Advanced Review

Assessment of language acquisition

Peter A. de Villiers* and Jill G. de Villiers

This review addresses questions of what should be assessed in language acquisition, and how to do it. The design of a language assessment is crucially connected to its purpose, whether for diagnosis, development of an intervention plan, or for research. Precise profiles of language strengths and weaknesses are required for clear definitions of the phenotypes of particular language and neurodevelopmental disorders. The benefits and costs of formal tests versus language sampling assessments are reviewed. Content validity, theoretically and empirically grounded in child language acquisition, is claimed to be centrally important for appropriate assessment. Without this grounding, links between phenomena can be missed, and interpretations of underlying difficulties can be compromised. Sensitivity and specificity of assessment instruments are often assessed using a gold standard of existing tests and diagnostic practices, but problems arise if that standard is biased against particular groups or dialects. The paper addresses the issues raised by the goal of unbiased assessment of children from diverse linguistic and cultural backgrounds, especially speakers of non-mainstream dialects or bilingual children. A variety of new approaches are discussed for language assessment, including dynamic assessment, experimental tools such as intermodal preferential looking, and training studies that assess generalization. Stress is placed on the need for measures of the process of acquisition rather than just levels of achievement.
WIREs Cogn Sci 2010 1 230–244

What makes for an effective assessment of language acquisition? In this paper, we address two fundamental questions that underlie the issue: first, 'What aspects of language should be assessed?' and second, 'How should language acquisition be assessed?' Although we will draw on examples from existing tests to illustrate points, our focus is on broader considerations on which any evaluation of existing materials should be based. In reviewing the principles that should guide the assessment of language acquisition we hope to lay the groundwork for a new generation of assessment instruments.

*Correspondence to: pdevilli@smith.edu

Smith College, Department of Psychology, Northampton, MA, USA

DOI: 10.1002/wcs.30

WHAT ASPECTS OF LANGUAGE SHOULD BE ASSESSED?

The Purposes of Assessment

First, the nature of an assessment of language acquisition must be informed by the purposes of the assessment.[1] Is it for diagnosis, planning therapy, monitoring progress, or research?

Diagnosis

A language assessment may be carried out by a speech and language pathologist (SLP) to *diagnose a child as having a language delay or disorder*. This assessment may be mandated to determine the child's *eligibility for intervention services* or for coverage by insurance programs. For these purposes, the child's performance on a variety of language measures is compared with the norms for children of the same age, and depending on the test, the cut off for identifying children as language impaired is usually placed at

1 to 1.5 standard deviations (sd) below the mean of the normative sample for their age.

Several considerations should enter into decisions about the diagnosis of language impairment. Most importantly, the comparison sample of children must be representative of the population group from which the child comes. Comparing the performance of a particular child against the norms for a population of different socioeconomic, ethnic, cultural or linguistic background is inappropriate (see Section *Multicultural Assessment, Dialects and Bilingualism Below on Multicultural Issues in Assessment*). The breadth and depth of the assessment will depend on the nature of the underlying problem that the child may display in his conversational interaction, or whether the child has a separate diagnosis of a neurodevelopmental disorder or syndrome. For example, the need for comprehensive assessment of phonological factors in speech may depend on whether the child presents with articulation difficulties on a screening test for such problems. Similarly, assessment of verbal children diagnosed with autism spectrum disorders (ASD) would benefit from a more extensive focus on their pragmatic skills because language impairments in high functioning children with ASD differentially impact their understanding of communication and conversation.[2]

Assessment for the purposes of diagnosis and placement in services should examine both production (expressive language) and comprehension (receptive language). In many children with language impairment, comprehension of language may be less impaired than production; whereas in cases of central processing problems, comprehension is characteristically more impaired than production.[3] Even if both receptive and expressive language problems are observed, the balance between them is an important consideration for planning effective intervention.

Determining the severity of the language impairment and its impact on the children's day-to-day social interaction and ability to function in a classroom setting is also important. Assessment in these cases would profitably include observation to establish their everyday level of functioning. An alternative (though less direct) way of obtaining similar information is through questionnaires and checklists filled out by the children's primary caregiver or teacher.

Language Profiles

Another purpose of assessment is to *establish the primary nature of the language impairment* and a *profile of the child's strengths and weaknesses* across the different domains and subdomains of language. This information can then shape the planning of the content and goals of language intervention and the establishment of an individual educational plan (IEP) for the child.

For these purposes, a criterion-referenced assessment is the most informative, where the individual's performance is compared not with a normative group but with a criterion for mastery of that feature of language. The criteria for mastery must refer to typical patterns and stages of development of those forms and functions. An example of such criterion-referenced information would be evidence of where the child was in the acquisition of a particular type of sentence structure, for example, asking and answering different types of questions, when the stages of development or characteristic developmental errors are known from the published research on typically developing children.

Unlike in an omnibus diagnostic test, if such an assessment is to be useful for the planning of language intervention, it must sample a large enough set of items evaluating the child's mastery of a particular subdomain of syntax to allow for a true profile of linguistic knowledge. If a test only samples one or two items exemplifying a particular grammatical construction, this is very difficult.

In practice, given the high caseloads of many SLPs and the restrictions placed on insurance funding for assessment and treatment, there is often a trade-off between the depth and comprehensiveness of the language assessment that can be carried out and the amount of time available. This problem can be mitigated somewhat by doing the assessment in two stages—a broader assessment with less depth to establish a diagnosis and identify areas in which the child seems to be particularly impaired, followed by more in-depth assessment of particular domains to inform the planning of therapy. To assess in depth, instruments with more specific focus on subdomains of language might be used (e.g., the Rice/Wexler test of morphosyntax[4]), or the SLP might use sets of probe items that concentrate on the areas of syntactic difficulty identified in the first phase. Probes might be generated by the therapist or be adapted from subtests on assessment instruments that have numerous items measuring a particular form or function (e.g., the question-asking subtest of the Diagnostic Evaluation of Language Variation—Norm-Referenced (DELV-NR)[5]). Age equivalency should be interpreted cautiously on criterion-referenced assessments, because most have not been standardized on a larger population sample but extrapolated from the research literature on language acquisition.

