ASSESSING LANGUAGE SKILLS IN YOUNG CHILDREN

Julie Dockrell
Chloë Marshall
Psychology and Human Development,
Institute of Education,
London

ACCEPTED VERSION OF PAPER PUBLISHED 2015

Background: Language and communication skills are central to children’s ability to engage in social relationships and access learning experiences. This paper identifies issues which practitioners and researchers should consider when assessing language skills. A range of current language assessments is reviewed. Key findings: Current screening measures do not meet psychometric prerequisites to identify language problems. There are significant challenges in the interpretation of language assessments, where socioeconomic status, language status and dialect, hearing impairment and test characteristics impact results. Conclusions: Psychometrically sound assessments of language are an essential component of developing effective and efficient interventions. The language trajectories of preschool children vary substantially; current screening measures have significant limitations. Composite measures of language performance are better indicators of language problems and disorders than single measures of component skills.

Key Practitioner Message
- Language is a complex system to assess, comprising a range of subsystems
- Regular monitoring of language is preferable; one-off screenings have limited power to predict later performance because children’s developmental trajectories vary
- Composite language measures provide more reliable and valid assessments of children’s language skills
- Many assessment tools are not suitable for the range of preschool children who experience language delays and problems
- Dynamic assessment is more culturally fair to those from different linguistic or cultural backgrounds and may be more sensitive for measuring change in language over time
1. Introduction

In the 13 years since the journal published a paper reviewing approaches to the assessment of preschool children's language skills (Dockrell, 2001), there have been significant advances in our understanding of language acquisition and the ways in which language impacts on later development. A much greater awareness now exists that language development reflects an interaction between the intrinsic capacities of the child and the context in which he is developing (Thomas, 2010). Even within the language system itself, complex interplays between the subcomponents of the language system have been identified (Dixon & Marchman, 2007; Tomblin & Zhang, 2006). Through a series of different longitudinal studies, the importance of early oracy skills in scaffolding learning in schools and supporting social relationships has been further articulated, both for typically developing children and those who experience developmental challenges (Catts, Bridges, Little, & Tomblin, 2008; Conti-Ramsden, Durkin, Simkin, & Knox, 2009; Dockrell, Lindsay, & Palikara, 2011; Lindsay & Dockrell, 2012; Schoon, Parsons, Rush, & Law, 2010; St Clair, Pickles, Durkin, & Conti-Ramsden, 2011; Tomblin, Zhang, Buckwalter, & Catts, 2000; Wellman et al., 2011). In tandem, there has been renewed concern about the oral language skills that children possess at school entry (Bercow, 2008), and an explicit realisation that in comparison to their peers some children will struggle to acquire oral language and this will have a long term impact on learning and attainments (Young et al., 2002).

These converging developments have highlighted the importance of early intervention in educational policy and practice (Allen, 2011) and the need for
practitioners to develop their skills in supporting children’s oral language competencies (McCartney & Ellis, 2013). To achieve these goals it is important that psychologists, speech and language therapists, specialist teachers and researchers are able to assess language competencies and understand the strengths and limitations of the tools they use. Moreover, effective and targeted interventions and the ability to monitor progress require tools that are reliable, valid and fit for purpose.

Researchers have risen to this challenge and over the recent years these issues have been the focus of a number of specific test reviews (Friberg, 2010; Hoffman, Loeb, Brandel, & Gillam, 2011; Law & Roy, 2008; Spaulding, Plante, & Farinella, 2006) and a key textbook focusing on the assessment of language disorders for speech and language therapists has been produced (McCauley, 2001). We do not aim to replicate or attempt to supersede these reviews. Instead our aim is to complement them by raising key issues which practitioners and researchers should consider when they examine young children’s language skills. In addition to problems with test interpretation per se, new challenges in understanding language performance are explored. We begin by considering the key dimensions of the language system and identifying factors which can impact on language development and the assessment process. We then provide an overview of areas to consider when assessing language competence and conclude by highlighting new developments and approaches to assessing preschool children’s language.

