

Education
Endowment
Foundation

Review of assessment measures in the early years

Language and literacy, numeracy, and social
emotional development and mental health

December 2017

Authors:

Julie Dockrell, Anna Llaurodo, Jane Hurry, Richard Cowan, Erini Flouri, Anneka Dawson



Institute of Education

CONTENTS

Executive summary	2
1. Research brief	3
2. Key questions	5
Does the measure reflect the target domain?	5
Is the measure appropriate for the target population?	5
Are criteria developmentally appropriate?	5
Does the test conform to minimum psychometric properties?	5
Does administration of the test or tool require any special considerations?	6
3. The assessment of young children’s competencies	7
Criteria for choice of assessments	9
4. Domains	10
4.1 Language and communication	10
Conventional approaches to assessing oral language	11
Ways forward in assessing language and communication	11
4.2 Literacy.....	12
Reading	12
Core reading skills	13
Conventional approaches to the assessment of literacy.....	15
Ways forward in assessing early literacy	15
4.3 Numeracy.....	16
Counting	17
Simple arithmetic.....	18
Conventional approaches to the assessment of numeracy	18
Ways forward in assessing early numeracy	19
4.4 Social-emotional development.....	19
Social competence	19
Emotional competence.....	20
Behaviour problems	21
Self-regulation	21
Conventional approaches to the assessment of social and emotional skills	22
Ways forward in assessing social and emotional skills.....	22
4.5 The learning environment	23
Home learning environment	23
Early years settings	24
Ways forward in assessing the learning environment.....	27
5. Methods	28
5.1 Child-based assessments	28
5.2 Environment assessments.....	30
6. Results	33
6.1 Omnibus tests	33
6.2 Language tests.....	34
6.3 Literacy tests	35
6.4 Numeracy tests	37
6.5 Social and emotional development.....	38
References	40
Appendix A: Assessments identified in search	52
Appendix B: Assessments by domain	60
Domain 1: Omnibus	60
Domain 2: Language.....	62
Domain 3: Literacy	64
Domain 4: Social-emotional	65
Domain 5: Numeracy	67

Executive summary

We completed a systematic search of measures to assess language, literacy, numeracy, and social-emotional development for children up to the age of six. These included individual assessments of children's competencies in these areas and assessments of the children's home and early years environment. A review of the relevant competencies for these four core domains provides the basis for identifying appropriate assessments.

Norm-referenced and criterion-referenced tests of children's performance were included. At this point in their development, the majority of direct assessments of children's competencies involve one-to-one assessments. Measures completed by proxy (by parents, care staff, or teachers) were also included and typically used for social-emotional aspects of development and assessments of the environment. All measures were evaluated for developmental appropriateness and psychometric features.

One hundred and forty-six individual child assessments were identified, of which 13 had no published information and were excluded from further evaluation. For the remaining 133 assessments, a qualitative synthesis was completed where there were appropriate U.K. norms for the measure and where data was provided on reliability and validity (N = 47). We also included criterion-referenced assessments that included the key domains. In addition, we identified nine published tools reporting psychometric information for evaluating home and early learning environments.

A wide range of language measures with established norms was identified. By contrast, for the domains of numeracy and literacy, fewer measures were available. However, for children in the target age range, criterion-referenced measures may be more appropriate for identifying early literacy and numeracy skills. Although there are many measures of social-emotional development available, the majority failed to meet the psychometric selection criteria. Four measures of social-emotional development captured a range of key elements of the domain.

A set of questions to consider when choosing a measure to evaluate children's abilities was developed from the review process.

1. Research brief and Glossary

Glossary

Centile: the percentile score that divides observations in a set into 100 intervals each containing an equal number of observations.

Construct validity: the extent to which a test measures what it claims to measure.

Criterion referenced assessment: measures a child's performance against a fixed set of predetermined criteria or learning standards—that is, concise, written descriptions of what students are expected to know and be able to do at a specific stage of their education. The driving test is an example of a criterion referenced assessment.

Domain: in the current review, domains are considered to be a set of interrelated principles, skills, and rules of application that apply to a developmental competency. Language, reading, numeracy, and socio-emotional development are considered domains.

Dynamic assessment: a method of conducting an assessment which aims to identify the skills that an individual child possesses as well as their learning potential. The dynamic assessment procedure emphasizes the learning process and accounts for the amount and nature of examiner investment.

Face validity: examines whether a test 'appears' to measure what it is supposed to measure.

Inter-rater reliability: the relationship between scorings of the same test or behaviour by two different raters (assessors or scorers) working independently.

Norm referenced assessment: measures a child's performance against the performance results of a statistically selected group of test-takers, typically of the same age or grade level and background who have already taken the assessment.

Outliers: data values that differ greatly from the majority of a set of data. These values fall outside of an overall trend that is present in the data.

Ranking: the process of ordering a set of results on a scale. There is no assumption that the intervals between the ranks are equivalent.

Raw data: results as recorded which have not been transformed statistically or recategorised.

Reliability: the extent to which a test gives consistent results if applied more than once to the same child under the same conditions. Reliability coefficients are, typically, correlations that assess reliability. Different reliability measures exist. Internal consistency demonstrates the extent to which a measure is consistent with itself. External reliability demonstrates the extent to which a measure varies from one use to another.

Responsiveness to change: dependent on the intended purpose and domain, measures can be used to identify 'meaningful change' (for example, above a threshold for intervention).

Standard scores: raw scores can be transformed into standardised scores to allow for comparisons between measures and pupils. These scores take into account a pupil's relative position in relation to a wider sample or 'established norm'.

Split-half reliability: a measure of reliability in which the test items are randomly divided into two equal sets and the correlation coefficient between the total scores on the two sets is used as the estimate of reliability.

Test-retest reliability: a sample of participants are tested on two separate occasions by the same test and the reliability is taken as the correlation between the scores.

Validity: a test is valid if it measures what it is supposed to measure. Both content validity (see 'face validity' and 'construct validity') and criterion validity of a test can be examined.

Variance: a measure of the spread or dispersion of a set of scores.

Definitions sourced from Miller, P. M. and Wilson, M. (1983) *A Dictionary of Social Science Methods*, New York: Wiley; Rust, J. and Golombok, S. (1999) *Modern Psychometrics: The Science of Psychological Assessment*, London: Routledge; 'The Glossary of Educational Reform', <http://edglossary.org/about/>

A review of the current assessments for children under the age of six was carried out. Assessments relating to both the individual child and to their home or early years environments were considered.

- In terms of child-based assessments, four domains were examined: language, literacy, numeracy, and social-emotional development.
- Where omnibus developmental measures existed that included one or more of these domains, they were also examined in the review.
- Both norm- and criterion-referenced measures were included.
- We also examined tools that were designed to assess the child's environment. Tools that aimed to provide an overview of the learning environment, and those relating to domain-specific skills relevant to the core domains, were evaluated.

The review was guided by research experts. It included a critical review of current measures and a review of measures used in published studies. For each of the core domains, key features of the domain were identified and implications for assessment and evaluation made at the end of each of the sections.

To identify inclusionary criteria for measures to be evaluated in qualitative synthesis, we drew on the research literature addressing psychometric features of assessments, specific requirements for evaluating changes in performance over time, and guidance from experts in the field. A series of inclusionary and exclusionary criteria for the measures designed to assess children under six years of age were developed and evaluated by the steering committee.

Both norm- and criterion-referenced assessments were included but subject to different inclusionary criteria. Only norm-referenced tests with U.K. standardisations were included in the final selection while criterion-referenced tests were included irrespective of their country of construction. A stepped inclusion criterion was used and this is described in the methods section.

A searchable database of all the measures was created, and for measures that met our minimum inclusionary criteria (see 'Criteria for choice of assessments', page 9), star ratings were developed. Stars were awarded for psychometric properties of the measure. Three stars were available and were awarded for:

- recent (post 2000) representative U.K. normed sample;
- published data on reliability; and
- published data on instrument validity.

There was also a maximum of three stars for implementation. Implementation stars were awarded for:

- unrestricted use, no prior professional accreditation required;
- ease of scoring; and
- administration time (less than 30 minutes).

As a first step, it is important to identify the key questions to ask when deciding on assessments for language, literacy, numeracy, social-emotional behaviour, and the home and learning environments for children in the early years. In the following section we outline these key questions. The subsequent report then develops these themes providing evidence and examples. A database was developed alongside this report which indicates which assessments meet the set criteria for inclusion.

2. Key questions

Does the measure reflect the target domain?

To answer the question it is necessary to understand what the key competencies are that underpin or measure the skills targeted. This understanding should be informed by current research findings. There is consistent evidence across both child and environmental measures that domain-specific measures, rather than general measures, are more sensitive to tapping key constructs and are more sensitive to change. Reviews of the domains are provided in Section 4.

Is the measure appropriate for the target population?

There are a number of challenges in assessing children under the age of six. Typically, they need to be assessed one-to-one and time for assessment is limited by their skills and their ability to attend to the task. Lack of skills in areas necessary for completing the task, such as oral language, written language, or motor skills, may result in unreliable results. For example, the majority of tests require children to follow oral instructions. In using tests, assessors must be alert to the possibility that children's failure of items in the target domain, for example literacy and numeracy, may reflect difficulties in understanding the test language rather than insecure or inadequate knowledge of the target domain.

For children in this age-range there will be some skills, such as numeracy and literacy, which are just beginning to develop. These early stages will be more appropriately assessed using criterion-referenced tests. When norm-referenced tests are used, norms should be provided for a relevant U.K. population so that appropriate conclusions can be drawn about the child's performance. Where children are at the top or bottom of the age-range recommended for the test, the assessment may be less sensitive. Section 3 addresses this issue.

Are criteria developmentally appropriate?

Criterion-referenced tests are fixed to a set of predetermined *criteria* which include concise, written descriptions of what students are expected to know or do at that specific point in development. By contrast, normed-referenced tests compare an individual to similar peers of the same age. Norm-referenced tests need to demonstrate they are valid measures of the construct.

Does the test conform to minimum psychometric properties?

Measures need to provide data that is robust and allows analyses over time. Because of the inherent psychometric problems associated with age and grade equivalents that seriously limit their reliability and validity, these scores should not be used for making diagnostic, placement decisions, or for tracking development (Bracken, 1988; Reynolds, 1981). Standard scores are a more accurate representation of an examinee's ability because they are based not only on the mean at a given age level but also on the distribution of scores. As such, there are minimum psychometric properties that tests must contain (Section 3 in the report provides details).

It is important to consider both the reliability and validity of the measures used. There are four main types of reliability: test-retest, inter-rater, parallel form, and split-half (see Rust and Golombok, 1991). All published tests should report, at a minimum, test-retest reliability, which demonstrates that a test gives consistent results over a short period (for example, one week). Where scoring is not completely objective—for example, using checklists or coding qualitative data—inter-rater reliability should be reported. This helps ascertain whether different raters (scorers or assessors) would provide the same scores.

The validity of a measure can also be examined in a number of ways (see Rust and Golombok, 1991) of which the most important are concurrent and predictive validity. 'Concurrent validity' establishes how well a test compares to another measure of the same construct; 'predictive validity'

concerns how well an assessment predicts future performance on the construct. We found few measures in our review that included predictive validity for performance in children under six. Therefore, in our star ratings, we focused on concurrent validity, but where measures of predictive validity were available, these are reported.

Does administration of the test or tool require any special considerations?

Some assessments, specifically one-to-one tests used by psychologists or speech and language therapists, require special accreditation before use. This can limit their utility in intervention studies and in classrooms.

For one-to-one assessments of young children, it is important to ascertain that the child can attend for the full duration of the assessment protocol. It may be necessary to break up testing into several sessions: the younger the child, the shorter each session should be.

Special consideration needs to be taken about training for observational schedules, and assessments should be made of the reliability of the observers' judgments. Finally, both transparency of coding and scoring should be considered. Scoring criteria should be explicit and unambiguous. Where raters need to score children's responses—such as narrative language—clear guidelines are required to ensure the robustness of the assessments.

3. The assessment of young children's competencies

A number of different types of tools may be used, either alone or in combination, to assist in assessing or profiling early skills: standardised norm-referenced tests, criterion-referenced tests, developmental profiles or checklists, observational procedures, teaching experiments, or bespoke experimental measures. Measures may either involve direct assessment of the child's skills with standardised tests or observations, or be collected by using a proxy such as a parent or teacher.

Norm-referenced tests provide information about where an individual lies on a particular ability (such as naming vocabulary) or attainment (such as reading or numeracy) in comparison to peers of the same age. The measurements produced are of varying degrees of sophistication: ranks, centiles, and standardised scores. Rank and centile scores are highly susceptible to over-interpretation (Tindal and Marston, 1986) and percentiles are not appropriate for statistical analyses of group data over time (Kline, 1986).

Norm-referenced tests should present evidence about the standardisation sample and measures of reliability and validity (Rust and Golombok, 1999). Assessments can report different types of reliability, but to ensure the robustness of the measure, reliability should be reported as it demonstrates the consistency of the measure. Of particular importance for measures concerning children are measures of external reliability that indicate whether the test varies in usage. The two main measures are inter-rater reliability and test-retest reliability. 'Inter-rater reliability' refers to the extent that different raters give consistent estimates of the child's performance or behaviour; 'test-retest reliability' refers to the extent to which the test will produce the same results when measurement occurs at a close point in time. Test-retest reliability coefficients (coefficients of stability) vary between 0 and 1, where 1 indicates perfect reliability and coefficients below 0.7 indicate unreliability. For tests in the qualitative synthesis of assessments, at least one measure of reliability was reported.

