View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Central Archive at the University of Reading

University of Reading

The measurement of bilingual abilities: central challenges

Book or Report Section

Accepted Version

Treffers-Daller, J. (2018) The measurement of bilingual abilities: central challenges. In: De Houwer, A. and Ortega, L. (eds.) The Cambridge Handbook of Bilingualism. Cambridge University Press, Cambridge, pp. 289-306. ISBN 9781316831922 doi: https://doi.org/10.1017/9781316831922.016 Available at http://centaur.reading.ac.uk/75364/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1017/9781316831922.016

Publisher: Cambridge University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

University of Reading

Central Archive at the University of Reading

Reading's research outputs online

Chapter 16. The Measurement of Bilingual Abilities: Central Challenges

Jeanine Treffers-Daller

In: A. de Houwer and L. Ortega (Eds.) (2019). *The Cambridge Handbook of Bilingualism* (pp.289-306). Cambridge: Cambridge University Press

1 Introduction

Scholars use the term "bilingual" to refer to individuals who learned their languages under very different circumstances and who may differ widely from each other in what they can do or usually do with these languages in everyday life. They do not necessarily define bilingualism based on how well a person speaks a language. This can be rather confusing to outside observers who are faced with the complexity of the bilingual experience and may ask themselves: "Do I have to be able to fully understand two languages and speak them fluently to count as a bilingual?", or: "Am I a bilingual if I can have an informal conversation in two languages?". Answering these questions is not easy, as researchers define bilingualism in a variety of ways.

The contributions to this Handbook show that the concept and definition of bilingualism go well beyond the type of language abilities that individuals have. Yet, the issue of the measurement of bilingual abilities is of central to importance to researchers in the field. The range of approaches to understanding and measuring bilingual abilities, however, makes it difficult to summarize the available literature and compare apparently contradictory results with each other. A key problem in tackling the issue of comparability among studies is that researchers use different terms for key concepts and that concepts used across the field only partly overlap (see also Section 2). Grosjean (1998, p. 48) points out that contradictory findings are often due to the lack of clarity of the concepts under investigation, including the concept of the bilingual individual itself, and to how concepts are operationalised in a particular study. Fortunately, the situation has somewhat improved over the years. This can be seen, for instance, in Hulstijn's (2012) analysis of 140 articles published in the journal *Bilingualism, Language and Cognition,* one of the leading journals in the field, in the period from its first publication in 1998 up to 2011. Hulstijn found that just 19% of the papers published in the first half of this period used an objective test to measure informants' language abilities, while this percentage had increased to 54% of the papers in the second half of the period investigated.

Many researchers now in principle subscribe to a holistic view of bilingualism (Grosjean, 1985), according to which all the languages spoken by bilinguals need to be

included in analyses and not just bilinguals' second or weaker language(s). This view is also shared by organisations of practitioners such as the Royal College of Speech and Language Therapists in the United Kingdom (Royal College of Speech & Language Therapists, 2006).

A more negative, fractional view of bilinguals is, however, still prevalent in much of the academic literature. In many studies comparisons are made between so-called native speakers (assumed to be monolingual) and bilinguals, which often result in a negative view of the abilities of the latter by comparison with the former. Differences between monolinguals and bilinguals are then labelled “the bilingual deficit”, and bilinguals who have a stronger and a weaker language are labelled “unbalanced bilinguals” or even “non-proficient bilinguals”. This deficit view is unfortunately still prevalent in both the field of second language acquisition and that of bilingualism, in spite of Skutnabb-Kangas's (1981, p. 194) warning many years ago that simple measurements on which monolinguals and bilinguals are compared do more harm than good. Considering the output of L2 learners as incomplete approximations of a purported complete monolingual native speaker target has aptly been labelled the target-deviation perspective by several scholars in second language acquisition (see Ortega, this volume). As it is high time to abandon this perspective, Ortega (2013) calls for a bi/multilingual turn in second language acquisition research in which the focus is not on what L2 learners cannot achieve but on what makes learning a language later in life unique. However, as Pearson (2010) notes, such a bilingual turn still needs to be accomplished even in the field of bilingualism itself.

Comparing L2 learners with presumably monolingual native speakers is also problematic because such native speakers’ abilities differ widely from each other. Alderson (1980) not only found that what he termed native speakers did not always restore grammatical gaps in a cloze test but also that the differences between native and what he called non-native (L2) speakers were very small, and that some non-native speakers outperformed the native ones. This findings led Alderson to conclude that “native speaker proficiency, even on lower-order tasks, varies” (1980, p.74) and he considered the use of native speakers as criteria-setting for non-native speakers on tests to be “misguided” (1980, p.75). In a similar vein, Hamilton, Lopes, McNamara and Sheridan's (1993) reported on the performance of presumably monolingual native speakers on a reading and writing test that is widely used with adults seeking to work or study in an English-speaking country (the International English Language Testing System, or IELTS, is a, see https://www.ielts.org/). Hamilton et al. found this performance to be “far from homogeneous” (1993, p. 348), and related to educational level and work experience: Native speakers with higher levels of

education or jobs which required them to engage in extensive high level reading obtained significantly higher scores than respondents with lower educational levels (see, e.g., Street & Dąbrowska, 2010, for similar findings). Such findings and his own study comparing large groups of native and non-native speakers of Dutch led Andringa (2014) to posit that choosing a sample of native speakers that is representative of the full range of educational and socioeconomic and occupational backgrounds in any society is crucially important if one wants to compare native speakers against non-native speakers. In many studies, native speaker samples have unfortunately been biased in favour of just highly educated speakers, often consisting of college students who are arguably from relatively affluent or middle class families, thus adding an important variable other than just native speaker status. Comparisons between so-called native and non-native speakers are much more valid if educational levels and socioeconomic and occupational backgrounds are controlled for.

Hulstijn (2015) points out that very little research has been done into the abilities shared by (monolingual) native speakers. In his model of language proficiency in native and non-native speakers, Hulstijn proposes that native speakers share the most frequent words and grammatical rules that are being used in oral/aural modes (Basic Language Cognition, BLC). However, they do not necessarily share any of the infrequent lexis and uncommon grammatical rules that are used in written (and occasionally in some oral/aural) modes (Higher Language Cognition, HLC). Therefore there is great variability in native speakers' knowledge, which makes it clear that using monolingual "norms" as a yardstick against which the performance of second language learners is measured is highly problematic.