2. THE LANGUAGE SYSTEM
It is important to distinguish between communication and language. Communication is the transmission of information and infants communicate from the minute they are born. Language, in contrast, is primarily a representational system that emerges as the child’s cognitive skills scaffold understanding and organisation of the world. Children are born with a rich set of cognitive skills that support language learning. These include perceptual skills, memory skills, the ability to focus attention, and other reasoning abilities such as the ability to draw analogies and to create and manipulate representations of objects, actions and the minds of others (Moll & Tomasello, 2010). Together these competencies help children to work out the rules of language, identify the meanings of new words and understand the demands of social situations and the views of others. These skills do not work in isolation; language development is a subtle interaction between the contexts in which children develop and the cognitive skills they bring to the language learning enterprise (for overviews, see Brookes & Kempe, 2012; Saxton, 2010).

The language system is itself composed of a number of subcomponents that are important for effective understanding and communication. These include the lexicon (vocabulary), syntax (the rules for combining words into phrases and sentences), morphology (the rules for constructing larger words out of smaller units of meaning), phonology (the sounds that make up words and the rules that combine sounds) and pragmatics (the rules of social communication). These subcomponents work together in a dynamic and developmental fashion. As children master language they produce extended and coherent oral narratives, which allow them to communicate with others effectively and efficiently.
Thinking of language as a dynamic system highlights the ways in which even
minor problems can affect a child’s ability to access the curriculum and
contribute to daily experiences. For example, delays in lexical acquisition can
impact on the fluency of a child’s expressive language, while failures to process
key syntactic components may impact on social interaction. These difficulties can
result in a variety of different developmental trajectories (Leonard, 1997),
although we are a long way from describing and explaining preschool language
trajectories accurately (Eadie et al., 2014; Law & Roy, 2008; Nelson et al., 2006 a,
b).

Models of the language system and the way it develops vary. Some researchers
have argued that much of the variance in standardized measures of language
appears to be attributable to a single common factor, although there is a
developmental trend during middle childhood for grammatical abilities and
vocabulary abilities to become differentiated. The use of tests to reliably
distinguish separate different vocabulary and grammar factors in the early years
is, therefore, problematic from this perspective. (Tomblin & Zhang, 2006). Others
have argued that it is problematic to identify only expressive language problems
as comprehension problems are usually also present but are undetected
(Leonard, 2009). As researchers continue to explore aspects of the language
system new assessments are developed to profile children’s competencies. As we
discuss below, many of these measures have not yet been rigorously evaluated,
but it is likely that practitioners will be exposed to these concepts and measures
and in the future they may offer reliable and valid ways of exploring children’s
language skills.
3. IDENTIFYING LANGUAGE PROBLEMS

Many children experience difficulties with acquiring language. For some children these challenges occur because they have had limited exposure to language learning opportunities, while for others it will be because of difficulties with hearing or vision, and there is also a significant group of children who experience difficulties for no obvious physical or psychological reason.

3.1 Social Disadvantage

There has been a long-standing acknowledgement that poor language skills are associated with social disadvantage (W. B. Dockrell, 1966), and more recent work has enriched our understanding about the extent of this association and the patterns of language difficulties which are shown (Roy & Chiat, 2013). Prevalence rates of language delays in disadvantaged populations are high, but rates of identification are often low (King et al., 2005). Moreover, the poorest outcomes are disproportionately associated with the most socially and economically disadvantaged (Washbrook & Waldfogel, 2010), although these differences are not evident in referrals to speech and language therapy services (Roy & Chiat, 2013). Whether the language problems in such children are qualitatively different from those in children with more specific language disorders remains a matter of debate. However, there is little evidence to suggest that such differences will be evident in the preschool period, although there may be subtle differences in speech processing (Roy & Chiat, 2013).

3.2 Different Dialects and Bilingualism

There is a growing concern that children from ethnic minority groups are over-represented in the caseloads of speech and language therapists and are over-
identified generally as having speech language and communication needs (Dockrell, Lindsay, Roulstone & Law, in press). There is also a growing awareness that non-standard varieties of English differ from the Standard English that language assessments are designed to test. Children should not be viewed as having a speech or language disorder because they speak a variety of English other than the standard dialect. Nevertheless, a proportion of children who speak non-standard dialects will have a developmental language disorder, just as is the case for a proportion of standard dialect speakers, and identifying those children requires appropriate assessments (Seymour, 2004).

