual similarity effects in Chinese and English language users: Implications for the use of cognitive resources in short-term memory. *Bilingualism: Language and Cognition*, 13, 499–512.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge & Strasbourg: Cambridge University Press & Council of Europe.
- Cummins, J. (1980a). The construct of language proficiency in bilingual education. In J. E. Alatis (ed.), *Current issues in bilingual education: Georgetown University Round Table on Languages and Linguistics (GURT) 1980*, pp. 81–103. Washington, DC: Georgetown University Press.
- Cummins, J. (1980b). The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *TESOL Quarterly*, 14, 175–187.
- Daller, M. H., Yldiz, C., & de Jong, N. H. (2011). Language dominance in Turkish–German bilinguals: Methodological aspects of measurements in structurally different languages. *International Journal of Bilingualism*, 15, 215–236.
- Dong, Y., Gui, S., & MacWhinney, B. (2005). Shared and separate meanings in the bilingual mental lexicon. *Bilingualism: Language and Cognition*, 8, 221–238.
- Dressler, C., Carlo, M. S., Snow, C. E., August, D., & White, C. E. (2011). Spanish-speaking students' use of cognate knowledge to infer the meaning of English words. *Bilingualism: Language and Cognition*, 14, 243–255.
- Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient bilingual dominance scale. *Bilingualism: Language and Cognition*, 12, 273–289.
- Engelen, J. (2009). First and second language listening ability in noise: A within-subject comparison. Master's thesis, University of Amsterdam.
- Felser, C., Sato, M., & Bertenshaw, N. (2009). The on-line application of binding Principle A in English as a second language. *Bilingualism: Language and Cognition*, 12, 485–502.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd edn.). Los Angeles, CA: Sage.
- Gauvin, H. S., & Hulstijn, J. H. (2010). Exploring a new technique for comparing bilinguals L1 and L2 reading speed. *Reading in a Foreign Language*, 22, 84–103.
- Goetz, P. J. (2003). The effects of bilingualism on theory of mind development. *Bilingualism: Language and Cognition*, 6, 1–15.
- Grosjean, F. (1989). Neurologists, beware! The bilingual is not two monolinguals in one person. *Brain & Language*, 36, 3–15.
- Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, 1, 131–149.
- Grosjean, F., Li, P., Munte, T. F., & Rodriguez-Fornells, A. (2003). Imaging bilinguals: When the neurosciences meet the language sciences. *Bilingualism: Language and Cognition*, 6, 159–165.
- Grotjahn, R. (ed.) (2002). *Der C-Test: Theoretische Grundlagen und praktische Anwendungen*. Bochum: AKS Verlag.
- Guion, S. G., Harada, T., & Clark, J. J. (2004). Early and late Spanish–English bilinguals' acquisition of English word stress patterns. *Bilingualism: Language and Cognition*, 7, 207–226.
- Hammill, D. D., Brown, V. L., Larsen, S. C., & Weiderholt, J. L. (1994). *Test of adolescent and adult language* (3rd edn.). Austin, TX: Pro-ed.
- Hernandez, A. E., & Meschyan, G. (2006). Executive function is necessary to enhance lexical processing in a less proficient L2: Evidence from fMRI during picture naming. *Bilingualism: Language and Cognition*, 9, 177–188.
- Hopp, H. (2009). The syntax–discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition*, 12, 463–483.
- Hulstijn, J. H. (2010). Measuring second language proficiency. In E. Blom & S. Unsworth (eds.), *Experimental methods in language acquisition research* (EMLAR), pp. 185–199. Amsterdam: John Benjamins.
- Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, 8, 229–249.
- Hulstijn, J. H., & Marchena, E. (1989). Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition*, 11, 241–255.
- Hyltenstam, K., Bylund, E., Abrahamsson, N., & Park, H. (2009). Dominant-language replacement: The case of international adoptees. *Bilingualism: Language and Cognition*, 12, 121–140.
- Kim, J., Montrul, S., & Yoon, J. (2010). Dominant language influence in acquisition and attrition of binding: Interpretation of the Korean reflexive caki. *Bilingualism: Language and Cognition*, 13, 73–84.
- Knopsky, A. C., & Amrhein, P. C. (2007). Phonological facilitation through translation in a bilingual picture-naming task. *Bilingualism: Language and Cognition*, 10, 211–223.

- Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, 11, 261–271.
- Kroll, J. F., Bobb, S. C., & Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, 9, 119–135.
- Lee, M., & Williams, J. N. (2001). Lexical access in spoken word production by bilinguals: Evidence from the semantic competitor priming paradigm. *Bilingualism: Language and Cognition*, 4, 233–248.
- Lehtonen, M., & Laine, M. (2003). How word frequency affects morphological processing in monolinguals and bilinguals. *Bilingualism: Language and Cognition*, 6, 213–225.
- Li, D., Zhang, X., & Wang, G. (2011). Senior Chinese high school students' awareness of thematic and taxonomic relations in L1 and L2. *Bilingualism: Language and Cognition*, 14, 444–457.
