Unbiased Language Assessment: Contributions of Linguistic Theory

Jill G. de Villiers

Department of Psychology, Smith College, Northampton, Massachusetts 01063;
email: jdevilli@smith.edu

Keywords

assessment, dialects, bilingualism, minority languages, multiculturalism

Abstract

This review addresses several situations of language learning to make concrete the issue of fairness—and justice—that arises in designing assessments. First, I discuss the implications of dialect variation in American English, asking how assessment has taken dialect into consideration. Second, I address the question of how to assess the distributed knowledge of bilingual or dual-language learners. The evaluation of the language skills of children growing up in poverty asks whether the current focus on the quantity of caregiver input is misplaced. Third, I address a special case in which the young speakers of a minority language, Romani, are judged to be unfit for schooling because they fail tests in the state language. Finally, I examine the difficult issue of language assessments in countries with multiple official languages and few resources. In each of these areas, the involvement and expertise of linguists are essential for knowing how the grammar works and what might be important to assess.

1. WHAT IS A CHILDHOOD LANGUAGE IMPAIRMENT?

Children's language acquisition is a robust process, heralded by contemporary linguists as proof of our biological preparedness, whether or not they follow Chomsky in arguing for an innate language acquisition device for acquiring Universal Grammar (Hoff 2013a). The age milestones for babbling, first words, early sentences, and even fairly complete grammars by age 6 are found universally and have led to the claims that most of language acquisition is biologically guided maturation, within broad environmental limits.

Yet there is a profession comprising more than 130,000 trained speech/language practitioners (SLPs) in the United States alone, many of whose caseloads include children suffering from a speech or language impairment sufficient to cause severe interruptions of normal interaction and schooling. Some of these difficulties are secondary to other disorders, such as genetic syndromes known to cause cognitive impairments or profound deafness in the absence of a signing model. Others arise from autism spectrum disorders (ASDs); a high percentage of such children have social-communicative problems and impaired language development. Still others are said to have specific language impairment, or SLI; that is, they do not learn language on a typical timetable despite having normal development in other ways.

The ontological status of this last category of SLI is hotly debated. After all, there are some individuals who have prodigious language ability even beyond the feat of typical children. Surely there is a range of skill and speed at language learning, and some children inevitably fall into the bottom 10% or top 10%? If children with SLI "caught up" over time, this would be less worrisome, but many do not, and lead circumscribed lives both educationally and socially (Leonard 2014). Early intervention can ameliorate some of these difficulties, but it must be done early (Fricke et al. 2013).

Some researchers (e.g., Rice 2013) argue that there is an underlying genetic disorder causing SLI, given that it tends to run in families and seems to have similar incidence worldwide. Furthermore, twin studies have shown that SLI co-occurs more often in identical twins who share a common genome than in fraternal twins who do not. Others (e.g., Reilly et al. 2014) remain unconvinced by the data that designate these children as a special group. Those who argue for the special nature of SLI have often been linguists who see a hope for precision in identifying what exactly has gone wrong, because that might be traceable to a set of genes. For example, Rice & Wexler (1996) argue that children with SLI show delayed maturation of obligatory tense, so that children continue longer in a stage that allows optional infinitives in place of a tensed matrix verb in their sentences. This delay then affects the rest of their development. In another example, van der Lely & Battell (2003) proposed a specific failure in the Movement domain, leading to failures with *wh*-questions, embedded complements, and relative clauses. But it remains to be determined whether these are distinctive properties of children with SLI or inevitable stages whenever a child is delayed in grammar.

2. ASSESSMENT

The background of language impairment helps to answer the question: What is the motivation for testing a child's language development? An increasing number of parents and teachers in the United States are alarmed by standardized testing, particularly for very young children. Why test at all? In the United States, English language skills are tested to ensure that a child has the oral language and comprehension skills necessary for schooling. Schooling adds new demands: A family may be able to converse with a child, but the child may appear unintelligible to outsiders. School represents a much more rigorous testing ground for the adequacy of children's language mastery.

However, identifying the source of a language problem is complicated by other factors. Assessments of language development are designed to test for a language delay. First, the cause of this delay could be SLI, hypothesized to be partly or mostly biological or genetic in origin if it is not secondary to deafness—if the child is speaking, not signing—or broader cognitive deficits (Tager-Flusberg & Cooper 1999, Rice 2013, Leonard 2014). Second, there may be other children who are growing up in neglect so profound that they have not had the chance to acquire the vocabulary and syntax typical of a 5-year-old entering school (Westby 2007). Third, there may be children whose home environment is sufficiently different from that of "mainstream" children such that they have an entirely different set of skills, vocabulary, and language usage (Heath 1982). Fourth, there may be children whose home dialect is different from the language of schooling, which may or may not entail cultural difference, too (Heath 1983, Wyatt 2002). Finally, there are children whose home language is not English, who may or may not have had enough exposure to English by age 5 to succeed in an English-speaking environment (Peña & Halle 2011). How can a language test sort out these possibilities?

An important consideration is the best age for children to be tested. One argument is that one should wait until it becomes necessary to know, namely kindergarten age. That timing would allow some slower learners to catch up, those in less-ideal circumstances to have greater experiences, and those with a different language or dialect background to have a chance for broader exposure outside of the home, perhaps in preschool. The remaining children who test poorly would then be more easily classified as having genuine language delay. However, that is not an optimum solution for children with language disorders, although in practice it is typical that only a minority are identified prior to age 5 (Tomblin et al. 1997). Although waiting to test until age 5 seems fair for typically developing children, it might be catastrophic for children who really have a language disorder not to have early intervention, while the brain is more malleable (Paul & Norbury 2012). If an assessment done earlier were accidentally to catch some children who simply need more intense input, that temporary mislabeling might be considered a relatively small cost relative to the benefits to the children receiving therapeutic assistance.

But is it fair to test and then label as delayed young children who have a different home dialect or language? The costs of labeling or segregating such children as needing speech/language therapy goes against the recommendation of the American Speech–Language–Hearing Association:

> A speaker of any language or dialect may exhibit a language disorder unrelated to his or her use of the native dialect. An essential step toward making accurate assessments of communication disorders is to distinguish between those aspects of linguistic variation that represent regular patterns in the speaker's dialect and those that represent true disorders in speech and language. (Am. Speech Lang. Hear. Assoc. 1983, p. 25)

3. THE QUESTION OF DIALECT

As recently as the 1960s, the dialect known as African American English (AAE) was said to be insufficient to support higher-level, logical thought (Bereiter & Engelmann 1966) and that, therefore, children who spoke it had a deficit in preparation for education. Since that time, linguists and educators have argued against that view and have carefully documented the properties of AAE (Labov 1972, Green 2002).

Despite some progress, many tests available for assessment of English do not account for the differences between English and AAE, and AAE-speaking children with speech/language problems are diagnosed as having language problems proportionately more often than mainstream English speakers (Seymour et al. 1998). SLPs are encouraged to modify the scoring schemes of standard

tests to enable dialect-type responses to be accepted. Accepting responses in either language may help the examinee perform better, but it invalidates the test's standardization and renders the norms inapplicable. At best, the results could be used descriptively (Wyatt 2002, Rhodes et al. 2005).

In the 1990s, the National Institutes of Health (NIH) recognized this conundrum and put out a call for grants to address the issue for children who spoke AAE. Seymour et al. (2005) began work on the test that would become the Diagnostic Evaluation of Language Variation (DELV). The DELV test was conceived under difficult political circumstances. While it was under review at NIH in 1996, the school board of Oakland, California resolved that the superintendent of schools should devise a program to improve the English language acquisition and application skills of African American students, of whom 70% were in special education. However, the resolution argued for recognition of the special properties of the language the board called "Ebonics," rather than AAE. The goal was to recognize the special nature of the language spoken by the majority of their students, and to emphasize its pan-African roots rather than consider it an English dialect or a form of slang. Congressman Peter T. King (Republican, New York) immediately introduced a House resolution denying federal funds to "any program that is based upon the premise that AAE is a legitimate language" (O'Neill 1998a).

The resulting furor is well documented in the book edited by Perry & Delpit (1998), who collected opinions from educators, linguists, community members, and teachers. The astonishing lack of attention to linguistics was widespread in the United States, with even the most liberal commentators and columnists failing to understand or even research the phenomena on which they had such definite negative opinions (O'Neill 1998a,b). AAE was characterized in the media as rule-less, slang, gibberish, or classroom rap, and the Oakland resolution was described as a plague, a cop-out, and even as "something that the Klan might propose,"[1]—perhaps as a form of apartheid, creating separate educational paths. Despite this highly charged atmosphere, the grant to develop the test that became the DELV was funded (Seymour et al. 2005).