There are also a number of different forms of test validity. Validity can be assessed by comparing a test with a similar test designed to measure the same functions at the same point in time ('concurrent validity')—so one reading test might be compared with another. Alternatively, the results of the test can be compared with another measure of the attribute ('construct validity'). For example, results on an expressive vocabulary test might be validated against a receptive vocabulary test. Test constructors may report different types of validity and reliability, but to ensure the robustness of the measure, such data should be available. The existence of such data has been incorporated into the star ratings for psychometric properties.

Tests can only provide appropriate norms if they are used for the population for which they were intended. The comparability between the standardisation sample used to determine normal values of the test and the sample that is being assessed are key for normative comparisons. The standardisation sample is the basis for establishing national norms. Although the sample is smaller than the population it represents, it cannot be too small or it will be biased. Test developers select a sample of participants they believe represent the relevant population. According to the Standards for Educational and Psychological Tests (American Psychological Association, 2014), the number of participants in each group should not be less than 100. The standardisation sample needs to be based on a systematic stratification of socio-economic status, geographic region, ethnicity, and gender (see for example Cicchetti, 1994) which allows population norms to be created. The standardisation sample must be large enough to provide stable estimates of the population average and variance. Very small normative samples are problematic because too much depends on the performance of the few participants tested, and in these cases, participants with very high or very low scores (outliers) have undue influence on averages and standard deviations.

The criteria outlined above are the basis for both national and international guidance for interpreting normative data from tests. In the U.K., care must be taken if U.S. norms are used: these are unlikely to be appropriate for U.K. samples for a number of contextual and psychometric reasons. For

example, as the U.S. standardisation sample will have included children from very different social contexts with different educational opportunities, ethnic composition, and geographical distributions, patterns of performance may vary. By corollary, norms based on a sample of monolingual children may not be appropriate for bilingual children or for children who are English language learners. Norms should also be as current as possible: norms become outdated by about three points a decade (Trahan, Stuebing, Hiscock and Fletcher, 2014), commonly referred to as 'the Flynn effect'. In addition, norms may well be influenced by changes in teaching practices. So for example, the current U.K. emphasis on phonics has led to children performing much better on phonics tests than was previously the case. As such, it is recommended that measures are regularly re-standardised. Data for test reliability and validity should be available in relevant technical appendices. Alternative ways of sourcing this information is through the peer-reviewed research literature, however, as we found, a significant number of assessments are published with little supporting psychometric information and with no access to technical appendices. This questions their utility for screening, assessment, and measuring change.

Criterion-referenced tests are concerned with the skills required to complete a task or set of tasks. The criteria are used to identify performance in relation to task-specific skills, such as, 'can count to ten orally, in order, and unaided'. Such tests help identify whether or not a child possesses some particular skill or competence and may allow for the analysis of error patterns. These tests are not designed to discriminate among individuals in the same way that norm-referenced tests do; rather, they are designed to provide a clear indication of what a child can and cannot do and thereby are a guide to what skills should be taught next, and if used in an intervention study, whether the target has been achieved.

A test can be both norm-referenced and criterion-referenced. Criterion-referenced tests are characterised by the deliberate attempt to yield measurements that are directly interpretable in terms of specified performance standards. Despite this emphasis, there is no single type of test that can be identified as the prototype criterion-referenced test. The assessment tool will depend on the nature of the domain and the focus of the scale. A well-defined domain is necessary to identify potential test items. Thus, to develop a criterion-referenced tool for reading we must know what the constituent skills are for reading at a specified point in development. As we shall show, both letter-naming and phonic skills are important in the initial stages of learning to read. However, the tool might also focus on different dimensions of the reading process and this would result in the construction of scales with different items, such as proficiency (how fluently reading proceeds) or the developmental prerequisites to reading such as oral language skills.

Although standardised, norm-referenced assessments are still the most widely used means of evaluating children's abilities, 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 and measuring change (Grigorenko, 2009). There is, therefore, an increasing interest in an interactive approach to conducting assessments that focuses on the child's ability to respond to intervention—the child's capacity for change or 'modifiability' (Cho, Compton, Fuchs, Fuchs and Bouton, 2014; Fuchs, Compton, Fuchs, Bouton and Caffrey, 2011; Seethaler, Fuchs, Fuchs and Compton, 2012; Tiekstra, Minnaert and Hessels, 2016). This approach is called 'dynamic assessment', and, unlike traditional testing, it employs a 'test, teach, retest' procedure to assess the child's learning processes.

Dynamic assessment is considered more culturally fair to those from different linguistic or cultural backgrounds (Lidz and Peña, 2009) and may be more sensitive for measuring change in language over time (Hasson and 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 and Gillam, 2007). While several types of dynamic assessment have been developed in recent years, including for vocabulary (Camilleri and Botting, 2013), syntax (Hasson, Dodd and Botting, 2012) and narrative (Peña, Spaulding and Plante, 2006), as well as more

omnibus tests (Hasson *et al.*, 2013), only one language test is commercially available (Miller, Gillam, and Peña, 2001).

Criteria for choice of assessments

The data derived from these different assessments has implications for analyses and interpretation. For example, the Education Endowment Foundation (EEF) recommends in its [analysis guidance](#) that evaluators control for prior attainment using a regression model. It is therefore important to consider which pre-test to use, and whether or not it is possible to use National Pupil Database data (please see the [EEF's paper comparing the NPD to standardised tests](#)). In addition, it is important to be aware that the nature of the assessment chosen to measure the outcome of an intervention has implications for the likely results and size of effects reported. For example, experimenter-designed assessments consistently result in higher effect sizes than standardized measures (Rosenhine and Meister, 1994).

To identify inclusionary criteria for measures to be evaluated in qualitative synthesis, we drew on the research literature addressing psychometric features of assessments, specific requirements for evaluating interventions, and guidance from experts in the field. A series of inclusionary and exclusionary criteria for the measures designed to assess children under six years of age was developed and evaluated by the steering committee. Both norm- and criterion-referenced assessments were included, but subject to different inclusionary criteria. Only norm-referenced tests with U.K. standardisations were included in the final selection, while criterion-referenced tests were included irrespective of their country of construction. A stepped inclusion criterion was used and this is described in the Methods section.

4. Domains

The domains of interest were specified in the tender: language and communication, literacy, numeracy, social-emotional development, and the home and early years learning environment. As a first step, we conducted an initial review of the research literature to establish the potential competencies to be considered within each domain. These are outlined below.

4.1 Language and communication

It is important to distinguish between 'language' and 'communication'. Overviews of the development of communication and language can be found in recent texts (see for example Saxton, 2010) and a recent EEF report also focuses specifically on early language development (Law, Charlton, Dockrell, Gascoigne, McKean and Theakston, 2017), providing an overview of typical language development.

Communication is the transmission of information, and infants communicate from the minute they are born. Early infant communication is a precursor to later language and social development and these early skills can be examined. Interactional competencies that develop in the first two years of life provide a foundation for further social and communicative developments. Included within social communication are, for example, joint attention, which is the ability to co-ordinate attention with a social partner around an event or object, and discourse management, which includes the appropriate use of behaviours such as turn-taking and the use of gestures. The use of gestures, for example, emerges between seven and 15 months (Liszkowski *et al.*, 2012) and intentional pointing from 12 months (Tomasello, Carpenter and Liszkowski, 2007). These skills are assessed both by observation and caregiver or professional reports. There is considerable variation in when children develop these skills thus specifying developmental milestones is difficult. Moreover, many of the current checklists in use have not been validated psychometrically in terms of reliability and validity (see for example Mok and Lam, 2011) making the assessment of these communicative skills challenging.

Language, in contrast, is primarily a representational system that emerges as the child's cognitive skills scaffold understanding and organisation of the world. These skills provide a foundation for learning and socio-emotional development and involve both understanding language ('receptive language') and using language ('expressive language'). The language system is itself composed of a number of subcomponents that are important for effective understanding and communication and differentially impact on later development. 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).

As children's language develops, they produce extended and coherent oral narratives which allow them to communicate with others effectively and efficiently. All of these subcomponents of the language system can be assessed independently, focusing on either receptive or expressive skills, or can be used to provide a holistic measure. However, some researchers have argued that much of the variance in standardised measures of language appears to be attributable to a single common factor pointing to the need for using language assessments that include a combination of language subcomponents and provide an overall language outcome (Language and Literacy Research Consortium, 2015; Tomblin and Zhang, 2006).

Conventional approaches to assessing oral language

Assessments of language development typically focus on three domains: vocabulary, grammar, and social communication. Vocabulary measures are commonly used to assess children's language skills. Both the child's understanding of words (receptive vocabulary) and their ability to use words (expressive vocabulary) can be assessed. Typically, forced-choice receptive measures are used but these provide limited information about the child's vocabulary skills and are not informative about language delays more generally. Research has consistently indicated that vocabulary scores cannot be used as if they were indicators of general language ability (Gray, Plante, Vance and Henrichsen, 1999; Spaulding, Hosmer and Schechtman, 2013). By contrast, more complex measures assessing depth and breadth of vocabulary often include oral definitions. While these measures provide a greater insight into the child's lexical representations, the measures can conflate other aspects of expressive language with vocabulary knowledge.

Grammatical knowledge can also be assessed both by receptive and expressive measures. Sentence repetition tests are quick and easy to administer, and also easy to score because, in contrast to most expressive language tests (such as tests of narrative), the targets are explicit and precisely specified (Chiat *et al.*, 2013). Depending on how they are constructed (what types of sentence structures they use), they can yield not only quantitative data (such as how many sentences a child repeats accurately) but also qualitative data (for example, what types of sentences a child finds particularly difficult, and what type of errors they make; Riches *et al.*, 2010; Komeili and Marshall, 2013).

Narrative assessments are less frequently used to assess children's language competence, but narrative is an important means of measuring language performance and is correlated with later written text production. Conventional language tests elicit production and test comprehension using artificial tasks; by contrast, narrative tasks provide a more natural 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 from three and a half to seven years, but the standardisation is old, scoring is complex, and errors are common (Haccoun, 2001). Thus, while it remains one of the most commonly used tools, training is required for administration and reliable scoring. 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 and Gillam, 2008).

The majority of assessments of oral language are not designed for, nor are they appropriate for, monitoring progress. They typically take a long time to administer, can be complicated to score, and have specifications about the length of time required between assessments due the psychometric properties of the measures. Measures considered gold standard by clinicians, such as the CELF (Clinical Evaluation of Language Fundamentals; Semel, Wiig and Secord, 2004), require clinical training and are not intended for frequent administration. Recently there have been attempts to address these challenges. For example, the Developmental Snapshot (Gilkerson and Richards, 2008) was designed to assess vocal and language skills in children aged from 2–36 months. It can be administered monthly, takes 15 minutes, and does not require any intensive training of the parents. There are currently no U.K. norms for the Developmental Snapshot, but data from the U.S. is promising, suggesting that Snapshot can work as a tool to provide parents with feedback on child language skills, reveal risk for delays, act as a progress monitor, measure intervention effectiveness, and increase parental awareness of language behaviours (Gilkerson, Richards, Greenwood and Montgomery, 2017).

Ways forward in assessing language and communication

The multifaceted nature of language and communication requires a decision about which subcomponents should be evaluated. Each component involves both receptive and expressive aspects and the assessment of one subcomponent without drawing on another is challenging. Social

communication skills are least likely to be evaluated from a standardised perspective, but are likely to be important elements in later social-emotional and behavioural development. The use of universal screening instruments—to assess pre-school children's language and communication abilities, including appropriate communication in context in order to target interventions—have not proved sufficiently reliable to date for their use to be recommended (Siu, 2015). The high degree of variability in the nature of children's language trajectories in the preschool years is particularly challenging (Reilly, McKean and Levickis, 2014; Reilly, Tomblin, *et al.*, 2014; Reilly *et al.*, 2010), and research highlights the need to match comparison samples using the most reliable and valid measures to address the efficacy of any intervention.

There are a number of sources that review language measures (Dockrell and Marshall, 2015; Friberg, 2010; Hoffman, Loeb, Brandel and Gillam, 2011; Law and Roy, 2008) or provide wider guidance on the assessment of language difficulties (McCauley, 2001). Swineford and colleagues (Swineford, Thurm, Baird, Wetherby and Swedo, 2014) have provided an overview of measures available to assess pragmatic language. There is increasing evidence to demonstrate that the exposures children have to language, both at home and in early years settings, have a significant impact on their language trajectories; assessments of language should, therefore, also capture the language-learning environment (Hoff, 2003; Law *et al.*, 2007; Pullen and Justice, 2003; Weisleder and Fernald, 2013).

4.2 Literacy

Literacy includes reading and writing. The only element of writing covered here is spelling as although as children mature grammar and composition become important, for the under sixes these dimensions are insufficiently developed to reliably assess.