Monitoring Progress

Language assessments are frequently carried out *to monitor and measure a child's progress* during therapy or educational intervention. In these cases, the assessment needs to target the structures that are the focus of intervention, with new items to assess generalization from the items and situations trained in therapy. It is also important to observe the child in more natural communication situations, to see whether mastery of the aspects of language targeted in therapy transfers to usage in a wider range of communication situations. For example, does a child who is taught to produce a range of wh-questions to gain information in a structured question elicitation game transfer that skill to their interactions in the classroom? Sometimes a child might be able to master the language tasks with elicitation or comprehension materials used in the testing situation, but not generalize that understanding to everyday conversations. This is often the case in children with ASD where trained skills may not generalize across situations.[3]

Research

Finally, a language assessment might be carried out for *research* purposes, where the researcher may be interested in examining the pattern of language acquisition in children who suffer from various neurodevelopmental disorders (e.g., ASD, or Williams syndrome), or in investigating the relationship between language acquisition and other aspects of children's development (e.g., the interaction between language and cognitive development). Language is frequently used as an outcome measure for educational interventions (e.g., curricular innovations in preschools) or as a measure of the effects of different child-rearing practices. The success of the human genome project has raised the possibility that with a sufficiently clear description of the phenotype, the genetic basis of particular language disorders might be clarified or even located.[6]

Again the aspects of language that should be assessed will vary with the purpose of the research. An evaluation of the impact of a preschool curriculum intervention, for instance, should tailor the domains and subdomains to be assessed to capture the children's acquisition of aspects of language that are likely to contribute to school readiness and educational success, such as vocabulary and narrative skills. Similarly, there is a tension between research attempting to determine the genetics of a particular kind of language impairment and the goal of articulating the linguistic nature of childhood language disorders across a number of possible phenotypes. The former might require a narrowly circumscribed domain of language to be assessed in-depth to specify a phenotype; the latter would require more comprehensive assessment of language across several domains. Mervis and Robinson[7] lay out several characteristics that an 'ideal' assessment instrument must have for genotype/phenotype studies of different types of language impairment. They argue that norm-referenced, precisely quantified profiles of children's strengths and weaknesses across several fundamental components of language are required for such studies to be successful. The extent to which such profiles distinguish between the language impairments associated with different disorders or syndromes might indicate common or separable genetic bases as well as interactions among genes that determine characteristic phenotypes.

It should be noted that there is potentially a built-in conservatism in research using language assessment instruments: even if the test is not an ideal instrument, researchers are likely to continue using it to make their results comparable with previously published work on the same population. For example, in many studies of the relationship between language acquisition and cognitive development in different populations of children, researchers have relied on the Peabody Picture Vocabulary Test (PPVT) as their major or even sole measure of language status. The PPVT is a standardized test of vocabulary comprehension in which children choose the picture that corresponds to the meaning of a word.[8] However, this is far too narrow a focus to provide an adequate characterization of the language skills of children and can considerably over- or underestimate children's language acquisition in some circumstances (see Section *Multicultural Assessment, Dialects and Bilingualism Below on Multicultural Issues in Assessment* below).

Empirical Validation of the Content of Assessment

Methods of assessment are empirically validated in two different ways. First, the content of the test is validated by research on the pattern of typical language acquisition and by research on the characteristics of the language of children with particular language impairments. Second, assessment is validated by its effectiveness in providing appropriate information for the purposes for which it is being carried out. Thus, for diagnostic purposes, assessments are typically evaluated by their *sensitivity* at identifying children with language impairment (i.e., does the test pick out all the children with language impairment from a sample of children), and their *specificity* (i.e., does the test pick out only language-impaired children and not

falsely identify typically developing children as being impaired) (see Section *Predictive Validity* for more on sensitivity and specificity).

Research on Typical Development

How is assessment informed by research on typical language acquisition? One way is by taking into account well-attested stages or patterns of acquisition of particular sentence constructions, the characteristic developmental errors that children make, and dependencies between the mastery of different forms in typical acquisition. The application of such information can be illustrated by children's acquisition of tense markers. It is well established that there is a coherent pattern of tense omission in the grammars of early English, and that this correlates with such phenomena as dropping the subject ('—going home'), using the wrong case on the subject ('Me go' instead of 'I go'), and so forth. In the absence of a theoretical account from normal acquisition, these might seem like unrelated phenomena and they would not be tested together.[4,9]

In another illustration, linguistic research on wh-questions shows that across many languages, wh-questions are exhaustive; that is, a response to a wh-question must enumerate all of the relevant items or events. This is particularly true in the case of double questions such as 'Who ate what?' where the reply must pair up all the eaters with all of the things that are eaten, and typically developing children know this at an early stage. However, without this empirical and theoretical background, one might not suspect that it is linked to how children deal with the semantic notion of a *variable,* and hence is connected to how the child treats quantifiers like *every*.[10] In English and many other languages, wh-questions also entail movement rules in which the wh-word must be extracted from different places in the underlying structure and moved to the front of the sentence. There are subtle restrictions on such movement, which the average person knows, but of which they are entirely unaware.

An assessment based on theoretically motivated research on typical children's grammatical development has the potential to provide important information on disorders. The question comprehension subtest of the DELV-NR, for example, is organized around items that test these two properties of wh-questions of English: exhaustivity and constraints on movement. The subtest is strongly discriminating of children with and without specific language impairment (SLI) without any bias against children who speak African American English (AAE) dialect. Language-impaired children perform significantly worse than age-matched typically developing children on this subtest and persist with errors characteristic of an earlier stage of development: however, they do not violate the structural constraints on movement rules considered to be part of Universal Grammar.[11]

Paying attention to theoretical accounts of language acquisition is justified if it has practical import. In the absence of a theory, a therapist might provide the wrong level of analysis of a problem, and approach its remediation the wrong way. Suppose, for example, that the child's failure to ask exhaustive questions is considered to be a symptom of a general tendency to be elliptical, or not saying 'enough'. Therapy might promote fuller responses and still not get closer to the fundamental problem. Or, case errors on the subject may be trained in isolation from mastery of the obligatory tense markers, when an indirect approach might in fact promote a faster reorganization of the grammar. At present, few ideas based on such theory-based linkages between grammatical features have made their way into mainstream practice.

However, as a number of other chapters in this series attest, theories of language acquisition also tend to change as theoretical approaches to linguistics and characterizations of the adult language evolve (or change in fashion). For example, there are substantial contrasts in the way that syntax acquisition is described by generative approaches that refer to underlying principles of Universal Grammar and the setting of linguistic parameters,[10,12] or by usage-based construction grammar approaches[13,14] that appeal to exemplar or formulaic learning and processes of distributional analysis and analogy across constructions. The implications of these differences have not been clearly worked out for the development of assessment methods and materials. One of the problems that a usage or item learning approach might have in assessing early language acquisition is that it would need to be determined that the children to be assessed have had roughly equivalent exposure to the forms to be assessed, otherwise differences between children could just be due to differential exposure rather than to underlying mastery or knowledge of the forms. On the other hand, the explanatory basis on the usage account is frequency of exposure, providing an attractive easy remedy for language delay if it is right.

Research on Atypical Populations

The content of assessment should also be informed by research on the language characteristics of children with specific language impairment or neurodevelopmental syndromes such as autism.

Considerable research has demonstrated that many children with SLI present with fundamental problems in the production and comprehension of inflectional morphology.[9,15] Many factors may contribute to the differential vulnerability of morphosyntax to language impairment. Inflectional suffixes are frequently unstressed in the speech input and may also be semantically redundant in the linguistic and situational context of the conversation. Thus, language processing problems may contribute to the problem.[15]

On the other hand, the problems that children with SLI have with morphosyntax may have a linguistic basis. To account for the differential difficulty of finiteness markers, Rice and Wexler[9,16] proposed an influential theory of SLI based on research on the acquisition of tense morphology in typical development: the extended optional infinitive hypothesis. On this theory, SLI can be characterized as a delay in maturation of the principle that tense marking is obligatory, a realization that typically developing children arrive at by around age 30 months. The consequences are widespread in the grammar, affecting subjects, case, negation and auxiliaries, as well as tense inflections.