3.1. The Diagnostic Evaluation of Language Variation Test

AAE has been characterized as a set of distinguishing features (Washington & Craig 2004), which poses a problem in that these features are viewed as a set of unconnected properties rather than a coherent grammar. Counting and doing a frequency analysis of distinct phonological and morphological features are useful ways to characterize whether speakers use patterns of AAE in their speech (Oetting & McDonald 2002), but Green (2011) emphasizes that it is better to characterize AAE as a full, coherent linguistic system, of which these features are superficial symptoms. AAE has a distinctive grammar that emphasizes aspect over tense, as just one example.

For this and other reasons, the methodological difficulties faced by the DELV team were as demanding as the political zeitgeist. The researchers wanted to create a test to evaluate the properties of AAE directly—an ideal goal. For example, if the tense/aspect system of AAE is quite radically different from standard or mainstream American English (MAE) (e.g., Green 2002), then perhaps items could be devised that would identify children who are at an age-appropriate point on a path to mastery of AAE. MAE-speaking children presumably would not show the same competence, because they have different grammars. Failure of AAE-speaking children on an MAE test could reflect a difference in the grammars, not a deficit in language skills. The same could be true for children who speak MAE but are tested on their competence in AAE grammar.

---

[1]A spokesman on Fox News in 1996, quoted by O'Neill (1998b).

The DELV team ultimately set aside that goal, with reluctance. There were three major obstacles to achieving it. First, AAE was frequently characterized as displaying "optional" morphology, for example, of past tense, of possessive, or of *to*-infinitives (Seymour et al. 1998, Washington & Craig 2004). This meant that it would be necessary to differentiate typically developing AAE speakers who show a certain degree of optionality even in the fully developed grammar from typically developing MAE speakers who show optionality only in earlier developmental stages. Omission frequency—a subtractive property—was never going to be viable as a test item.

Second, the distinctive additive properties of AAE, such as the rich aspectual system (Green 2002), proved elusive to careful testing. In fact, examples of this intransigence have accumulated as researchers in other languages have tried to develop careful tests of aspect morphology, such as telicity or habituality (Wagner 2002, Hacohen 2006, Van Hout 2008). Most language assessment tests for children involve showing the child pictures of objects or events, and either asking them questions about the events or eliciting descriptions of them. But the temporal quality of events is not easily depicted, especially to children who may not know certain conventions about how picture sequences show time. The morphosyntactic distinctions are hard to make precise, as they are frequently conditioned by multiple semantic factors. Problematically, on pilot items speakers of MAE often succeeded as well as speakers of AAE, possibly reflecting exposure to multiple dialects or overlap in meanings with similar-sounding forms, such as *bin* in AAE and *been* in MAE. Linguistic progress was made, but not in time for the development of items for the DELV (Green & Roeper 2007).

Third, a sociological problem emerged. The team successfully piloted a task using habitual *be* in AAE, in which the child had to pick a picture in which *a boy be riding a bike*—that is, the pictures showed one boy who rode a bike every day, and another boy who was riding it only once. When the initial data came back from clinicians around the country who had been recruited to test children in their caseload, the results were very puzzling. Investigation revealed that several clinicians—almost all white, MAE-speaking SLPs—had assumed *be* was a mistake and had said instead *a boy is riding a bike*. As the team of developers contemplated asking MAE speakers to pronounce AAE forms such as negative concord, multiple aspectual forms, or even AAE phonological forms, it became clear that this was not a viable choice for a test in a country without enough AAE-speaking SLPs to test children authentically.

The obvious path to take was to use so-called noncontrastive features, properties of the language that do not differ between the dialects in question. For example, Stockman (1996) took language samples from AAE from which to choose items that would be equally valid for both AAE and MAE speakers. Craig & Washington (2000) and Washington & Craig (2004) developed a battery of noncontrastive items for screening and testing. The DELV researchers then identified environments in which the grammatical features behaved similarly. For example, although copulas are frequently omitted in AAE, they are never omitted in past-tense environments (as in Arabic). That is, it is acceptable to use "zero copula" in AAE in a third-person present-tense environment, as in *She right*, but not in a past-tense context, as in *Yesterday she right*.

Proceeding in this way, the researchers used extensive preliminary research on AAE to identify linguistic items that were neutral between the dialects in well-defined environments, and employed them for testing (Benedicto et al. 1998, Seymour & Pearson 2004). The solution to fair testing on the DELV was to identify circumstances in which the playing field was level for speakers of either dialect—and perhaps a few more, such as Southern White English or Appalachian English (Oetting & McDonald 2002).

The group found grammatical circumstances that conditioned the optionality, so that the morphemes were obligatory in each dialect. Additionally, in several areas of syntax the two dialects

This boy was leaving school one day when he saw a sign about a special art class. He really wanted to go, but he had to sign his name on the sheet, and he didn't have a pen. When he was walking home he saw a pen lying on the sidewalk. He ran back to school anvd signed up for the art class.

Where did the boy pick up the pen to sign his name?
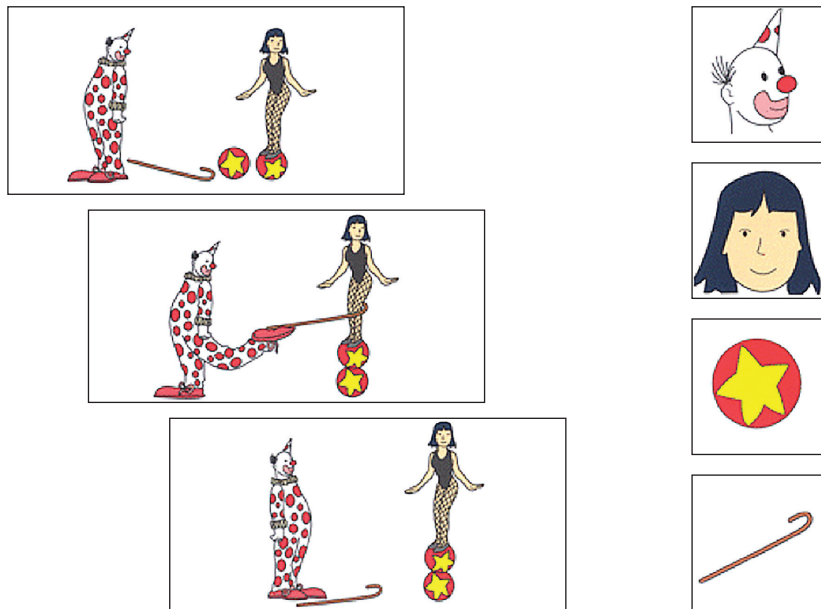
Figure 1

An example *wh*-question from the pilot version of the *Diagnostic Evaluation of Language Variation* (*DELV*) test. Copyright 2003, NCS Pearson, Inc. Modified with permission. All rights reserved. *Diagnostic Evaluation of Language Variation* and *DELV* are trademarks, in the United States and/or other countries, of Pearson Education, Inc. or its affiliates(s).

behave equivalently, as in certain passives and complex *wh*-questions (de Villiers et al. 2008). For example, both dialects forbid *wh*-question extraction from an *in order to* clause (see **Figure 1** for an example). In semantics, the group chose to avoid areas in which cultural differences might result in different vocabulary, instead focusing on phenomena where there was commonality, such as syntactic bootstrapping of novel verbs from argument frames (see **Figure 2** for an example) (Johnson & de Villiers 2009). The final test proved equivalent for AAE and MAE speakers matched in social class, yet it still discriminated in terms of identifying children with language impairments from either dialect (Seymour et al. 2005).

The practice by which SLPs had been advised to give credit for absent morphology or alternative dialect responses had been a source of frustration and doubt, and probably resulted in children being underdiagnosed as having language difficulties—that is, it overcorrected for the initial problem (Campbell 1996). The researchers who developed the DELV were convinced that they had to set a high-enough bar for language skills to ensure that children with real disorders were identified. All of the subtests make great demands on children's linguistic skills, more so than many other tests that may underestimate children's ability.

It may be important that linguistic assessments challenge children's skills for cultural reasons, too. Wyatt (2002) emphasized that in African American culture, parents and caregivers rarely give their children "test" questions like middle-class white parents do, such as "What's on the wall?" or "What color is this?". Instead, questions to the child are often authentic ones, asked in circumstances where the parent does not know the answer, such as "Where did you put your shoes?". The risk of asking children easy questions to which the questioner knows the answer is that the child could draw the implicature that there must be more to it than the obvious answer.

The DELV was designed to resolve inequities in testing children with dialects that differ from that of the mainstream. Even the items were drawn to depict primarily black children. Its reception, however, has been less positive than anticipated. First, the test became identified as "the test for black children." Often, school systems cannot afford to administer more than one test, especially if only a small number of their students are black or AAE speakers. Second, SLPs may be criticized if they give one test to white children and the DELV to black children. Parents have occasionally taken issue with this practice, asking that their AAE-speaking child be treated like everyone else

The clown is syping the woman.