Reading

The broad underlying dimensions of reading have been usefully captured in the Simple Model of Reading (Gough, 1996), which proposes that reading comprehension (RC), in lay terms what we typically mean by 'reading', is a combination of word reading efficiency (WRE) and language comprehension (LC)—or $RC = WRE \times LC$. In other words, in order to read, the child needs to be able to read single words and also to understand what they are reading at sentence and text level. The understanding aspect is equated with language comprehension in the Simple Model of Reading. This model is very widely accepted and provides an overarching conceptualisation of the domain. (Language comprehension is discussed in Section 4.1.) Reading assessments typically assess not only WRE but also the child's understanding of the text, always referred to as 'comprehension', and it can be argued that there is more to the comprehension side of reading than is captured in language assessment (Hurry and Doctor, 2007). However, the assessment of comprehension in the medium of reading is reliant on children's ability to read words. Children under five typically are not efficient word readers, and even children under six are not sufficiently fluent word readers to enable the effective measurement of reading comprehension. The focus here will therefore be substantially on WRE, also referred to as early or emergent literacy (McCachlan and Arrow, 2014; Norwalk, Di Perna and Pui-Wa, 2014).

Children read words either through sight word recognition or through some aspect of alphabetic decoding that converts letters to sounds. Opinions differ as to the extent of the separation of sight word reading and other methods of decoding, but there is a widely accepted developmental sequence: some aspects of sight word recognition are normally considered to come first (the pre-alphabetic phase), followed by increasing use of the alphabetic decoding strategies (the partial alphabetic stage), followed by the full alphabetic stage (Ehri, 1995). Sight word recognition makes use of the visual features of words; alphabetic decoding requires children to understand and apply the link between letters and sounds, referred to as the alphabetic principal (Byrne, 1998). Ehri's 'partial alphabetic phase' is an acknowledgement that understanding the connection between letters

and sounds is not an all-or-nothing event and that children take time to merge these different strategies for reading, initially making use of partial connections between letters and sounds. According to Ehri (1994), the pre-alphabetic stage is typically observed in preschool; partial alphabetic reading begins normally at the start of schooling, but may be developed beforehand; and full alphabetic reading happens in the first two years of school (Ehri's 'full-alphabetic phase'), developing in later years into secure and fluent word reading (Ehri's 'consolidated-alphabetic phase', also referred to by Frith (1985) and others as the 'orthographic phase').

Core reading skills

The core skills of WRE are:

- early or emergent skills, comprising letter knowledge and phonological skills;
- word reading accuracy; and
- word reading efficiency—combined speed and accuracy (McClachlan and Arrow, 2014).

These are described below.

Letter knowledge

Letter knowledge has been consistently found to predict reading skills (see Foulin, 2005 for a review), evident in children up to age 9–10 years (Senechal, 2006). There are at least two aspects to letter knowledge: letter name knowledge (LNK) and letter sound knowledge (LSK). LNK is implicated in Ehri's pre-alphabetic phase and has been shown to have a role in early reading and spelling (Ehri and Wilce, 1985), perhaps, in part, providing a link between orthographic and phonological cues before children have securely acquired the alphabetic principle (Treiman and Rodriguez, 1999). LNK aids LSK (Kim, Petscher, Foorman and Zhou, 2010). LSK helps children attempt to read unfamiliar words and make phonemically correct attempts at spelling (Arrow, 2007). However, in order to break the alphabetic code, children need not only LSK but also phonological awareness (Foorman *et al.*, 2003). Schatschnieder, Fletcher, Francis, Carlson, and Foorman (2004), in a longitudinal study of typically-developing readers in the U.S.A., found that LNK was predictive of first grade (6–7 years of age) reading outcomes only at the beginning of kindergarten (5–6 years of age) whereas LSK was predictive at the beginning and end of kindergarten. By first grade, neither LNK nor LSK predicted subsequent reading outcomes because most students knew virtually all letter names and sounds at the beginning of the year. The acquisition of these skills will be influenced both by nationally recommended teaching guidelines and age of school entry.

Phonological awareness

Phonological awareness is the ability to hear and manipulate the sounds in language and follows a rough developmental sequence (Anthony and Francis, 2005) starting with an awareness of, and ability to manipulate, the larger sound chunks of words and syllables, before developing the ability to recognise the smaller units in different positions in the word, with an awareness of rimes, alliteration, and phonemes. This entails an ability to identify the various units of sound and then to manipulate them in operations of sound deletion and blending. Assessments of phonological awareness typically include measures of the awareness of rhyme and alliteration, and the ability to manipulate sounds—as in spoonerisms, phoneme segmentation, deletion, and blending:

- Spoonerism is the transposition of initial sounds—for example, *cat* with a /f/ gives *fat*.
- Phoneme segmentation is the separation of spoken words into their phonemes—for example, segmenting *cat* into /c/, /a/, and /t/.
- Deletion involves saying what a word will sound like on the removal of a sound, for example, *stop* remove /t/ is *sop*.
- Blending is identifying a word from its constituent sounds (phonemes or syllables) in isolation—for example, /c/, /a/, /t/ spells *cat*.

In addition to assessing the child's ability to perform these operations, phonological tests may assess the speed at which they can do this (fluency) through Rapid Automatic Naming. Children can be assessed on their ability to recognise rhyme and alliteration as young as three years of age and these skills are incorporated in a number of language and literacy assessments, for example, the Phonological Assessment Battery (PhAB—see Table 5).

Word reading

Word reading accuracy is typically measured through word reading and non-word reading tests. The former usually involves a list of words of increasing difficulty, both phonetically regular and irregular—the latter usually a list of non-words. Word reading tests do not typically differentiate between sight word reading and non-lexical decoding since children can use either method to read the words. It is for this reason that non-word reading tests are useful as children must use non-lexical strategies to decode these unfamiliar words. Word reading efficiency, or fluency, is a measure of how quickly the child can accurately read words and non-words, obtained by timing the speed at which the child correctly reads a specified list of words. Measures of fluency are really most useful for children older than six years as they assess the more mature, automatized reading which characterises the consolidated-alphabetic (Ehri, 1994) or orthographic (Frith, 1985) phase. Typically, children will begin to learn to read words from the beginning of formal schooling and to read non-words using phonic strategies slightly later. This reflects the developmental sequence referred to above, where sight word reading typically precedes phonological decoding.

Conventions of print

While it is not possible to test preschoolers' reading comprehension, it is proposed by Clay (2002) that measuring children's understanding of the concepts about print provides a measure of their knowledge of the conventions of print—such as: print and not pictures carry the message, directionality, spoken to written word correspondence, and the meaning of various aspects of punctuation.

Spelling

Spelling skills are typically quite highly correlated with reading—for example, Ehri (1987) reports correlations ranging from 0.66–0.90. This is unsurprising because the two aspects of literacy both rely on similar skills. In the initial stages of spelling, which concern us here, those skills rely substantially on letter knowledge and phonological awareness. Treiman and Cassar (1997) summarise the principle developmental theories of spelling development (Ehri, 1986; Gentry, 1982; Henderson, 1985). The various theories differ in some details, but they all recognise that early spelling development is characterised by an increasing use of phonic strategies. As spellers mature, they build a greater repertoire of sight spelling and begin to understand the meaning relations among words (morphemes) which they use to help them spell. Ehri (1986) identifies three broad stages of spelling development, similar to her stages of reading development mentioned above:

1. semi-phonetic (typically 4–6 years): spellers begin to distinguish sounds or syllables, usually the first or last, with letters that match their letter names—R (are), U (you), J-F (giraffe);
2. phonetic (typically 5–7 years): the child can symbolise the entire sound structure of words in their spellings, and the letters they choose are more conventional, but the letters are assigned strictly on the basis of sound—GERAF (giraffe) (Treiman and Cassar, 1997); and
3. morphemic (6–11 years): the child becomes more aware of conventional spelling, employing visual and morphological information in spelling—for example, children learn to represent the 't' sound at the end of past tense regular verbs with 'ed' (Nunes, Bryant and Bindman, 1997).

Prior to the start of schooling (age 3–5 years) children will make up writing, beginning with scribbles and increasingly encompassing letters or letter-like symbols, but with little evidence of LNK or LSK.

This is referred to as 'emergent writing' (Clay, 2002), the 'pre-alphabetic stage' (Ehri, 1986), or the 'precommunicative stage' (Gentry, 1982). The semi-phonetic phase only normally commences with the beginning of formal schooling and is slower to take off (in terms of numbers of words correctly spelled) than reading (in terms of numbers of words correctly read) which is to be expected as it requires the child to produce a spelling rather than recognise a word. Thus assessing spelling below the age of six is very limited, with an absence of appropriately standardised norms of word spelling, or even ability to write letter sounds and blends (as in the WIAT-II, Wechsler D., 2005). Tests that allow children to choose their own words to spell (as in the Written Vocabulary Test, Clay, 2013) or to write the sounds in words (as in Hearing Sounds in Words, Clay, 2013) allow children under six years a greater opportunity to show what they can do.

Conventional approaches to the assessment of literacy

By far the most common and reliable form of assessment of children's literacy skills is through standardised tasks, both visual and oral. For children aged two to six years, assessment needs to be individual as they will be typically unable to record their responses (for example, the administration of Performance Indicators in Primary Schools (PIPS), Demetriou, Merrell and Tymms, 2017).

There have been assessments of young children's literacy that are completed by teachers using observation and a checklist approach, notably, in England, the Early Years Foundation Stage Profile (EYFSP). While these criterion-referenced tests may be useful for teachers managing children's learning, they are not recommended as a primary measure for the evaluation of interventions, though collecting this data may have value for links to policy and practice. In the view of the Department for Education, 'The EYFS Profile is not intended to be used for ongoing assessment or for entry-level assessment for Early Years settings or Reception classes' (DfE, 2014, p. 5). The known psychometric properties of the test confirm this position. The current version of the EYFS Profile, introduced in 2012, comprises 17 items measured on a three-point scale. Children's performance on this profile are recorded on the NPD, which may be generally useful for evaluators to consider, as outlined above, however, only two items address literacy (reading and writing) making this a very crude scale, and there is no data publicly available on the reliability or validity of the items.

Ways forward in assessing early literacy

Besides the need to conduct individual assessment, tasks need to be framed in a context which is meaningful to children. Assessing children younger than three years old on these tasks is inadvisable as they are unlikely to have any of the necessary skills. Because the tests involve children listening to sounds, it is very important that they are tested in a quiet location and that appropriate actions are taken to ensure the participants have no hearing problems.

A further challenge is negotiating floor and ceiling effects. Most children are unlikely to be able to read words or non-words until they have been in school for a term or two and therefore testing reading is likely to result in a floor effect. However, during their first year in formal schooling they will improve dramatically in aspects of the emergent reading skills outlined above, notably in LNK, so a test which is appropriate for a four-and-a-half-year-old may fail to discriminate between children a year later due to a ceiling effect. A test of word or non-word reading which shows a floor effect at age five might be useful a year later—and as the child matures. While tests of phonological awareness have a larger age-range than conventional reading tests for this young age group, depending on the purpose of the assessment, measures of reading (both WRE and comprehension) are likely to be desirable for tests of effectiveness spanning the early school years. Baseline assessment of reading can therefore be useful—even where there is little discrimination between five-year-olds due to floor effects—as it establishes a common platform for the assessment of later progress.

Criterion-referenced assessments will provide greater clarity in exactly what the child can and cannot do.

Children's skills in different aspects of 'emergent reading' and reading will be influenced by the instruction they have experienced. Children exposed to phonics instruction will apply those strategies; children taught using sight words techniques will be more adept in sight word reading. Assessments should be selected with an awareness of the instructional context.

There is a fairly straightforward array of relevant literacy skills to be assessed in three- to six-year-olds. Tests should include assessment of:

1. LNK and LSK: these skills are tested very simply and quickly, typically in under five minutes.
2. Phonological awareness: a full battery of phonological tests can take 20–30 minutes to conduct, however, if time is limited, selecting key subtests provides a useful strategy, probably most suitably involving awareness and manipulation of larger sound chunks (words and syllables) and awareness of smaller sound chunks (rhyme and alliteration).
3. If the testing is a baseline for a longitudinal study going into Year 1, it might be useful to assess word and non-word reading at baseline, even though many children may score nothing or very little, as it will provide a measure which can be re-used at follow-up, demonstrating progress on the same measure over time. However, such measures should be used in addition to, rather than instead of, the tests of letter knowledge and phonological awareness as they are unlikely to capture the variation between children at baseline. The phonics screening check—recorded on the NPD—comprises 20 words and 20 non-words and was specifically designed for Year 1 children (5–6-year olds): in 2016, 81% of children (32 out of 40) met the expected standard at the end of Year 1 (DfE, 2016) suggesting a good level of sensitivity for that point and therefore a useful outcome measure at six years old. Lack of data about performance of younger children makes it problematic as a baseline measure for early years interventions.

4.3 Numeracy

There is general agreement that in the preschool years many children begin developing the symbolic number skills of counting, transcoding, comparing number magnitudes, and simple arithmetic that are required for more advanced numeracy (Geary, 1994; Siegler and Braithwaite, 2017). The development of these skills is a function of the product of instruction and the child's general cognitive skills: with no instruction there is no development, however cognitively skilled the child. Symbolic number skills and general cognitive skills continue to develop during primary school.