Whether the explanation for the problem is in terms of speech processing or a more linguistic account, the relative difficulty of morphosyntax for children with SLI is a robust finding in research characterizing their language acquisition (at least for English and several other European languages). As a result, many widely used tests of English acquisition (e.g., the Comprehensive Assessment of Spoken Language (CASL),[17] the Clinical Evaluation of Language Fundamentals (CELF)-4),[18] the Test of Early Language Development (TELD)-3),[19] and the Test of Language Development: Intermediate (TOLD):I-4[20]) provide subtests of morphosyntax. As we discuss more fully below, this can become problematic when we are dealing with children whose dialect of English differs from mainstream English precisely with respect to the most frequently tested elements of morphosyntax. For these children, low scores on tests of morphosyntax could reflect either language impairment or correct use of their dialect.

Other lines of research on children with SLI have suggested that there is at least a subgroup who suffers from fundamental grammatical problems with embedded forms and movement rules. These children with 'grammatical SLI' experience differential difficulty with complex syntax—in complex questions, complement clauses, and relative clauses.[21,22] Still other approaches to SLI highlight the speech processing problems of the children.[23,24] Children with SLI perform poorly on nonword speech repetition tasks.

Multicultural Assessment, Dialects, and Bilingualism

Most normative information about language acquisition and language impairment comes from European American, middle class children. However, a significant number of children requiring services in schools come from culturally and linguistically diverse backgrounds. Some speak dialects of English that differ from the mainstream English taught in school (e.g., AAE, Caribbean English, Appalachian English, or various forms of Southern English dialect). These dialects represent rule-governed varieties of English that differ in phonology, syntax, morphology, vocabulary, and pragmatics from the mainstream form of English used in formal of educational settings. They may vary regionally and across social class and are frequently stigmatized and regarded as incorrect or inappropriate. Other children enter school with varying degrees of bilingualism or are acquiring English as at least a second language. Bilingualism is best considered as a continuum with monolinguals in each language at each end and degrees of bilingualism in between varying across situations. Cases of non-mainstream dialect and bilingualism (or multilingualism) raise particular problems for assessment of language acquisition.

The goal is to assess language acquisition with the least bias against culturally and linguistically diverse children. The assessment must distinguish *language difference*, that is, expected variation in domains of language due to the language community from which the child is drawn, from *language disorder*, deviations in the child's language from the norms of his own community.[25–28] Biases in assessment can arise in several ways. Items on formal tests may assess aspects of grammar for which the child's dialect provides a different correct answer than the one credited as correct according to Mainstream American English (MAE), or they may require access to information and experiences that are culturally conditioned.

For example, focus on the morphosyntax of MAE in many widely used assessment instruments frequently tests those very areas of the grammar of English in which there is considerable dialectal variation. Children with language impairment frequently omit tense markers and other grammatical morphemes for an extended period of time,[9,15,16] but these morphemes are also omitted in AAE as part of the normal rule system of the dialect. Thus, if the procedures and norms of the MAE language tests are used with African American children, it becomes difficult to tell whether a child scores low on the test because of mismatches with his dialect, or because he is delayed.[29] In the light of this problem, many of the available

tests include a warning about use with dialect speakers and even identify items that should be omitted for such speakers. This invalidates the use of the norms for those parts of the tests. Not surprising then, African American children are overrepresented in the caseloads of speech and language therapists and especially in special education classes, where eligibility is often determined on the basis of language assessments.[25] Some of these children are misdiagnosed as language impaired because of the mismatch between their dialect and the language assessment instruments used. On the other hand, it is possible that other African American children who do have genuine language impairments are not receiving language intervention services because the deviations in their language from the mainstream patterns is simply 'assumed' to be dialectal.

One way to address the issue of bias against AAE dialect speakers is to distinguish between *contrastive* items that systematically vary between AAE and MAE, and *noncontrastive* items in which both AAE and MAE require the same response from the child.[27,30] *Contrastive* items include various aspects of the morphosyntax and phonology of English (e.g., third person present tense '–s', or reduction of consonant clusters in word-final position). These items are then only used for identifying deviation of the child's language from MAE toward AAE dialect, not for diagnosing possible language impairment. *Noncontrastive* items include morphosyntactic features obligatory in both MAE and AAE (e.g., past tense copula 'was/were') as well as fundamental aspects of syntax that are common across different dialects of English. This strategy was used to produce a dialect neutral assessment instrument in the DELV Screener and the DELV-NR that could identify the child as a non-mainstream dialect speaker as well as appropriately diagnose language impairment.[5,31]

Similar concerns arise in the assessment of lexical development. Acquired vocabulary is particularly subject to cultural variation across language communities and dialects of English, dependent as it is on exposure to words in particular domains of experience (e.g., names for animals, foodstuffs, tools, etc.). Many standardized vocabulary tests may therefore be measuring access to experiences, reading, and picture materials as much as ability to learn words. This is not a problem if the goal is to assess whether children know the vocabulary needed for reading materials used in their classrooms, but it is a major problem if the goal is to diagnose the children as typically developing or language impaired. To determine the latter, it is the children's ability to learn words when exposed to the learning opportunities that must be assessed (see below in Section *Vocabulary Assessment*).

Assessing the language of bilingual children raises similar issues as well as new ones.[32–34] Language assessment for a bilingual child needs to be carried out in both languages and across several communication situations, because each language may be used with different interaction styles with different family and community members depending on the language being spoken. Bilingual children may also have differential mastery of syntax across their two languages. In the case of vocabulary assessment in particular, the tester needs to determine the referents for which the child has an appropriate label in either language, as well as in both languages.[33] Research has shown that bilingual children acquire both overlapping and nonoverlapping lexicons in the languages they are learning.[33,35,36] Assessment in only one of the child's languages therefore underestimates their lexical knowledge. Measures of the number of concepts for which the bilingual child has a label in either language (termed 'conceptual scoring'[33,35]) seems to more closely approximate monolingual vocabulary acquisition norms than measures of vocabulary achievement in one of their languages.

Standardized tests developed for English cannot simply be translated into another language, for example, Spanish; separate materials and instruments need to be developed specifically for the language they are intended to assess.[34] The complexity of different syntactic forms and constructions, as well as what gets obligatorily marked, varies across languages. Thus, profiles of strengths and weaknesses may differ for a given child across languages even if he/she is a balanced bilingual with roughly equal linguistic competence in the two languages.

Most importantly, a clinician or researcher carrying out a culturally sensitive assessment of language acquisition must have both knowledge and positive attitudes about different dialects and cultural practices. The testing environment must be one in which the child feels comfortable to code switch between dialects or between languages, or to mix them, so that she reveals her complete linguistic repertoire. In the case of a bilingual assessment, the tester must either have near native competence in both languages, or be able to work with an interpreter and appropriate assessment materials and procedures to evaluate the child's mastery and functional use of both languages.