Finally, as Rothman and Treffers-Daller (2014) have argued, it is in fact not appropriate to reserve the label "native speaker" for monolinguals only. Given that a native language is "one that is acquired from naturalistic exposure, in early childhood and in an authentic social context/speech community" (Rothman & Treffers-Daller, 2014, p. 95), bilinguals who have acquired two or more languages in early childhood have more than one native language.

Avoiding the deficit view inherent in much of the work in the field is extremely important. Therefore the key aim of this chapter is to identify what makes individuals with bilingual abilities unique speaker-hearers in their own right and how these abilities can be measured. In Section 2, I first explore the construct of bilingual abilities, how it relates to some other associated constructs and how it can be modelled in different theoretical frameworks. Identifying the uniqueness of the bilingual experience will necessarily entail referring to code-switching and translanguaging and to the mutual influences between the

languages spoken by bilinguals, which have been described with a range of different terms, including crosslinguistic influence and transfer (the older negative term interference has fallen into disuse). In Section 3 I move on to the operationalisation and measurement of bilingual abilities in relation to how appropriate they are for different types of bilinguals. Attention will be paid to the measurement of bilingualism across the lifespan, and wherever appropriate assessments for children are discussed as well as assessments for adults. The final section focuses on directions for future research in the field.

2 Bilingual Abilities

2.1 Defining the Construct

Before we can discuss the different ways in which bilingual abilities can be operationalised or measured, it is important to clarify what the construct means and how it relates to other constructs. The construct of bilingual (language) abilities is an expansion of the notion of language ability, which was developed in the field of second language testing by Bachman and Palmer (2010). Bachman and Palmer (2010) see the construct as covering a wide range of abilities, including linguistic and sociolinguistic abilities and strategic knowledge. The term bilingual abilities comprises all these different components, but the choice of a plural form is more appropriate for bilinguals, as it reflects the fact that bilinguals possess these abilities in more than one language.

The construct of bilingual abilities is not limited to the domain of syntax. This is signalled by the terminological choice of “abilities” over “competence”, as the latter term is strongly associated with the generative framework and in that context often interpreted to be the ideal speaker-hearer’s underlying rule based knowledge of her/his language (Chomsky, 1965). In the current chapter I conceive of bilingual abilities as encompassing a wide range of uses of language and many different dimensions, including both traditional notions of proficiency that are narrow as well as sociolinguistic, pragmatic or strategic competencies. This understanding contrasts with other existing delimitations of L2 ability in the field. On the one hand, it is broader than the construct of language proficiency proposed by Hulstijn (2015), who mainly refers to lexical, grammatical, phonotactic and prosodic elements, but does not include sociolinguistic, pragmatic or strategic competencies. On the other hand, it is not as comprehensive as the widely used construct of multi-competence, which originally referred to the knowledge of more than one language in the same mind (Cook, 1991), but was later expanded to cover not just language knowledge but also wider cognitive processes and has recently been redefined as “the overall system of a mind or a community that uses more

than one language" (Cook, 2016, p. 3). In discussing bilingual abilities below, I will identify different components, dimensions, domains or skills within this construct from a range of different perspectives. It is not possible to review all of these here, and the points made are necessarily selective.

Importantly, the construct of bilingual abilities also covers the ability to process two languages. However, modelling bilingual processing abilities is a complex task because, as Larsen-Freeman, Schmid and Lowie (2011) point out, models of non-static scenarios, which include bilingualism, need to account for systems that are under development as is the case, for example, for children who are learning two languages from birth or in adult L2 learners. Another key issue for models of lexical processing is how to account for the ways in which bilinguals map meaning onto form (see Paap, this volume, for a review of bilingual processing models).

Particularly challenging for models of bilingual language processing is the case of bimodal bilingualism as found among bilinguals who use a spoken and a signed language. Contrary to unimodal bilinguals, bimodal bilinguals can produce simultaneous output in two languages through oral and visual channels (see Tang & Sze, this volume, for extensive discussion). This simultaneous production through oral and visual channels presents challenges for models of language production designed to explain the production of single language elements at the time. In addition, while both bimodal and unimodal bilinguals can switch between languages, bimodal bilinguals' production of code-blends (simultaneously produced words and signs) reveals that coactivation and inhibition processes in bimodal bilinguals are different from those found in unimodal bilinguals (Emmorey, Giezen & Gollan, 2016).

A few words also need to be said about comprehension versus production abilities, and the traditional division into four skills: speaking, listening, reading and writing. For a given language, there can be vast differences between a bilingual's level of comprehension and production. As discussed in De Houwer (2007), bilingual children can develop oral production skills in the languages they are exposed to. However, whether or not they actually do so depends to a large extent on the input patterns in the families in which they grow up. In her analysis of a dataset of 1,356 bilingual families in Flanders in which parents spoke both a minority language and the majority language, Dutch, at home, De Houwer (2007) showed that success rates in the intergenerational transmission of the ability to speak the minority language differed according to the distribution of the languages in the parent pair, and ranged from 34% to 93%. The most successful transmission occurred in families where one parent

spoke the minority language and the other parent spoke the minority as well as the majority language. While bilingual families are understandably keen to have their children develop production skills in all languages spoken at home, receptive multilingualism (Bahtina & ten Thije, 2012) may be an asset both for individuals and society at large, because bilinguals who have comprehension skills across different languages can communicate in settings in which they would otherwise face insurmountable barriers.

Although many researchers have focused on the development of oral/aural skills, the number of studies focusing on bilingual children's reading skills is steadily increasing. These studies show that some of the skills that underpin the ability to read, such as phonological awareness (e.g., being aware of the sounds the words *hit* and *bit* have in common) and morphological awareness (e.g., being aware of the meaning of individual morphemes such as *–er* in *worker*) are transferable between languages. This is particularly the case if the language in which children have learned to read first has a transparent orthography, that is, is characterised by a regular or close to one-to-one correspondence between graphemes and phonemes (Kirby, Deacon, Bowers, Izenberg, Wade-Woolley & Parrila, 2012).

The fact that certain skills transfer between languages can also be explained in the framework of Cummins' (1979, p. 222) interdependence hypothesis, which states that "the development of competence in a second language (L2) is partially a function of the type of competence already developed in L1 at the time when intensive exposure to L2 begins". In their study involving 1,062 Spanish L1 children, Winsler, Kim and Richard (2014) have offered very strong evidence for the validity of the interdependence hypothesis: Children who had stronger skills in Spanish at age 4 made faster progress in learning English a year later.