What is less agreed is the relationship between symbolic number skills and non-symbolic skills. Non-symbolic number skills include discriminating between sets—for example, identifying the more numerous of two sets of objects without counting—and being surprised by impossible results of transformations, such as $1 + 1 = 1$. Some have inferred these non-symbolic number skills from studies of the looking times of human babies (Feigenson, Libertus and Halberda, 2013; Wynn, 1992) but the inference of skills from looking times is claimed to be unsound and challenged by evidence from the study of toddlers who lack the skills that have been credited to infants (Huttenlocher, Jordan and Levine, 1994; Mix, Levine and Huttenlocher, 1997; Moore and Cocas, 2006).

Non-symbolic skills deficits have been claimed to underlie mathematical learning difficulties, or dyscalculia (Mazzocco, Feigenson and Halberda, 2011), but this, too, is challenged (Iuculano, Tang, Hall and Butterworth, 2008). Similarly, there are mixed results about the prediction of later symbolic number skills from earlier non-symbolic number skills (Göbel, Watson, Lervåg and Hulme, 2014; Starr, Libertus and Brannon, 2013). Perhaps part of the problem about non-symbolic skills is the result of measurement problems (Gilmore, Attridge and Inglis, 2011).

Correlational studies are invariably limited in determining the influence of particular skills on others. They need to be supplemented with plausible and empirically supported theories of how the skills affect others. Evidence from properly conducted intervention studies has the potential to strengthen

the causal inferences made. Currently, there is no clear support for the view that training young children to improve their non-symbolic skills benefits their symbolic number skills (Räsänen, Salminen, Wilson, Aunio and Dehaene, 2009) whereas several studies show the benefits of symbolic number skill training (Siegler, 2009).

Counting

Children are first introduced to symbolic numbers through being taught the number word sequence and how to use it to count sets of objects. In counting a set of objects, the child co-ordinates the sequence of symbolic number words, starting with one, with the items in the set so that each item has one—and only one—number word. When all the items in the set have been co-ordinated with number words, the last symbolic number word represents the cardinal value, or exact numerosity, of the set.

Preschool children vary in how far they can consistently and accurately recite the number word sequence, how skilful they are in co-ordinating it to determine numerosity, and how flexible they are in counting. Flexible counting can be manifested by the ability to count objects in nonlinear arrays and by the understanding of the irrelevance of the order of counting individual items, as long as each object is counted once and only once. Gelman and Gallistel (1978) proposed that counting involved three how-to-count principles: 'the one-one principle'—that each item must be associated with one, and only one, number tag; 'the stable-order principle'—that the sequence of number tags must be consistent from one count to the next; and 'the cardinal principle'—that the last number tag represents the number of items in the set. They also proposed the 'order-irrelevance principle'—that the order in which a set is counted does not matter as long as the counting respects the three how-to-count principles and the 'abstraction principle'—that any discriminable entities can be counted.

Gelman and Gallistel (1978) asserted that children must be innately equipped with skeletal versions of the counting principles. Also, Gelman and Meck (1983) found that even preschoolers who did not abide by these principles in their own counting were adept at judging puppet counts, in particular discriminating correct counts from counts that violated the how-to-count principles.

By contrast, Briars and Siegler (1984) asserted that there was no need to credit children with skeletal principles. The widespread early emergence of counting was due to the widespread attempts by adults to teach preschoolers to count. Furthermore, their studies of children judging puppet counts indicated that even at five, many children were unclear about the difference between essential and optional features of counting.

It is typical of early number assessments to include some measure of knowledge of the number word sequence, and some items assessing the ability to count sets of objects or subsets, such as the red objects in an array of red and yellow objects. Some assessments include items requiring children to assess the correctness of a puppet's counting.

Transcoding

Transcoding is the ability to translate between the Hindu-Arabic system for representing number using digits and place value, and the verbal numeration system, spoken and written. Skill in transcoding is essential for written arithmetic. It can be assessed by tasks involving reading numbers in one system or writing them in the other. Early number assessments typically include some items assessing reading Hindu-Arabic numbers with one or two digits.

Comparing numerical magnitudes

Although the number word sequence embodies differences in numerical magnitude (for example, later number words, such as seven, refer to greater magnitudes than earlier number words, such as six) many children lack insight into the number word sequence. It is some time after being able to recite the number sequence consistently and accurately that children are able to judge whether six is more or less than seven.

In research on individual differences in the number skills of older children, both speed and accuracy of numeral comparison have been found to correlate with variation in more general arithmetic skills (Cowan and Powell, 2014). Early number assessments typically include just accuracy measures.

Simple arithmetic

Facility in simple arithmetic—addition of single digits and complementary subtractions—has long been recognised as a foundation for more complex arithmetic (Cowan *et al.*, 2011). Knowledge of combinations so that retrieval is used to solve all simple arithmetic problems has been considered to be fundamental for mathematical development (Baroody, 2006) and primary school systems have sought to develop it in pupils. The foundational role for such knowledge is, however, challenged by the variation shown by undergraduates (LeFevre, Sadesky and Bisanz, 1996), and the limited use of retrieval by some primary school children who are doing well in mathematics (Cowan *et al.*, 2011; Siegler, 1988). Accurate solution of simple arithmetic problems may be fundamental, but accurate solution of these problems by retrieval is not. There are various formats for presenting simple arithmetic problems: nonverbal arithmetic (Jordan, Huttenlocher and Levine, 1994), number-fact problems, and story problems.

Nonverbal arithmetic items involve children observing the tester represent a number with objects on a mat. The objects are then screened. The experimenter then adds or subtracts a number of objects to the set under cover and the child responds by saying, or showing on their mat, the number of objects now under the experimenter's cover. Number-fact problems are presented verbally, for example, 'How much is two and four?' Children are allowed to solve these problems by any method. Some will state their answer. Others may count, with or without using their fingers.

Story problems embed simple arithmetic problems in verbal contexts. Story problems involving addition and subtraction can vary substantially in complexity (Riley and Greeno, 1988). Most children from kindergarten onward succeed on problems where the result is the unknown, but it is not until Year 4 that similar levels of success are achieved on problems with unknown initial quantities. More complex story problems make greater demands on both mathematical and language understanding because the child has to understand the story to be able to identify the corresponding arithmetic problem. Persistent weakness in solving story problems by children with mathematical difficulties has frequently been reported (for example, Russell and Ginsburg, 1984). Early years assessments just use the simplest form of story problem that is 'result unknown' as in: 'Sally has four crayons. Stan gives her three more crayons. How many crayons does Sally have now?' Most early years assessments feature number-fact problems. Some have all three types of simple calculation problems.

Conventional approaches to the assessment of numeracy

Until recently, research expenditure on understanding typical number development has lagged considerably behind expenditure on understanding language and literacy development (Dowker and Sigley, 2010). This research is now yielding new ideas about what skills are core and how they develop. These new ideas have yet to influence the blueprints for designing or revising U.K.-standardised tests such as BAS III and WIAT II UK. This lack of influence of contemporary ideas on tests is a failing because it means that the tests do not reflect the target domain.

Another failing of existing standardised tests is that they have too few items assessing important number skills. Knowledge is not all or none: skills in counting, transcoding, and simple arithmetic develop over several years. Using too few items to assess a skill risks insufficiency in initial assessment and insensitivity to change.

If standardised tests are currently inadequate because they lack a grounding in contemporary numeracy research, then it may seem better to take tests directly from academic research. This has other kinds of risk. In academic research there is rarely any obligation on researchers to establish the reliability of the procedures they use beyond internal reliability, or to establish validity beyond

face validity. There is no expectation that the results with any individual will be used to make decisions about them or that the results with groups will be used to judge the effectiveness of programmes.

Instruments such as the Number Knowledge Test (Griffin, 2005) and the Number Sense Screener (Jordan, Glutting, Ramineni and Watkins, 2010) are developing into assessments with good psychometric characteristics although their applicability to the study of very young children is less established. The advantages of these measures over existing standardised and norm-referenced tests lie in their foundation in current knowledge of number development and the number of items per skill.

Ways forward in assessing early numeracy

In common with other domains, children below five years will vary considerably in their experience of having their number skills assessed. More than other domains, early years teachers may lack expertise in assessing number skills. The implications of these are that number skills assessments of very young children are likely to be inaccurate and so genuine improvement may be difficult to determine.

Assessing cognitive skills like numeracy requires children's understanding of what they are being asked to do, willingness to co-operate, and effort. Researchers are free to exclude children who are unwilling or unable to comply with assessment, but this is not a good solution for intervention studies.

Good psychometric practice requires that all measures used should avoid zero scores (Grigorenko and Sternberg, 1999), however, scores of zero may be genuine reflections of the child's knowledge if valid measures of preschool numeracy are used with children as young as two. If very young children are to be assessed then it may be worthwhile to develop adult informant checklists to supplement child-based measures (see section on 'home learning environment', page 25).

A good assessment of children's number skills would assess competence in each number skill. It would include at least one item in each skill that all children pass. It would include another five items that vary in complexity

4.4 Social-emotional development

Early social and emotional development is broadly defined as the capacity of the young child to form close and secure adult and peer relationships, experience and regulate emotions in socially and culturally appropriate ways, and explore the environment and learn. Much research suggests that skills within the social and emotional domain play a critical role for later life success across a wide variety of outcomes (Heckman *et al.*, 2006; Jones *et al.*, 2015). The challenge, however, despite many years of research on early socio-emotional skills, is how they should be measured. This is due, in part, to a weak consensus about which constructs should be used to capture them (Halle and Darling-Churchill, 2016). Perhaps the most widely used classification of these constructs is into four, theoretically distinct, domains: social competence, emotional competence, behaviour problems, and self-regulation. For children under five, these are related to the three broad [SPECTRUM](#) domains of 'motivation, goal orientation and perseverance', 'social and emotional competency', and 'mental health and wellbeing'. SPECTRUM was commissioned by the EEF and is an online database of non-academic skills and essential skills-measures based on a review by the University of Manchester.

Social competence

'Social competence' refers to the extent to which young children are effective in their social interactions with others, both peers and older children or adults, including being able to make and

sustain social connections, demonstrate co-operation with peers and adults, and adjust their behaviour to meet the demands of different social contexts such as playgroups and families. Peer relationships and prosocial behaviour have received much attention, especially in social developmental psychology.

Peer relationship problems, such as exclusion or rejection, are very strongly related to mental health problems (Arseneault, Bowes and Shakoor, 2010; Arseneault *et al.*, 2006; Fekkes *et al.*, 2006; Gini and Pozzoli, 2009). Via their association with stress and depressive symptoms, they can also impair cognitive functioning (Holmes, Kim-Spoon and Deater-Deckard, 2016; Vaillancourt *et al.*, 2011) and academic achievement (Buhs, Ladd and Herald, 2006). Exclusion or rejection by peers is typically studied in the context of bullying (Godleski, Kamper, Ostrov, Hart and Blakely-McClure, 2015; Wolke, Woods, Bloomfield and Karstadt, 2000). Bullying is a distinct type of aggression, characterised by a repeated and systematic abuse of power (Olweus, 2013). It can encompass both physical and verbal aggressive actions, and can be direct (for example, hitting) or indirect (spreading rumours). There is much research into the relationship between bullying involvement (whether bully, victim, bully-victim, or 'neutral' status) and health and behaviour (for example, Arseneault *et al.*, 2006; 2010; Olweus, 2013).

Prosocial behaviour is a cluster of traits and behaviours including helping, sharing, comforting, and co-operating. Prosociality benefits others and promotes positive social relationships, but is also related to subjective well-being in both children (Eisenberg, Fabes and Spinrad, 2006) and adults (Weinstein and Ryan, 2010). Its relationship with mental health, especially in children, is more complex. While prosocial behaviour is negatively related to 'acting-out' behaviours, it co-occurs with both high and low anxiety and depressive symptoms (Nantel-Vivier, Pihl, Côté and Tremblay, 2014). It does appear, however, that engagement in prosocial behaviour is linked to subsequent social information processing, an important determinant of children's social adjustment (Crick and Dodge, 1994), such that children engaging in prosocial behaviour show benign attribution biases and socially competent response strategies (Laible, McGinley, Carlo, Augustine and Murphy, 2014; Nelson and Crick, 1999). Laible *et al.* (2014) showed that engaging in prosocial behaviour likely evokes positive responses from others, which, in turn, cements children's positive internal working models and trust in the goodness of others.

Emotional competence

'Emotional competence' refers to the ability to understand the emotions of self and others, read emotional cues, react to others' emotions, regulate one's own emotions, and understand the consequences of one's own emotional expressiveness. Thus, social and emotional competence are strongly inter-related. Most research on emotional competence and its development in young children has focused on the development and consequences of young children's ability to understand others' feelings ('emotion understanding') and to manage their own ('emotion regulation'). With respect to emotion understanding, much research has shown that children who understand emotions are more prosocial, have better relationships with their peers, are more liked by them, and are rated as more socially competent by their teachers. Emotion understanding allows young children to react appropriately to others, thus enabling them to function appropriately and effectively in a social world (Denham *et al.*, 2003). Emotion regulation, which 'consists of the extrinsic and intrinsic processes responsible for monitoring, evaluating, and modifying emotional reactions [...] to accomplish one's goals' (Thompson, 1994, pp. 27–28) is critical in initiating, motivating, and organising adaptive behaviour, but also in preventing maladaptive behaviours (Cole *et al.*, 2004). For example, much research has shown that both children with anxiety and depressive symptoms and those with externalising symptomatology (see below for definitions) show poor emotion regulation (Eisenberg *et al.*, 2001).