HOW SHOULD LANGUAGE ACQUISITION BE ASSESSED?

Formal Versus Informal Assessment Techniques

A distinction is often made between *formal* and *informal* methods of language assessment. Formal assessment instruments consist of structured sets of items designed to elicit production or to test the comprehension of a particular sentence structure, morpheme, or word. They are usually supported by pictures or by the manipulation of objects to act out the meaning of an utterance. Formal assessment instruments for English are mostly standardized on a representative sample of the UK or US population and provide norms in the form of standard scores for language domains (syntax, semantics, phonology, and/or pragmatics), expressive versus receptive language, or an overall language quotient. If norms are provided for language domains or for specific language subtest scores, a profile of a child's strengths and weaknesses can also be derived.

Informal or 'alternative' assessment refers to the collection and analysis of more open-ended spontaneous language samples. These assessment methods use more authentic (real-life) tasks and sample the child's communicative behaviors in naturalistic settings.[37] The fundamental requirement for an adequate language sample is that it be representative of the child's communication skills in natural settings—in function, in content, and in complexity. To ensure that samples can be taken across communicators (tester, parent/caregiver, or peers) and across topics and situations (e.g., talking about personal experiences, playing a game or performing a cooperative task, describing an activity or a location, telling a story). The child may be observed in several settings to see if the language sample collected in a clinical setting is representative of their language skills, or checked against teacher or caregiver's reports about the child's communicative skills at home or at school. In the case of children from diverse cultural and linguistic backgrounds, it is also important to consider cultural practices that may govern patterns of conversation for the child, for example, the extent to which the child should disagree with or question adults, and so on. Expectations for talking and interacting with adults versus peers often differ across different cultural communities and that could affect language samples.[32,34,38]

Often, the term 'informal' assessment is something of a misnomer because it typically involves quite controlled elicitation procedures and quantitative as well as qualitative measures of behavior/performance. To ensure that a range of language forms and functions are sampled, there is often a protocol to guide the topics and discourse genres elicited by the tester. For example, with preschoolers, there might be free conversation around a set of toys, pictures for the child to describe or tell a story about, and questions about both events and personal experiences to generate both narrative and expository discourse (e.g., 'Tell me about a time when you got into trouble.' Or 'What is your favorite sandwich. Tell me how to make it.'). The samples are recorded on video or audiotape, and then transcribed and subjected to a variety of syntactic, lexical, and functional analyses depending on the purpose of the assessment. There are several systems of analysis available in the literature that have been shown to be discriminating of children with language impairment[37,39] and some of these can now be calculated by computer programs once some basic coding of sentence structures has been carried out by hand and entered into a database.[40,41] Most of these analysis systems focus on morphosyntax production and accuracy (e.g., past tenses, plurals, definite and indefinite articles, etc.); the range of sentence construction types produced (e.g., wh-questions, coordinate sentences, complement clauses, passive voice sentences, relative clauses, etc.); the phrasal and clausal complexity of the sentences produced; vocabulary variety across the sample; or analyses of the pragmatic functions that the speech serves (e.g., questions asked and answered, negatives produced, etc.). Most systems for the analysis of language samples are not standardized to provide population norms by age like those given by formal language tests. Instead, they provide information about the differential difficulty and stages of acquisition of the different sentence types being measured, or criterion-referenced guidelines about what a typical 3, 4, 5, or 6 year old looks like on these measures based on published research.[37,42] The child being assessed is then compared with the characteristic pattern of development of typically developing children of their age in each of the areas of language being analyzed. An exception is the Systematic Analysis of Language Transcripts (SALT) system,[41] which now provides age norms for monolingual English speakers and bilingual English/Spanish speakers on some basic measures of language such as mean length of utterance and number of different words per 50 utterances.

General Strengths and Weaknesses of Formal and Informal Language Assessment

Formal language tests have the advantage of controlling both the language input to the child and to a significant extent the output from the child. The former allows closer comparison across children because all are exposed to the same eliciting conditions. The

latter greatly simplifies scoring and analysis of the child's responses because there is at most a small set of correct or appropriate answers for each test item. However, the cost of that control is typically a limiting of the pragmatic appropriateness or communicative authenticity of the language. Many of the most widely used tests of language use sentence repetition (either immediate or delayed imitation) or completion of sentences begun by the examiner to assess expressive syntax. Similarly, language comprehension is assessed by the child choosing a picture that depicts the closest meaning to a sentence uttered by the tester in isolation from any discourse context. Each of these procedures lacks ecological validity relative to the child's use and understanding of the language in the context of real conversation. Imitation and sentence repetition tasks interact with the child's speech perception and auditory working memory and do not directly measure language formulation and appropriate usage. Thus, they can over- or under-estimate functional language level. Indeed, some children with differential pragmatic impairments may pass formal tests of syntax and semantics by solving the 'problem' set by the test items, but not know when and how to use that syntactic and vocabulary knowledge in genuine ongoing discourse.[3] It might be necessary to compare children's contextualized understanding of language with all the scaffolding and support of familiar interactors, routines, and context with their decontextualized understanding of the language in a testing situation, if the goal is in-depth assessment for intervention. Most formal language assessments sample a wide range of grammatical forms but at a very shallow depth, that is, they test very few examples of any given construction type or syntactic rule. This makes them more useful for diagnosing a general language impairment than for determining profiles of language skills that can be used for intervention or educational planning, or for most research purposes.

Informal assessments of spontaneous language have the strength of a relatively authentic communicative setting and appropriate pragmatic motivation for the language that is produced. This form of assessment is particularly appropriate for assessing children's conversational interaction skills (such as responsiveness to verbal and nonverbal cues for turn taking, and appropriate topic initiation, maintenance or switching) or their narrative skills. The time needed to collect, reliably transcribe, code, and then analyze language samples varies with the elicitation protocol, the intelligibility of the child, and the complexity of the features of language to be analyzed. Less time for transcription is needed if an elicited narrative is a retelling by the child of a story told by the examiner or is based on a picture sequence. In cases where the content is not known, as in personal anecdotes, accurate transcription is more difficult and takes longer. Some basic measures of language development such as mean utterance length in words or number of different words in a sample require little or no coding for computer analysis, but most sophisticated measures of discourse cohesion or syntactic complexity require coding of linguistic features. This process can be labor and time intensive. Furthermore, linguistic analyses of language samples have to be interpreted by skilled practitioners. These are the attendant costs of the language sampling method.

WHAT ARE THE PROPERTIES OF A GOOD ASSESSMENT MEASURE?

Several properties characterize an effective assessment instrument. Some apply at the level of individual items and subtests; others apply to the assessment as a whole. Here, we highlight a few of these that we consider to be most critical (see Kennedy[1] for a more complete review).