In the USA, dialectical variations such as African American English (AAE) have, in the past, often been considered “incorrect”. The Diagnostic Evaluation of Language Variance (DELV; Seymour, Roeper, & de Villiers, 2005) takes both Mainstream American English (MAE) and AAE into account in the range of allowable responses. For example, one item that assesses use of prepositions is, “He’s not climbing with the cat, he’s climbing....” One acceptable response is “by hisself.” This response is not representative of MAE, but it is a response frequently given by typically developing children who speak AAE. Hence the item assesses the target structure, which is inclusion and correct use of the preposition (i.e. by), but does not penalise the child for using a different form of the object pronoun (i.e. himself, hisself). Allowing a range of responses reduces the bias against non-standard dialect speakers that is inherent in many standardised tests.

There has also long been a concern about the impact of bilingualism on children’s language development: bilingual children have been considered at risk
for, variously, impaired language development, impaired cognitive development, risk of academic delay/failure, and social-cultural exclusion (Genesee & Nicoladis, 2009). Professionals’ views of bilingualism are now more positive, and, indeed, the majority of the world’s population is bilingual (Mueller Gathercole, 2013). Nevertheless, children who are learning English as an additional language (EAL) are very heterogeneous with respect to when they learn English and the quantity and quality of English input that they receive, and there is only limited normative data on the developmental trajectory of these learners (Bedore & Peña, 2008).

The heterogeneity of bilingual learners means that it is problematic to assess language in children with EAL using norms derived from monolingual children (Mueller Gathercole, 2013). For a child who scores significantly below age expectations, the question remains as to whether this low score represents evidence for a language disorder or whether it instead reflects that the child has had less exposure to English than the norming population. One way of trying to identify whether the child has a language disorder is to assess him in his other language(s), under the assumption that a language disorder would also show up in the child’s dominant language (Mueller Gathercole, 2013). However, reliable language assessments exist for only a few languages, and even if they do exist, assessors fluent in those languages may not be available. The result is a situation where children from bilingual backgrounds are sometimes over-identified with a language disorder, because assessors have inappropriate developmental expectations, and are sometimes under-identified because assessors wait for
children to learn more English before identifying difficulties (Bedore & Peña, 2008).

### 3.3 Hearing Impairment

Children who experience deafness, and even mild or unilateral hearing impairment, typically experience delays in receptive and expressive language development. Just as we described for children with EAL, test norms for hearing children may not be appropriate when assessing deaf children. In this case, the issue is whether a deaf child achieves low scores because of difficulties accessing spoken language or whether the child has a language learning impairment above and beyond that. Furthermore, it is an empirical issue as to whether tests designed for hearing children are appropriate for assessing language in deaf children: this cannot be assumed.

In an example of the type of study that is needed, Webb and Lederberg (2014) tested 108 American 5-year-old deaf children with cochlear implants or hearing aids using the phonological awareness subtests of the Phonological Awareness Test-2 (Robertson & Salter, 2007) and the Test of Preschool Early Literacy (Lonigan, Wagner, Torgesen, & Rashotte, 2007). Item analyses showed that both tests had good psychometric properties (e.g. high item discriminations and internal consistencies), and that scores on them correlated both concurrently and predictively with early literacy (as is the case for hearing children). Hence although deaf children score low on these two standardised tests, they are still valid for use with deaf children who have functional aided hearing.
Assessments are gradually being developed to assess the sign language abilities of deaf children who sign. The population of deaf signers is small. The majority of deaf children are born to hearing parents and so are not exposed to sign language from birth, but instead learn sign language when they go to school. The incidence of additional educational needs amongst deaf children is high (Consortium for Research in Deaf Education, 2012). Consequently, it is difficult to get together a large enough sample for norming, and the normative sample is typically heterogeneous (Mann, Roy & Marshall, 2013). For example, the most widely used test of British Sign Language (BSL) skills, the BSL Receptive Skills Test, which tests grammatical comprehension using a picture-pointing format, was standardised on only 135 children between the ages of 3 and 13, who included deaf children of deaf parents, hearing children with deaf parents, and selected deaf children from hearing families whose exposure to BSL was known. Nevertheless, the test has robust psychometric properties. The same authors have since developed an assessment of BSL narrative skills using the same principles of standardisation (Herman et al, 2004). A third test of BSL, which targets both receptive and expressive vocabulary, has also been developed, but is not yet standardised (Mann et al, 2013).