Cummins (1979) furthermore proposed a fundamental distinction between Basic Interpersonal Communicative Skills (BICS) and Cognitive Academic Language Proficiency (CALP), where the former refers to everyday conversational language, and the latter to language used in educational contexts. This dichotomy bears some similarity to Hulstijn's (2015) distinction between Basic Language Cognition and High Language Cognition (see the Introduction), although Cummins' distinction was not proposed as a theory of language but mainly intended to offer guidance for policy and practice. Cummins' distinction between BICS and CALP has been very influential in educational circles, in particular in those with a particular interest in the academic development of minority children, although it has also met with criticisms and modifications along the way (see Cummins, 2016). Going well beyond Cummins' notion of interdependence, the relation between a bilingual's languages has been studied in great detail.

2.2 Dominance and Balance Problematized

Many researchers assume that bilinguals often have a stronger and a weaker language, and that it is rare for bilinguals to be balanced (Silva-Corvalán & Treffers-Daller, 2016), that is, to have equal abilities in their two languages. In fact, Romaine (1989) is particularly strongly opposed to the idea of balance, which to her is largely an artefact of a deficit view of bilinguals, and does not do justice to the specific abilities of bilinguals. Scholars' general preference for "balance" in bilinguals and treating it as a kind of ideal possibly also results from a general bias in favour of symmetry across a wide range of fields, including architecture, biology and music (Treffers-Daller, 2016). However, as both Birdsong (2016) and Treffers-Daller (2016) pointed out, perfect balance in bilinguals is probably as rare as ambidexterity, a term used to describe the skills of a person who can carry out a variety of tasks with equal skill with either hand. Most individuals are predominantly right-handed or predominantly left-handed, and no one would qualify right-handed individuals as "unbalanced" just because they can carry out some tasks better with their right hand. For speakers of three or more languages it is even less likely that the distribution of labour between the languages or the levels of ability achieved in each are completely balanced (Curdt-Christiansen, 2016). De Houwer and Bornstein (2016) have furthermore pointed out that bilinguals can indeed perhaps have similar levels of, for instance, word knowledge in each language, but that these levels are not necessarily high. In other words, "balance" is not necessarily a positive thing.

Bilinguals who have a stronger and a weaker language are often labelled "unbalanced", with all the negative connotations a term with *un*- implies. However, as many scholars have noted, different abilities in each language are the normal result of the fact that bilinguals learn and use their languages for different purposes in everyday life, for example, one language at home and one at work or school. Different abilities in each of their languages are simply linked to the specific purposes for which bilinguals need each of them. Different learning and usage profiles for each language make it extremely unlikely that bilinguals will develop exactly the same skills in each language. In fact, Grosjean (2016) estimates that bilinguals often have translation equivalents for only 30-37% of their words in the other language, which clearly shows that the communicative needs and abilities in one language are not necessarily mirrored by those in another language. Grosjean (2016) notes there is a need to dig much deeper into the different activities and functions for which a bilingual's

languages are being used as the skills and knowledge that a bilingual develops in each language are strongly domain specific.

2.3 Uniquely Bilingual Abilities

The task of fully defining bilingual abilities calls for a consideration of abilities that are uniquely available to bilinguals. The uniqueness of the bilingual experience is perhaps most visible in the ability of bilinguals to seamlessly switch from one language to another, which is generally called code-switching, code-mixing, or, within some perspectives, translanguaging (for further discussion, see the chapters by De Houwer and García & Tupas, this volume). Poplack (1980) was the first to show that code-switching is not a sign of lack of mastery of languages but is, in fact, highly skilled behaviour, and that the most intimate forms of code-switching are found among bilinguals with high levels of ability in both languages. A detailed model of code-switching is proposed in Muysken (2013), who distinguishes four main types of code-switching and shows how the occurrence of these different types depends on (a) typological factors, (b) processing constraints linked to the bilingual abilities of the speakers and (c) societal factors, such as the duration of language contact and power relations within society (see further Aalberse & Muysken, this volume). Psycholinguistic approaches to code-switching which focus on switching from one language to another that is externally induced in a laboratory setting (Gullberg, Indefrey & Muysken, 2009) are becoming more prominent. Such studies can provide more insights into the simultaneous processing of two languages in bilinguals. Work in this field shows in particular that bilinguals can never completely switch off any of their languages, and that words from different languages are probably not stored separately. When speakers need to retrieve a particular word from one of their languages, they need to search both lexicons, a phenomenon often referred to as non-selective access (see further Paap, this volume).

Spontaneous code-switching is different from crosslinguistic influence (CLI), which does not generally involve the importation of lexical material (words) from one language into another, but only a sound, a meaning or a grammatical pattern (Treffers-Daller, 2009). Examples of CLI can be found in the devoicing of final voiced obstruants such as [g] in English words such as *bag*, which are often pronounced [bæk] instead of [bæg] by L1 German users of English as an L2, or in the choice of the preposition *on* in *on school* by L1 Dutch users of English as an L2, where *at school* would be the conventional choice. The forms used by L2 users of English here are likely the result of crosslinguistic influence from German and Dutch, respectively. The existence of CLI across all levels of linguistic analysis

(phonetics, phonology, morphology, lexicon, syntax, pragmatics, etc.) also means that bilinguals can never be two monolinguals in one person, and monolingual-likeness in either of the two languages is generally an impossibility (see Biedroń and Birdsong, this volume).

CLI may be reinforced if speakers with the same L1 hear such L2-influenced pronunciations frequently as members of the same bilingual community of, for example, Dutch-English or German-English bilinguals in the United Kingdom. The longer the duration of language contact between two communities, the more such language contact phenomena can become integrated into the language varieties spoken by individuals growing up in bilingual communities (in both directions), probably as a result of the constant co-activation of both languages in the minds of the speakers (these are studied in the field of contact-induced language change; see Aalberse & Muysken, this volume). For researchers in the field who take the notion of multi-competence (Cook, 2016) seriously it is important to be aware of the long term effects of language contact on the varieties spoken by bilinguals. This is crucial because the language varieties spoken and heard by bilinguals are often rather different from the varieties that are spoken outside the bilingual communities under study. This has important implications for qualifying the input to, for example, heritage speakers who learned their L1 (the heritage language) in childhood, but became more proficient or dominant in another language later in childhood (Benmamoun, Montrul & Polinsky, 2013; see also Armon-Lotem & Meir, this volume). Speakers of Turkish in Germany, for example, grow up with a variety of Turkish that is considerably different from that spoken in Turkey, and thus develop abilities in Turkish that are different too. Evaluations of speakers' bilingual abilities must take into account the fact that bilinguals may not have had the chance to learn monolingual varieties of a particular language.