First, there is the issue of content validity that we raised above under empirical validation. As far as is possible, items on a valid assessment should measure underlying linguistic competence, not some other performance limitation. There should not be nonlinguistic strategies that allow the child to arrive at the correct answer by guessing, and items should not depend on specific acquired knowledge (such as vocabulary) that requires exposure to particular cultural experiences, unless it is clear that all of the children taking the assessment have had sufficient exposure to those experiences. Thus, items should not be biased against speakers coming from linguistically and culturally different backgrounds. Within these constraints, good items and subtests should produce strong growth curves across the range of ages of the children that the test is assessing and target those aspects of language that are important for the purpose of the assessment.

Predictive Validity

Two measures that are frequently provided to evaluate the predictive validity of an assessment are the *sensitivity* and *specificity* of the measure.[1] *Sensitivity* refers to the extent to which some diagnostic criterion for identifying a child as having a language disorder or delay (usually 1 to 1.5 sd below the mean standard score on the measure) picks out the children who on *a priori* grounds (an existing diagnosis by a clinician using 'best practices') are determined to be language

impaired, that is, a test is sensitive if it identifies as language impaired a very high percentage of the children who are language impaired (true positives). *Specificity* refers to the extent to which a test correctly determines that a typically developing child does not have a language disorder or delay (true negatives). There is a trade-off between the criteria that are set for diagnosing impairment with each of these measures. Applying a stricter criterion for impairment (e.g., increasing the cut off from 1 to 1.5 sd below the mean standard score) increases the specificity of the test but at the cost of reduced sensitivity. Applying a weaker criterion increases sensitivity, but reduces specificity. The best criterion to use depends on the costs (to the child and family as well as to the community) of over- or under-diagnosing language impairment. In addition, the measures of sensitivity and specificity interact with the underlying percentage of language-impaired children in the sample. The smaller the proportion of *a priori* impaired children in the sample of children being assessed, the harder it is to detect them and the lower the sensitivity and specificity on a test for the same cut off criterion. Normed assessment instruments usually provide sensitivity and specificity scores for different cut off criteria (typically 1.0, 1.25 and 1.5 sd below the mean standard score) and for different proportions of *a priori* language-impaired children. Clinical assessment for purposes of diagnosis of impairment is usually conducted on groups of children already suspected of having a language impairment on the basis of observed language difficulties at home and at school, so the proportions of the children being assessed who are likely to have an actual disorder are considerably higher than the general population.

High sensitivity and specificity sound like excellent properties for an assessment instrument to have, but it is not nearly as simple as that. Both of those measures use the 'gold standard' of some *a priori* determination of the children as language impaired or typically developing as the categorization to be predicted. But that makes the measure of sensitivity and specificity only as good as that 'gold standard'. If there are reasons to question the accuracy of the *a priori* categorization of the children being used in the measures, then determination of the sensitivity and specificity of a new test is compromised. Typically, the children in the norming samples for a new test have been diagnosed by SLPs on the basis of existing older tests of syntax or vocabulary, so the 'gold standard' becomes determined by the most widely used instruments. Why is this problematic? First, it places a restriction on the development of new assessment instruments. Tests that measure the same things in the same way as widely used older tests will correlate more highly with the outcomes of those tests and so with the 'gold standard'. Thus, it is not surprising that several of the most widely used language assessment instruments look almost the same, measuring the same aspects of syntax and common vocabulary achievement and using the same techniques of sentence repetition and sentence completion in production and picture choice in comprehension.

More importantly, if there are reasons to question the validity of prior tests on theoretical grounds, or if those tests assess aspects of syntax or vocabulary or pragmatics that are culture or dialect specific, then the 'gold standard' (the *a priori* classification of the children) becomes tarnished. This is particularly relevant when children may be getting misdiagnosed (either over- or under-diagnosed or both) because their dialect differs from the mainstream dialect. A new test that is not biased against those children and does a better job of accurately assessing their underlying language skills actually *should not* correlate too highly with existing, biased tests. However, if clinicians are using the older flawed instruments to determine the *a priori* language status of these populations, then the sensitivity and specificity of a new unbiased test measured in the traditional way could be compromised. In these circumstances, the content validity of the new instrument, its relationship to empirical research on language acquisition, becomes a crucial index of validity.

Consider an example of this issue in practice. When the DELV-NR was developed to provide an unbiased test of the language acquisition of children who are speakers of AAE and similar dialects, language samples were collected by certified SLPs from 78 5- and 6-year old African American children as part of the validation of the test. Twenty of the children had been diagnosed by speech and language clinicians as LI on the basis of previous testing and were receiving language intervention services. Language samples (mean length = 175 utterances) were elicited according to a structured protocol designed to produce narratives based on a picture sequence and on personally experienced life events, expository discourse in the form of explanations and directions, and some general conversation around pictures and problem solving activities.

Two different types of analyses were performed on the language samples. First, the transcripts were coded and entered into Computerized Profiling[40] and SALT[41] for analyses of a variety of quantitative aspects of the transcripts. These programs provided measures of mean length of utterance (MLU) in words

and morphemes, variety of vocabulary (number of different words per 50 utterances), and mean clausal complexity across all the utterances. In addition, Scarborough's[39] Index of Productive Syntax (IPSyn) was calculated for each sample. Second, more detailed syntactic and discourse analyses were carried out on the language samples. All uses of the complex syntax characteristic of written text ('literate language forms')[43,44] were coded. This included adverbial clauses of cause and time, complement clauses with verbs of communication and mental state, and relative clauses. The fluency (amount of prompting needed) and coherence of the children's narratives and expository discourse was also rated.

All of the children also completed all the DELV-NR syntax, semantics and pragmatics subtests, which provided a total language standard score. Pearson et al.[45] took a criterion of >1.5 sd below the normative mean on the DELV-NR as the cut off indicating likely language impairment. Twenty-one of the children were picked out by the DELV-NR as presenting with language impairment according to this criterion. Most importantly, seven children who had been diagnosed by the SLPs as having LI passed the DELV-NR, whereas eight children who were not receiving language intervention services fell below this 1.5sd DELV-NR cut off for identifying language problems. A central question addressed by the study was: For the 15 children for which there was a discrepancy between their scores on the DELV test and their *a priori* clinical classification by the testing SLPs, did the language sample analyses support the DELV or the clinician's diagnosis?

The results showed clear support for the DELV-NR from the language sample analyses. First, there were significant correlations between all the children's DELV-NR scores and the language sample measures (Table 1).

Second, the 21 children who fell below the DELV-NR cut off score produced significantly lower scores than the other 57 children on both the quantitative language sample measures (especially on variety of vocabulary and measures of sentence complexity) and on the literate language and discourse analysis (Table 2).

Across the same measures there were fewer significant differences between the 20 *a priori* LI children (as designated by the SLPs) and the other TD children in the sample (Table 3).

In short, for the 15 children for whom there was a discrepancy between the SLP and DELV classification, the profiles of language strengths and weaknesses derived from the language samples supported the DELV classification in 13 cases (86.7%). This finding raises questions about the application of the *a priori* 'gold standard' to determine the sensitivity and specificity of new assessment instruments, especially when there are concerns about the accuracy of existing methods of diagnosis for particular populations of children. It also suggests that the content and construct validity of a test and its connection with current research on language acquisition are paramount.