Here are the things in the pictures. I want you to show me...
Which one was the syper?
Which one got syped?
Which one was the woman standing on?

To help you remember, I'm going to say it again,
"The clown is syping the woman."
Which one was sypable?
Which one was syping?

Figure 2

An example item from the Fast Mapping subtest of the *Diagnostic Evaluation of Language Variation* (*DELV*) test. Copyright 2003, NCS Pearson, Inc. Modified with permission. All rights reserved. *Diagnostic Evaluation of Language Variation* and *DELV* are trademarks, in the United States and/or other countries, of Pearson Education, Inc. or its affiliates(s).

and given the "proper" test. Now, 10 years later, DELV is only slowly gaining acceptance as a fair assessment for children with nonmainstream dialects as well as MAE speakers (Oetting et al. 2016). New studies of AAE-speaking children are endeavoring to probe more deeply into how a disorder manifests within AAE grammar—that is, how language delay affects the grammar itself, as opposed to the noncontrastive properties on which the DELV focuses (Green & Roeper 2007, Stockman 2010, Oetting et al. 2013, Pearson et al. 2013).

Third, the test has sometimes been anticipated as a panacea for a wider social and educational problem. Some SLPs who have used the test with their populations of low-income AAE-speaking children have still found a distressing failure rate among those tested. Did the test set too high a bar? An alternative assumption is that, whereas an assessment can attempt to level the playing field for testing, no test can erase the circumstances that create inequities in prenatal care, access to health care, nutrition, quality of schools, parental education, and the many other documented disparities in the lives of black and white children in the United States, even matched for socioeconomic status (SES) (e.g., Fryer & Levitt 2004, Flores & Tomany-Korman 2008).

3.2. How Can a Test Be Validated?

The DELV faced an obstacle in acceptance for methodological reasons. Recent papers about language assessments (Plante & Vance 1994, Spaulding et al. 2006) have made a strong case for empirical validity, using as the best criteria the specificity and sensitivity of a test. Specificity refers to the ability of a test to distinguish children with the trait in question (in this case, language delay, but it could be, e.g., diabetes in a medical domain) from those who lack the trait. An ideal test should capture only those who have the trait. Sensitivity refers to the ability of the test to capture all the children who have the trait, and not miss any who could benefit from treatment. An ideal test should capture all of those who have the trait. In practice, there is a trade-off between being sensitive enough to catch everyone with a language delay and being specific enough not to catch some of those who do not.

The careful reader might have noticed an assumption in the definitions of these terms. How do we know who actually has a language delay, independent of some test? In the field, an existing test that has been broadly accepted and applied is used as the "gold standard" by which to sort children into the two groups, and the test in question has its sensitivity and specificity assessed against that a priori grouping. But tests like the DELV were created out of suspicion that previous tests were unfair to dialect speakers, so how can a gold standard be established (de Villiers & de Villiers 2010)? In fact, the final norm-referenced DELV achieved acceptable levels of sensitivity and specificity. Pearson et al. (2014) offer a detailed discussion of this topic. These authors compared complex indices of spoken speech analysis as a validity index to complement clinicians' judgment, and achieved a specificity level of 0.91 and a sensitivity level of 0.97.

The problem continues to haunt attempts to break new ground with innovative tests in the United States. And these criteria create a major problem for the validation of tests in countries where no tests yet exist (Liu et al. 2016). The paradox is this: If there is an accurate way to determine who does and does not have a language disorder, then why make a test? If not, how is a test ever going to gain acceptance?

Fortunately, tests need not be judged only on specificity and sensitivity. Test makers employ other criteria for the reliability and validity of assessments that can be used in the absence of a comparison standard. These indices include internal reliability, namely whether the items intercorrelate, suggesting they are all tapping into a central construct. An important type of reliability is assessed by a procedure called Rasch analysis, which is used to determine whether items rank the same way across children, and whether children can be ranked the same way across items. Predictive validity asks whether performance on a test can predict how the child will do at reading in elementary school, or in social conversation as judged by a teacher.

Most interesting to linguists and language acquisition experts is the concept of construct validity: Are there good reasons for putting these items on a test, given what is known about child language development? For example, it might never occur to a lay person that a question such as "Who ate what?" could be revealing: It is a short, rare type of question, but one that requires a precise sort of answer, one involving paired sets. The person questioned must respond by saying, for instance, "John ate the chips, Jane ate the cookies, and Carlos ate the burgers." It simply wouldn't do as an answer to list just the people or just the foods, or to give one example pair. It has the property of exhaustivity, which is highly discriminating of language impairments, even across languages (Schulz & Roeper 2011, de Villiers et al. 2016). And what is an important milestone? Rice & Wexler (1996) argued that a crucial milestone is one involving the obligatoriness of tense. Young children in many languages believe tense is optional; children with SLI are said to be systematically delayed in that achievement, with a cascade of effects on their grammars. And which items might be different across dialects or ethnolects? The AAE project and many like it

have revealed the importance of considering grammatical variation, for example, in using *be* copula forms, variation that untrained people may take to be immaturity or laziness (Oetting et al. 2016). These are the kinds of contributions that linguists must make to create better tests, whether in English or other languages.

4. FAIR ASSESSMENT AND BILINGUALISM

In 1968, an important piece of educational legislation was passed in the United States: the Bilingual Education Act, spearheaded largely by Spanish speakers. Its goal was to encourage school districts to incorporate native language instruction in order to ease the transition to English for dual-language learners. Doing so is compatible with research suggesting that it takes between 5 and 8 years for a child to master two languages fully (Hoff 2013a). Furthermore, an abrupt transition to English can lead to loss of the native language, a failure to achieve literacy in the first language, and unrealistic expectations that a child can achieve native-like fluency equivalent to that of their monolingual peers in a short period of time (Cha & Goldenberg 2015, Zepeda & Rodriguez 2014).

In 2002, the Bilingual Education Act was overturned in favor of the federal English Language Acquisition Act, part of the No Child Left Behind legislation. The focus shifted entirely to the acquisition of English, as fast as possible, with neither a provision for retention of a first language nor support of dual-language-learning. The outcome was that many Spanish-speaking children lost the opportunity to retain or develop literacy in Spanish, ironically at the same time that non-Latinos were being encouraged to learn Spanish as a second language to improve their marketability in the workplace (Ortman & Shin 2011).

4.1. The Bilingual English Spanish Assessment

In 1997, against this background, researchers began to develop the Bilingual English Spanish Assessment (BESA), also with NIH funding (Peña et al. 2013). The prevailing opinion, still common today among SLPs, is that a child who has language delays or specific language impairment could only be disadvantaged by hearing two languages. Generations of parents have been told to stop talking to their children in their first and often most fluent language, and instead to speak only English. Recent empirical research has contradicted that opinion (Peña & Bedore 2009), yet it persists.

Bilingual children are also at risk of being characterized as deficient in language skills if they do not display the same skills that monolingual children do in the language of the state when they start schooling. In consequence, many dual-language learners come to the attention of SLPs for assessment as being language delayed.

In comparison with the appropriate assessment of children who speak a dialect other than MAE, the methodological and theoretical difficulties of creating a test for bilingual children are enormous. The problem is that bilinguals lie on a multidimensional scale. In children who have no language problems, the profile of low English skills and high Spanish skills may lie on one end of the scale, and high English and low Spanish skills on the other end. But the children most in need of detection have low English and low Spanish skills. That would seem difficult enough to assess, were it not for the age factor. What counts as "high English skills" if a child is bilingual at a certain age? And is there a difference between sequential bilingualism, where the child learns one language after the other, and simultaneous bilingualism, where the child learns two languages at once (Zepeda & Rodriguez 2014, Iglesias 2015)?

Note that not all dual-language learners' experiences are alike. Children's English language knowledge can be very different depending on the practices of their families and communities. Older siblings might bring English into the home, or one parent may be fluent in English and

the other not. The communities in which the children live vary hugely in the amount of English used daily. Peña & Halle (2011) argue that children who hear a second language less than 20% of the time are "functionally monolingual." Cumulative exposure is what counts, so functionality cannot be judged only by the number of years of exposure, but rather by the percentage of time in each language multiplied by the total number of years (Unsworth 2013).

It is evident that dual-language learners must be assessed in both of their languages to determine whether they have a genuine language difficulty. In addition, such testing would be fair only if the norms against which these children were judged were norms taken from bilingual children, not monolingual children. As Peña & Halle (2011, p. 30) state,

> Test validity for DLL [dual-language-learning] children is threatened when available norms are based on monolingual children (either children who speak English as a native language or those who speak the home language as the native and sole language), when the child's cultural experiences do not match test expectations, or when the items are not presented in a way that allows the child to demonstrate competence.