3.4 Unexplained difficulties with the language system

There is a large group of children who experience language delays for no obvious reason. Practitioners, policy makers, and researchers in the UK use a range of different terms to describe this population (Dockrell, Lindsay, Letchford, & Mackie, 2006), and different terms are used internationally: in Europe,
dysphasia; in the USA, specific language impairment; and in parts of Canada, dysphasia or primary language disorder (Tomblin et al., 2003). The new DSM-5 uses the term ‘language disorder’. These labels all refer to children who have difficulties with the acquisition and processing of oral language. Two questions have troubled practitioners and researchers: the requirement for children’s language skills to be discrepant from other aspects of development (sometimes called cognitive referencing), and the distinction between delayed and disordered language.

Discrepancy criteria have frequently been used in attempts to identify children with specific language disorder, where a contrast is drawn between language skills and non-verbal ability. The criteria used across research studies and speech and language therapy services vary considerably, and the distinction is problematic both conceptually and practically. At a practical level there are concerns about measurement and the determination of the appropriate formula for the discrepancy (Aram, Morris & Hall, 1992; Plante, 1998). Problems include the methods for measuring the discrepancy, the meaningfulness of the distinction, the accuracy of the normative data being used to conclude that a discrepancy exists and the tests used to establish non-verbal ability (Krassowski & Plante, 1997). Language problems may also impact on children’s performance on non-verbal tasks, thereby affecting assessments of non-verbal ability. The new DSM-5 does not include a discrepancy criterion for language disorders. Further, it is important to note that no differences in response to oral language intervention have been found for children with and without discrepancies between their verbal and non-verbal performance (Bowyer-Crane et al., 2011;
Friel-Patti, 1999). Practitioners may need to place less reliance on simplistic models of discrepancy, and make greater attempts to characterise the child’s performance on different tasks and situations, resulting in a profile of skills and needs.

The delay/disorder distinction continues to permeate the literature. This distinction needs to be contextualized in terms of the aspect of speech, language and communication considered. It is also dependent on an understanding of what are typical patterns of development and typical error patterns. A child would be considered delayed if he were performing in a manner similar to much younger children but disordered if his performance was dissimilar to that found in the typical range of performance. By contrast a child’s language would be described as disordered if he was performing in a way not seen in typical development or the patterns and extent of his language learning needs were severe. Standardized tests (see section below) are frequently used by speech and language therapists and psychologists to identify severity and discrepancies within the language system. This is problematic as cut-off points are arbitrary and unvalidated and there is significant variation between the tests (Spaulding et al., 2006). Moreover, researchers and clinicians should be cautious in determining the severity of children’s language disorder using norm-referenced test performance, given the inconsistency in guidelines and lack of empirical data within test manuals to support this use (Spaulding, Szulga, & Figueroa, 2012). Response to intervention models are an alternative approach to considering levels of severity (see for example the ASHA statement, http://www.asha.org/slp/schools/prof-consult/RtoI/).
4. **WHY ASSESS CHILDREN’S LANGUAGE SKILLS?**

Language skills are often assessed in preschool children, either as part of the curriculum to monitor progress, or through screening. Many of these assessments are carried out by teachers and Special Educational Needs coordinators. With the current drive to enhance early language skills, assessments are also used as pre-intervention and post-intervention measures to evaluate the impact of oral language interventions. Although repeated standardised testing can be a useful way of tracking progress over time, the standardised scores in those situations must be interpreted with caution. Regression to the mean can occur if a score is extreme on its first measurement, it will tend to be closer to the average on its second measurement, and this “improvement” may be down to chance or measurement error rather than to any meaningful improvement in performance (Dockrell & Law, 2007). It is most likely to occur when measures are less reliable and with samples that are selected to have extremely low scores (Zhang & Tomblin, 2003). Regression to the mean is a particular concern in the evaluation of language interventions, where a control group who is not receiving the intervention is required for comparison.

4.1 **Screening**

Identification and assessment are not the same processes (McCauley, 2001). The purpose of identification is to verify the existence of a problem, that is to distinguish between children whose language skills are below expectation and those who are currently performing in the average range. In contrast,
assessment aims to characterise the nature and extent of the child’s difficulties in terms of differing language skills

In theory language problems in children under five can be identified through indirect or direct screening. Indirect screening occurs when a parent or professional notes a problem with a child’s language development which they feel warrants further investigation. By contrast direct screening involves the use of a screening instrument on a population to identify children whose language levels are below expectation. Direct screening for language delay raises methodological challenges (Eriksson, Westerlund, & Miniscalco, 2010). Of key importance are the properties of the screening measure, the measure used as the benchmark for language problems (Van Agt, van der Stege, de Ridder-Sluiter, & de Koning, 2007), and who completes the screening test e.g. parent, professional or teacher (Antoniazzi, Snow, & Dickson-Swift, 2010).

