C A M H	12072	WILEY	Dispatch: 12.7.14	CE: Sumithra
Journal Code	Manuscript No.		No. of pages: 10	PE: Durga

Child and Adolescent Mental Health Volume **, No. *, 2014, pp. **–** doi:10.1111/camh.12072

Measurement Issues: Assessing language skills in young children

Julie Dockrell & Chloë Marshall

Psychology and Human Development, Institute of Education, London, UK. E-mail: Julie.Dockrell@ioe.ac.uk

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. **Conclusion:** 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

Keywords: Language; assessment; psychometrics

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 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, for both 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 realization that in comparison with 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 used. 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.

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 organization 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 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 Brooks & 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 the fluency of a child's expressive language, while failures to process key syntactic components may impact 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, Nygren, Walker, & Panoscha, 2006a,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 undetected, often because of the limitations in the measures available (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 future, they may offer reliable and valid ways of exploring children's language skills.

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.

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).

Different dialects and bilingualism

There is a growing concern that children from ethnic minority groups are overrepresented in the caseloads of speech and language therapists and are overidentified generally as having speech language and communication needs (Dockrell, Lindsay, Roulstone, & Law, 2014). There is also a growing awareness that nonstandard varieties of English differ from 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 nonstandard 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 United States, 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 penalize the child for using a different form of the object pronoun (i.e. *himself*, *hisself*). Allowing a range of responses reduces the bias against nonstandard dialect speakers that is inherent in many standardized 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 are only limited normative data on the developmental trajectories 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 overidentified with a language disorder, because assessors have inappropriate developmental expectations, and are sometimes underidentified because assessors wait for children to learn more English before identifying difficulties (Bedore & Peña, 2008).

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 standardized 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, 2013). 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 (Herman, Holmes, & Woll, 1999), which tests grammatical comprehension using a picture-pointing format, was standardized 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 standardization (Herman et al., 2004). A third test of BSL, which targets both receptive and expressive vocabulary, has also been developed, but is not yet standardized (Mann et al., 2013).

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 United Kingdom use a range of different terms to describe this population (Dockrell, Lindsay, Letchford, & Mackie, 2006), and different terms are used internationally: in Europe, dys-

phasia; in the United States, 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 nonverbal 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 nonverbal ability (Krassowski & Plante, 1997). Language problems may also impact children's performance on nonverbal tasks, thereby affecting assessments of nonverbal ability. The new DSM-5 does not include a discrepancy criterion for language disorders. Furthermore, 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 nonverbal 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 characterize 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. 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/).

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 preintervention and postintervention measures to evaluate the impact of oral language interventions. Although repeated standardized testing can be a useful way of tracking progress over time, the standardized 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 that is not receiving the intervention is required for a comparison that will allow conclusions about the 'effect' of the intervention.

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 characterize 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 trade-off 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, overidentification 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. Clinical Evaluation of Language Fundamentals (CELF) screening test (Semel, Wiig, & Secord, 2004a,b). 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 overreferral for diagnostic assessments. These conclusions are further supported by systematic reviews (Law, Boyle, Harris, Harkness, & Nye, 2000; Nelson et al., 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 24–47 months. It is noteworthy that both 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 future is fraught with difficulties. Studies that have attempted this have been unsuccessful in identifying language factors that predict future performance (Law, Rush, Anandan, Cox, & Wood, 2012; Nelson et al., 2006a,b; Wilson, McQuaige, Thompson, & McConnachie, 2013). As Snowling, Hulme, Bailey, Sthothard, and Lindsay (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.

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 characterize 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 – standardized tests, dynamic assessments and parental or teacher checklists. These approaches are not mutually exclusive.

Standardized 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 to decide whether a test is fit for purpose (Friberg, 2010; Hutchinson, 1996).

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 have 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.

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; however, 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 standardized 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 standardized, 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 et al., 2004a,b), 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. Komeili & Marshall, 2013; Riches et al., 2010).

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 language performance. 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 five 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 interrater 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 the test is not

reliable in these circumstances (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 burgenoing 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).

Dynamic assessment

Although standardized, 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 pretest, 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 are 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 & 17 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, Camilleri, Jones, Smith, & Dodd, 2013; Peña, Reséndiz, & 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).