3 Measuring Language Abilities in Two or More Languages

3.1 Measuring a Multifaceted Phenomenon and Dealing with Variability

As the construct of bilingual abilities is a very wide ranging one, no single measure can assess a bilingual's abilities in any comprehensive manner. Most researchers nowadays prefer to choose a particular domain, skill or component that they consider to be important, and measure informants’ abilities in that area, in the knowledge that the measure represents only a fraction of the multifaceted nature of any bilingual individual's language abilities.

Measuring a wide range of abilities is generally not an option. This is why in some cases the measurements that are used to assess specific phenomena (e.g., scores on vocabulary tests) are interpreted as a proxy for the construct of bilingual abilities itself.

Especially vocabulary tests, and in particular tests which measure bilinguals' vocabulary size in their L2, are frequently used as a proxy for general language proficiency. However, Kremmel and Schmitt (2016) point out that researchers who use such tests usually do not offer reflections on what exactly the tests measure. This means that the rationale for the choice of this particular component as a proxy for general language proficiency is rather weak.

Researchers agree that bilinguals' abilities differ widely from each other. Those who learn two languages from birth are different from those who learn an L2 in childhood (De Houwer, 2013), and these are again different from those who learn an L2 in adulthood or those who are in the process of losing one of their languages (compare, for instance, Singleton & Pfenninger and Keijzer & de Bot, this volume). As is apparent from many of this Handbook's contributions, the variability in language abilities even within a single type of bilinguals can be very large. The variability also increases with the number of languages that people know: Each of their languages can start at different points in time and the scenarios in which individuals become multilingual may be very different from each other (see Quay & Montanari, this volume).

In order to gain an understanding of the variability in bilingual abilities, researchers often use other- and self-reports. It is common for researchers measuring children's bilingual abilities during the first years of life to use parent report forms, where parents are asked to check lists that provide estimates of the words their children understand or produce (to be discussed below). Researchers working with adults, too, have used reports, but in this case in the form of self-assessments, that is, they ask individuals to indicate their own abilities in some or all of the four skills, that is, gauging their own understanding, speaking, reading and writing, often using a simple rating scale. Interestingly, however, researchers hardly ever ask bilinguals to assess their own proficiency in both languages in terms of more detailed components of language use (vocabulary, grammar, phonology, pragmatics, etc.). As Hulstijn (2012) points out, assessing one's own vocabulary, phonological or grammatical knowledge is quite difficult for non-experts, and therefore not really an option for lay self-assessments.

Scholars who use self-assessments often correlate these with actual measures of performance such as a test of speaking, reading, or writing. In their meta-analysis of correlations between self assessments and performance outcomes across a range of domains (e.g., "language competence", sports, or academic ability) Zell and Krizan (2014) show that such correlations tend to be of moderate strength only. They also claim, however, that for

language competence correlations between subjective perceptions and actual performance can be higher than in most other domains they investigated. Their claim is based on a high correlation coefficient of *r*=.63 found in the only study of language self-assessment they were able to include, namely an earlier meta-analysis of 60 studies carried out by Ross (1998), who looked at self assessments and objective measurements of L2 abilities across the four traditional language skills. Zell and Krizan (2014) attribute the relatively high correlation for language competence to the existence of "ubiquitous" feedback and to language skills being "objectively defined" (p. 117). While it may be true that L2 learners get regular feedback on their performance if they are in a classroom setting, this is not normally the case for bilinguals who have learned their languages in uninstructed settings. Also, the separability of the four skills has been heavily criticized (Bachman & Palmer, 2010): Writers need to read over what they have written, and speakers need to listen to their interlocutors. Bachman and Palmer (2010) instead suggest that more fine-grained approaches which specify what they call Target Language Use Domains (such as making a compliment or a complaint) offer potentially more fruitful avenues for the (self-) assessment of learners' skills.

As bilingualism is such a multifaceted phenomenon, the key question is how we can measure bilingual abilities in such a way that the measurements are appropriate for the target groups and do not reinforce the deficit view.

There is a large variety of language instruments and tests that have been developed for monolingual children that are increasingly being used with bilingual children. The norms associated with these tests are generally appropriate for only monolinguals and not for bilinguals (Gathercole, Thomas, Roberts, Hughes & Hughes, 2013). Defining norms for bilingual children is particularly complex because of the great variability among them (Cruz-Ferreira, 2010). As bilingual children can be exposed to their two languages in very different proportions, and may have started from birth or later in childhood, Gathercole et al. (2013) propose that levels of exposure need to be taken into account in the measurement of bilingual abilities (see the contributions by Armon-Lotem & Meir and Scheffner-Hammer & Edmonds, this volume, for further discussion).

Considerable progress has been made in the development of new measurements focused on young bilinguals' abilities. Some of these are summarized below. Vocabulary has been a central area in the measurement of bilingual abilities, particularly in children. This may relate to the fact that in the first years of life the comprehension and production of words vocabulary are the first clearly language related skills to emerge, and that only later grammatical development comes in. In the first school years vocabulary becomes central to

the acquisition of literacy and it is central to overall academic development. As discussed in the next subsection, measuring vocabulary well (across the lifespan) has its own methodological challenges.

3.2 Measuring Vocabulary

An important new way to measure bilingual children's vocabulary size across two languages was developed by Pearson, Fernández and Oller (1993). In order to measure vocabulary knowledge in Spanish and English in children who were growing up with two languages from the beginning, Pearson et al. used the American English and the Spanish versions of the MacArthur-Bates Communicative Development Inventory (CDI), a standardized parent report form through which (monolingual) children's comprehension and production vocabularies are noted (Fenson, Dale, Reznick, Thal, Bates, Hartung, Pethick & Reilly, 1993). Pearson et al. (1993) introduced the notion of Total Conceptual Vocabulary (TCV), which refers to the counting of all concepts a child understands, regardless of whether s/he knows the concept in either one or both of her/his languages. More precisely, TCV refers to the total number of lexicalised concepts (in fact, form-meaning pairings) known by bilingual children. In monolinguals, the TCV score is assumed to equal total word knowledge, called the Total Vocabulary Score. In bilinguals, the TCV is calculated by counting cross language synonyms which are both known by a child only once, and by accepting the child's knowledge of the word in just a single language as knowledge of the relevant concept. Thus, supposing a child understood the two words Spanish *perro* as well as its English translation *dog*, TCV would be 1; it would remain 1 if the child only knew Spanish *perro* or only English *dog*. Thus, the TCV differs from a simple total score of all words known summed across both languages (the Total Vocabulary Score) as in that calculation *perro* and *dog* would each be counted.