NEW DIRECTIONS FOR ASSESSMENT

With these principles in mind, there are several recent developments in the study of language acquisition that we think are likely to influence the creation of the next generation of assessment instruments.

Assessing the Process of Language Acquisition Rather than Static Achievement

Vocabulary Assessment

As we pointed out above, acquired vocabulary is particularly subject to cultural as well as linguistic variation across dialects of English and across language communities because it depends on exposure to words for objects in particular domains of experience (e.g., names for animals, foodstuffs, tools, etc.). Early verb learning may vary less across language and cultural groups because the actions and activities engaged in with children may be more common, although verbs may still display significant cross-linguistic differences in the meanings they encode.[46] New assessments of lexical development may therefore focus more on verb learning than on nouns and set out to assess children's ability to learn new words from context rather than the number of words that they know.

TABLE 1 | Correlations Between the DELV-NR Total Standard Scores and Various Measures of the Children's Language Samples

Number of Different Words/50 Utterances	MLU in Words	LARSP Clausal Complexity	IPSyn Sentence Structure	Literate Language Composite	Extended Discourse Fluency
0.474	0.372	0.322	0.266	0.368	0.497
$p = 0.000$	$p = .001$	$p = 0.004$	$p = 0.019$	$p = 0.001$	$p = 0.000$

TABLE 2 | Comparison of Several Language Sample Measures for TD Versus LI Children Diagnosed According to the DELV-NR Total Standard Scores

	TD Mean	LI Mean	*F*	*df*	*p*	eta^2
Words/50 utterances	107.0	81.6	17.24	1,75	0.000***	0.187
MLU in words	4.48	3.52	11.85	1,75	0.001***	0.136
LARSP clausal complexity	2.77	2.29	9.09	1,75	0.004**	0.108
IPSyn sentence structure	29.3	25.9	5.01	1,75	0.028*	0.063
Literate language composite	7.28	5.10	8.97	1,75	0.004**	0.107
Discourse fluency	2.53	1.74	23.32	1,75	0.000***	0.237

Analysis of covariance (ANCOVA) analyses with clinical status as a fixed effect and age as a covariate.

TABLE 3 | Comparison of Several Language Sample Measures for TD Versus LI Children Diagnosed by the Speech and Language Clinicians (*A priori* Categorization)

	TD Mean	LI Mean	*F*	*df*	*p*	eta^2
Words/50 utterances	104.1	88.7	6.10	1,75	0.016*	0.075
MLU in words	4.38	3.75	6.48	1,75	0.025*	0.065
LARSP clausal complexity	2.73	2.40	3.85	1,75	0.054	0.049
IPSyn sentence structure	28.6	27.8	0.35	1,75	0.556	0.005
Literate language composite	6.9	6.10	0.86	1,75	0.357	0.011
Discourse fluency	2.45	1.93	18.68	1,75	0.001***	0.132

ANCOVA analyses with clinical status as a fixed effect and age as a covariate.

This procedure forms the basis of the DELV-NR word learning subtest.[5,47,48] The subtest assesses children's ability to 'fast map' the meaning of a word from the linguistic context in which it appears. Most of children's vocabulary after the first year or two of language acquisition is not acquired through caregivers pointing out and labeling referents for the child. In fact, most words are learned from the conversational context in which they are embedded. Children must use the sentence structure as well as the referential situation and communicative context to determine what part of speech the word might be as well as some of its semantic properties. Normally developing children between the ages of three and five seem to be able to do this from very few exposures to a word in context, hence the notion of 'fast mapping' of at least partial word meanings.[49] Linguistic context is particularly significant for the learning of verb meanings. Gleitman[50] coined the term 'syntactic bootstrapping' to characterize the process of narrowing down the likely meaning of a verb by the sentence structures in which it occurs. Children with language impairment seem to have special difficulties in the process of fast mapping new meanings from linguistic context,[51] and they are prone to have particularly limited verb lexicons relative to their knowledge of nouns.[52]

The fast mapping test on the DELV shows the children a sequence of three pictures of two characters performing unusual actions. The events are then described in a single sentence containing a novel (nonsense) verb. The stimulus sentences vary as to the type of argument structure provided, with transitive, intransitive, dative, and complement structure frames. Each picture sequence depicts at least two actions, sometimes involving action of one character on another, sometimes not. Hence, without the syntax of the stimulus sentence being taken into account, they are referentially ambiguous. Following the stimulus sentence, the children are asked a series of questions to see what meaning they attribute to the unknown verb.

The data from the standardization of the DELV showed that the argument structures and the test questions varied in their difficulty for the children. Transitive/intransitive frames were easier than dative ones, and the latter were much easier than the complement structures, although they include the same number of arguments of the verb. This allowed the task to work well as a semantic assessment for 8 and 9 year olds as well as for younger children. There was steady developmental growth across the entire age range of 4 through 9, and no significant dialect differences were observed at any of the ages.

However, the task was clearly discriminating of language impairment.[47,48]

Syntax Acquisition Training Studies

In the study of syntax acquisition, several researchers have successfully used training studies to understand children's acquisition of syntactic forms that are of low frequency or not properly mastered yet in the child's speech. Two examples are teaching the passive voice[53] and sentence complements under verbs of communication.[54] After modeling and providing corrective feedback with a limited set of exemplars, the child is then asked to generalize the structure to describe new referential materials and events. The generalizations that the children do and do not make provide evidence for their representation of the syntactic form, for example, whether it has inherent semantic or lexical restrictions. These training and generalization procedures may provide a more dynamic assessment of children's language acquisition, focused on the process of acquisition and the child's language learning ability rather than on specific acquired knowledge.

Dynamic Assessment

Dynamic assessment is a related case, involving interactive procedures during language assessment in which different amounts of scaffolding and feedback are given to the child to gauge the child's responsiveness to intervention and their potential for change. In general, the sequence is to assess performance on some linguistic task (e.g., narrative production), then to teach or facilitate some linguistic skills involved in that task (e.g., narrative production with supportive prompts and feedback), and then to assess independent performance again.[55,56] The procedure provides information about how much the child can self-correct or respond to prompts or other supportive scaffolding of the task to change their performance. In this way, it makes a more direct link between the assessment and intervention phases of treatment for children with language impairments or delays. Because it focuses on language learning ability rather than achieved linguistic knowledge, dynamic assessment may also provide an unbiased evaluation of the language acquisition of children from different cultural and linguistic backgrounds.[56–58]

New Techniques to Assess Language Production and Comprehension

Language Production

In an earlier section of this review, we highlighted some of the strengths and weaknesses of imitation and sentence completion tasks widely used in formal tests of syntax production and of much more open-ended, informal language sample collection and analysis. Future directions in the development of effective assessment procedures for language production may include more use of what might be termed 'semi-formal' *elicited production* tasks that structure the interactive situation to constrain the form and meaning to be produced.[59,60] Such procedures include referential communication (barrier) games to elicit locative terms, adjectives, prepositional phrases, and relative clauses.[59] A promising technique that has been used in the research on language acquisition to elicit questions and relative clauses requires the child to address another character (a puppet or second experimenter) to ask for information and to pass on information.[60] The goal of these techniques is to preserve the pragmatic motivation and communicative components of the task (a strength of language sampling approaches), while limiting the range of appropriate responses for each item so that they can be easily scored at the time of testing (an advantage of formal testing).