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, Kristoffersen, Bleses, Wehberg, & Jørgensen, 2013) and also a version in British Sign Language (Woolfe, Herman, Roy & Woll, 2010). 18

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).

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 as advocated 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 assess-

ments 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.

Acknowledgements

No external funding was received for this work. The authors have declared that they have no competing or potential conflicts of interest.

References

Allen, G. (2011). Early intervention: the next steps. London: HM Government.

Antoniazzi, D., Snow, P., & Dickson-Swift, V. (2010). Teacher identification of children at risk for language impairment in the first year of school. *International Journal of Speech-Language Pathology*, *12*, 244–252.

Bedore, L.M., & Peña, E.D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, *11*(1), 1–29.

Bercow, J. (2008). The Bercow report: A review of services for children and young people (0–19) with speech, language and communication needs: DCSF. Available from: http://www.dcsf.gov.uk/bercowreview/docs/7771-DCSF-BERCOW.PDF.

Betz, S.K., Eickhoff, J.R., & Sullivan, S.F. (2013). Factors influencing the selection of standardized tests for the diagnosis of specific language impairment. *Language Speech and Hearing Services in Schools*, *44*, 133–146.

Bishop, D.V.M. (2003). Test for reception of grammar (2nd edn). London: Psychological Corporation.

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, *18*(1), 1–21.

Brooks, P.J., & Kempe, V. (2012). Language development. Chichester: BPS Blackwell.

Caesar, L.G., & Kohler, P.D. (2009). Tools clinicians use: A survey of language assessment procedures used by school-based speech-language pathologists. *Communication Disorders Quarterly*, *30*, 226–236.

Camilleri, B., & Botting, N. (2013). Beyond static assessment of children's receptive vocabulary: A dynamic assessment of word learning ability. *International Journal of Language and Communication Disorders*, *48*, 565–581.

Catts, H.W., Bridges, M.S., Little, T.D., & Tomblin, J.B. (2008). Reading achievement growth in children with language impairments. *Journal of Speech Language and Hearing Research*, *51*, 1569–1579.

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

Consortium for Research in Deaf Education (2013). CRIDE report on 2013 survey on educational provision for deaf children in England. Online. Available from: http://www.ndcs.org.uk/professional_support/national_data/uk_education_.html [last accessed 14 June 2014].

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry*, *42*(6), 741–748.

Conti-Ramsden, G., Durkin, K., Simkin, Z., & Knox, E. (2009). Specific language impairment and school outcomes. I: Identifying and explaining variability at the end of compulsory education. *International Journal of Language & Communication Disorders*, *44*, 15–35.

Dixon, J.A., & Marchman, V.A. (2007). Grammar and the lexicon: Developmental ordering in language acquisition. *Child Development*, *78*, 190–212.

Dockrell, W.B. (1966). Secondary education, social class and development of abilities. *British Journal of Educational Psychology*, *36*, 7–14.

Dockrell, J.E. (2001). Assessing language skills in preschool children. *Child and Adolescent Mental Health*, *6*, 74–86.

Dockrell, J.E., & Law, J. (2007). Measuring and understanding patterns of change in intervention studies with children. *Implications for evidence-based practice*, *1*, 86–97.

Dockrell, J.E., Lindsay, G., Letchford, B., & Mackie, C. (2006). Educational provision for children with specific speech and language difficulties: perspectives of speech and language therapy service managers. *International Journal of Language & Communication Disorders*, *41*, 423–440.

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., Lindsay, G., Roulstone, S., & Law, J. (2014). Supporting children with speech language and communication needs: An overview of the results of the better communication research programme. *International Journal of Language and Communication Disorders*, *?????*, ?????–??????.

Dockrell, J.E., & Messer, D. (2004). Later vocabulary acquisition. In R. Berman (Ed.), Language development across childhood and adolescence: Psycholinguistic and crosslinguistic perspectives (pp. 35–52). Trends in Language Acquisition Research 3. ????????: John Benjamins.

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.

Dunn, L.M., & Dunn, L.M. (1997). Examiner's manual for the PPVT-III: Peabody picture vocabulary test (3rd edn). Circle Pines, MN: American Guidance Service.

Dunn, M., Flax, J., Sliwinski, H., & Aram, D. (1996). The use of spontaneous language measures as criteria for identifying children with specific language impairment: An attempt to reconcile clinical and research incongruence. *Journal of Speech and Hearing Research*, *39*, 643–654.