Behaviour problems

This domain has attracted the most research interest—partly because of the strong empirical links between early behaviour problems and many adverse life outcomes later on. Among young children, ‘behaviour problems’ is sometimes used interchangeably with ‘emotional and behavioural problems’ (‘difficulties’) and ‘social, emotional and behavioural difficulties’. Research on the presentation, nosology, and epidemiology of clinically significant behaviour problems in preschool children from age two years on (when, typically, behaviour problems can first be measured) tends to focus on the five most common groups of childhood psychiatric disorders: attention deficit hyperactivity disorder, oppositional defiant disorder, conduct disorder, anxiety disorders, and depressive disorders. There are several approaches to classifying behaviour difficulties in preschoolers and determining the boundaries between normative variation and clinically significant presentations. Most emphasise the importance of using empirically-derived checklists, one of which, the Strengths and Difficulties Questionnaire (SDQ, Goodman), is very widely used in the U.K. In these approaches, behaviour problems are typically classified as ‘internalising’ (such as worry, fear, anxiety, sadness, shyness, or social withdrawal) or ‘externalising’ (such as hostility, disruptiveness, noncompliance, or aggression). It is important to note that behaviour problems and social or emotional competence are not mutually exclusive: some children will be emotionally or socially competent but also exhibit behaviour problems; similarly, children who are ‘problem-free’ may still show low emotional or social competence. It is also important to note that many problems evident early are transient, reflecting, perhaps, a difficult developmental transition, a reaction to a stressful life event, or family turmoil. Concerns have been raised about both over-identification of problem behaviours that are time-limited, and under-identification of serious problems. Longitudinal studies are necessary to differentiate between transient adjustment difficulties and more persistent and potentially severe problems (Campbell *et al.*, 2016).

Self-regulation

‘Self-regulation’, sometimes also called self-control or self-discipline by personality researchers, can be considered from both behavioural and neural systems perspectives and it concerns (a) emotion-related processes of reactivity and regulation, and (b) cognitive processes, including working memory and attention deployment. The consensus is that children’s self-regulation indexes their ability to manage or modulate positive and negative emotions—to inhibit or control their behaviour, and to shift and focus their attention. Behaviour problems and self-regulation are therefore closely inter-related, not least because difficulties in self-regulation manifest themselves as behaviour problems and a lack of social and emotional competence. In preschoolers, competence in self-regulation is inferred sometimes from performance in tasks measuring the temperament dimensions of emotion regulation and ‘effortful control’, and sometimes from performance in tasks measuring ‘executive function’ and emotion regulation. Executive function and effortful control share considerable similarities and overlaps in definitions, core components, and measurement. To simplify, effortful control is primarily the focus of temperament research (Kochanska, Murray and Harlan, 2000), and executive function, that of cognitive neuroscience and clinical psychology (Garon, Bryson and Smith, 2008). Calls for a model of self-regulation integrating the theory and methodology of effortful control and executive function, not new, increasingly point to the benefits of such an integration for research and theory but also educational and clinical practices (Zhou, Chen and Main, 2012).

Because self-regulation is so broadly defined, its measurement can range from laboratory or experimental tests and observation measures to subjective reports (Duckworth and Kern, 2011; McClelland and Cameron, 2012). We note here three recent advances with respect to the measurement of executive function and emotional regulation in preschoolers. The first is the NIH Toolbox project whose NIH Toolbox Early Childhood Cognition Battery, recommended for ages three to six, measures executive function.¹ The battery includes the DCCS (Dimensional Change Card

¹ <http://www.nihtoolbox.org/Resources/Documents/NIH%20Toolbox-Summary%20description.docx>

Sort), Flanker (measuring attention and inhibitory control), Picture Sequence Memory, and Picture Vocabulary measures. The second is the development of a quick, easy-to-use, direct measure of self-regulation for children aged four to six, the Head-Toes-Knees-Shoulders test (HTKS; Ponitz, McClelland, Matthews and Morrison, 2009). However, neither of these are published, nor are relevant psychometric data reported and therefore not examined in the database in the second stage. The third, that emotional dysregulation, can be reliably measured with the Child Behavior Checklist (CBCL) in preschoolers (Geeraerts *et al.*, 2015). The Dysregulation Profile (DP) of the CBCL is derived from the Anxious/Depressed, Aggressive Behavior, and Attention Problems scales (or AAA scales), in turn pointing to the overlap, discussed above, between behaviour problems and emotional dysregulation.

Conventional approaches to the assessment of social and emotional skills

In the U.K., the SDQ (Strengths and Difficulties Questionnaire) is typically used to assess social and emotional skills in children, especially when broad assessments of such skills are required. As discussed in the next section, however, emotional and social skills need to be considered in context. For example, children with disabilities often show elevated levels of socio-emotional problems, especially when they have co-occurring communication difficulties (ASD: Maskey *et al.*, 2013; SLCN: Lindsay, Dockrell and Strand, 2007). Another important point to note is that emotional and social skills in children, especially in the early years, can show moderate or even weak stability over time (Wichstrøm *et al.*, 2017). This, in turn, means that measuring them cross-sectionally, frequently the only possibility, can lead to erroneous conclusions.

Ways forward in assessing social and emotional skills

As shown above, social and emotional skills are captured by both indices of difficulties and measures of strengths. Difficulties and strengths do not fall on a continuum, and the absence of problems does not guarantee the presence of competencies. In fact, many children follow atypical joint trajectories of difficulties and strengths (Nantel-Vivier *et al.*, 2014). It is therefore important to measure, wherever possible, both. Another important point to note is that there may be significant challenges associated with capturing these competencies across different child populations, such as children with disabilities or those from diverse cultural backgrounds. For example, socialisation practices that differ across cultural groups can result in corresponding differences in parent-child or peer interactions. This can lead to seemingly erroneous conclusions concerning the normality of certain behaviours, especially before children become integrated in the state education system. There is a need, therefore, for measures of social and emotional development that are sensitive to cultural variations in social behaviour and that accurately capture the diversity of young children's competencies. Another important challenge is that the borders between social and emotional development, and other important domains of young children's development, are not always clear or agreed upon. For example, there is disagreement within and across the fields of psychology as to whether executive function skills are part of children's self-regulation or are a distinct set of skills that contribute to children's self-regulation. There is clearly a need for more work to provide conceptual clarity. The last two decades have generated much useful information about the importance of early social and emotional competencies for success across developmental periods and in a variety of life domains. However, the area of social and emotional development still suffers from the 'jingle and jangle fallacies' (Jones *et al.*, 2016): the 'jingle fallacy' refers to the use of a single term to represent a wide variety of skills, and the 'jangle fallacy' refers to the use of different terms to refer to the same skill. Nonetheless, there is increasing recognition that building consensus as to the definition of social and emotional skills and ensuring depth and precision in measurement must be a priority (Darling-Churchill and Lippman, 2016). Perhaps the measurement of social and emotional skills could benefit from the development of a clear taxonomy of criteria on which measures can be evaluated, in line with the taxonomies found in other areas of research (for example, <http://www.cosmin.nl/>). As discussed in the previous section, it will also benefit from recognising the

importance of undertaking a careful consideration of a child's characteristics and contexts and a longitudinal assessment of his or her skills.

4.5 The learning environment

This section moves from examining assessments of the child to focus on the child's environmental context. The rationale for considering environmental factors is outlined and approaches to assessing the child's home and early years environment considered. Unlike language, literacy, numeracy, and social-emotional development, there are few well-developed measures in the area. This is reflected in this report in that we provide, in this section, a narrative overview of the measures available but they are not included in the measures database. Effectively, assessments of the environment depend either on parental report, through questionnaires, or observations of the home and early years contexts—or a combination of the two.

Home learning environment

Attempts to examine home learning environments have been in use for over half a century reflecting the understanding that children develop within a wider social context (Bronfenbrenner and Morris, 2006). These measures aim to go beyond describing children's social context as determined by socio-economic status and to identify and profile environmental processes that can drive development. It is now well established that the quality of a child's home learning environment in the first years of life is associated with cognitive performance at school entry and later educational achievement. To provide a differentiated view of the child's early learning opportunities, attempts have been made to develop measures which assess both the emotional support and cognitive stimulation in the home environment.

One of the earliest measures developed—and most commonly used—is the HOME ('Home Observation for Measurement of the Environment', Elardo and Bradley, 1981). The primary goal of the instrument is to measure, within a natural context, the quality and quantity of stimulation and support available to a child in the home environment—the child as an active recipient of inputs from objects, events, and transactions occurring in connection with the family surroundings. The infant and toddler version of HOME (birth to three years of age) has five subscales:

- emotional and verbal responsiveness of the primary caregiver;
- avoidance of restriction and punishment;
- organisation of the physical and temporal environment;
- provision of appropriate play materials;
- parental involvement with the child and
- variety in daily stimulation.

The Early Childhood HOME (for children from three to six) has eight subscales including: learning materials, academic stimulation, physical environment, responsiveness, academic stimulation, modelling, variety, and acceptance.

The HOME scale is administered in the form of an interview about the child while observations of the mothers' behaviour are also made. Its strength is that it is relatively simple to administer, has good psychometric properties, and can be used to predict cognitive and linguistic development (for a review, see Totsika and Sylva, 2004). Over the years, researchers have altered the format in a range of different ways and there have been numerous debates about both which scales and which formats should be used (Linver, Brooks-Gunn and Cabrera, 2004). For example, father presence has been isolated in some studies so that its unique effects can be examined (Mott, 1993). Although not as frequently utilised as the overall scale and the two main subscales, individual items and researcher-constructed subscales representing a variety of specific concepts are also studied as predictors and as outcomes. There is also an argument that large scale studies can use a limited

number of HOME items to create conceptually meaningful measures of the child's environment (Fulgini, Han and Brooks-Gunn, 2004).

There are both practical and conceptual limitations with the measure. From a practical perspective, a major limitation is a lack of standardised administration procedure; this can lead to inconsistent results, especially if researchers are not trained. Conceptually, the HOME, like other similar measures, is a global assessment. There is increasing evidence to indicate that models of the home learning environment should consider different domains of support, both general and domain-specific (Kluczniok, Lehl, Kuger and Rossbach, 2013). As we describe below, domain-specific approaches have considered both the home literacy environment (HLE) and the home numeracy environment (HNE). The HLE has typically been conceptualised as both home literacy practices and parental reading beliefs. Data is generally collected by bespoke questionnaires (that is, not in the public domain) or single questions. The majority of work has focused on mothers' practices but the role of fathers becomes increasingly important when mothers have lower levels of education (Foster, Froyen, Skibbe, Bowles and Decker, 2016). Studies have examined a wide range of HLEs, including children in areas of disadvantage (Farver, Xu, Lonigan and Eppe, 2013), children with learning challenges (Breit-Smith, Cabell and Justice, 2010; Dolezal-Sams, Nordquist and Twardosz, 2009), and children in a range of cultural contexts (Kalia and Reese, 2009; Marjanovic Umek, Podlesek and Fekonja, 2005). While there is some overlap in the questions asked, there is a lack of consistency in the parental completion checklists of the HLE. Typical items include frequency of reading to the child, the number of children's books in the home, the frequency of parental teaching of literacy skills, and the frequency of library visits. Studies vary in the number of items used, with some studies reporting only three items, and often there is no evidence of the psychometric rationale for the choice of items. In addition, research studies have demonstrated that aspects of the HLE environment are differentially related to literacy and language development (Hood, Conlon and Andrews, 2008). As such, intervention studies need to be specific about *a priori* targets and HLE domains.

In contrast to work on the HLE, there are fewer studies on the HNE, yet there is evidence that the HNE influences children's mathematical abilities (Kleemans, Peeters, Segers and Verhoeven, 2012), with different aspects of the home environment predicting different numeracy skills (Skwarchuk, Sowinski and LeFevre, 2014). Similar to the literacy measures, both beliefs and practices are assessed by parents completing checklists. No standard measures were found. Measures ranged from as few as three questions (Niklas and Schneider, 2014) to more complex tools developed for a specific study. For example, the Early Maths Questionnaire contains 36 questions on maths-related activities and 13 on maths-related beliefs (Missall, Hojnoski, Caskie and Repasky, 2015).

Despite the numerous bespoke questionnaires to examine HLE and HNE, there are few tested observation schedules that directly focus on both language and literacy. The Child/Home Environmental Language and Literacy Observation (CHELLO), is a notable exception. The CHELLO has been developed to examine the quality of the language and literacy environment in home-based childcare settings (Neuman, Koh and Dwyer, 2008). The tool provides professionals with the ability to examine key instructional and affective supports known to promote children's learning. The publishers state that the measure can be used to provide benchmarks to measure change in language and literacy and the initial research study identifies a number of psychometric strengths.

Early years settings

As with the home learning environment, there is incremental data about the importance of the 'quality' of the early years settings that children attend (Sylva, 2014). A number of unpublished measures exist which assess both the classroom context and teachers responsiveness. Our focus is on measures in the public domain that have been used in research, ideally in the U.K. (see Bryant, 2010, for more details).