Dual-language-learning children use their languages in different contexts and with different interlocutors, and as a result their knowledge in each language is distinct. Research has shown that in the first few years the vocabulary the child knows is often unique to each language rather than common to both languages (Pearson et al. 1995; Peña et al. 2002a,b). Therefore, dual-language learners' knowledge is distributed knowledge. Although dual-language learners often know fewer words in each of their two languages relative to monolinguals (Bialystok et al. 2010), their overall vocabulary, including all known words, is on a par with that of monolinguals (Pearson et al. 1993). Researchers have adopted different methods of counting bilinguals' vocabulary (de Villiers 2015) that give children credit for knowing a concept in one language or the other, or combine the two vocabularies (Core et al. 2013). Distributed knowledge in bilinguals means relatively slower growth in each language but comparable overall knowledge.

In the design of a fair linguistic test for bilingual children, Peña & Quinn (1997) emphasize cultural differences in language use. As mentioned above, some cultures may not address test questions to children, questions for which the adult already knows the answer. In some cases, children may know the functions for many different objects—that is, how they are used, or their description—but may not know the particular labels.

In sum, even with regard to their English skills, dual-language learners should not be assessed on an instrument that was normed only on monolingual speakers. Moreover, translation or interpretation of an English test for testing Spanish skills is a very poor practice. The use of a translator or interpreter for administering a test can help overcome a language barrier, but it completely undermines the score validity, even when the interpreter is highly skilled. This is not the way tests are normed; that is, the group being tested must be drawn from the same population that was used for the standardization of the test, or the comparison is invalid (Rhodes et al. 2005).

The design of a fair test for bilingual children must take all of these considerations into account. Peña et al. (2013) exemplified these best practices in their creation of the BESA. Like the DELV, the BESA is an omnibus test assessing semantics, syntax, morphology, and phonology in each language. The developed norms are for US Spanish–English bilinguals. The BESA team carefully considered the various dialects of Spanish that are represented in the US population and either chose noncontrastive items or allowed alternative responses to accommodate them. Rather than test the same skills in each language, the subtests were designed to reflect linguistic differences between Spanish and English and how each is acquired. A screening test called the BESOS consists of the best-discriminating items from the BESA, and it has high sensitivity and specificity against

the BESA (Peña et al. 2011). The problem of validating the BESA—the gold standard problem—has been treated analogously to the DELV, using spontaneous speech samples and short narratives, with linguistically trained personnel to score them (Bedore et al. 2010).

Can tests accommodate the distributed nature of knowledge also observed in grammar (Fiestas & Peña 2004, Oller et al. 2007)? It has been argued that a best score should be taken in subdomains of language (de Villiers 2015). For example, in a new screening test for preschoolers, the Quick Interactive Language Screener (QUILS), dual-language learners of Spanish and English take tests in each of their languages on parallel subtests, such as fast mapping of novel adjectives, extending passive to a novel verb, and understanding multiple prepositional phrases (Golinkoff et al. 2016). The children then receive standard scores based on their (dual-language-learning) peers in Spanish and English, as well as a best score representing the sum of scores from the best performance in either language on each subtest. A similar scoring procedure was incorporated into the BESA test for dual-language learners.

4.2. Bilingualism Elsewhere

Outside the United States, migration also creates inequities for children with a home language different from that of the state, and assessment can lead to misdiagnoses of language impairment. As a result, a project named COST Action was begun in Europe in 2010. The organizers argued that recent history had led to dramatic increases in the number of children being raised with two or more languages in multilingual communities. COST (2010) suggested that the number of children entering school knowing more than one language increased by more than 300% in Ireland, Italy, and Spain in the year 2000, and by 50% in 2008 and 2009 in the United Kingdom (see **http://ec.europa.eu/education**). One can imagine even more dramatic increases following the recent influx of refugees into Europe.

The COST researchers recognized that the serious lack of diagnostic instruments presented a difficulty. Their goal was to distinguish bilingual migrant children with language impairments from typically developing dual-language learners who are still in the process of learning the state language, but will eventually catch up with their monolingual peers. In the Netherlands, for example, bilingual children constitute 14% of the mainstream school population, yet in special schools for children with SLI they make up 24% of the population (COST 2010). The researchers assumed that if bilingual children are not being misdiagnosed, the percentage of bilingual children should be identical in mainstream and special schools.

The COST Action project spawned much additional research to discover how best to differentiate bilingual development from SLI (Armon-Lotem et al. 2015), and to investigate the manifestation of SLI—when it is found to be a correct diagnosis—in bilingual versus monolingual children (Friedmann & Novogrodsky 2011, Armon-Lotem 2012, Gagarina et al. 2016). The topic is too vast for coverage in this short review, but research is badly needed in order to reduce bias in assessment for bilingual children everywhere.

5. FAIR TESTING AND POVERTY

Bernstein (1971) referred to the language of poor children and families in England as the "restricted code," contrasting it with the "elaborated code" of middle-class families. The restricted code was characterized by formulaic discourse, fast and condensed, with low levels of familiar and local vocabulary and syntax, all in the service of communication among close intimates. In contrast, the elaborated code was characterized as more explicit and specific in its reference, and as portraying a wide variety of syntactic devices.

Groundbreaking research in the United States by Heath (1983) showed that children from the three distinct communities of Maintown (middle class and white), Roadville (working class and white), and Trackton (working class and black) had very different uses of language, especially with regard to literacy and narrative. This study demonstrated that the lower-class children entered school ill prepared for its demands, as their different style and skills did not match teachers' expectations. The children from the poorer black and white communities had not been exposed to the rules of how to use books to connect to their experiences, had not rehearsed the alphabet, and in many cases had not been exposed to the rituals of, say, bedtime reading (Heath 1982, Neuman 2000, Neuman & Celano 2001). Yet Heath emphasized the skills that these children did have, such as oral storytelling in the case of black families, a talent untapped in school. More recently, Hoff (2013b) argued that, even as we acknowledge different strengths and values of different cultures of child-rearing, the differential impact on readiness for US mainstream schools has to be recognized. In her review of research on SES differences in language, Hoff (2013b, p. 5) defined SES in terms of maternal education rather than wealth, which emphasizes the style of talk and interaction:

> It is important also to note that neither low- nor high-SES children are monolithic in their language skill, that substantial individual differences exist within both populations, and that the distributions of skills among lower and higher SES children overlap . . . . Nonetheless, the effect of SES on children's early language skills is large, pervasive, and robust.

5.1. Quantity and Quality

In a study on parental language input to a small sample of children ($N = 42$) in three social classes in the United States, Hart & Risley (1996) found massive differences in the amount of talk directed at children in different families. They extrapolated from their data that, by the age of 5, the poorest group of children in six families subsisting on social assistance (all of whom were African American) heard 30 million fewer words than the middle-class children. This 30-million-word gap between poor children and their middle-class peers is now widely accepted, and interventions to close it are under way. These findings have become a major focus of educators, policy makers, and commercial enterprises. The city of Providence, Rhode Island (**http://www.providencetalks.org/about/**) has made available to interested families a kind of "word pedometer" (Ford et al. 2008) to give them feedback about how many adult words a child is exposed to daily in order to persuade poorly performing households to speak more to their children, and thus close the gap.

However, these efforts have met with considerable resistance from community members who argue that word learning is not a fair measure of children's communicative experiences, from activists who argue that teaching more words is the least of a poor family's problems and is instead a symptom whose root causes need to be fixed, and from linguists who believe that it is the wrong index of language growth. The topic is vast and interconnected with social and political issues that are outside the scope of this review, but I discuss some important research by psycholinguists that has raised questions about the fundamental impact of the quantity of talk addressed to children as an index of a rich linguistic environment.

Reporting on a much larger sample of children and parents, Hirsh-Pasek et al. (2015) emphasize that a more important factor than the number of words a child hears is the quality of interaction, meaning the child's participation in "conversational duets" in which a caregiver can carefully orchestrate the ways that the child joins a conversation and learns. Cartmill et al. (2013) used 50 longitudinal samples of parent–child interaction to explore which particular circumstances are "high information" for conveying word reference. In this study, adults viewed videotapes with the soundtrack removed, and the researchers asked them at certain points what an overlaid beep or nonsense word might refer to. Even adults score far below 100% on these questions, but they

are most successful in circumstances where an adult in the video is highlighting an object by pointing or ensuring joint eye contact. Some families did this more effectively than others, and the families who did it most effectively had children who showed the fastest vocabulary growth on a standardized test later on. Most importantly, the researchers claim, these rare, high-information "perfect" conditions for learning word meaning are equally frequent across different SES groups in a sample as large as that of Hart & Risley (1996), and reasonably diverse. This research suggests that a careful analysis of quality of interaction, and perhaps an extended definition that incorporates microcultural variation in that definition, will be more useful than counting words.