Eadie, P., Nguyen, C., Carlin, J., Bavin, E., Bretherton, L., & Reilly, S. (2014). Stability of language performance at 4 and 5 years: Measurement and participant variability. *International Journal of Language & Communication Disorders*, *49*, 215–247. http://dx.doi.org/10.1111/1460-6984.12065.

Edwards, S., Letts, C., & Sinka, I. (2011). The new Reynell Developmental Language Scales. GL Assessment.

Eigenbrood, R. (2007) Test review of the clinical evaluation of language fundamentals preschool. In K.F. Geisinger, R.A. Spies, J.F. Carlson & B.S. Blake. The 17th Mental measurements year book, 2nd edn. Available from http://marketplace.unl.edu/buros/.

Eriksson, M., Westerlund, M., & Miniscalco, C. (2010). Problems and limitations in studies on screening for language delay. *Research in Developmental Disabilities*, *31*, 943–950.

Friberg, J.C. (2010). Considerations for test selection: How do validity and reliability impact diagnostic decisions? *Child Language Teaching & Therapy*, *26*, 77–92.

Genesee, F., & Nicoladis, E. (2009). Bilingual first language acquisition. In E. Hoff & M. Shatz. Blackwell handbook of language development (pp. 324–342). Chichester, UK: Blackwell.

Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language Speech and Hearing Services in Schools*, *30*, 196–206.

Greenslade, K.J., Plante, E., & Vance, R. (2009). The diagnostic accuracy and construct validity of the structured photographic expressive language test – preschool: second edition (SPELT-P2). *Language Speech, and Hearing Services in Schools*, *40*, 150–160.

Haccoun, R.R. (2001). Test review of the Renfrew Bus Story. In B.S. Plake & J.C. Impara (Eds.), The 14th Mental measurements year book. Available from http://marketplace.unl.edu/buros/ [last accessed 20 April 2014].

Hall, N.E., & Segarra, V.R. (2007). Predicting academic performance in children with language impairment: The role of parent report. *Journal of Communication Disorders*, *40*, 82–95.

Hasson, N., Camilleri, B., Jones, C., Smith, J., & Dodd, B. (2013). Discriminating disorder from difference using dynamic assessment with bilingual children. *Child Language Teaching and Therapy*, *29*, 52–70.

Hasson, N., Dodd, B., & Botting, N. (2012). Dynamic Assessment of Sentence Structure (DASS): Design and evaluation of a novel procedure for assessment of syntax in children with language impairments. *International Journal of Speech Language and Communication Disorders*, *47*, 285–299.

Herman, R., Grove, N., Holmes, S., Morgan, G., Sutherland, H., & Woll, B. (2004). Assessing BSL development – Production test. London: City University London.

Herman, R., Holmes, S., & Woll, B. (1999). Assessing BSL development – Receptive skills test. Coleford, UK: Forest Bookshop.

Hoffman, L.M., Loeb, D.F., Brandel, J., & Gillam, R.B. (2011). Concurrent and construct validity of oral language measures with school-age children with specific language impairment. *Journal of Speech Language and Hearing Research*, *54*, 1597–1608.

Hutchinson, T.A. (1996). What to look for in the technical manual: Twenty questions for users. *Language Speech and Hearing Services in Schools*, *27*, 109–121.

King, T.M., Rosenberg, L.A., Fuddy, L., McFarlane, E., Sia, C., & Duggan, A.K. (2005). Prevalence and early identification of language delays among at-risk three year olds. *Journal of Developmental and Behavioral Pediatrics*, *26*, 293–303.

Komeili, M., & Marshall, C.R. (2013). Sentence repetition as a measure of morphosyntax in monolingual and bilingual children. *Clinical Linguistics and Phonetics*, *27*, 152–162.

de Koning, H.J., de Ridder-Sluiter, J.G., van Agt, H.M.E., Reep-van den Bergh, C.M.M., van der Stege, H.A., Korfage, I.J., … & van der Maas, P. J. (2004). A cluster-randomised trial of screening for language disorders in toddlers. *Journal of Medical Screening*, *11*, 109–116.