Pearson et al. (1993) compared the bilingual children's Total Vocabulary and TCV scores to those of a demographically matched group of mostly English-speaking monolingual children with reference to (English) monolingual-based percentile norms, regardless of whether individual bilingual children knew more words in one language than the other. No bilingual-monolingual differences were found. Note, however, that at the time of the Pearson et al. (1993) study, monolingual percentile norms existed only for the American English CDI, but Pearson et al. took these percentile norms as a proxy for Spanish as well. As more recent work has shown, CDI-based monolingual percentile norms may in fact differ across languages. In addition, in order to compare children's abilities in a single language only (for

which Total Vocabulary and TCV scores are irrelevant), Pearson et al. determined which language bilingual children knew more words in (as they say, were more dominant in). They then compared only the bilingual children's dominant language to the monolinguals' only language, again with reference to monolingual-based percentile norms. The bilingual children's production lexicon sizes in their dominant language were not significantly different from those of the mostly English-speaking monolinguals. Only for the Total Vocabulary Scores and the TCV score comparisons, then, were Spanish and English dominant children combined and compared as a group to monolinguals. The methodology followed in this study reveals that a deficit approach to the measurement of bilingual abilities can be avoided by taking into account within group differences among bilingual children.

Different results emerge from a meta-analysis of studies involving 1,738 children (996 bilinguals and 772 monolinguals) between the ages of 3 and 10 by Bialystok, Luk, Peets and Yang (2010). These researchers showed that bilingual children have smaller comprehension vocabularies in English as measured with the Peabody Picture Vocabulary Test (Dunn & Dunn, 2007) than their monolingual counterparts across the age range investigated. The authors found a nine-point difference between the results of the monolinguals and the bilinguals in their sample. Unfortunately, however, the children were not divided into dominance groups as in Pearson et al. (1993). The large discrepancies between the two groups seems to indicate that English was actually the weaker language of the two for many children, and probably a language they had started learning later in life (De Houwer, Bornstein & Putnick, 2014). It is understandable that precise measurements of language dominance are difficult to make for a sample which includes such a wide range of languages other than English. However, the children or the parents could have been asked what the child's stronger or weaker language(s) were. Separate analyses could then have been made of the English vocabulary of English-dominant children and children for whom the other language(s) was/were dominant. It is likely that at least for a subgroup of children in the sample, the English-dominant ones, knowledge of English vocabulary was similar to that of monolingual children. Interestingly, for words related to the school domain, no significant differences were found between bilinguals and monolinguals, which shows that looking at within group differences is likely to provide a more positive, differentiated picture of the abilities of bilinguals.

De Houwer et al. (2014) took up the challenge to control for the many sociolinguistic variables that affect vocabulary development in their study of the comprehension and production vocabularies of 31 Dutch-French bilingual children and 30 monolingual children

Dutch-speaking children, aged 13 and 20 months, using Dutch and French versions of the CDI. The authors ensured the two groups were matched in age, birth status, birth order, family socioeconomic status (SES), and gender. In addition, the bilingual children were exposed to each language for the same length of time as the monolingual children in the sample because they had grown up with both languages from birth. Thus, the authors avoided potential confounds in the study that could be the result of variables which are known to affect vocabulary development. After completing detailed analyses of monolinguals and bilinguals at the two different ages, De Houwer et al. (2014, p. 1209) conclude that there is "no evidence of consistent differences between young bilinguals' and monolinguals' vocabulary sizes" and recommend that researchers look at causes other than bilingualism to explain variability in lexical knowledge within and between groups.

A key issue that makes it difficult to assess both languages of a bilingual is the fact that there is a lack of standardized assessments for a wide range of languages spoken by bilinguals. It is not possible to simply translate an English test or instrument into another language as the vocabulary items and the grammatical structures that need to be tested do not have a one-to-one correspondence to those in another language. Also there are intercultural differences that make it likely that children or adults use very different vocabulary items in different languages. In fact, Gathercole (2013) suggests that translating a test from one language into another would be similar to giving a piano player a musical test that was created for the drums.

While creating tests that are completely equivalent in two languages will remain an elusive goal, comparable assessment instruments are urgently needed for researchers and practitioners alike. This is why researchers have adapted existing measures to other languages. The CDI, for example, now exists for close to 100 language varieties, including several sign languages, and new adaptations are being constantly developed (see https://mb-cdi.stanford.edu/adaptations.html). The PPVT has been adapted for a number of other languages too, including British English, French, German, Spanish and Welsh, although concerns about possible cultural biases in its use are reported for some populations (Haitana, Pitama & Rucklidge, 2010).

An innovative way of measuring vocabulary knowledge among children across different languages can be found in Haman, Łuniewska and Pomiechowska (2015), who started the development of Crosslinguistic Lexical Tasks (CLTs) for five-year-old children as part of the Bilingualism-Specific Language Impairment (BiSLI) Cost Action IS0804 (http://www.cost.eu/COST_Actions/isch/IS0804). Because bilingual children as well as

children with Specific Language Impairment (SLI) sometimes present with smaller vocabularies in a particular language compared to monolingual typically developing children, it is important for researchers and practitioners to have tools to be able to measure vocabulary in pre-school children, so that those with delays in vocabulary development and those with SLI can be distinguished from each other (see also Scheffner-Hammer & Edmonds, this volume). The novelty of the Haman et al. (2015) approach is that they did not develop CLTs by adapting an existing English test, as is being done for the CDI and the PPVT, but that they initiated a new crosslinguistic assessment tool. They created a database of over 1,000 pictures of actions and objects which groups of adult speakers of 34 different languages were asked to name. Participants were also asked to evaluate whether these could be named easily and unequivocally. Subsequently, stringent criteria were developed for the selection of 30 nouns and 30 verbs associated with the pictures to ensure that the words that were selected for the CLTs in different languages were comparable with respect to, for example, the complexity of the candidate words. A cultural bias was avoided because the tests were developed in parallel for different languages, and there was no single language that was taken as the model for the CLTs. Variants of pictures (e.g., women with or without a head scarf) were created for different languages to ensure that picture styles were appropriate for the culture of the language under investigation. While the aim of the CLTs was to differentiate between typically developing and language-impaired monolingual and bilingual children, in future they can probably also be used as a baseline measure of bilingual abilities and/or language dominance in bilingual children, precisely because the measurements will be comparable across languages. The availability of reaction times for this task makes it an attractive option for researchers interested in lexical processing too.