Language Comprehension

Comprehension of syntactic structures or word meanings is typically assessed by requiring children to perform an action, to pick out one from a set of pictures, to point to an object, to act out some event, or to respond verbally. However, failure to respond appropriately can occur for several reasons, including issues of compliance and motor skills, as well as lack of language comprehension. The past 20 years has seen a huge increase in the measurement of eye gaze and the preferential looking paradigm[61–63] to study semantic and syntactic comprehension in very young typically developing children. These procedures show children's discrimination of semantic and grammatical differences prior to them being able to perform differential actions or respond verbally in response to verbal input. Hence, application of preferential looking measures to assessment of children with suspected disorders may provide a more sensitive assessment of prerequisite language processing or early comprehension of syntactic and semantic features. Swenson, Kelley, Fein, and Naigles[64] have successfully used intermodal preferential looking measures to assess aspects of the early syntax acquisition of children with ASD.

CONCLUSION

In conducting an effective assessment of language acquisition, a major consideration should be what

the testing is for, because that dictates the depth and the breadth required. In our view, content validity, linked to theories and empirical findings about both typical children and those with impaired language, is paramount. The test items need to reveal significant clues about the nature of the child's grammar, or word learning, or communicative functioning. Predictive validity linked to existing tests is a questionable standard if new assessments are to make progress over old, especially when issues of linguistic or cultural bias are suspected. Both formal and informal methods have benefits and costs, between ease and ecological validity, but each may be suited to different domains of language. The most informative assessments will combine both norm-referenced testing and sampling of the child's language in conversation and extended discourse. New procedures that come closer to mimicking the demands of the process of language learning are worth further exploration.

REFERENCES

1. Kennedy M. Principles of assessment. In: Paul R, Cascella PW, eds. *Introduction to Clinical Methods in Communication Disorders*. Baltimore (MD): Paul H. Brooks Publishing; 2007, 39–83.
2. Tager-Flusberg H, Paul R, Lord C. Language and communication in autism. In: Volkmar F, Paul R, Klin A, Cohen DJ, eds. *Handbook of Autism and Pervasive Developmental Disorders*. 3rd ed. New York, NY: John Wiley & Sons; 2005.
3. Paul R. *Language Disorders from Infancy Through Adolescence: Assessment and Intervention*. 3rd ed. St. Louis, MO: Elsevier; 2007.
4. Rice ML, Wexler K. *Rice/Wexler Test of Early Grammatical Impairment*. San Antonio, TX: The Psychological Corporation; 2001.
5. Seymour H, Roeper T, de Villiers JG, de Villiers PA. *Diagnostic Evaluation of Language Variation—Norm-Referenced (DELV-NR)*. San Antonio, TX: The Psychological Corporation; 2005.
6. Rice ML, Warren SF, eds. *Developmental Language Disorders: From Phenotypes to Etiologies*. Mahwah, NJ: Erlbaum; 2004.
7. Mervis CB, Robinson BF. Designing measures for profiling and genotype/phenotype studies of individuals with genetic syndromes or developmental language disorders. *Appl Psycholinguistics* 2005, 26:41–64.
8. Dunn LM, Dunn LM. *Peabody Picture Vocabulary Test*. 4th ed. Circle Pines, MN: American Guidance Service; 2006.
9. Rice ML, Wexler K. Toward tense as a clinical marker of specific language impairment in English-speaking children. *J Speech Lang Hear Res* 1996, 39:1239–1257.
10. Roeper T. *The Prism of Grammar*. Cambridge, MA: MIT Press; 2007.
11. de Villiers JG, Roeper T, Bland-Stewart L, Pearson B. Answering hard questions: wh-movement across dialects and disorder. *Appl Psycholinguistics* 2008, 29:67–103.
12. Snyder W. *Child Language: The Parametric Approach*. New York, NY: Oxford University Press; 2007.
13. Tomasello M. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press; 2003.
14. Tomasello M. Acquiring linguistic constructions. In: Siegler R, Kuhn D, eds. *Handbook of Child Psychology: Cognition, Perception, and Language*. 6th ed., Vol. 2. New York, NY: John Wiley & Sons; 2006, 255–298.
15. Leonard L. *Children with Specific Language Impairment*. Cambridge, MA: MIT Press; 1998.
16. Rice ML, Wexler K, Cleave P. Specific language impairment as a period of extended optional infinitive. *J Speech Lang Hear Res* 1995, 38:850–863.
17. Carrow-Woolfolk E. *Comprehensive Assessment of Spoken Language (CASL)*. Circle Pines, MN: AGS Publishing; 1999.
18. Semel E, Wiig E, Secord W. *Clinical Evaluation of Language Fundamentals*. 4th ed. (CELF-4). San Antonio, TX: PsychCorp; 2003.
19. Hresko WP, Reid K, Hammill D. *Test of Early Language Development*. 3rd ed. (TELD-3). San Antonio, TX: PsychCorp; 1999.
20. Hammill D, Newcomer P. *Test of Language Development: Intermediate*. 4th ed. (TOLD:I-4). Austin, TX: ProEd; 2007.
21. van der Lely H. Domain-specific cognitive systems: insight from Grammatical-SLI. *Trends Cogn Sci* 2005, 9:53–59.
22. van der Lely H, Battell J. Wh-movement in children with grammatical SLI: a test of the RDDR hypothesis. *Language* 2003, 79:153–181.
23. Campbell T, Dollaghan C, Needleman H, Janosky J. Reducing bias in language assessment: processing-dependent measures. *J Speech Hear Res* 1997, 40:519–525.