Krassowski, E., & Plante, E. (1997). IQ variability in children with SLI: Implications for use of cognitive referencing in determining SLI. *Journal of Communication Disorders*, *30*(1), 1–9.

Laing, G.J., Law, J., Levin, A., & Logan, S. (2002). Evaluation of a structured test and a parent led method for screening for speech and language problems: Prospective population based study. *British Medical Journal*, *325*, 1152–1154B.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: Findings from a systematic review of the literature. *International Journal of Language & Communication Disorders*, *35*, 165–188.

Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the communicative development inventories. *Child and Adolescent Mental Health*, *13*, 198–206.

Law, J., Rush, R., Anandan, C., Cox, M., & Wood, R. (2012). Predicting language change between 3 and 5 years and its implications for early identification. *Pediatrics*, *130*, E132–E137.

van der Lely, H.K.J., Gardner, H., Froud, K., & McClelland, A. (2007). Grammar and phonology screening (GAPS) test. St Mabyn: STASS Publications.

Leonard, L.B. (2009). Is expressive language disorder an accurate diagnostic category? *American Journal of Speech-Language Pathology*, *18*, 115–123.

Lidz, C.S., & Peña, E.D. (2009). Response to intervention and dynamic assessment: Do we just appear to be speaking the same language? *Seminars in Speech and Language*, *30*(2), 121–133.

Lindsay, G., & Dockrell, J.E. (2012). Longitudinal patterns of behavioral, emotional, and social difficulties and self-concepts in adolescents with a history of specific language impairment. *Language Speech and Hearing Services in Schools*, *43*, 445–460.

Longo, A.P. (2005). Test review of the British Picture Vocabulary Scale, Second edition. In R.A. Spies & B.S. Blake (Eds.), The sixteenth mental measurements yearbook. Retrieved from http://marketlace.unl.edu/buros/ [last accessed 20 April 2014].

Mann, W., Roy, P., & Marshall, C.R. (2013). A look at the other 90% – Investigating vocabulary knowledge in British Sign Language in deaf children from different language learning backgrounds. *Deafness and Education International*, *15*, 91–116.

Mason, K., Rowley, K., Marshall, C.R., Atkinson, J.R., Herman, R., Woll, B., & Morgan, G. (2010). Identifying SLI in deaf children acquiring British Sign Language: Implications for theory and practice. *British Journal of Developmental Psychology*, *28*, 33–49.

McCartney, E., & Ellis, S. (2013). The linguistically aware teacher and the teacher-aware linguist. *Clinical Linguistics & Phonetics*, *27*, 419–427.

McCauley, R. (2001). Assessment of language disorders in children. Mahwah, NJ: Lawrence Erlbaum Publishers.

Merrell, A.W., & Plante, E. (1997). Norm referenced test interpretation in the diagnostic process. *Language Speech and Hearing Services in Schools*, *28*, 50–58.

Miller, M.D. (2010) Test review of the WIAT-3. In R.A. Spies, J.F. Carlson & K.F. Geisinger (Eds.). The 18th Mental measurements year book. Available from http://marketplace.unl.edu/buros/ [last accessed 20 April 2014].

Miller, L., Gillam, R., & Peña, E. (2001). Dynamic assessment and intervention: Improving children's narrative abilities. Austin, TX: PRO-ED.

Mok, M.M.C., & Lam, H.M.Y. (2011). Assessment of language development of preschoolers: Validating Morrow's Checklist for assessing early literacy development. *Early Child Development and Care*, *181*, 203–220.

Moll, H., & Tomasello, M. (2010). Infant cognition. *Current Biology*, *20*, R872–R875.

Mueller Gathercole, V.C. (2013). Assessment of multi-tasking wonders: Music, olympics and language. In V.C. Mueller Gathercole (Ed.), Issues in the assessment of bilinguals (pp. 1–19). Bristol, UK: Multilingual Matters.

Nelson, H.D., Nygren, P., Walker, M., & Panoscha, R. (2006a). Screening for speech and language delay in preschool children: Systematic evidence review for the US preventive services task force (vol 117, pg 298, 2006). *Pediatrics*, *117*, 2336–2337.

Nelson, H.D., Nygren, P., Walker, M., & Panoscha, R. (2006b). Screening for speech and language delay in preschool children: Systematic evidence review for the US preventive services task force. *Pediatrics*, *117*, E298–E319.