The ECERS-R (for ages two years and five months to five years) is one of the most widely used assessments of the quality of environment in early childhood classrooms. There are also more

extended versions (ECERS-E) and versions for infants and toddlers (ITERS-R), and now the ECERS-3. The latter has not yet been evaluated in the U.K. context. The ECERS-R is a global measure of quality that includes both structural aspects of the environment (such as the physical environment) and process quality (for example, the kind of language that staff direct to children). The scale was first published in 1980 and a revised version was published in 1998 (Harms and Clifford, 1980; Harms *et al.*, 1998). The measure is designed to reflect developmentally-appropriate practice and includes seven subscales: space and furnishings (eight items), personal care routines (six items), language reasoning (four items), activities (ten items), interaction (five items), programme structure (four items), and parents and staff (six items). All versions of the ECERS are time-consuming, ranging from two hours to over three hours for each assessment period.

Research on the validity of the ECERS-R and the psychometric properties of the tool is limited given how widely it is used. The studies which have been conducted produced mixed results with some attesting to the validity of the scales (Cassidy, Hestenes, Hegde, Hestenes and Mims, 2005), but others are more critical (for example, Gordon, Fujimoto, Kaestner, Korenman and Abner, 2013). Recent research questions the scale's structural validity and suggests that it does not identify a single factor. Overall, there is a general lack of domain specificity within the items and a lack of fit for items within the scale of the items included. Criterion validity has also been questioned with mixed results and small effect sizes in predicting child outcomes and moderate effect sizes for teacher-reported quality (Gordon *et al.*, 2013). Psychometric limitations with the order specified in the measure have been highlighted with few of the indicators capturing moderate or high levels of quality (Gordon *et al.*, 2015). Nonetheless, there is indicative evidence that the tool is sensitive to changes in response to intervention (Bryant, 2010). Moreover, the new ECERS-3 has improved the scaling and, importantly, enhanced the focus on teaching and interactions.

A recent U.K. development building on the ECERS-R and ECERS-E/ITERS-R is the Sustained Shared Thinking and Emotional Well-Being (SSTEW) Scale for two- to five-year-olds (Siraj, Kingston and Melhuish, 2015). The scale focuses on the adult's role in supporting learning and development. The SSTEW consists of five subscales containing 14 items. Inter-observer reliability is reported by the authors to be 0.79 to 0.92 for subscales. The newness of the publication of the scale means that it has not been subject to wider evaluation and trial. The authors advise that the scale needs to be used by someone with knowledge of child development and appropriate practice. Moreover, the authors advise that some staff or settings may not be ready for all of the subscales, particularly those relating to critical thinking, assessment for learning, and supporting and extending language and communication.

By contrast, the CLASS measures have received evaluation in a range of contexts. CLASSPreK, K-3 and CLASS-Toddler were developed as tools to assess process quality in the classroom focused on teacher-child interactions (La Paro, Pianta and Stuhlman, 2004; Pianta, Hamre and La Paro, 2008). The CLASS operationalises difficult constructs such as scaffolding and contingent responding for early year education. There is evidence that these are important dimensions in supporting children's development. The CLASS consists of three domains: 'emotional support', 'classroom organisation', and 'instructions support'. Each domain is made up of several dimensions:

- emotional support includes negative climate, positive climate, teacher sensitivity, and regard for student perspective;
- classroom organisation includes behaviour management, productivity, and instructional learning formats; and
- instructional support includes concept development, quality of feedback, and language modelling, along with literacy focus.

Factor analysis of the earlier version of the CLASS yielded two domains: 'emotional climate' and 'instructional climate' (Perlman *et al.*, 2016) where 'emotional climate' included all of the dimensions of the emotional support domain and the behaviour management dimension that later became part

of the classroom organisation domain. 'Instructional climate' was a composite of concept development and quality of feedback.

The CLASS-Toddler is structured into two domains: emotional and behavioural support (EBS) and engaged support for learning (ESL). The first domain, EBS, includes five dimensions: positive climate, negative climate, teacher sensitivity, regard for child perspective, and behaviour guidance. These dimensions focus on the emotional connection between the teachers and children, including teachers' responsiveness to children and awareness of children's developmental and individual needs, the degree to which classroom activities and interactions reflect the interests of the children and encourage child autonomy, and the use of effective methods to prevent and redirect problem behaviour. The ESL domain is assessed through the dimensions of 'facilitation of learning and development' and 'language modelling'. These dimensions consider teachers' ability to facilitate classroom routines, materials, and activities to support children's learning and development opportunities. They also focus on teachers' use of language stimulation and language facilitation techniques to encourage children's language development. Each dimension is rated on a continuous scale from one (low) to seven (high) based on behaviours observed across the 20-minute segment. Dimension scores are averaged across the cycles to yield a classroom score for each dimension. Reliability estimates are good (Thomason and La Paro, 2009).

CLASS-Toddler ratings were positively associated with fewer behaviour problems; specifically, children in classrooms with higher levels on the CLASS-Toddler domains of 'emotional and behavioural support' as well as 'engaged support for learning' were reported to have fewer behaviour problems (La Paro, Williamson and Hatfield, 2014). The CLASS domain that was most closely linked to child outcomes was 'instructional support' (Perlman *et al.*, 2016).

There are also two tools which target the language (and literacy) environment of early years settings. Both measures are grounded in the research literature underpinning language and literacy development, share similar features, and take about one hour to complete. Of particular importance is the emphasis on strong oral language environments including conversations, vocabulary, and opportunities to use language for authentic purposes. The Early Language and Literacy Classroom Observation (ELLCO, Smith, M. and Dickson D., 2002) has two versions: 'Pre to K' (four to five years) and 'K to 3' (five to eight years). It comprises three elements, including: a characteristics of literacy environment checklist (24 items), classroom observation (14 items), and literacy activities (9 items). The checklist has good inter-rater reliability, and detailed statistics—including features related to stability and change—are available, but no validity data is reported.²

The Communication Supporting Classrooms Tool focuses solely on the oral language environment (Dockrell, Bakopoulou, Law, Spencer and Lindsay, 2015). Based on the premise that the development of oral language skills is fundamental to learning and academic progress, this checklist was designed to capture the language learning environment, opportunities, and interactions. The checklist has demonstrated good inter-rater reliability and discriminates between different classrooms, but lacks evidence demonstrating its association with child language outcomes. The tool is freely available from the Communication Trust.³

Finally, it is also worth noting the Early Childhood Observation Measure (EECOM, Stipek & Byler, 2004) which is currently unpublished. The EECOM was created as a global classroom observation tool primarily to be used in research for studying the effects of teacher training and other interventions on classroom practices, and the effect of classroom practices on child outcomes. The tool includes scales describing two different approaches to instruction, constructivist and traditional learning theory. There is evidence that the scale is reliable (Stipek & Byler, 2005) and the scale has been used in a number of research studies. However, no published manual exist so we do not include it in any further discussion.

² <http://archive.brookespublishing.com/documents/ellco-pre-k-technical.pdf>

³ <https://www.thecommunicationtrust.org.uk/resources/resources-for-practitioners/communication-supporting-classroom-observation-tool.aspx>

Ways forward in assessing the learning environment

This review highlights a number of ways in which assessments of the learning environment raise different types of questions for research evaluation. Measures tend to focus on the quality of the environment rather than the quantity of learning opportunities. Results also show that domain-specific measures of the home and preschool quality were more strongly related to children's development than global quality measures (Votruba-Drzal and Miller, 2016). There is a lack of comparative British data in a standard form for the majority of the measures.

Reliability and validity data are not reported for many of the domain questions, and items in the generic measures—such as the HOME—are dated. This is clearly an important area to examine in research studies but careful choice of measures and an understanding of what they do, and do not, capture is essential. Findings suggest the need for additional research related to very young children's experiences in centre-based childcare to examine associations between quality and child experiences and outcomes. Teacher-reported child outcomes measures also have limitations. To ultimately understand children's experiences in childcare, additional measures of child outcomes should be considered to provide a profile of the children's learning and social-emotional experiences.

Choosing measures should be based on the specific domains of interest (Bryant, 2010). Given the potential in all of these measures to be subject to personal opinions, research studies should ensure that appropriate training with the scoring guidelines occurs and assessments of inter-rater reliability are collected. Specific care needs to be taken to ensure that the tools are culturally valid and sensitive. Associations between classroom quality and outcomes are typically significant but modest (Burchinal *et al.*, 2009).

5. Methods

5.1 Child-based assessments

In order to carry out a systematic evaluation of children’s skills, three conditions must be met:

- tests must provide information about reliability and validity (see, for example, Friberg, 2010) and should meet minimum psychometric standards (Rust and Golombok, 1999);
- practitioners must have the appropriate qualifications; and
- tests must be used correctly.

Our first strand of activity focused on assessment evaluation with an examination of the psychometric properties of the child-based assessments.

Our search did not include bespoke baseline assessments devised by local authorities in response to the government’s requirement (1999) that all children should be assessed within seven weeks of entering school (see Lindsay and Desforges, 1998). This decision was based on:

- the variation of elements included in the assessments;
- the differences in the ways in which the relevant domains were operationalized;
- the lack of basic psychometric data on the measures to support their use;
- the lack of consistency in assessment approaches; and
- challenges in scoring (Lindsay and Lewis, 2003; Lindsay, Martineau and Lewis, 2004).

However, we did include recent baseline test developments which are included in the section covering omnibus measures.

Table 1 outlines our strategy for evaluating the potential measures. The table delineates our search criteria. All tests identified are in Appendix A. Our first set of exclusionary criteria and our final coding of the measures. Detailed coding of all included measures can be found, by domain, in Appendix B.

Table 1: Search strategy for child based assessments

Identification of tests	First filter	Second filter
	Tests were excluded if they failed to meet criteria:*	If tests passed the first filter criteria, additional information coded:
<p>Search key databases and web sources to identify tests used or cited in papers and reviews.</p> <p>Define search terms related to each of the four domains.</p> <p>For the purpose of this review, tasks or non-validated questionnaires were excluded.</p> <p>Only post-2000 published tests were retained.</p>	<ol style="list-style-type: none"> 1. Appropriate for children under 6 years 2. U.K.-based standardisation sample for norm-referenced tests 3. Standard scores or norm-referenced raw scores for U.K. population 4. Criterion/construct validity correlations with at least one other measure 5. Data on test reliability 6. Administration guidelines 7. Scoring criteria 	<ol style="list-style-type: none"> 1. Freedom from biases or caveat 2. Prior knowledge required 3. Testing time required 4. Evidence of sensitivity to intervention 5. Workload/administrative requirements

* Data to corroborate findings must have been available either in test manuals or published in peer reviewed outlets.

Each identified test was logged in the database and coded (where data was available) for: purpose, publication date, age range/population, standardisation sample, scores, reliability coefficients, face and content validity, and linkages with other measures. This data was sourced from technical reports and appendices of the measures, where available, from publishers or test designers. Where no data was available in technical reports, we searched the peer review literature for any research studies which had aimed to evaluate the psychometric properties of the measure. Those that met the specified minimum criteria were included for a second strand of more detailed analyses including, for example, practicality, restrictions on use, and evidence of sensitivity to intervention. Figure 1 presents the flow diagram for the identification of child assessments.

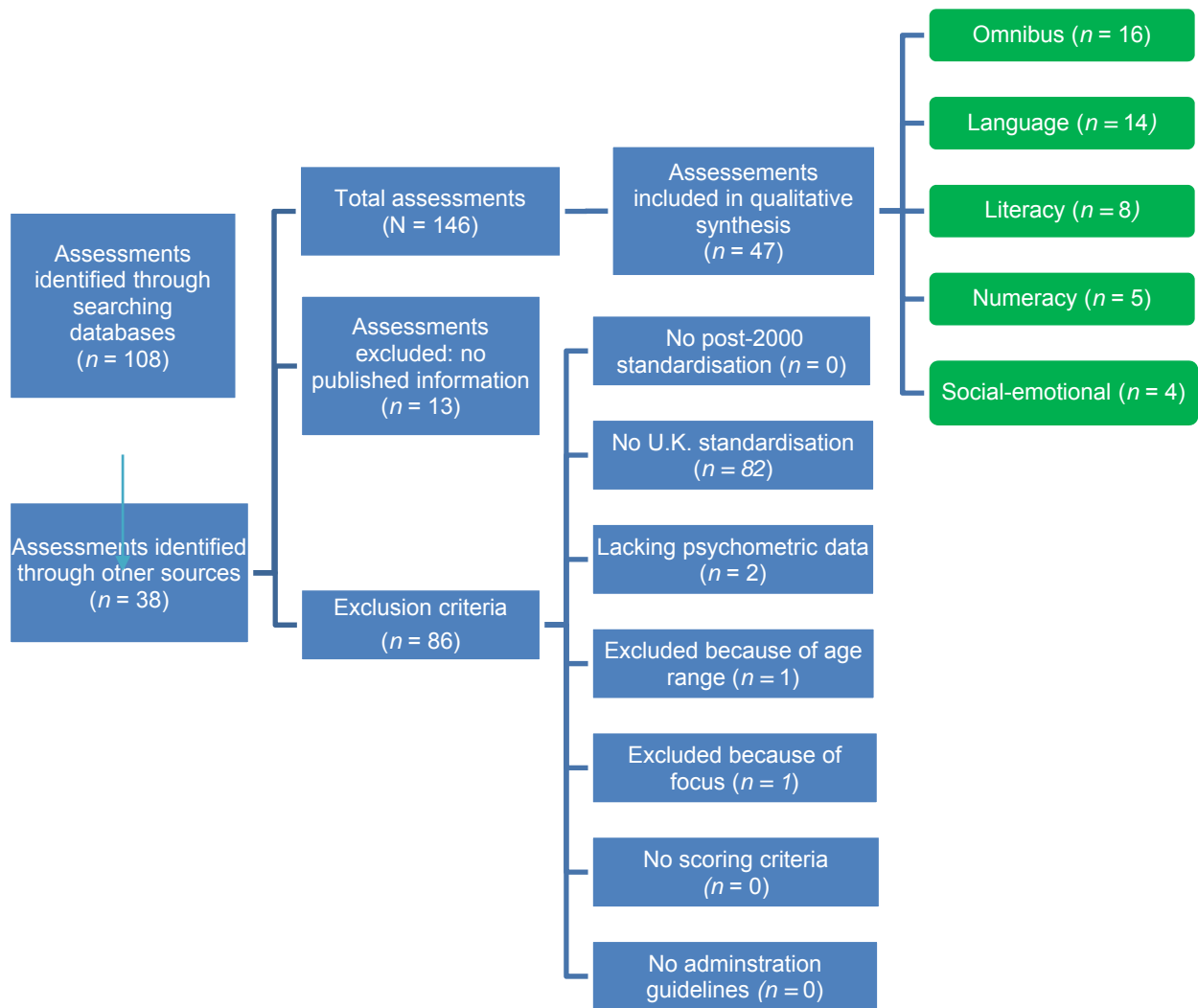


Figure 1: Flow diagram of assessments identified

Note: Omnibus tests include measures which allow evaluation of domain performance on subscales. When a measure was not published, and also did not have U.K. norms, it is only counted under ‘no U.K. standardisation’ in Figure 1. Eleven measures were not published and had no U.K. standardisation.