5.2. Process-Type Measures

But what have these arguments got to do with assessment? Poor children who may not have many books in the home from which to learn about, say, types of animals or machines (favorites on vocabulary tests), or for whom a visit to a zoo, museum, or library is a rare event, might be just as adept at learning a new word from context if given the chance (Hirsh-Pasek et al. 2005). In assessment, process-type indices might level the playing field for children whose limited experiences have not led to increased noun accumulation. Several researchers have explored an alternative called dynamic assessment, in which a child learns words through a set of guided activities and this skill is then measured (Peña et al. 1992, Hasson & Joffe 2007). Dynamic assessment focuses on the quantity and quality of intervention necessary to produce a positive change and, one may argue, is closer to natural pedagogy. It is important to discriminate between children who have not yet learned something because of experiential differences and children who struggle to learn and therefore might need therapy. Johnson (2001) explored whether phenomena such as syntactic bootstrapping, or learning a new verb from its associated argument structure, could be a viable complement to more-standard vocabulary measures, especially for speakers of AAE. Indeed, this subtest has proved to be a valuable addition to the overall profile on the DELV (Johnson & de Villiers 2009).

Although the value of process measures as a diagnostic tool is not in dispute, the initial promise that process measures might ameliorate the effects of social class has not been fulfilled. Many studies have shown that a child's ability to learn a new word from linguistic context is well predicted from how many other words the child knows, especially in dealing with so-called dense lexical neighborhoods, in which fine distinctions in meaning and sound contrast have to be made (Storkel & Hoover 2011). In addition, facility with morphology and syntactic structure, the clues needed for syntactic bootstrapping, is contaminated by SES effects. There are two responses to such findings, and they suggest quite different solutions that are hotly debated. On one hand, some linguists argue that these differences are real, and that fair assessment cannot necessarily overcome the effects of unequal distribution of social and material resources. On the other hand, others argue that a fair test might never exist:

> Probably no test can be created that will entirely eliminate the influence of learning and cultural experiences. The test content and materials, the language in which the questions are phrased, the test directions, the categories for classifying the responses, the scoring criteria, and the validity criteria are all culture bound. (Sattler 1992, p. 579)

6. STIGMATIZED LANGUAGES

Roma children in Europe grow up in poverty, marginalization, and stigma. Kyuchukov et al. (2016) claim that these children face three distinct challenges:

1. Their families are often poor, and most of their parents have had very little formal education (Kyuchukov 2005).

2. From a very young age, they learn more than one dialect of Romani and/or other minority languages such as Turkish, but in most cases they do not learn the state language from birth (Kyuchukov 2010).

3. They are part of a stigmatized group that is often subject to discrimination in housing, employment, and education (ERRC 2014, Amnesty Int. et al. 2015). Because the status of a language is perfectly predicted by the status of the people who speak it, Romani is often regarded as a deficient or degraded language. And because its roots are in Indian languages such as Hindi, it is also unlike the European languages that surround it (Hancock 1993).

For many decades, the question of how best to educate Roma children has been debated, with major political groups issuing edicts to try to address the problem (Council of Europe 2000, 2007). Some European researchers have argued that Roma children are handicapped by their knowledge of Romani as a mother tongue, which is considered a barrier for acquisition of the state language. But others have published papers claiming inherent deficiencies in the IQ of Roma (Bakalar 2004). In several European countries, if Roma children have not mastered the language of the state by the age of 6, when they enter primary school, they are sent to special schools for the cognitively impaired (Fremlova & Ureche 2011, ERRC 2014, Amnesty Int. et al. 2015). In Slovakia, up to 65% of Roma children are placed in special schools because of their lack of mastery of Slovak at the start of schooling (Friedman et al. 2009, Lajčakova 2013). Because few teachers or officials know Romani or acknowledge it as a mother tongue, Roma children are in dire need of linguistic justice.

Hristo Kyuchukov, a Roma psycholinguist from Bulgaria, conducted ethnographic research on Roma language development, emphasizing the cultural differences between the ways Roma engage in linguistic pedagogy with their children and the mainstream practices in the West. As also suggested by much research on AAE speakers and on children growing up in Hispanic American families, both styles of interaction and expectations about young children's conversational roles in Roma families may clash with the culture of schooling (Reger 1999, Kyuchukov 2009, Teasley 2013).

For all these reasons, traditional language assessments might be doomed to fail. Roma children are exposed to very few books at home, and they lead lives of great material restriction, limiting their potential school-type vocabularies. However, a change in cultural practices—either in the families or in the schools—takes longer than a lifetime, and the needs of Roma children are pressing. A fair test of Romani language skill could contribute to a solution to this dilemma. At a minimum, it would indicate to educators that some Roma children, probably the majority, have a broad knowledge of the properties of grammar and vocabulary in a full natural language and therefore should not be considered candidates for segregated schools that treat them as cognitively impaired. In 2008, attendees at a conference organized by the Council of Europe in Bratislava discussed educational issues faced by Roma communities and families, specifically the role of Romani in the educational process of Roma children (Kyuchukov 2008). One of the recommendations was that Roma children should be tested in their mother tongue before being placed in special schools.

As a native speaker of Romani who is widely acquainted with Romani dialects across Europe and a skilled observer of young Roma children, Kyuchukov was in a good position to develop a language assessment with items that had appropriate construct validity. He began the process of developing such an assessment instrument in 2013.

Kyuchukov chose items that considered the products of learning, such as the passive voice or the exhaustivity of *wh*-questions, as well as items designed to test the process of learning (Hirsh-Pasek et al. 2005). For these items, *wug*-type tests (Berko 1958) assessed whether a child could extend past
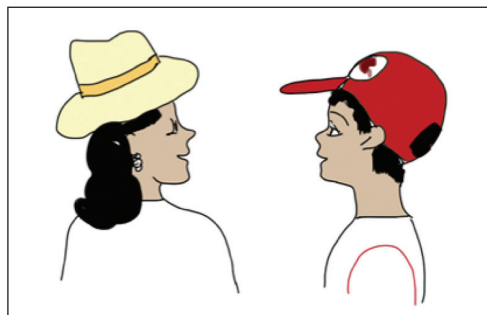
Kon so si ko šoro?
who what have-3sg. on head
"Who is wearing what?"

Figure 3

Multiple *Wh*-Item from Test of Romani (under development) (Kyuchukov & de Villiers 2014).

tense to a novel verb in production, and others asked whether the child could infer the meaning of a novel word from gender and other cues in the linguistic context. These kinds of items stretch children's linguistic knowledge and prove that what they know is flexible and creative, not learned by rote or lexically specific. Furthermore, the test had to take into account dialect variations across Romani, of which there are many (Matras 2002), and it had to be amenable to minor adaptations (such as slight variants of lexical items or morphemes). Ultimately, Kyuchukov developed and illustrated nine different subtests, and testing has been performed with samples of 3- to 5-year-olds in Bulgaria, Croatia, the Czech Republic, and Slovakia. A comparative study was conducted in Sweden and the countries from which Swedish Roma migrate, namely Macedonia and Serbia. Sweden is trying hard to integrate young Roma by making provision for Romani in preschool and hiring native-speaking Roma aides and teachers.

The overall results demonstrate the resilience of these young learners when they are tested in their home language. For example, on a test of exhaustive *wh*-movement (Schulz & Roeper 2011), a test that has been used across many different European languages to identify semantic difficulties, the Roma children gave 80%-exhaustive readings at age 5, exceeding the 60% average at that age found in other research (see **Figure 3** for an example) (Schulz & Roeper 2011, de Villiers et al. 2016). Concerning the intricate possessive Romani morphology, with its 16-way paradigm of bidirectional agreement, even the 3-year-olds got most of the questions right, including those involving novel nouns (see **Figure 4** for an example) (Kyuchukov & de Villiers 2014, Kyuchukov et al. 2016).

The research is still a long way from proving the utility of these tests for identifying children with language impairments in the acquisition of Romani. The prospect of a satisfactory a priori classification against which to validate the test seems distant indeed. The purpose here is more educational: If children have mastered a mother tongue, as demonstrated via rigorous tests that prove the creative nature of their grammar and their capacity to learn, then these children are neither language delayed nor cognitively impaired.