24. Dollaghan C, Campbell T. Nonword repetition and child language impairment. *J Speech Lang Hear Res* 1998, 41:1136–1146.
25. Seymour H. The challenge of language assessment for African American English-speaking children: a historical perspective. *Semin Speech Lang* 2004, 25:1–12.
26. Seymour H, Bland-Stewart L, Green L. Difference versus deficit in child African American English. *Lang Speech Hear Serv Sch* 1998, 29:96–108.
27. Seymour H, Pearson BZ. Steps in designing and implementing an innovative assessment instrument. *Semin Speech Lang* 2004, 25:27–32.
28. Taylor O, Payne K. Language and communication differences. In: Shames G, Wiig E, Secord W, eds. *Human Communication Disorders: An Introduction*. 4th ed. New York, NY: Merrill; 1994, 136–173.
29. Wyatt T. Assessing the communicative abilities of clients from diverse cultural and language backgrounds. In: Battle D, ed. *Communication Disorders in Multicultural Populations*. Boston, MA: Butterworth-Heinneman; 2002, 415–459.
30. Seymour H, Seymour C. A therapeutic model for communication disorders among Black English speaking children. *J Speech Hear Disord* 1977, 42:247–256.
31. Seymour H, Roeper T, de Villiers JG. *Diagnostic Evaluation of Language Variation—Screener (DELV-SC)*. San Antonio, TX: The Psychological Corporation; 2003.
32. Goldstein B, Iglesias A. Issues of cultural and linguistic diversity. In: Paul R, Cascella PW, eds. *Introduction to Clinical Methods in Communication Disorders*. Baltimore, MD: Paul H. Brooks Publishing; 2007, 283–301.
33. Pearson BZ. Assessing lexical development in bilingual babies and toddlers. *Int J Biling* 1998, 2:347–372.
34. Peña E. Lost in translation: methodological considerations in cross-cultural research. *Child Dev* 2007, 78:1255–1264.
35. Peña E, Bedore L, Zlatic-Giunta R. Category generation performance of young bilingual children. *J Speech Lang Hear Res* 2002, 45:938–947.
36. Bedore L, Peña E, Garcia M, Cortez C. Conceptual versus monolingual scoring: when does it make a difference. *Speech Lang Hear Serv Sch* 2005, 36:188–200.
37. Paul R, Tetnowski J, Reuler E. Communication sampling procedures. In: Paul R, Cascella PW, eds. *Introduction to Clinical Methods in Communication Disorders*. Baltimore, MD: Paul H. Brooks Publishing; 2007, 111–154.
38. Stockman I. The promises and pitfalls of language sample analysis as an assessment tool for linguistic minority children. *Lang Speech Hear Serv Sch* 1996, 27:355–366.
39. Scarborough H. Index of productive syntax. *Appl Psycholinguistics* 1990, 11:1–22.
40. Long S, Fey M, Channel R. *Computerized Profiling (CP: Version 9.26)*. Cleveland, OH: Case Western Reserve University; 2000.
41. Miller J, Iglesias A. *Systematic Analysis of Language Transcripts (SALT), V8*. Madison, WI: Language Analysis Laboratory, University of Wisconsin; 2005.
42. Miller J. *Assessing Language Production in Children: Experimental Procedures*. Boston, MA: Allyn & Bacon; 1981.
43. Perera K. Language acquisition and writing. In: Fletcher P, Garman M, eds. *Language Acquisition*. 2nd ed. Cambridge: Cambridge University Press; 1986.
44. Westby C. Learning to talk, talking to learn: oral-literate language differences. In: Simon C, ed. *Communication Skills and Classroom Success*. San Diego, CA: College-Hill Press; 1991.
45. Pearson BZ, Magaziner K, Sunderland K, de Villiers PA. *Assessing language growth and delay between 5 and 7*. In: *Poster Presented at: International Association for the Study of Child Language Conference*; 2008 July 28—Aug 1; Edinburgh.
46. Hohenstein J, Eisenberg A, Naigles L. Is he floating across or crossing afloat? Cross-influence of L1 and L2 in Spanish-English bilingual adults. *Biling Lang Cogn* 2006, 9:249–261.
47. de Villiers JG. Cultural and linguistic fairness in the assessment of semantics. *Semin Speech Lang* 2004, 25:73–90.
48. Johnson V, de Villiers JG. Syntactic frames in fast mapping verbs: the effects of age, dialect and clinical status. *J Speech Lang Hear Res*. 2009, 52:610–622.
49. Dollaghan C. Child meets word: "fast mapping" in preschool children. *J Speech Hear Res* 1985, 28:449–454.
50. Gleitman L. The structural sources of verb meaning. *Lang Acquis* 1990, 1:1–17.
51. Rice ML, Buhr JC, Nemeth M. Fast mapping word-learning abilities of language delayed preschoolers. *J Speech Hear Disord* 1990, 55:33–42.
52. Rice ML, Bode JV. GAPS in the verb lexicons of children with specific language impairment. *First Lang* 1993, 13:113–131.
53. de Villiers JG. The process of rule-learning in child speech—a new look. In: Nelson K, ed. *Child Language*, Vol. 2. New York, NY: Gardner Press; 1980.
54. Hale CM, Tager-Flusberg H. The influence of language on theory of mind: a training study. *Dev Sci* 2003, 6:346–359.
55. Peña E. Dynamic assessment: the model and its language applications. In: Cole KN, Dale PS, Thal DJ, eds. *Communication and Language Intervention Series: Vol. 6. Assessment of Communication and Language*. Baltimore, MD: Paul H. Brooks Publishing Company; 1996, 281–307.

56. Peña E, Gillam R, Malek M, Ruiz-Felter R, Resendiz M, et al. Dynamic assessment of children from culturally diverse backgrounds: applications to narrative assessment. *J Speech Lang Hear Res* 2006, 49:1–21.
57. Lidz CS, Peña ED. Dynamic assessment: the model, its relevance as a non-biased approach, and its application to Latino American preschool children. *Lang Speech Hear Serv Sch* 1996, 27:367–372.
58. Peña E, Iglesias A, Lidz CS. Reducing test bias through dynamic assessment of children's word learning ability. *Am J Speech Lang Pathol* 2001, 10:138–154.
59. de Villiers PA. Assessing pragmatic skills in elicited production. *Semin Speech Lang* 2004, 25:57–71.
60. Thornton R. Elicited production. In: McDaniel D, McKee C, Cairns HS, eds. *Methods for Assessing Children's Syntax*. Cambridge, MA: MIT Press; 1996, 77–102.
61. Fernald A, Pinto J, Swingley D, Weinberg A, McRoberts G. Rapid gains in speed of verbal processing by infants in the second year. *Psychol Sci* 1998, 9:228–231.
62. Hirsh-Pasek K, Golinkoff RM. The intermodal preferential looking paradigm: a window onto emerging language comprehension. In: McDaniel D, McKee C, Cairns HS, eds. *Methods for Assessing Children's Syntax*. Cambridge, MA: MIT Press; 1996, 105–124.
63. Naigles L. Children use syntax to learn verb meanings. *J Child Lang* 1990, 17:357–374.
64. Swenson L, Kelley E, Fein D, Naigles L. Children with autism display typical language learning characteristics: evidence from preferential looking. *Child Dev* 2007, 78:542–557.

FURTHER READING

Iglesias A. Communicative assessment of the Hispanic child. *Semin Speech Lang* 2001, 22: 3–16.
Losardo A, Notari-Syerson A. *Alternative Approaches to Assessing Young Children*. Baltimore, MD: Paul H. Brooks Publishing; 2001.
Paul R. *Language Disorders from Infancy Through Adolescence: Assessment and Intervention*. 3rd ed. St Louis, MO: Mosby Elsevier; 2007.
Paul R, Cascella PW, eds. *Introduction to Clinical Methods in Communication Disorders*. 2nd ed. Baltimore, MD: Paul H. Brooks Publishing; 2007.
Seymour H, Pearson BZ, eds. Evaluating language variation: distinguishing development and dialect from disorder. *Semin Speech Lang* 2004, 25.