The vocabulary measurements considered here so far mainly tap into bilinguals' knowledge of single words, which are often tested out of context. Such a focus on single words is typical of the research carried out with young bilingual children. Of course, vocabulary knowledge encompasses much more than knowledge about the link between a form and a meaning of an isolated word, as it primarily involves knowledge about the use of words in context (Nation, 2001). The roles of context, polysemy, and phraseology in shaping vocabulary knowledge increase and become particularly central to the measurement of the adult lexicon. Valid assessments of adults' bilingual lexical abilities should therefore include measurements of bilinguals' knowledge of the ways in which words are used. This also involves knowledge of different linguistic registers, as in, for example, knowing what the difference is between the use of *help* and *aid*. Such nuances are more difficult to measure

than comprehension or production knowledge about the link between the form and meaning of isolated words. Much of the research that has been conducted in Corpus Linguistics and Second Language Acquisition in recent years has focused on the patterns in which words are used in first-language corpora (e.g., Hoey, 2005) and on the acquisition of such patterns and fixed expressions, often called formulaic sequences, by second language learners (e.g., Martinez & Schmitt, 2012). While tests of such fixed expressions have been developed for English and for some other languages, we are not yet in a position to assess bilingual abilities in the use of formulaic sequences. As is well known, there is no one-to-one translation of formulaic sequences in two languages, so that an expression such as *pay attention* translates as *faire attention* "make attention" in French, and *comparing apples and oranges* becomes *Äpfel und Birnen vergleichen* "comparing apples and pears" in German. These examples serve to underscore the fact that measurements of bilingual abilities that are just based on comparisons of the knowledge of single words do not do justice to the highly complex abilities that are involved in using words in context in general and in fixed expressions in particular. Analyses of formulaic language could be made on the basis of transcripts of oral language, particularly if speech samples are elicited with a standardised tool, as free speech would be too variable to make analyses of formulaic sequences possible.

3.3 Select Approaches to Measuring Syntactic Abilities

This brings us to the measurement of syntactic abilities in two languages. There are many grammar tests, some standardized and some home-made, that are used by researchers in bilingualism and second language acquisition. Space allows only a brief discussion of some selected approaches here.

Measuring grammar across a bilingual's languages is a complicated undertaking because of the large typological differences between languages. The Mean Length of Utterance (MLU) is a widely used measure of language development in young children. Some researchers (e.g., Schmeißer, Hager, Arnaus Gil, Jansen, Geveler, Eichler, Patuto & Müller, 2016) opt for comparisons of MLU in two languages, and compute language dominance indices on the basis of MLU differentials. However, many authors have pointed out that a key issue here is that for some languages such as Chinese, MLU is best computed in words, but for others, such as English, it is best computed in morphemes. Typological differences between languages make it therefore very difficult to compare MLUs across languages. Other measures of complexity are the percentage of multimorphemic utterances or the Upper Bound (that is, the number of morphemes or words in a child's longest utterance),

but, again, these are difficult to compare across languages. Moreover, these measures do not really tap into the complexities of grammar as such and MLUs are not a valid indicator of syntactic ability beyond a tally of about four.

Interesting approaches to the measurement of syntactic abilities in bilingual children have been developed by Chiat, Armon-Lotem, Marinis, Polišenká, Roy and Seeff-Gabriel (2013), who use sentence repetition tasks to measure sequential bilingual children's ability to produce specific sentence structures such as passive constructions or relative clauses in their L2. This initiative was further developed by Marinis and Armon-Lotem (2015), who show that sentence repetition tasks can be used to diagnose language delay or language impairment in bilingual children. Thus, if children have difficulties in both languages, this would be an indication of language delay or language impairment, whilst for typically developing children, uneven development across languages is entirely normal and to be expected, that is, children may be more advanced in one language than the other, but with one language generally showing no signs of any kind of delayed development (De Houwer, 2009; see also Scheffner-Hammer & Edmonds, this volume). Again as part of the Bilingualism-SLI COST Action mentioned in subsection 3.2, Chiat et al. (2013) developed sentence repetition tasks for a wide range of languages, which makes these potentially interesting for future measurements of bilingual abilities and language dominance. For adults too, elicited imitation tasks have been shown to be useful as a shortcut to measure L2 proficiency for research purposes, for example, in order to distinguish foreign and heritage language speakers at two proficiency levels of Mandarin (Wu & Ortega, 2013) or foreign language speakers of L2 Spanish at three different levels of proficiency and experience with the language (Bowden, 2016).

Currently, grammatical tests or instruments focusing on all the languages a bilingual knows are still lacking. The focus is usually just on a single language. This necessarily underestimates bilinguals' grammatical abilities.

4 Conclusion

In this chapter we have seen, first of all, that the construct of bilingual abilities is a broad, multidimensional one that cannot be adequately captured by assessing one skill or knowledge domain only. While many rese.archers choose to measure bilinguals' comprehension or production vocabulary knowledge in both languages as a proxy for the assessment of overall bilingual abilities during the early years of bilingual development, the choice of such a measure is not always well motivated. If we take the multidimensionality of

bilingualism seriously, and we acknowledge the importance of context, polysemy, and registers most particularly present in adult language usage, a far wider range of skills and knowledge that bilinguals possess should be taken into account in the assessment of individuals' bilingual abilities, such as their sociolinguistic and pragmatic knowledge. It has hardly been investigated, for example, to what extent child or adult bilinguals develop distinct politeness routines in two languages, or are able to use or understand standard varieties as well as regional varieties that are found in different parts of the countries (but see Chevrot & Ghimenton, this volume). The difficulty is, of course, that measuring such knowledge is much more complicated than measuring knowledge of the form and meaning of isolated forms.

It is difficult to specify which components of the construct are most relevant for the measurement of bilingual abilities as all depends on the research questions being asked. For some projects, lexical knowledge is most important, but for others grammatical, phonological or sociolinguistic knowledge might be more relevant. Few researchers offer principled reasons for their choice of a particular measurement. The emphasis on vocabulary in much child bilingualism research may be understandable, as measures of vocabulary have often been shown to be successful at discriminating between groups. In addition, by controlling for vocabulary, disadvantages that bilinguals may experience in lexical retrieval by comparison with monolinguals can disappear (Bialystok, Craik & Luk, 2008). Nevertheless, in most cases, the choice of the PPVT as a proxy measure of language ability seems to have been made mainly for reasons of convenience (see subsection 3.2).