7. MULTIPLE OFFICIAL LANGUAGES

People in the United States may find a country with 11 official state languages unusual, although South Africa is certainly not an outlier in this regard. As might be expected, there are no normed,
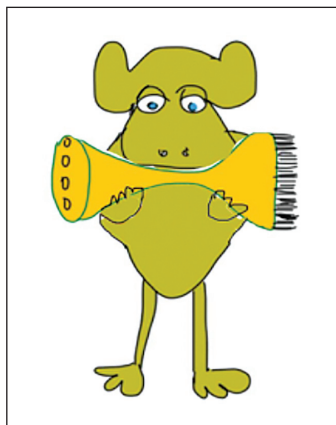
This is a boho. He has a suki.
Point to object:
This is?
Kaka ….(o bokoskori suki)
this...(the boho(masc.sg)-GEN suki(fem.sg))
This is...(boho's suki).

Figure 4

Novel Possessive Item from Test of Romani (under development) (Kyuchukov & de Villiers 2014).

validated standardized tests in South Africa, except in English. The DELV has been translated and adapted for use with Afrikaans speakers (van Dulm & Southwood 2008), but only for informal assessments and research, because there are no norms and the market was not large enough to interest the publisher in support. Yet Afrikaans is the home language of more than 6.5 million people, and most of these speakers are nonwhite. The late linguist Susan Suzman (e.g., Demuth & Suzman 1997, Suzman 2002) made major efforts on behalf of children speaking isiZulu, a language with more than 11.5 million speakers. She once witnessed an English-speaking SLP trying to test a boy in isiZulu by enlisting the help of the isiZulu-speaking school custodian to translate her questions. The SLP was trying to assess the boy's comprehension of prepositions, in ignorance of the fact that there are no prepositions in isiZulu. Other trained SLPs and linguists have developed suitable materials to test linguistic phenomena unique to Bantu languages, such as noun-class systems in which nouns can fall into one of 15 classes and trigger agreement on their verbs and adjectives (Smouse 2013). Bortz (1994) undertook an extensive effort to develop morphosyntactic tests, not via translation, in isiZulu, with the goal of extension to other Bantu languages. Local norms were developed on her test, the ZERLA, for more than 300 isiZulu-speaking preschool children.

Despite the relatively large sample and its attention to the unique morphological aspects of the language, ZERLA has gone unused. The cultural context may help us understand why. However linguistically appropriate a test may seem, if it does not have acceptance from the community of speakers, the educators, and the native-speaking SLPs, it will not serve its purpose. But what if there are no trained native-speaking linguists from the community in question? Recently, researchers took a different approach that considered not only the constructs particular to the languages but also the cultural context and community acceptance. In this approach, known as Delphi, the researchers included a carefully selected focus group to ensure cultural familiarity and suitability (Pascoe et al. 2013). A group in the Western Cape of South Africa has constructed and refined

a set of appropriate words for a vocabulary and speech discrimination test in isiXhosa. Some groundwork has been laid, but the prospect of standardized tests in the Bantu languages is still a distant one.

More has to change, educationally and societally, for standardized testing even to be meaningful. In South Africa, children who do not speak English or Afrikaans as a first language face a different infrastructure than those who do. Pascoe (2010) reports from a survey of 150 SLPs in the Western Cape that only one of them spoke isiXhosa, for a population of more than eight million native speakers. As in many of the cases discussed above, most children there are multilingual, with a Bantu first language, followed by Afrikaans and English, plus usually another Bantu language. Yet some are struggling with a first language, and my colleagues and I witnessed this in a township outside of Cape Town about a decade ago. There, 30 kindergarteners used as a classroom a shipping container with no windows, only an open side, which made the room unbearably hot or cold, depending on the season. The 50 or 60 first-graders, who spoke mostly isiXhosa, had graduated to a large, industrial-style classroom. There were no books in these children's native languages, but there were several dated, imported English books about Dick and Jane living in leafy English suburbs with their dog. Most of the children were industriously copying English words into small notebooks with dotted lines, but a group of perhaps 8 or 10 pupils were seated at a side table and seemed to be outside the activities. The teacher explained that these children could not participate with the rest, and were just being kept busy. At the time, we had been actively engaged with native-speaking linguists at the University of Cape Town, imagining how we could conduct research and help advance the fair assessment of children, but this brief experience at the school in the township was a sudden brake on our optimism. Would it help to have these children officially diagnosed, if there was nothing yet in place for intervention or services? Clearly, the role of early language assessment in a cultural context like this one must be carefully weighed against priorities; the complexities are staggering. But linguists need to have a voice in these deliberations.

8. CONCLUSION

Good assessment requires collaboration among people with varying knowledge and expertise. Fair assessment is still an elusive goal. The challenge requires bringing the best of empirical research to bear on devising good tests that go beyond translation and recognize the diverse paths in multicultural societies. However, cultural context is also a significant factor, and bringing on board a larger community to advise on suitability is necessary for an assessment to work. Furthermore, assessments have a place only in a society that has a plan, a way forward, for children in need of intervention. Linguists are beginning to play a role in persuading governments to invest in services and advocate for children with language delays. Children should not sit out their childhoods on the sidelines.

DISCLOSURE STATEMENT

I am an author of the DELV test and therefore receive royalties. I refer in the text to QUILS, a new screening test on which I will receive royalties. I am an author of DREAM, the Mandarin test mentioned in the text, and therefore may receive royalties in the future.

LITERATURE CITED

Am. Speech Lang. Hear. Assoc. 1983. Social dialects (position statement). *Asha* 25:25–27. **http://www.asha.
    org/policy/PS1983-00115.htm**

Amnesty Int., ERRC (Eur. Roma Rights Cent.), Open Soc. Fund, COSIV (Czech Soc. Incl. Educ.). 2015. *Czech Republic: Eight years after the D.H. judgment a comprehensive desegregation of schools must take place*. Report, index number EUR 71/2863/2015, Amnesty Int., London. **https://www.amnesty.org/en/documents/ eur71/2863/2015/en/**

Armon-Lotem S. 2012. Introduction. Bilingual children with SLI—the nature of the problem. *Biling. Lang. Cogn.* 15:1–4

Armon-Lotem S, de Jong J, Meir N, ed. 2015. *Communication Disorders Across Languages*, vol. 13: *Assessing Multilingual Children. Disentangling Bilingualism from Language Impairment.* Bristol, UK: Multiling. Matters

Bakalar P. 2004. The IQ of Gypsies in Central Europe. *Mankind Q.* 44:291–300

Benedicto E, Abdulkarim L, Johnson D, Garrett D, Seymour HN. 1998. Overt copulas in African American English speaking children. In *Proceedings of the 22nd Boston University Conference on Language Development*, ed. A Greenhill, M Hughes, H Littlefield, H Walsh, pp. 50–57. Somerville, MA: Cascadilla

Bedore LM, Peña ED, Gillam RB, Ho T. 2010. Language sample measures and language ability in Spanish English bilingual kindergarteners. *J. Commun. Disord.* 43:498–510

Bereiter C, Engelmann S. 1966. *Teaching Disadvantaged Children in the Preschool*. Englewood Cliffs, NJ: Prentice Hall

Berko J. 1958. The child's learning of English morphology. *Word* 14:150–77

Bernstein B. 1971. *Class, Codes and Control: Theoretical Studies Towards a Sociology of Language*. London: Routledge/Kegan Paul

Bialystok E, Luk G, Peets KF, Yang S. 2010. Receptive vocabulary differences in monolingual and bilingual children. *Biling. Lang. Cogn.* 13:525–31

Bortz M. 1994. *Piloting a language assessment for preschool Zulu speakers*. Presented at Natl. Speech. Lang. Hear. Conf., Durban, S. Afr. **http://www.play-it-forward-toys.co.za/Downloads/presclevahdev.pdf**

Campbell LR. 1996. Issues in service delivery to African American children. In *Communication Development and Disorders in African American Children: Research, Assessment, and Intervention*, ed. AG Kamhi, KI Pollock, JL Harris, pp. 73–93. Baltimore: Brookes

Cartmill EA, Armstrong BF, Gleitman LR, Goldin-Meadow S, Medina TN, Trueswell JC. 2013. Quality of early parent input predicts child vocabulary 3 years later. *PNAS* 110:11278–83

Cha K, Goldenberg C. 2015. The complex relationship between bilingual home language input and kindergarten children's Spanish and English oral proficiencies. *J. Educ. Psychol.* 107:935–53

Core C, Hoff E, Rumiche R, Señor M. 2013. Total and conceptual vocabulary in Spanish–English bilinguals from 22 to 30 months: implications for assessment. *J. Speech Lang. Hear. Res.* 56:1637–49

COST. 2010. *Language impairment in a multilingual society: linguistic patterns and the road to assessment. Monitoring progress report.* COST action IS0804, COST, Brussels

Council of Europe. 2000. *Recommendation No R (2000) 4 of the Committee of Ministers to Member States on the Education of Roma/Gypsy Children in Europe*. Strasbourg, France: Council of Europe. **https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId= 09000016805e2e91**

Council of Europe. 2007. *A Curriculum Framework for Romani*. Strasbourg, France: Council of Europe