Newcomer, P.L., & Hamill, D.D. (2008). Test of language development-primary (4th edn). Austin, TX: Pro-Ed.

Nordahl-Hansen, A., Kaale, A., & Ulvund, S.E. (2013). Inter-rater reliability of parent and preschool teacher ratings of language in children with autism. *Research in Autism Spectrum Disorders*, *7*(11), 1391–1396.

Pan, B.A., Rowe, M.L., Spier, E., & Tamis-Lemonda, A. (2004). Measuring productive vocabulary of toddlers in low-income families: Concurrent and predictive validity of three sources of data. *Journal of Child Language*, *31*, 587–608.

Pankratz, M.E., Plante, E., Vance, R., & Insalaco, D.M. (2007). The diagnostic and predictive validity of the Renfrew bus story. *Language Speech and Hearing Services in Schools*, *38*, 390–399.

Peña, E., Gillam, R., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C., & Sabel, T. (2006). Dynamic assessment of children from culturally diverse backgrounds: Applications to narrative assessment. *Journal of Speech, Language, Hearing Research*, *49*(5), 1–21.

Peña, E.D., Reséndiz, M., & Gillam, R.B. (2007). The role of clinical judgements of modifiability in the diagnosis of language impairment. *Advances in Speech-Language Pathology*, *9*, 332–345.

Peña, E.D., Spaulding, T.J., & Plante, E. (2006). The composition of normative groups and diagnostic decision making: Shooting ourselves in the foot. *American Journal of Speech-Language Pathology*, *15*, 247–254.

Pesco, D., & O'Neill, D.K. (2012). Predicting later language outcomes from the language use inventory. *Journal of Speech Language and Hearing Research*, *55*(2), 421–434.

Petersen, D.R., Gillam, S.L., & Gillam, R.R. (2008). Emerging procedures in narrative assessment – The index of narrative complexity. *Topics in Language Disorders*, *28*, 115–130.

Riches, N.G., Loucas, T., Baird, G., Charman, T., & Simonoff, E. (2010). Sentence repetition in adolescents with specific language impairments and autism: An investigation of complex syntax. *International Journal of Language and Communication Disorders*, *45*, 47–60.

Rodd, L.J., & Braine, M.D.S. (1971). Children's imitations of syntactic constructions as a measure of linguistic competence. *Journal of Verbal Learning and Verbal Behaviour*, *10*, 430–443.

Roy, P., & Chiat, S. (2013). Teasing apart disadvantage from disorder: the case of poor language. In C. Marshall (Ed.), Current issues in developmental disorders. London: Psychology Press.

Saxton, M. (2010). Child language: Acquisition and development. London: Sage Publications.

Schoon, I., Parsons, S., Rush, R., & Law, J. (2010). Children's language ability and psychosocial development: A 29-year follow-up study. *Pediatrics*, *126*, E73–E80.

Schwartz, A.H., & Daly, D.A. (1978). Elicited imitation in language assessment: A tool for formulating and evaluating treatment. *Journal of Communication Disorders*, *11*, 25–35.

Seeff-Gabriel, B., Chiat, S., & Roy, P. (2008). Early repetition battery. London: Pearson Assessment.

Semel, E.S., Wiig, E.H., & Secord, W.A. (2004a). Clinical evaluation of language fundamentals-preschool. San Antonio, TX: The Psychological Corporation.

Semel, E., Wiig, E.H., & Secord, W.A. (2004b). Clinical evaluation of language fundamentals, fourth edition—Screening test (CELF-4 screening test). ????????: The Psychological Corporation/A Harcourt Assessment Company.

Seymour, H.N. (2004). The challenge of language assessment for African American English-speaking children: A historical perspective. *Seminars in Speech and Language*, *25*, 3–12.

Seymour, H.N., Roeper, T.W., & de Villiers, J. (2005). Diagnostic evaluation of language variance-norm referenced. London, UK: Pearson Assessment.

Simonsen, H., Kristoffersen, K., Bleses, D., Wehberg, S., & Jørgensen, R. (2013). The Norwegian communicative development inventories: Reliability, main developmental trends and gender differences. *First Language*, doi:10.1177/0142723713510997.