Central to screening are the notions of sensitivity and specificity. A test with high levels of sensitivity accurately identifies children as cases who have language problems, when a benchmark test is used, whereas high levels of specificity means that the measure does not identify as cases children who do not have a language problem. Any measure must meet minimal standards for both sensitivity and specificity, but there will inevitably be a tradeoff between the two, depending on the purpose of the screening. It has been argued that a higher standard should be met for sensitivity, as poor sensitivity might have a more profound effect on the child and future academic progress. However, over-identification of children has resource implications and may cause unnecessary distress.
Many tests do not meet these basic criteria for screening purposes. Even tests that are specifically described as screening tests may not be adequate e.g. CELF (Clinical Evaluation of Language Fundamentals) screening test (Semel, Wiig & Secord, 2004). More generally studies have consistently raised concerns about the ability of screening tests to detect children with concurrent language problems, that is problems at the time of testing (de Koning et al., 2004; Laing, Law, Levin, & Logan, 2002). In Laing et al’s (2002) study, both parent questionnaire and the screen failed to accurately identify those with language problems which led to an over-referral for diagnostic assessments. These conclusions are further supported by systematic reviews (Law, Boyle, Harris, Harkness, & Nye, 2000; Nelson, Nygren, Walker, & Panoscha, 2006a,b).

Nonetheless given the importance of identifying children who struggle with oral language there have been continued attempts to devise psychometrically sound screening devices. For example Greenslade, Plante, and Vance (2009) demonstrated high levels of specificity (100%) and sensitivity (90%) for the Structured Photographic Expressive Language Test-Preschool 2, which assesses a range of language features, and Pesco and O’Neill (2012) reported 93% specificity and 81% sensitivity for the Language Use Inventory for children aged 24 to 47 months. It is noteworthy that both of the measures sample a range of language skills. However, as both studies are North American it cannot be assumed that these results will generalize to other countries and languages.

While we can be reasonably accurate about identifying who is not language delayed there is continued disagreement about who is experiencing a language difficulty (Law et al., 1998). It is, therefore, not surprising that using screening
measures to predict the likelihood of a child experiencing language difficulties in the future is fraught with difficulties. Studies that have attempted this have been unsuccessful in identifying language factors which predict future performance (Law, Rush, Anandan, Cox, & Wood, 2012; Nelson, Nygren, Walker, & Panoscha, 2006a; Wilson, McQuaige, Thompson, & McConnachie, 2013). As Snowling et al (2012) concluded, regular monitoring is preferable because one-off screenings of aspects of development, including language and reading, have limited power to predict later performance because children's developmental trajectories vary.

4.2 ASSESSMENT

Identifying the existence of a problem is the first step in the language assessment process. For children who are identified it is then important to characterise the nature and extent of their difficulties in terms of differing language skills. Measures that meet screening criteria may not be acceptable in profiling a child's difficulties (Merrell & Plante, 1997) so it is important to evaluate the appropriateness of the measure for its intended purpose.

A broad range of information-gathering activities are available to meet this goal. The assessment process itself will be guided by the initial evaluation of the child, the theoretical orientation held by the assessor and practical constraints related to time and resources. Three broad approaches to the assessment of language problems can be identified – standardised tests, dynamic assessments, and parental or teacher checklists. These approaches are not mutually exclusive.

4.2.1 Standardised tests of oral language

Standardized tests are amongst the primary assessment tools used by speech and language therapists and psychologists to diagnose a child’s language
problems. They provide objective information in a standardized setting that allows performance to be related to normative data. Many child language tests are commercially available and these tests broadly fall into two categories, those that aim to provide oral language composite scores (omnibus measures) and those which target specific components of the language system. In this section we will consider these different approaches to assessment and provide details about some commonly used tests. Readers who wish to find out about other tests can consult the Buros test review centre (http://marketplace.unl.edu/buros/)

As with screening measures, it is important to ascertain the reliability and validity of the test and to consider the standardization sample of the test. Tests standardized on specific or limited samples have implications for whom the test should be used with and the appropriateness of generalizing from the results. Construct and concurrent validity of measures may also vary from the normative sample (Hoffman et al., 2011). By corollary if children with language disorder are included in the normative sample identification accuracy can be reduced (Peña, Spaulding, & Plante, 2006). It is of concern that the quality of a standardized test, as measured by the test's psychometric properties, does not appear to influence how frequently a test is used by practitioners (Betz, Eickhoff, & Sullivan, 2013), despite the fact that guidelines do exist about what to look for in test manuals in order to decide whether a test is fit for purpose (Friberg, 2010; Hutchinson, 1996).