Craig HK, Washington JA. 2000. An assessment battery for identifying language impairments in African American children. *J. Speech Lang. Hear. Res.* 43:366–79

Demuth K, Suzman S. 1997. Language impairment in Zulu. In *Proceedings of the 21st Annual Boston University Conference on Language Development*, ed. EM Hughes, A Green, 1:124–35. Somerville, MA: Cascadilla

de Villiers JG. 2015. Taking account of both languages in the assessment of dual language learners. *Semin. Speech Lang.* 36:120–32

de Villiers JG, Ning CY, Liu L, Rolfhus E, Hutchings T, et al. 2016. *Acquisition and delay of exhaustivity in multiple* wh-*questions: evidence from children acquiring Mandarin*. Poster presented at 35th Annu. Symp. Res. Child Lang. Disord., Madison, WI, June 16–18

de Villiers JG, Roeper T, Bland-Stewart L, Pearson BZ. 2008 Answering hard questions: *wh*-movement across dialects and disorder. *Appl. Psycholinguist.* 29:67–103

de Villiers PA, de Villiers JG. 2010. Assessment of language acquisition. *Cogn. Sci.* 1:230–44

ERRC (Eur. Roma Rights Cent.). 2014. *Overcoming barriers: ensuring that the Roma children are fully engaged and achieving in education*. Ref. number 140195, Off. Stand. Educ. Child. Serv. Skills, Manchester, UK. **https://www.gov.uk/government/publications/ensuring-roma-children-achieve-in-education**

Fiestas CE, Peña ED. 2004. Narrative discourse in bilingual children: language and task effects. *Lang. Speech Hear. Serv. Sch.* 35:155–68

Flores G, Tomany-Korman SC. 2008. Racial and ethnic disparities in medical and dental health, access to care, and use of services in US children. *Pediatrics* 121:e286–98

Ford M, Baer CT, Xu D, Yapanel U, Gray S. 2008. *The LENATM language environment analysis system: audio specifications of the DLP-0121*. Tech. rep. LTR-03-2. Lena Found., Boulder, CO

Fremlova L, Ureche H. 2011. *From Segregation to Inclusion: Roma Pupils in the United Kingdom. A Pilot Research Project*. Budapest: Roma Educ. Fund

Fricke S, Bowyer-Crane C, Haley AJ, Hulme C, Snowling MJ. 2013. Efficacy of language intervention in the early years. *J. Child Psychol. Psychiatry* 54: 280–90

Friedman E, Gallová Kriglerová E, Kubánová M, Slosiarik M. 2009. *School as Ghetto: Systemic Overrepresentation of Roma in Special Education in Slovakia*. Budapest: Roma Educ. Fund

Friedmann N, Novogrodsky R. 2011. Which questions are most difficult to understand? The comprehension of *wh* questions in three subtypes of SLI. *Lingua* 121:367–82

Fryer RG Jr., Levitt SD. 2004. Understanding the black–white test score gap in the first two years of school. *Rev. Econ. Stat.* 86:447–64

Gagarina G, Klop D, Tsimpli I, Walters J. 2016. Narrative abilities in bilingual children. *Appl. Psycholinguist.* 37:11–17

Golinkoff R, de Villiers JG, Hirsh-Pasek K, Iglesias A, Wilson M. 2016. *QUILS: Quick Interactive Language Screener*. Baltimore: Brookes

Green LJ. 2002. *African American English: A Linguistic Introduction*. Cambridge, UK: Cambridge Univ. Press

Green LJ. 2011. *Language and the African American Child*. Cambridge, UK: Cambridge Univ. Press

Green L, Roeper T. 2007. The acquisition path for tense–aspect: remote past and habitual in child African American English. *Lang. Acquis.* 14:269–313

Hacohen A. 2006. The psychological reality of the telic/atelic distinction: evidence from adult and child Hebrew. In *Proceedings of the 21st Annual Meeting of the Israel Association for Theoretical Linguistics* (*IATL 21*), ed. YN Falk. Jerusalem: IATL. 19 pp. **http://linguistics.huji.ac.il/IATL/21/Hacohen.pdf**

Hancock I. 1993. *A Grammar of Vlax Romani*. London/Austin, TX: Romanestan

Hart B, Risley TR. 1996. *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore: Brookes

Hasson N, Joffe V. 2007. The case for dynamic assessment in speech and language therapy. *Child Lang. Teach. Ther.* 23:9–25

Heath SB. 1982. What no bedtime story means: language skills at home and school. *Lang. Soc.* 11:49–76

Heath SB. 1983. *Ways with Words: Language*, *Life and Work in Communities and Classrooms*. Cambridge, UK: Cambridge Univ. Press

Hirsh-Pasek K, Adamson LB, Bakeman R, Owen MT, Golinkoff R, et al. 2015. The contribution of early communication quality to low-income children's language success. *Psychol. Sci.* 26:1071–83

Hirsh-Pasek K, Kochanoff A, Newcombe N, de Villiers JG. 2005. Using scientific knowledge to inform preschool assessment: making the case for empirical validity. *Soc. Policy Rep.* 19:3–19

Hoff E. 2013a. *Language Development*. Boston: Cengage

Hoff E. 2013b. Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Dev. Psychol.* 49:4–14

Iglesias A. 2015. Language impairment in bilingual children: from theory to practice. *Semin. Speech Lang.* 3687–88

Johnson VE. 2001. *Fast mapping verb meaning from argument structure*. PhD thesis, Univ. Mass., Amherst

Johnson VE, de Villiers JG. 2009. Syntactic frames in fast mapping verbs: the effects of age, dialect and clinical status. *J. Speech Lang. Hear. Res.* 52:610–22

Kyuchukov H. 2005. Early socialization of Roma children in Bulgaria. In *Bilingualism and Education: From the Family to the School*, ed. XP Rodriguez-Yanez, AM Lorenzo Suarez, F Ramallo, pp. 161–68. Munich, Ger.: Lincom

Kyuchukov H. 2008. Language comprehension of Roma children from special schools in Bulgaria, Czech Republic and Slovakia. In *Equal Access to Quality Education for Children from Socially Disadvantaged Settings*, ed. D Kopčnova, pp. 35–41. Bratislava: Res. Inst. Child Psychol. Pathopsychol./Educ. Sect., Slovak Comm. UNESCO

Kyuchukov H. 2009. *Подготовка за обучение на ромските деца в домашни условия*. [*Preparation for Education of Roma Children in Home Environment*.] Sofia, Bulg.: Wini

Kyuchukov H. 2010. Romani language competence. In *Situation of Roma Minority in the Czech Republic, Hungary, Poland and Slovakia*, ed. J Balvin, L Kwadrants, pp. 427–65. Wrocław, Pol.: Prom

Kyuchukov H, de Villiers J. 2014. *Addressing the rights of Roma children for a language assessment in their native language of Romani*. Poster presented at 35th Annu. Symp. Res. Child Lang. Disord., Madison, WI, June 16–18

Kyuchukov H, de Villiers J, Hasan B. 2016. Language complexity, narratives and theory of mind of Romani-speaking children. In *Södertörn Series on Romani Studies*, vol. 2: *Proceedings of the 2015 Nordic Conference on Romani Studies*. In press

Labov W. 1972. *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia: Univ. Pa. Press

Lajčakova J. 2013. *Civil Society Monitoring Report on the implementation of the National Roma Integration Strategy and Roma Decade Action Plan in 2012 in Slovakia*. Decade Roma Incl. Secr. Found., Budapest

Leonard LB. 2014. *Children with Specific Language Impairment*. Cambridge, MA: MIT Press. 2nd ed.