Snowling, M., Hulme, C., Bailey, A., Sthothard, S., & Lindsay, G. (2012). Better Communication Research Programme: Language and literacy attainment of pupils during Early Years and through KS2: Does teacher assessment at five provide a valid measure of children's current and future educational attainment. DFE-RR177a.

Spaulding, T.J., Hosmer, S., & Schechtman, C. (2013). Investigating the interchangeability and diagnostic utility of the PPVT-III and PPVT-IV for children with and without SLI. *International Journal of Speech-Language Pathology*, *15*, 453–462.

Spaulding, T.J., Plante, E., & Farinella, K.A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language Speech and Hearing Services in Schools*, *37*(1), 61–72.

Spaulding, T.J., Szulga, M.S., & Figueroa, C. (2012). Using norm-referenced tests to determine severity of language impairment in children: Disconnect between US Policy Makers and Test Developers. *Language Speech and Hearing Services in Schools*, *43*, 176–190.

St Clair, M.C., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2011). A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of specific language impairment (SLI). *Journal of Communication Disorders*, *44*, 186–199.

Thomas, M.S.C. (2010). Language acquisition in developmental disorders. In M. Kail & M. Hickmann (Eds.), Language acquisition across linguistic and cognitive systems (pp. 67–87). Philadelphia, PA: John Benjamins.

Tomblin, J.B., & Zhang, X.Y. (2006). The dimensionality of language ability in school-age children. *Journal of Speech Language and Hearing Research*, *49*, 1193–1208.

Tomblin, J.B., Zhang, X.Y., Buckwalter, P., & Catts, H. (2000). The association of reading disability, behavioral disorders, and language impairment among second-grade children. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*, 473–482.

Van Agt, H.M.E., van der Stege, H.A., de Ridder-Sluiter, J.G., & de Koning, H.J. (2007). Detecting language problems: Accuracy of five language screening instruments in preschool children. *Developmental Medicine and Child Neurology*, *49*, 117–122.

Washbrook, E., & Waldfogel, J. (2010). Low income and early cognitive development in the UK. London: Sutton Trust.

Webb, M.-Y.L., & Lederberg, A.R. (2014). Measuring phonological awareness in deaf and hard-of-hearing children. *Journal of Speech, Language and Hearing Research*, *57*, 131–142.

Wellman, R.L., Lewis, B.A., Freebairn, L.A., Avrich, A.A., Hansen, A.J., & Stein, C.M. (2011). Narrative ability of children with speech sound disorders and the prediction of later literacy skills. *Language Speech and Hearing Services in Schools*, *42*(4), 561–579.

Wilson, P., McQuaige, F., Thompson, L., & McConnachie, A. (2013). Language delay is not predictable from available risk factors. *Scientific World Journal*, doi:10.1155/2013/947018.

Young, A.R., Beitchman, J.H., Johnson, C., Douglas, L., Atkinson, L., Escobar, M., & Wilson, B. (2002). Young adult academic outcomes in a longitudinal sample of early identified language impaired and control children. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *43*, 635–645.

Zhang, X.Y., & Tomblin, J.B. (2003). Explaining and controlling regression to the mean in longitudinal research designs. *Journal of Speech Language and Hearing Research*, *46*, 1340–1351.

Accepted for publication: 19 June 2014

Author Query Form

Journal: CAMH

Article: 12072

Dear Author,

During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers on the query sheet if there is insufficient space on the page proofs. Please write clearly and follow the conventions shown on the attached corrections sheet. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.

Many thanks for your assistance.

Query reference	Query	Remarks
1	AUTHOR: The set of keywords is extracted from the supplied PDF source file. Please check if the keywords are okay.	
2	AUTHOR: Brookes & Kempe, 2012 has been changed to Brooks & Kempe, 2012 so that this citation matches the Reference List. Please confirm that this is correct.	
3	AUTHOR: Leonard, 1997 has not been included in the Reference List, please supply full publication details.	
4	AUTHOR: More than one reference shares the same author and year-of-publication details. Please check that they have been correctly differentiated between using a, b, etc. after the year of publication.	
5	AUTHOR: Robertson & Salter, 2007 has not been included in the Reference List, please supply full publication details.	
6	AUTHOR: Lonigan, Wagner, Torgesen, & Rashotte, 2007 has not been included in the Reference List, please supply full publication details.	
7	AUTHOR: Tomblin et al., 2003 has not been included in the Reference List, please supply full publication details.	
8	AUTHOR: Aram, Morris & Hall, 1992 has not been included in the Reference List, please supply full publication details.	
9	AUTHOR: Plante, 1998 has not been included in the Reference List, please supply full publication details.	