4.2.2 Composite language measures

A number of composite language measures exist. These are often based on differing theoretical frameworks and, as such, the correlations between tests
may not be strong (Hoffman et al., 2011). The CELF instruments are perhaps the most commonly used measures of assessing language internationally and, as such, have been the focus of a number of psychometric studies (Eadie et al., 2014; Spaulding et al., 2006). The CELF-4 was one of the five instruments, of the 43 examined, with acceptable specificity and sensitivity data in Spaulding et al.’s (2006) study. Test-retest measures for both the CELF-4 and CELF-P2 are good. However, there is concern over the reliability of some of the subscales, in particular Sentence Structure in the CELF-P2 (Eigenbrood, 2007). Validity data has also been reported for the CELF-P2 and the CELF-4, where correlations range from 0.60 to 0.85 for different subtests. However, recent Australian research has indicated that the CELF-P2 does not demonstrate adequate levels of sensitivity (64%) to identify children with language disorders at age 5 (Eadie et al., 2014).

Other omnibus tests include the Preschool Language Scales (Boucher & Lewis, 1997) and the New Reynell Developmental Language Scales (Edwards, Letts & Sinka, 2011); both tests require further evaluation. In addition some more comprehensive psychometric assessments include oral language scales. For example the Wechsler Individual Achievement Test (3rd edition: WIAT-III) includes both a listening comprehension and oral expression scale. The WIAT-III subtests have strong psychometric properties and there is evidence to support the use of subtests with special populations (Miller, 2010). The WIAT-III has the added advantage of comparing across composite scores such as oral language and reading comprehension.

4.2.3 Tests of single elements of the language system
Subcomponents of the language system can be examined to further explore children's strengths and weaknesses. Single measures of language are consistently inadequate for determining whether a child is developing typically or is experiencing a delay at any age, and they become less reliable the younger the child (Thal & Katch, 1996). Nonetheless when the measures are reliable and valid they can, when combined with other forms of assessment, provide a profile of a child's strengths and needs.

**Vocabulary**

Vocabulary measures are commonly used to assess children's language skills. Typically, forced-choice receptive measures are used but these provide limited information about the child's vocabulary skills (Dockrell & Messer, 2004). By contrast, more complex measures assessing depth and breadth of vocabulary often ask children to provide oral definitions. While these measures provide a greater insight into the child's lexical representations, the measure conflates other aspects of expressive language with vocabulary knowledge. The most commonly used vocabulary measure in the UK is the British Picture Vocabulary Scale III (BPVS-III; based on the Peabody Picture Vocabulary, Dunn & Dunn, 1997). The BPVS-III provides norms for individuals aged 3-16 years with excellent reliability, reported as 0.91. Concurrent validity with other language measures is not high. Research has consistently indicated that vocabulary scores cannot be used as though they were indicators of general language ability (Gray, Plante, Vance, & Henrichsen, 1999; Spaulding, Hosmer, & Schechtman, 2013). For some children with language disorders vocabulary scores can be well within the
norm, despite wider problems with receptive and expressive language (Friberg, 2010; Spaulding et al., 2013). So while vocabulary tests have the advantage of being easy to administer and score, they should not be used as the sole measure to identify children with language difficulties (Longo, 2005).

**Receptive Grammar**

The Test for Reception of Grammar (Bishop, 2003) is, like the BPVS, a test of receptive language that employs a picture-pointing paradigm. The TROG targets sentence comprehension. It contains 80 stimulus items, arranged in blocks of 4, which test 20 grammatical contrasts (e.g. the prepositions “in” and “on”, pronouns, relative clauses). Blocks are arranged in a presumed ascending order of relative developmental difficulty.

The first version was standardised for children aged 4-12, but the latest edition, TROG-2, extends the age up to 16 and is also available electronically. The TROG-2 has a reported internal consistency of 0.88. Concurrent validity with other language measures, at around 0.5-0.6, is not high. The TROG is widely used clinically by speech and language therapists and in research studies, and it has been translated, although not necessarily standardised, into several other languages.