Some researchers include a range of measures of bilingual abilities in the same study. Those who opt for this solution often find that bilingual abilities are task-dependent and that performance differs on each measure (Bahrick, Hall, Goggin, Bahrick & Berger, 1994; Kupisch & van de Weijer, 2016). This complicates the task of choosing which measure is to be used as a proxy for overall bilingual abilities and highlights once more the arbitrariness of decisions taken in some studies where the PPVT or MLU are adopted as measures of overall bilingual abilities.

A more principled view of the aspects of the construct of bilingual abilities that should take priority particularly when investigating the abilities of late-timed sequential adults can be found in Hulstijn (2015). He suggests that the measurement of bilingual abilities in such individuals should cover Basic Language Cognition (BLC; see the Introduction) in two languages only, as this is the dimension that is shared by all, irrespective of bilinguals' levels of education, while language users differ in their mastery of Higher

Language Cognition (HLC). As BLC is exclusively oral, it means that measures of bilingual abilities should also necessarily be oral rather than written. While detailed operationalisations of BLC and HLC are still under development, it is intuitively appealing that priority is given to oral/aural skills over written ones, as many bilinguals are literate in only one of their languages, and some do not possess any literacy skills (see Simpson, this volume). An interesting question is also whether the distinction between BLC and HLC can be mapped on to Cummins' (1979) lower and higher thresholds of language competence. Knowledge of BLC in two languages could, for example, correspond to Cummins' lower threshold and knowledge of (parts of) HLC to the higher threshold. Future research into these issues will need to show if operationalising bilingual abilities in this way can shed new light on the impact of bilingualism on non-linguistic cognition.

An issue that will be important for the future is the measurement of domain-specific language use. While recently new questionnaires have been developed which aim to help researcher obtain further detail about bilinguals' use of languages (Gertken, Amengual & Birdsong, 2014; see also Armon-Lotem & Meir, this volume), Grosjean (2016) argues that much more detailed information is needed about the different domains in which bilinguals use their languages. In order to make this possible, Grosjean (2016) proposes the Complementarity Index, which is an important new tool that can help researchers quantify the distribution of labour between languages across different domains of use.

Although researchers interested in the concept of bilingual abilities or in the measurement of this construct have not often looked at perceptions of bilingual abilities by listeners, a focus on such perceptions could be highly interesting as it could provide further information about the dimensions of bilingual abilities that matter to listeners (whether expert or non-expert) and on the relative importance of different features of a bilinguals' speech that play a role in how that person is perceived.

There are many avenues to be explored and it is hoped that some of the challenges mentioned in this chapter will be taken up by researchers interested in further improving the measurement of bilingual abilities.

References

Alderson, J. C. (1980). Native and nonnative speaker performance on cloze tests. *Language Learning, 30*(1), 59-76.

Andringa, S. (2014). The use of native speaker norms in critical period hypothesis research. *Studies in Second Language Acquisition, 36*(3), 565.

Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.

Bahrick, H. P., Hall, L. K., Goggin, J. P., Bahrick, L. E., & Berger, S. A. (1994). Fifty years of language maintenance and language dominance in bilingual Hispanic immigrants. *Journal of Experimental Psychology: General, 123*(3), 264.

Bahtina, D., & ten Thije, J. D. (2012). Receptive multilingualism. In C. A. Chappelle (Ed.), *The encyclopedia of applied linguistics* (pp. XX-XX). Hoboken, N. J.: John Wiley. DOI: 10.1002/9781405198431.wbeal1001.

Benmamoun, E., Montrul, S., & Polinsky, M. (2013). Defining an "ideal" heritage speaker: theoretical and methodological challenges. Reply to peer commentaries. *Theoretical Linguistics, 39*, 259–294.

Bialystok, E., Craik, F. I., & Luk, G. (2008). Lexical access in bilinguals: Effects of vocabulary size and executive control. *Journal of Neurolinguistics, 21*(6), 522-538.

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

Birdsong, D. (2016). Dominance in bilingualism: Foundations of measurement, with insights from the study of handedness. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: issues of measurement and operationalization* (pp. 85-105). Cambridge: Cambridge University Press.

Bowden, H. W. (2016). Assessing second-language oral proficiency for research: The Spanish Elicited Imitation Task. *Studies in Second Language Acquisition, 38*(4), 647-675.

Chiat, S., Armon-Lotem, S., Marinis, T., Polišenká, K., Roy, P., & Seeff-Gabriel, B. (2013). Assessment of language abilities in sequential bilingual children: The potential of sentence imitation tasks. In V. C. Mueller Gathercole (Ed.), *Issues in the assessment of bilinguals* (pp. 56-89). Bristol: Multilingual Matters.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Cook, V. (1991). The poverty-of-the-stimulus argument and multi-competence. *Second Language Research, 7*(2), 103–117.

Cook, V. (2016). Premises of multicompetence. In V. Cook & W. Li (Eds.), *The Cambridge handbook of linguistic multi-competence* (pp. 1-25). Cambridge: Cambridge University Press.

Cruz-Ferreira, M. (Ed.) (2010). *Multilingual norms*. Frankfurt: Peter Lang.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222-251.

Cummins, J. (2016). Reflections on Cummins (1980), “The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue.”. *TESOL Quarterly, 50*(4), 940-944.

Curdt-Christiansen, X.-L. (2016). Conflicting language ideologies and contradictory language practices in Singaporean multilingual families. *Journal of Multilingual and Multicultural Development, 37*(7), 694–709.

De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics, 28*(3), 411-424.

De Houwer, A. (2009). *Bilingual first language acquisition*. Bristol: Multilingual Matters.

De Houwer, A. (2013). Early bilingualism. In C. A. Chappelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1822-1830). Hoboken, N. J.: John Wiley.

De Houwer, A., & Bornstein, M. H. (2016). Balance patterns in early bilingual acquisition: A longitudinal study of word comprehension and production. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: issues of measurement and operationalization* (pp. 134-155). Cambridge: Cambridge University Press.

De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual–monolingual comparison of young children’s vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics, 35*(6), 1189–1211.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test. Fourth edition (PPVT™-4)*. San Antonio: Pearson.