10	AUTHOR: Bowyer-Crane et al., 2011 has not been included in the Reference List, please supply full publication details.	
11	AUTHOR: Friel-Patti, 1999 has not been included in the Reference List, please supply full publication details.	
12	AUTHOR: More than one reference shares the same author and year-of-publication details (Semel, Wiig, & Secord, 2004). Please check that they have been correctly differentiated between using a, b, etc. after the year of publication.	
13	AUTHOR: Law et al., 1998 has not been included in the Reference List, please supply full publication details.	
14	AUTHOR: Boucher & Lewis, 1997 has not been included in the Reference List, please supply full publication details.	
15	AUTHOR: Thal & Katch, 1996 has not been included in the Reference List, please supply full publication details.	
16	AUTHOR: Renfrew, 1997 has not been included in the Reference List, please supply full publication details.	
17	AUTHOR: Hasson & Botting, 2010 has not been included in the Reference List, please supply full publication details.	
18	AUTHOR: Woolfe, Herman, Roy & Woll, 2010 has not been included in the Reference List, please supply full publication details.	
19	AUTHOR: Issue numbers are deleted in few Journal references, by considering those journal references carry a continuous page range throughout the volume, which is the required style for APA reference styling. Please check and confirm whether it is okay.	
20	AUTHOR: Please provide page range for reference Dockrell 2011.	
21	AUTHOR: Please provide the volume number, page range for reference Dockrell et al. (2014).	
22	AUTHOR: Please provide the publisher location for reference Dockrell and Messer (2004).	
23	AUTHOR: Dunn et al. (1996) has not been cited in the text. Please indicate where it should be cited; or delete from the Reference List.	

24	AUTHOR: Please provide missing details if any for reference Edwards 2011.	
25	AUTHOR: Please provide the publisher location for reference Semel et al. (2004b).	
26	AUTHOR: Please provide volume number and page range for reference Simonsen 2013.	
27	AUTHOR: Please provide volume number and page range for reference Wilson 2013.	

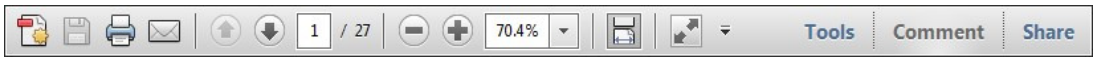
MARKED PROOF

Please correct and return this set

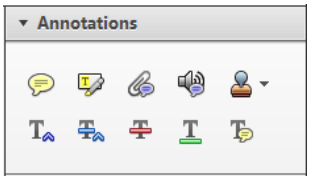
Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	*Textual mark*	*Marginal mark*
Leave unchanged	··· under matter to remain	New matter followed by
Insert in text the matter indicated in the margin		or (x2)
Delete	through single character, rule or underline or through all characters to be deleted	or (x2)
Substitute character or substitute part of one or more word(s)	through letter or through characters	new character or new characters
Change to italics	under matter to be changed	
Change to capitals	under matter to be changed	
Change to small capitals	under matter to be changed	
Change to bold type	under matter to be changed	
Change to bold italic	under matter to be changed	
Change to lower case	Encircle matter to be changed	
Change italic to upright type	(As above)	
Change bold to non-bold type	(As above)	
Insert 'superior' character	through character or where required	or under character e.g. or
Insert 'inferior' character	(As above)	over character e.g.
Insert full stop	(As above)	
Insert comma	(As above)	
Insert single quotation marks	(As above)	or and/or or
Insert double quotation marks	(As above)	or and/or or
Insert hyphen	(As above)	
Start new paragraph		
No new paragraph		
Transpose		
Close up	linking characters	
Insert or substitute space between characters or words	through character or where required	
Reduce space between characters or words	between characters or words affected	

Once you have Acrobat Reader open on your computer, click on the Comment tab at the right of the toolbar:



This will open up a panel down the right side of the document. The majority of tools you will use for annotating your proof will be in the Annotations section, pictured opposite. We've picked out some of these tools below:

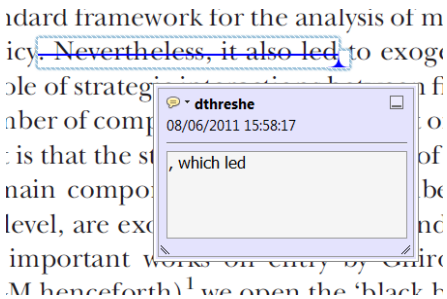


1. Replace (Ins) Tool – for replacing text.

Strikes a line through text and opens up a text box where replacement text can be entered.