**Sentence repetition**

Sentence repetition, also known as sentence imitation and sentence recall, is a technique with a long history in psycholinguistic research (Rodd & Braine, 1971) and language assessment (Schwartz & Daly, 1978). The assumption is that
children will only be able to repeat structures that are part of their language system. Indeed, many studies have shown that sentence repetition is significantly less accurate in children with developmental language disorders (see Conti-Ramsden, Botting & Faragher, 2001; Riches, Loucas, Baird, Charman & Simonoff, 2010, and references therein) and children who are not native speakers of the test language (Komeili & Marshall, 2013).

Many language assessments include a sentence repetition subtest, for example the CELF-P2 (Semel, Wiig & Secord, 2004), Test of Language Development (Newcomer & Hamill, 2008), Grammar and Phonology Screening Test (van der Lely, Gardner, Froud & McClelland, 2007) and the Early Repetition Battery (Seeff-Gabriel, Chiat & Roy, 2008). Sentence repetition subtests can have high reliability, but test administrators are advised to read the manuals carefully before deciding which one is going to best suit their needs.

Sentence repetition tests are quick and easy to administer, and also easy to score, because in contrast to most expressive language tests (e.g. tests of narrative), the targets are explicit and precisely specified (Chiat et al, 2013). Depending on how they are constructed (i.e. what types of sentence structures they use), they can yield not only quantitative data (i.e. how many sentences does a child repeat accurately?) but also qualitative data (i.e. what types of sentences does a child find particularly difficult, and what type of errors do they make?) (e.g. Riches et al, 2010; Komeili & Marshall, 2013).

**Narrative**

Narrative assessments are less frequently used ways of assessing children's language competence but as Botting (2002) argues, narrative is an important
means of measuring communicative competence. Conventional language tests
elicit production and test comprehension using artificial tasks. By contrast,
narrative tasks provide a more naturalistic setting to examine children’s
language skills.

Perhaps the most commonly used assessment of narrative is the Bus Story
(Renfrew, 1997), designed to be used for children between 3;6 and 7. The Bus
Story is a test of narrative recall whereby the assessor tells a story about a
naughty bus and the child is asked to repeat it. The child’s narrative is scored on
three dimensions: sentence length (calculated from the 5 longest sentences),
information and the use of subordinate and relative clauses. There are standard
scores for both sentence length and information. There is moderate test-retest
reliability but lower inter-rater reliability. Information about validity is limited.
Scoring of the Bus Story is complex and errors are common (Haccoun, 2001).
The measure should not be used when English is a second language, children
have sensory impairments or where there are diagnosed learning difficulties as
it is not reliable (Haccoun, 2001). In terms of identifying disorders during the
preschool years, the test is not sensitive and results in the over-identification of
typically developing children, and in particular minority children, as having poor
narrative skills (Pankratz, Plante, Vance, & Insalaco, 2007). The assessment of
narrative skill is a burgeoning area of research and the identification of narrative
assessments which can be scored consistently and are sensitive to change after
intervention (index of narrative complexity) is a promising development
(Petersen, Gillam, & Gillam, 2008).
4.2.4 Dynamic assessment

Although standardised, norm-referenced language assessments of the type we have discussed thus far are still the most widely-used means of evaluating children’s language abilities (Caesar & Kohler, 2009), these static tests only give a snapshot in time and do not reveal why children perform poorly. Children might perform poorly on a test for a variety of reasons, and these differences could potentially be important for intervention. There is therefore an increasing interest in an interactive approach to conducting assessments that focuses on the child’s ability to respond to intervention, i.e. his capacity for change or “modifiability”. This approach is called “dynamic assessment”, and unlike traditional testing, it employs a test, teach, retest procedure to assess the child’s learning processes. For example, the Dynamic Assessment of Word Learning (Camilleri & Botting, 2013), employs the BPVS as a pre-test, and then targets words from that test that the child failed to identify by using additional composite pictures featuring that vocabulary. Through a series of verbal prompts, including open questions (“What can you see in this picture?”) and probes (“Where is the woman?”), followed by opportunities to make direct links between the spoken word and the correct referent in the picture (“A balcony is part of a house, but it’s outside”), the assessor allows the child to demonstrate their knowledge of the word or to learn it in context. The amount and type of assistance that the child requires is recorded.