5.2 Environment assessments

To identify measures of the environment, we systematically searched research databases (WOS, SCOPUS, and ERIC) and the grey literature. Any publication referring to an evaluation of either the preschool home or early years environment was scrutinised and the measure used examined. In the majority of cases these were single questions and these approaches to capturing the environment were not considered further. Assessment of the generic environment and domains therein, which involved a systematic approach to evaluating the environment, were examined for inclusion. Table 2 provides details of the identified measures to assess the learning environment. These include both observation measures and checklists. These measures complement assessments of individual children by providing data on both the structure of the environment and the opportunities available for the children.

Table 2: Environment measures

HOME (Home Observation for Measurement of the Environment)	
Author, publisher, date	Caldwell and Bradley (2001).
Standardisation	U.S.A.; infancy/toddlers (IT); preschool/early childhood (EC).
Administration	IT: 45 binary items; EC: 55 binary items; completed through 1-hour observation and parental report.
Subscales	<ol style="list-style-type: none"> 1. Emotional and verbal responsivity of mother. 2. Avoidance of restriction and punishment. 3. Organisation of physical and temporal environment. 4. Provision of appropriate play materials. 5. Maternal involvement with child. 6. Opportunities for variety in daily stimulation.
Reliability	IT inter-observer agreement, at least 0.80; internal consistency of scales, 0.3–0.8.
Validity	Both concurrent and predictive with cognitive measures.
HOME – short form	
Author, publisher, date	Modification of the HOME inventory.
Standardisation	U.S.A. national longitudinal study of youth with norms derived from the study.
Administration	The HOME-SF is about half as long as the original HOME inventory.
Subscales	More than half of the HOME-SF's items are multi-response maternal reports that were reworded, with the assistance of the instrument's designers, from the original HOME Inventory's dichotomous observer ratings.
Reliability	Not reported
Validity	Not reported
Family Environment scale	
Author, publisher, date	Moos and Moos (1986), Palo Alto CA: Consulting Psychologists Press.
Standardisation	U.S.A., adolescents Australia.

Administration	20 minutes to complete.
Subscales	Measures social environmental aspects of the family; 90 items in 10 subscales.
Reliability	Mixed.
Validity	Conceptual strength but needs revision (Boyd, Gullone, Needleman and Burt, 1997).
Family Child Care Environment rating scale – revised	
Author, publisher, date	Harms, T., Cryer, D. and Clifford, R., Teachers College Press (2007).
Standardisation	Not reported.
Administration	Not reported.
Subscales	Scale consists of 37 items organised into 7 subscales: space and furnishings, personal care routines, listening and talking, activities, interaction, programme structure, parents and provider.
Reliability	Not reported.
Validity	Not reported.
CHELLO (Child/Home Early Language and Literacy Observation Tool)	
Author, publisher, date	Neuman, S. B., Dwyer, J. and Koh, S., Brookes Publishing Co. (2007): http://products.brookespublishing.com/ChildHome-Early-Language-and-Literacy-Observation-Tool-CHELLO-P499.aspx
Standardisation	Home-based providers.
Administration	1½–2-hour observation recommended to be used by professionals with a ‘strong background’ in early childhood education.
Subscales	Items rated on a scale of 1–5; comprises literacy environment checklist, family observation, and provider interview.
Reliability	Weighted kappa was highly substantial at 0.84; for the group/family observation, the weighted kappa was moderately high at 0.54.
Validity	Not reported
ECERS-R (Early Childhood Environment Rating Scale – revised)	
Author, publisher, date	Cryer, D., Harms, T. and Riley, C. (2003) <i>All about the ECERS–R</i> , Lewisville NC: Kaplan.
Standardisation	Not reported but see review for questions about validity and appropriateness of items.
Administration	Not reported
Subscales	Forty-three-item observational instrument; each item is rated from 1 (inadequate) to 7 (excellent) based on indicators—descriptions of quality listed below the 1, 3, 5, and 7 ratings. The scale contains seven subscales: space and furnishings (8 items), personal care routines (6 items), language reasoning (4 items), activities (10 items), interaction (5 items), programme structure (4 items), and parents and staff (6 items). Subscale scores are created by averaging across each of the items within a subscale, and the overall score is created by taking an average of all the items.
Reliability	Percentage of agreement across all 470 indicators, 86.1%, with all indicators having a percentage of agreement over 70% (Harms, Clifford and Cryer, 1998). Item level: the percentage of exact agreement was 48%; Pearson product

	moment correlation, 0.92; Spearman rank order correlation, 0.87. The interclass correlation total score was 0.92 (Clifford, Reszka and Rossbach, 2010).
Validity	Not reported
CLASS toddler (2–3 years, Pre K—K)	
Author, publisher, date	Toddler: La Paro, Hamre and Pianta (2011); Pre K: Pianta, La Paro and Hamre (2008).
Standardisation	U.S.A. with confirmatory work in the Netherlands for the Toddler version (Slot <i>et al.</i> , 2017).
Administration	Toddler: four 15–20-minute cycles; Pre-K: 2–3 hours, 30-minute cycles of observe-code; intensive training available.
Subscales	Toddler: 8 dimensions in 2 domains; Pre-K and K: teacher-child interactions in 3 domains: instructional support, emotional support, and classroom organisation, 10 items, 7-point scale.
Reliability	Inter-rater agreement between 73% and 86% across the scales.
Validity	Toddler positively related to children’s social emotional and language development.
Early Language and Literacy Observation (ELLCO)	
Author, publisher, date	Smith, M and Dickson D. (2002), available Brookes Publishing.
Standardisation	U.S.A.
Administration	Toolkit including checklist and rating scale.
Subscales	Literacy environment checklist, classroom observation, and literacy activities rating scale.
Reliability	Inter-rater reported as 80–90% agreement depending on scale.
Validity	Not reported
Communication Supporting Classrooms Observation Tool	
Author, publisher, date	Dockrell <i>et al.</i> (2015), available from the Communication Trust for download: https://www.thecommunicationtrust.org.uk/resources/resources/resources-for-practitioners/communication-supporting-classroom-observation-tool.aspx
Standardisation	U.K.
Administration	Checklist completed in schools, 1 hour.
Subscales	Three domains: environment, opportunities, and interactions. Profile of the setting’s oral language support.
Reliability	Inter-rater reliability between 74% and 80% for individual items.
Validity	Not reported

6. Results

Assessment tools that were either domain-specific or assessed the broad array of early childhood skills were sourced.⁴ We identified 146 assessment tools appropriate for children under six that included norm- and criterion-referenced tests which provided some psychometric information on their validity and reliability. Information on whether the test was free from biases and/or presented evidence of sensitivity to intervention was very rarely reported. In addition, although most tests included details concerning the administration time and the administration procedures, tests did not always indicate certification requirements. Table 3 shows the measures identified.

Table 3: Total figures identified in search (N = 146)

	Norms (U.K.)	Criterion	Both	Not published
Omnibus (n = 36)	30 (10)	5	1	0
Language (n = 42)	31 (14)	1	1	11
Literacy (n = 23)	15 (8)	1	3	4
Social-emotional (n = 29)	26 (3)	1	1	7
Numeracy (n = 16)	12(3)	2	0	4

6.1 Omnibus tests

We identified 36 omnibus measures that tapped skills related to at least two of the covered domains, of which, 16 could be used with children under the age of two. Of the 30 norm-referenced assessments, seven had U.K. norms collected within the last 18 years:

- Baseline and Baseline Progress (two-and-a-half to seven years);
- Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III) (1 month to 42 months);
- British Ability Scales (BAS III) (three to eight years and eleven months)
- Griffiths Mental Development Scales-Extended revised (GMDS-ER) (two to eight years)
- the WPPSI-IV (two years six months to seven years three months)
- the WIAT IIIUK (four years to 25 years and 11 months); and
- the WRAT-4 (5 to 94 years).

The criterion-referenced measures included a range of assessments which captured early arithmetic, reading, and learning indicators including phonological awareness.

All norm-referenced assessments for this age group needed to be administered individually, requiring between 12 to over 60 minutes (Early Learning Scale Preschool Assessment may be completed over several sessions) and the majority required relevant certification. Overall, these omnibuses were time-intensive and dependent on professional expertise. By contrast, the criterion-referenced measures provide clear benchmarks for key skills underpinning language, literacy, and numeracy in the early years from ages two to six. Six of these measures did not require specialist training to administer the measures.

⁴ We have used the version of tests available when completing the review. However, new versions of the Wechsler tests and the CELF will be published in 2017.

We also identified a range of recently published baseline assessment measures covering language literacy, numeracy, and behaviour. The EExBAR is one of three tests developed for the DfE to assess reception children on a basis similar to the previous Early Years Foundation Profile. The other two are the NFER Baseline Assessment and the CEM Baseline (which is a baseline equivalent of InCAS/PIPS). These three tests were compared by the Standards and Testing Agency in 2015 (Standards and Testing Agency, 2016) which concluded that they were not comparable, although correlation on literacy between the CEM and NFER assessments was 0.78, and on numeracy, 0.81. The EExBAR was the weakest of the three. The NFER uses teacher judgments, such as tasks on phonics segmenting similar to a range of tests recommended for literacy. Data on reliability and validity is not publically available but the Standards and Testing Agency report provides evidence of concurrent validity in literacy and the website reports that it has been 'standardised with a nationally representative sample of over 65,000 children'. The reports available suggest that it is age-normed, and for this reason meets the inclusion criteria applied in this report. The CEM test also covers numeracy and literacy and the programme produces age-standardised reports. There is no data on its reliability, and as it is an online test drawing on 247 items, comparisons across children and times of testing cannot be associated with specific indicators of performance.

6.2 Language tests

The domain review (Section 4.1) identified the importance of examining both structural (vocabulary and grammar) and pragmatic aspects of language. These aspects were included in one or more measures and 42 different measures were identified. Of the 31 norm-referenced tests, 14 had U.K. norms. Both omnibus language tests and tests of separate aspects of the language system (vocabulary and grammar) met the inclusionary criteria. Omnibus tests typically required specialist certification. Two tests (Wellcomm and Gaps) were open to general use.

Limited assessments of social communication were available and where these existed were part of a checklist completed by either parent or practitioner. The majority of the assessments involved individual testing of the child, but three—the MacArthur-Bates Communicative Development Inventories (CDIs), Children's Communication Checklist, and the Preschool Language Scale (under 2 years and 11 months)—were completed by other informants (typically carers).

Table 4: Assessments meeting criteria for language

	Global language*	Receptive vocabulary	Expressive vocabulary	Receptive grammar	Expressive grammar	Narrative language	Social communication
Omnibus							
British Abilities Scale III			√	√			
Brigance	√						√
WPPSI-IV		√	√	√			
WIAT-IIIUK		√	√	√	√	√	
Language Tests							
Action Picture Test revised			√		√		
Bus Story					√	√	
British Picture Vocabulary test (BPVS-3)		√					
CELF-Preschool 2	√						
Children’s Communication Checklist	√						√
DEAP			√		√		
Early Repetition Battery					√		
GAPS					√		
MacArthur Bates Communicative Development Inventories—words and sentences		√	√	√	√		
New Reynell Developmental Language Scales (NRDLS-4)	√						
Preschool Language Scales (PLS-4)	√						
Test of Reception of Grammar (TROG-2)				√			
Wellcomm: The Complete Speech Language Tool Kit	√						
Word Finding Vocabulary Test			√				

• Global language: measure including receptive and expressive vocabulary and grammar.

Measures of both the home and preschool oral language environment required training but not specialist accreditation. These measures speak to the language learning opportunities available for children.