How to use it

- Highlight a word or sentence.
- Click on the Replace (Ins) icon in the Annotations section.
- Type the replacement text into the blue box that appears.

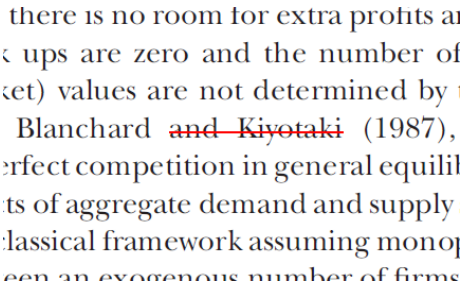


2. Strikethrough (Del) Tool – for deleting text.

Strikes a red line through text that is to be deleted.

How to use it

- Highlight a word or sentence.
- Click on the Strikethrough (Del) icon in the Annotations section.

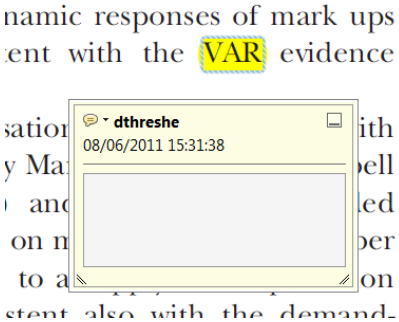


3. Add note to text Tool – for highlighting a section to be changed to bold or italic.

Highlights text in yellow and opens up a text box where comments can be entered.

How to use it

- Highlight the relevant section of text.
- Click on the Add note to text icon in the Annotations section.
- Type instruction on what should be changed regarding the text into the yellow box that appears.

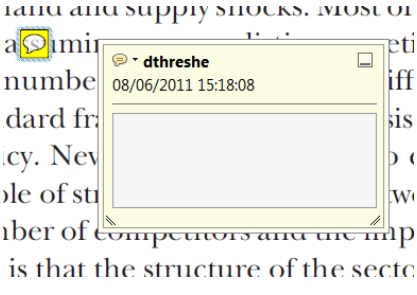


4. Add sticky note Tool – for making notes at specific points in the text.

Marks a point in the proof where a comment needs to be highlighted.

How to use it

- Click on the Add sticky note icon in the Annotations section.
- Click at the point in the proof where the comment should be inserted.
- Type the comment into the yellow box that appears.

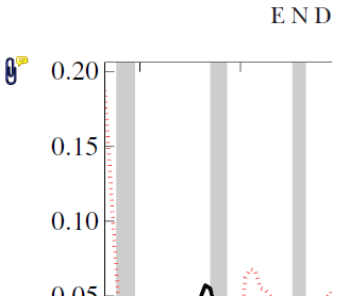


5. Attach File Tool – for inserting large amounts of text or replacement figures.

Inserts an icon linking to the attached file in the appropriate place in the text.

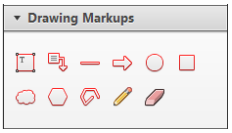
How to use it

- Click on the Attach File icon in the Annotations section.
- Click on the proof to where you'd like the attached file to be linked.
- Select the file to be attached from your computer or network.
- Select the colour and type of icon that will appear in the proof. Click OK.



6. Drawing Markups Tools – for drawing shapes, lines and freeform annotations on proofs and commenting on these marks.

Allows shapes, lines and freeform annotations to be drawn on proofs and for comment to be made on these marks.



How to use it

- Click on one of the shapes in the Drawing Markups section.
- Click on the proof at the relevant point and draw the selected shape with the cursor.
- To add a comment to the drawn shape, move the cursor over the shape until an arrowhead appears.
- Double click on the shape and type any text in the red box that appears.