Dynamic assessment is considered more culturally fair to those from different linguistic or cultural backgrounds (Lidz & Peña, 2009) and may be more sensitive for measuring change in language over time (Hasson & Botting, 2010).
Furthermore, this approach has been shown to distinguish between children whose language is delayed, but whose capacity for learning language is not impaired, versus those with a language disorder, be they monolingual or bilingual (Hasson et al., 2013; Peña, Resendiz & Gillam, 2007).

Whilst several types of dynamic assessment have been developed in recent years, including for vocabulary (Camilleri & Botting, 2013), syntax (Hasson, Dodd & Botting, 2012) and narrative (Peña et al, 2006), as well as more omnibus tests (Hasson et al, 2013), they are nevertheless not widely used by practitioners. To the best of our knowledge, only one is commercially available (Miller, Gillam & Peña, 2001).

4.2.5 Checklists

A number of checklists of early language skills, to be completed by parents and professionals, exist. Many of these have not been validated psychometrically in terms of reliability and validity (see for example Mok & Lam, 2011). The best researched are the Communication Development Inventories (see Law & Roy, 2008 for a review). These have been translated into a number of different languages, there is a Norwegian web-based version (Simonsen et al., 2013), and also a version in British Sign Language (Woolfe, Herman, Roy & Woll, 2010).

Using parental report data can be helpful to gain a broader perspective of a child’s language skills and when children are difficult to assess. Furthermore, checklists are inexpensive to use, and additional training is not required (Hall & Segarra, 2007; Nordahl-Hansen, Kaale, & Ulvund, 2013). However, as differences between respondents, such as their background, may affect how they report their children’s language skills, care needs to be taken in
interpreting the results (Pan, Rowe, Spier, & Tamis-Lemonda, 2004). Importantly, given the variability in language trajectories, checklists are not reliable in identifying children who will go on to experience language delays (Law & Roy, 2008).

5 CONCLUSIONS AND FUTURE DEVELOPMENTS

The language system is complex, composed of a number of subcomponents and the language trajectories of preschool children vary substantially. As such, the development of reliable and valid assessments is challenging, but they are of central importance for studying typical and atypical development. For clinical and diagnostic purposes it is clear that composite measures are more likely to identify language disorders, but even with such measures predicting future language skills is problematic in the preschool period. Using regular monitoring procedures advocated as by Snowling et al (2012) within a context where young children are provided with effective language learning environments may provide an important step in identifying children with language disorders (see Dockrell, Ricketts & Lindsay, 2012).

As we have shown, research studies continue to enhance our understanding of the language development process and aid our identification of children who experience persistent language disorders and the factors that are associated with these. In the early review (Dockrell, 2001) the assessment of narrative skills and dynamic assessment were highlighted as new developments. As we have shown they now play a much more important role in our assessment of children’s language performance. Of particular importance in recent years has been evidence from large-scale longitudinal studies which capture representative
populations. By corollary an understanding is also developing of language disorders in different populations and associated risk factors. For example, recent findings of language disorders in deaf children who sign (Mason et al, 2010) make it imperative that good quality assessment tools be developed for deaf signers: it cannot be assumed that all deaf children who are exposed to good models of sign language will learn the language without difficulty. Finally there is a need to develop assessments that are available (and standardized) across a variety of languages given our increasingly multicultural society.

Language and communication are essential to learning and development. Psychometrically-sound assessments are an essential component of developing effective and efficient interventions to support children’s attainments and aspirations. The current review has aimed to provide the necessary information to make informed decisions about assessing the language competencies of preschool children.

References


Dockrell, J. E., Lindsay, G., Roulstone, S. & Law, J. (in press). Supporting children with speech and language communication needs: an overview of the results of the better communication research programme. *International Journal of Language and Communication Disorders*.

Dockrell, J. E., Lindsay, G., & Palikara, O. (2011). Explaining the academic achievement at school leaving for pupils with a history of language impairment: Previous academic achievement and literacy skills. *Child Language Teaching & Therapy, 27*(2). doi: 10.1177/0265659011398671


Dockrell, J. E., Ricketts, J., & Lindsay, G. (2012). Understanding speech, language and communication needs: Profiles of need and provision. Thematic report Better Communication Research Programme. DFE-RR247-BCRP4