Liu X, de Villiers J, Ning C-Y, Rolfhus E, Lee W, et al. 2016. Research to establish the validity, reliability and clinical utility of a comprehensive language assessment of Mandarin. *J. Speech Lang. Hear. Res.* In press

Matras Y. 2002. *Romani: A Linguistic Introduction*. Cambridge, UK: Cambridge Univ. Press

Neuman SB. 2000. Social contexts for literacy development: a family literacy program. In *Play and Literacy in Early Childhood: Research from Multiple Perspectives*, ed. KA Roskos, JF Christie, pp. 153–68. Mahwah, NJ: Erlbaum

Neuman SB, Celano D. 2001. Access to print in low-income and middle-income communities: an ecological study of four neighborhoods. *Read. Res. Q.* 36:8–26

Oetting JB, Gregory KD, Rivière A. 2016. Changing how speech–language pathologists think and talk about dialect variation. *Perspect. ASHA Spec. Interest Groups* 16:28–37

Oetting JB, Lee R, Porter KL. 2013. Evaluating the grammars of children who speak nonmainstream dialects of English. *Top. Lang. Disord.* 33:140–51

Oetting JB, McDonald JL. 2002. Methods for characterizing participants' nonmainstream dialect use in child language research. *J. Speech Lang. Hear. Res.* 45:505–18

Oller DK, Pearson BZ, Cobo-Lewis AB. 2007. Profile effects in early bilingual language and literacy. *Appl. Psycholinguist.* 28:191–230

O'Neill W. 1998a. Education in the media. *Radic. Teach.* 54:13–17

O'Neill W. 1998b. If Ebonics isn't a language, then tell me, what is? See Perry & Delpit 1998, pp. 38–48

Ortman J, Shin H. 2011. *Language projections: 2010 to 2020*. Presented at Annu. Meet. Am. Sociol. Assoc., Las Vegas, Aug. 20–23

Pascoe M. 2010. Children with speech difficulties: an exploratory survey of clinical practice in the Western Cape. *S. Afr. J. Commun. Disord.* 57:66–75

Pascoe M, Rogers C, Norman V. 2013. Are we there yet? On a journey towards more contextually relevant resources in speech–language therapy and audiology. *S. Afr. J. Commun. Disord.* 60:2–9

Paul R, Norbury C. 2012. *Language Disorders from Infancy Through Adolescence: Listening, Speaking, Reading, Writing, and Communicating*. Amsterdam: Elsevier

Pearson BZ, Conner T, Jackson JE. 2013. Removing obstacles for African American English–speaking children through greater understanding of language difference. *Dev. Psychol.* 49:31–44

Pearson BZ, Fernández SC, Oller DK. 1993. Lexical development in bilingual infants and toddlers: comparison to monolingual norms. *Lang. Learn.* 43:93–120

Pearson BZ, Fernández S, Oller DK. 1995. Cross-language synonyms in the lexicons of bilingual infants: one language or two? *J. Child Lang.* 22:345–68

Pearson BZ, Jackson JE, Wu H. 2014. Seeking a valid gold standard for an innovative, dialect-neutral language test. *J. Speech Lang. Hear. Res.* 57:495–508

Peña ED, Bedore LM. 2009. Bilingualism in child language disorder. In *Handbook of Child Language Disorders*, ed. RG Schwarts, pp. 281–307. New York: Psychology

Peña ED, Bedore LM, Zlatic-Giunta R. 2002a. Category-generation performance of bilingual children: the influence of condition, category, and language. *J. Speech Lang. Hear. Res.* 45:938–47

Peña ED, Bedore LM, Zlatic-Giunta R. 2002b. Use of a category generation task to assess vocabulary skills of Spanish–English bilinguals. *J. Speech Lang. Hear. Res.* 46:938–47

Peña ED, Gillam RB, Bedore LM, Bohman T. 2011. Risk for poor performance on a language screening measure for bilingual preschoolers and kindergarteners. *Am. J. Speech Lang. Pathol.* 20:302–14

Peña ED, Gutiérrez-Clellen VF, Iglesias A, Goldstein BA, Bedore LM. 2013. *Bilingual English Spanish Assessment* (*BESA*) *Manual*. San Rafael, CA: AR-Clinical Publ.

Peña ED, Halle TG. 2011. Assessing preschool dual language learners: traveling a multiforked road. *Child Dev. Perspect.* 5:28–32

Peña E, Quinn R. 1997. Task familiarity: effects on the test performance of Puerto Rican and African American children. *Lang. Speech Hear. Serv. Sch.* 28:323–32

Peña E, Quinn R, Iglesias A. 1992. The application of dynamic methods to language assessment: a nonbiased procedure. *J. Spec. Educ.* 26:269–80

Perry T, Delpit L, ed. 1998. *The Real Ebonics Debate: Power*, *Language and the Education of African American Children*. Boston: Beacon

Plante E, Vance R. 1994. Selection of preschool language tests: a data-based approach. *Lang. Speech Hear. Serv. Sch.* 25:15–24

Reger Z. 1999. Teasing in the linguistic socialization of Gypsy children in Hungary. *Acta Linguist. Hung.* 46:289–315

Reilly S, Tomblin B, Law J, McKean C, Mensah FK, et al. 2014. Specific language impairment: a convenient label for whom? *Int. J. Lang. Commun. Disord.* 49:416–51

Rhodes R, Ochoa SH, Ortiz SO. 2005. *Comprehensive Assessment of Culturally and Linguistically Diverse Students: A Practical Approach*. New York: Guilford

Rice ML. 2013. Language growth and genetics of specific language impairment. *Int. J. Speech Lang. Pathol.* 15:223–33

Rice ML, Wexler K. 1996. Toward tense as a clinical marker of specific language impairment in English-speaking children. *J. Speech Lang. Hear. Res.* 39:1239–57

Sattler JM. 1992. *Assessment of Children*. San Diego, CA: Jerome Sattler. 3rd ed.

Schulz P, Roeper T. 2011. Acquisition of exhaustivity in *wh*-questions: a semantic dimension of SLI? *Lingua* 121(spec. issue):383–407

Seymour HN, Bland L, Green LJ. 1998. Difference versus deficit in child African-American English. *Lang. Speech Hear. Serv. Sch.* 29:96–108

Seymour HN, Pearson BZ, ed. 2004. Evaluating language variation: distinguishing development and dialect from disorder. *Semin. Speech Lang.* 25(1, spec. issue). New York: Thieme

Seymour HN, Roeper T, de Villiers J. 2005. *The DELV-NR: The Diagnostic Evaluation of Language Variation*. San Antonio, TX: Psychol. Corp. Norm-ref. ed.

Smouse MR. 2013. Uninterpretable features in comprehension: subject–verb agreement in isiXhosa. *S. Afr. J. Afr. Lang.* 33: 65–74

Spaulding TJ, Plante E, Farinella KA. 2006. Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Lang. Speech Hear. Serv. Sch.* 37:61–72

Stockman IJ. 1996. The promises and pitfalls of language sample analysis as an assessment tool for linguistic minority children. *Lang. Speech Hear. Serv. Sch.* 27:355–66

Stockman IJ. 2010. A review of developmental and applied language research on African American children: from a deficit to difference perspective on dialect differences. *Lang. Speech Hear. Serv. Sch.* 41:23–38

Storkel HL, Hoover JR. 2011. The influence of part-word phonotactic probability/neighborhood density on word learning by preschool children varying in expressive vocabulary. *J. Child Lang.* 38:628–43

Suzman SM. 2002. Morphological accessibility in Zulu. *Amsterdam Stud. Theory Hist. Linguist. Sci. Ser.* 4:155–74

Tager-Flusberg H, Cooper J. 1999. Present and future possibilities for defining a phenotype for specific language impairment. *J. Speech Lang. Hear. Res.* 42:1275–78

Teasley C. 2013. Educating against the cultural politics and complicities of containment. In *Roma Education in Europe: Practices, Policies and Politics*, ed. M Miskovic, pp. 29–41. New York: Routledge

Tomblin B, Records NL, Buckwalter PR, Zhang X, Smith E, O'Brien M. 1997. Prevalence of specific language impairment in kindergarten children. *J. Speech Lang. Hear. Res.* 40:1245–60

Unsworth S. 2013. Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: the case of Dutch gender. *Biling. Lang. Cogn.* 16:86–110

van der Lely HK, Battell J. 2003. *Wh*-movement in children with grammatical SLI: a test of the RDDR hypothesis. *Language* 79:153–81

van Dulm O, Southwood F. 2008. Toward a dialect-neutral Afrikaans-medium child language assessment instrument: test item development. *Lang. Matters* 39:300–15

Van Hout A. 2008. Acquiring perfectivity and telicity in Dutch, Italian and Polish. *Lingua* 118:1740–65

Wagner L. 2002 Children's comprehension of completion cntailments in the absence of agency cues. *J. Child Lang.* 29:109–25

Washington JA, Craig HK. 2004. A language screening protocol for use with young African American children in urban settings. *Am. J. Speech Lang. Pathol.* 13:329–40

Westby CE. 2007. Child maltreatment: a global issue. *Lang. Speech Hear. Serv. Sch.* 38:140–48

Wyatt TA. 2002. Assessing the communicative abilities of clients from diverse cultural and language backgrounds. In *Communication Disorders in Multicultural Populations*, ed. DE Battle, pp. 415–59. Boston: Butterworth-Heinemann. 3rd ed.

Zepeda M, Rodriguez JL. 2014. Bilingual development in early childhood. In *Mexican American Children and Families: Multidisciplinary Perspectives*, ed. Y Caldera, E Lindsey, pp. 122–34. New York: Routledge

Annual Review of Linguistics

Volume 3, 2017

Contents

Errata

An online log of corrections to *Annual Review of Linguistics* articles may be found at
http://www.annualreviews.org/errata/linguistics