Emmorey, K., Giezen, M. R., & Gollan, T. H. (2016). Psycholinguistic, cognitive, and neural implications of bimodal bilingualism. *Bilingualism: Language and Cognition, 19*(2), 223-242.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J., Pethick, S., & Reilly, J. (1993). *The MacArthur communicative development inventories: User's guide and technical manual*. San Diego: Singular Publishing Group.

Gathercole, V. C. M. (2013). Assessment of bilinguals: Innovative solutions. In V. C. M. Gathercole (Ed.), *Solutions for the assessment of bilinguals* (pp. 1-9). Bristol: Multilingual Matters.

Gathercole, V. C. M., Thomas, E. M., Roberts, E., Hughes, C., & Hughes, E. K. (2013). Why assessment needs to take exposure into account: Vocabulary and grammatical

abilities in bilingual children. In V. C. Mueller Gathercole (Ed.), *Issues in the assessment of bilinguals* (pp. 20-55). Bristol: Multilingual Matters.

Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the bilingual language profile. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 208-225). Bristol: Multilingual Matters.

Grosjean, F. (1985). The bilingual as a competent but specific speaker-hearer. *Journal of Multilingual & Multicultural Development, 6*(6), 467-477.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition, 1*(02), 131-149.

Grosjean, F. (2016). The Complementarity Principle and its impact on processing, acquisition, and dominance. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 36-65). Cambridge: Cambridge University Press.

Gullberg, M., Indefrey, P., & Muysken, P. (2009). Research techniques for the study of code-switching. In B. Bullock & A. J. Toribio (Eds.*), The Cambridge handbook of linguistic code-switching* (pp. 21-39). Cambridge: Cambridge University Press.

Haitana, T., Pitama, S., & Rucklidge, J. J. (2010). Cultural biases in the Peabody Picture Vocabulary Test-III: Testing Tamariki in a New Zealand sample New Zealand. *Journal of Psychology, 39*(3), 25-34.

Haman, E., Łuniewska, M., & Pomiechowska, B. (2015). Designing Cross-Linguistic Lexical Tasks (CLTs) for bilingual preschool children. In S. Armon-Lotem, J. de Jong, & M. Meir (Eds.), *Methods for assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 194-238). Bristol: Multilingual Matters.

Hamilton, J., Lopes, M., McNamara, T., & Sheridan, E. (1993). Rating scales and native speaker performance on a communicatively oriented EAP test. *Language Testing, 10*(3), 337-353.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.

Hulstijn, J. (2012). The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition, 15*(02), 422-433.

Hulstijn, J. (2015). *Language proficiency in native and non-native speakers: theory and research*. Amsterdam: John Benjamins.

Kirby, J. R., Deacon, S. H., Bowers, P. N., Izenberg, L., Wade-Woolley, L., & Parrila, R. (2012). Children's morphological awareness and reading ability. *Reading and Writing, 25*(2), 389-410.

Kremmel, B. & Schmitt, N. (2016) Interpreting vocabulary test scores: What do various item formats tell us about learners' ability to employ words?, *Language Assessment Quarterly, 13*(4), 377-392.

Kupisch, T., & Van de Weijer, J. (2016). The role of the childhood environment for language dominance: A study of adult simultaneous bilingual speakers of German and French. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: Issues of operationalization and measurement* (pp. 174-194). Cambridge: Cambridge University Press.

Larsen-Freeman, D., Schmid, M., & Lowie, W. (2011). Introduction: From structure to chaos: Twenty years of modeling bilingualism. In M. Schmid & W. Lowie (Eds.), Modeling bilingualism: From structure to chaos. In honor of Kees de Bot (pp. 1-12). Amsterdam: John Benjamins.

Marinis, T., & Armon-Lotem, S. (2015). Sentence repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 95-122). Bristol: Multilingual Matters.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics, 32*(3), 299-320.

Muysken, P. (2013). Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition, 16*(04), 709-730.

Nation, I. S. P. (2001). *Learning vocabulary in another langua*ge. Cambridge: Cambridge University Press.

Ortega, L. (2013). SLA for the 21st century: Disciplinary progress, transdisciplinary relevance, and the bi/multilingual turn. *Language Learning, 63*(1), 1-24.

Pearson, B. Z. (2010). We can no longer afford a monolingual norm. *Applied Psycholinguistics, 31*(2), 339-343.

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43*(1), 93-120.

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español. Toward a typology of code-switching. *Linguistics, 18*(7), 581-618.

Romaine, S. (1989). *Bilingualism*. Oxford: Basil Blackwell.

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing, 15*(1), 1-20.

Rothman, J., & Treffers-Daller, J. (2014). A prolegomenon to the construct of the native speaker: heritage speaker bilinguals are natives too! *Applied Linguistics, 35*(1), 93-98.

Royal College for Speech & Language Therapists (2006). *Communicating quality 3. RCSLT's guidance on best practice in service organisation and provision*. London: Royal College of Speech & Language Therapists.

Schmeißer, A., Hager, M., Arnaus Gil, L., Jansen, V., Geveler, J., Eichler, N., Patuto, M., & Müller, N. (2016). Related but different: The two concepts of language dominance and language proficiency. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilinguals: issues of operationalization and measurement*. Cambridge: Cambridge University Press.

Silva-Corvalán, C., & Treffers-Daller, J. (Eds.). (2016). *Language dominance in bilinguals: issues of operationalization and measurement*. Cambridge: Cambridge University Press.

Skutnabb-Kangas, T. (1981). *Bilingualism or not: The education of minorities.* Clevedon: Multilingual Matters.

Street, J., & Dąbrowska, E. (2010). More individual differences in Language Attainment: How much do adult native speakers of English know about passives and quantifiers? *Lingua, 120*, 2080-2094.

Treffers-Daller, J. (2009). *Code-switching and transfer: an exploration of similarities and difference*s. Cambridge: Cambridge University Press.

Treffers-Daller, J. (2016). Language dominance: the construct, its measurement and operationalization. In C. Silva-Corvalán & J. Treffers-Daller (Eds.), *Language dominance in bilingualism: Issues of measurement and operationalization* (pp. 235-265). Cambridge: Cambridge University Press.

Winsler, A., Kim, Y., & Richard, E. (2014) Socio-emotional skills, behavior problems, and Spanish competence predict the acquisition of English among English language learners in poverty. *Developmental Psychology, 50*(9), 2242-2254.

Wu, S. L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals, 46*(4), 680-704.

Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science, 9*(2), 111-125.