- Luk, G., de Sa, E., & Bialystok, E. (2011). Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, 14, 588–595.
- Meara, P. (1994). *LLEX: Lingua vocabulary tests version 1.4*. Swansea: University of Wales, Centre for Applied Language Studies.
- Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: A case of morphosyntactic convergence. *Bilingualism: Language and Cognition*, 7, 125–142.
- Montrul, S., & Ionin, T. (2010). Transfer effects in the interpretation of definite articles by Spanish heritage speakers. *Bilingualism: Language and Cognition*, 13, 449–473.
- Nicoladis, E. (2006). Cross-linguistic transfer in adjective–noun strings by preschool bilingual children. *Bilingualism: Language and Cognition*, 9, 15–32.
- Pienemann, M. (2005). *Cross-linguistic aspects of Processability Theory*. Amsterdam: John Benjamins.
- Pienemann, M., Keßler, J-U., & Itani-Adams, Y. (2011). Comparing processability across languages. *International Journal of Bilingualism*, 15, 128–146.
- Pika, S., Nicoladis, E., & Marentette, P. F. (2006). A cross-cultural study on the use of gestures: Evidence for cross-linguistic transfer? *Bilingualism: Language and Cognition*, 9, 319–327.
- Ping, S. W., & Rickard Liow, S. J. (2011). Morphophonemic transfer in English second language learners. *Bilingualism: Language and Cognition*, 14, 423–432.
- Purpura, J. E. (2008). Assessing communicative language ability: Models and their components. In E. Shohamy & N. H. Hornberger (eds.), *Encyclopedia of language and education* (2nd edn., vol. 7): *Language testing and assessment*, pp. 53–68. Dordrecht: Kluwer.
- Reyes, I., & Hernández, A. E. (2006). Sentence interpretation strategies in emergent bilingual children and adults. *Bilingualism: Language and Cognition*, 9, 51–69.
- Salamoura, A., & Williams, J. N. (2006). Lexical activation of cross-language syntactic priming. *Bilingualism: Language and Cognition*, 9, 299–307.
- Salamoura, A., & Williams, J. N. (2007). The representation of grammatical gender in the bilingual lexicon: Evidence from Greek and German. *Bilingualism: Language and Cognition*, 10, 257–275.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.
- Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, 11, 245–260.
- Slabakova, R. (2005). What is so difficult about telicity marking in L2 Russian? *Bilingualism: Language and Cognition*, 8, 63–77.
- Snape, N. (2008). Resetting the nominal mapping parameter in L2 English: Definite article use and the count–mass distinction. *Bilingualism: Language and Cognition*, 11, 63–79.
- Soares, C., & Grosjean, F. (1984). Bilinguals in a monolingual and a bilingual speech mode: The effect of lexical access. *Memory & Cognition*, 12, 380–386.
- Sopata, A. (2005). Optionality in non-native grammars: L2 acquisition of German constructions with absent expletives. *Bilingualism: Language and Cognition*, 8, 177–193.
- Sundara, M., Polka, L., & Baum, S. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition*, 9, 97–114.
- Taube-Schiff, M., & Segalowitz, N. (2005). Within-language attention control in second language processing. *Bilingualism: Language and Cognition*, 8, 195–206.
- Tokowicz, N., Michael, E. B., & Kroll, J. F. (2004). The roles of study-abroad experience and working-memory capacity in the types of errors made during translation. *Bilingualism: Language and Cognition*, 7, 255–272.
- Toribio, A. J. (2001). On the emergence of bilingual code-switching competence. *Bilingualism: Language and Cognition*, 4, 203–231.
- Treffers-Daller, J. (2011). Operationalizing and measuring language dominance. *International Journal of Bilingualism*, 15, 147–163.
- Tremblay, A. (2011). Learning to parse liaison-initial words: An eye-tracking study. *Bilingualism: Language and Cognition*, 14, 257–279.
- Trenkic, D. (2008). The representation of English articles in second language grammars: Determiners or adjectives? *Bilingualism: Language and Cognition*, 11, 1–18.
- Uchikoshi, Y. (2006). English vocabulary development in bilingual kindergarteners: What are the best predictors? *Bilingualism: Language and Cognition*, 9, 33–49.
- Von Studnitz, R. E., & Green, D. W. (2002a). Interlingual homograph interference in German–English bilinguals: Its modulation and locus of control. *Bilingualism: Language and Cognition*, 5, 1–23.
- Von Studnitz, R. E., & Green, D. W. (2002b). The cost of switching language in a semantic categorization task. *Bilingualism: Language and Cognition*, 5, 241–251.

- Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, 9, 71–88.
- Yan, B. (2007). Behavior of *wh*-words in English speakers' L2 Chinese *wh*-questions: Evidence of no variability, temporary variability and persistent variability in L2 grammars. *Bilingualism: Language and Cognition*, 10, 277–298.
- Yan, S., & Nicoladis, E. (2009). Finding le mot juste: Differences between bilingual and monolingual children's lexical access in comprehension and production. *Bilingualism: Language and Cognition*, 12, 323–335.