6.3 Literacy tests

Twenty-two omnibus tests which sampled items from the literacy domain and 23 literacy tests were identified which addressed at least one of the key, age-relevant literacy competencies identified in the domain review (Section 4.2)—letter knowledge, phonological awareness, decoding (word and non-word reading), conventions of print, and spelling. Of these, 17 met the study criteria (see Table

5). Eight were literacy-specific tests, but one of these, PIPA, has only partial U.K. norms. The DEST, ERDA, and OS can be administered by a suitably prepared classroom teacher; the others have training requirements for test administrators, for example, speech and language therapists, psychologists, or teacher experts in speech and language. Nine omnibus tests addressed at least one of the key, age-relevant literacy competencies. Six of the nine omnibus tests have U.K. norms (EARLI, BAS III, Baseline and Baseline Progress, NFER Baseline, WRAT-4, and WIAT-III UK) but the WRAT-4 and the WIAT-III UK literacy elements are not normed for children under six. The BAS-III must be administered by a suitably trained psychologist or equivalent; the other tests may be administered by suitably prepared teachers or equivalent.

There is a reasonably good choice for tests measuring letter knowledge and phonological awareness, often including a range of sub-tests in the case of phonological awareness. Individual tests or sub-tests are typically fairly quick to administer (under ten minutes). The more sub-tests selected, the longer the test time, so it is sensible to think about which aspects of the skill are most important, for example because they are age appropriate or reflect the nature of the intervention. Tests of word and non-word reading are challenging for under-sixes and spelling is very challenging. For example, the BAS-III Technical Manual advises that its word reading and spelling tests are 'intended to be given to children aged 6:00 to 17:11, but can also be administered to children aged 5:00 to 5:11 of above average literacy skill'. The phonics screening check is suitable for children at the end of Year 1.

Table 5: Assessments meeting criteria for literacy

	Letter knowledge	Phonological awareness	Word reading*	Conventions of print	Spelling
Omnibus					
BAS III			✓		✓
Baseline and baseline progress	✓		✓		
Brigance	✓	✓	✓	✓	✓
CIRCLE	✓	✓		✓	✓ ¹
EARLI	✓	✓			
ELS		✓		✓	
NFER Baseline Assessment		✓	✓		✓ ²
WIAT-III UK			✓		✓
WRAT-4			✓		✓
Literacy					
DIBELS	✓	✓	✓		
Diagnostic Test of Word Reading Processes		✓	✓		

ERDA		✓			
Observation Survey (OS)	✓	✓	✓	✓	✓
PALS	✓	✓			✓ ²
PIPA	✓	✓			
PhAB		✓	✓		
YARC	✓	✓	✓		

* Word reading—word and non-word.

¹ Name-writing and observation.

² Name-writing.

6.4 Numeracy tests

We identified 23 omnibus tests and 16 numeracy tests assessing number concepts and skills. Thirteen assessments met the study criteria. The domain review (Section 4.3) identified four main areas of preschool number development: counting, transcoding, comparing numerical magnitudes, and simple arithmetic. Counting is assessed by three types of item: knowledge of number sequence (C1), counting objects in arrays (C2), and judging puppet counts (C3). Transcoding is assessed by reading numerals (T1), writing numerals (T2), and matching numerals to numbers (T3). Comparing numerical magnitude is assessed with numeral comparisons (M1) and spoken number comparisons (M2). Simple arithmetic is assessed with nonverbal arithmetic (S1), number combinations (S2), and story problems (S3).

The table below summarises the coverage of these in the selected tests. The Early Learning Scale Preschool Assessment (ELS) is the only measure based on teacher observation.

Table 6: Assessments meeting criteria for numeracy

	Comparing numerical magnitudes										
	Counting			Transcoding					Simple arithmetic		
	C1	C2	C3	T1	T2	T3	M1	M2	S1	S2	S3
Omnibus tests											
Baseline and baseline progress		✓									✓
Brigance	✓	✓		✓	✓	✓				✓	✓
BAS III	✓	✓				✓	✓				✓
CIRCLE	✓	✓		✓						✓	
EARLI	✓	✓		✓							
ELS	✓	✓		✓	✓						
WIAT III UK	✓	✓			✓	✓				✓	✓
WRAT-4										✓	
Numeracy tests											
Keymath-3	✓	✓								✓	✓
MaLT- CAP	✓	✓			✓	✓	✓			✓	✓
PNI	✓	✓		✓							
The Preschool Numeracy Scales	✓	✓		✓			✓			✓	✓
PUMA	✓	✓		✓	✓	✓	✓			✓	✓

As this table indicates, only four of the selected tests (BAS III, MaLT-CAP, Preschool Numeracy Scales, and PUMA) feature items in all four areas and none include all types of item. As Dowker (2005) points out, there is no single mathematics ability. Hence the correlations between the different areas are not substantial. This makes choosing a test which does not cover all four areas a risky strategy for evaluating interventions. An intervention which produces gains in the areas assessed may not produce gains in the untested areas. An intervention which does *not* produce gains in the areas assessed may yet affect the development in untested areas.

6.5 Social and emotional development

We identified 29 assessments of social-emotional development, typically all completed either by parents or teachers. Four measures met the study criteria and captured competencies identified in the domain review. The four selected measures (see Table 7), all well-known and widely-used to assess social and emotional skills, are described below.

Table 7. Assessments meeting criteria for social and emotional development

Tests	Domains			
	Social Competence	Emotional Competence	Behaviour problems	Self-regulation
ASQ: SE-2	✓	✓	✓	✓
HBQ	✓	✓	✓	
SDQ	✓	✓	✓	
SCQ	✓			

The Ages and Stages Questionnaire, third edition (ASQ-3)⁵

The ASQ is a brief measure in which parents rate their child's current skills and development from the ages of one month to five and a half years. Twenty-one questionnaires are available within this age range. Parents answer 30 questions covering five domains of development, including communication, gross motor, fine motor, problem-solving, and adaptive skills. Parents are instructed to try activities with their child to facilitate accurate assessment. A pass/fail score is assigned for each area of development. The presence of any domain screened less than two standard deviations below the mean area score is considered a positive screen. The ASQ-3 has been validated by using large, standardised samples of children from diverse ethnic and socioeconomic backgrounds. The ASQ-3 has been validated in many countries and translated into numerous languages. In most publications, it has been reported that the ASQ is accurate in detecting true problems in apparently healthy children and even in children with biological risk factors, and has been widely used in primary care settings. The ASQ Social-Emotional second edition (ASQ: SE-2) is exclusively focused on children's social and emotional development from the ages of one month to six years. Domains covered include self-regulation, compliance, communication, adaptive behaviours, autonomy, affect, and interaction with people. It has nine questionnaires and about 30 items per questionnaire. It is valid and reliable, with reported test-test reliability about 89%, internal consistency about 84%, and validity about 83%.

⁵ <http://www.brookespublishing.com/resource-center/screening-and-assessment/asq/>

The MacArthur Health and Behavior Questionnaire (HBQ)⁶

This test was designed to collect data from adult reporters (parents and teachers) about children between the ages of four and eight years. The HBQ comprises multiple scales that measure children's mental health symptoms (for example, internalising and externalising symptoms), physical health, social functioning (for example, peer acceptance and peer rejection, prosocial behaviours), and, for older children, school functioning (such as academic competence, school engagement). The parent version (suitable for four- to eight-year-olds) includes 140 items. The items included in the HBQ are derived from existing measures. The HBC is not normed for the U.K. population but provides criterion-referenced items.

The Strengths and Difficulties Questionnaire (SDQ)⁷

The SDQ is a 25-item measure completed for the preschool years by the parent. It is a screening tool for 2–17-year-olds. The SDQ has four scales of 'difficulties' (emotional symptoms, conduct problems, hyperactivity, and peer problems), and one 'strength', prosocial behaviour. Each SDQ scale has five items (scored from 0, 'not true', to 2, 'certainly true') and the four difficulties scale scores are added to compute a 'total difficulties' score, measuring overall level of problem behaviour. The SDQ offers the following mapping:

- Internalising:
 - emotional symptoms—for example, 'many worries', 'often unhappy, downhearted', and
 - peer problems—for example, 'rather solitary, tends to play alone', 'picked on or bullied'.
- Externalising:
 - hyperactivity—for example, 'restless, overactive', 'easily distracted, concentration wanders', and
 - conduct problems—for example, 'often has temper tantrums or hot tempers', 'often fights with other children'.
- The prosocial behaviour scale has items such as, 'considerate of other people's feelings', and 'shares with other children'.

The SDQ's internal consistency is 'very good' or 'excellent' (Stone, Otten, Engels, Vermulst and Janssens, 2010). The SDQ is the most widely used measure of children's socio-emotional development in the U.K.

The Social Communication Questionnaire (SCQ)⁸

The SCQ is a brief instrument that can be used to evaluate anyone over four years age as long as his or her mental age exceeds two years. It is a 40-item parent questionnaire. The responses that caregivers provide about their children's social communication behaviours yield a total score. Scores above 15 points indicate social-communication difficulties that warrant further assessment for the presence of an autism spectrum disorder. The SCQ is meant to serve as an efficient method of identifying children with communication and social delays.

⁶ <https://macarthurbq.wordpress.com/>

⁷ <http://www.sdqinfo.com/>

⁸ <http://www.wpspublish.com/store/p/2954/social-communication-questionnaire-scq>

References

- American Psychological Association (2014) *The Standards for Psychological and Educational Testing*, Washington DC: AERA Publications.
- Anthony, J. L. and Francis, D. J. (2005) 'Development of phonological awareness', *Current Directions in Psychological Science*, 14, pp. 255–259.
- Arrow, A. W. (2007) 'Potential precursors to the development of phonological awareness in preschool children' (unpublished doctoral thesis), University of Auckland, New Zealand.
- Arseneault, L., Bowes, L. and Shakoor, S. (2010) 'Bullying victimization in youths and mental health problems: "Much ado about nothing"?'', *Psychological Medicine*, 40 (05), pp. 717–729.
- Arseneault, L., Walsh, E., Trzesniewski, K., Newcombe, R., Caspi, A. and Moffitt, T. E. (2006) 'Bullying victimization uniquely contributes to adjustment problems in young children: a nationally representative cohort study', *Pediatrics*, 118 (1), pp. 130–138.
- Baroody, A. J. (2006, August) 'Why children have difficulty mastering the basic number combinations and how to help them', *Teaching Children Mathematics*, 13, pp. 22–31.
- Bracken, B. A. (1988) 'Ten psychometric reasons why similar tests produce dissimilar results', *Journal of Psychology*, 26, pp. 155–166.
- Breit-Smith, A., Cabell, S. Q. and Justice, L. M. (2010) 'Home Literacy Experiences and Early Childhood Disability: A Descriptive Study Using the National Household Education Surveys (NHES) Program Database', *Language Speech and Hearing Services in Schools*, 41 (1), pp. 96–107, DOI: 10.1044/0161-1461(2009/08-0048).
- Briars, D. J. and Siegler, R. S. (1984) 'A featural analysis of preschoolers' counting knowledge', *Developmental Psychology*, 20, pp. 607–618.
- Bronfenbrenner, U. and Morris, P. (2006) 'The bioecological model of human development', in R. Lerner (ed.), *Handbook of Child Development, Vol 1: Theoretical Models of Human Development*, Hoboken NJ: Wiley (6th ed., p. 35).
- Bryant, D. (2010) 'Observational Measures of Quality in Center-Based Early Care and Education Programs', OPRE Research-to-Policy, Research-to-Practice Brief OPRE2011-10c, Washington DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Buhs, E. S., Ladd, G. W. and Herald, S. L. (2006) 'Peer exclusion and victimization: Processes that mediate the relation between peer group rejection and children's classroom engagement and achievement?', *Journal of Educational Psychology*, 98 (1), pp. 1–13.
- Burchinal, P., Kainz, K., Cai, K., Tout, K., Zaslow, M., Martinez-Beck, I. and Rathbeg, C. (2009) 'Early care and education quality and child outcomes', OPRE Research to Policy Brief, Washington DC: Office of Planning, research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Byrne, B. (1998) *The foundation of literacy: The child's acquisition of the alphabetic principle*, Hove: Psychology Press.
- Caffrey, E., Fuchs, D. and Fuchs, L. S. (2008) 'The Predictive Validity of Dynamic Assessment: A Review', *Journal of Special Education*, 41 (4), pp. 254-270, DOI: 10.1177/0022466907310366.
- Caldwell, B. M. and Bradley, R. H. (2001) *HOME inventory and administration manual* (3rd edn), University of Arkansas for Medical Sciences and University of Arkansas at Little Rock.
- Camilleri, B. and 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 (5), pp. 565-581, doi: 10.1111/1460-6984.12033.

- Campbell, S. B., Denham, S. A., Howarth, G. Z., Jones, S. M., Whittaker, J. V., Williford, A.P., ... and Darling-Churchill, K. (2016) 'Commentary on the review of measures of early childhood social and emotional development: Conceptualization, critique, and recommendations', *Journal of Applied Developmental Psychology*, 45, pp. 19–41.
- Cassidy, D. J., Hestenes, L. L., Hegde, A., Hestenes, S. and Mims, S. (2005) 'Measurement of quality in preschool child care classrooms: An exploratory and confirmatory factor analysis of the early childhood environment rating scale-revised', *Early Childhood Research Quarterly*, 20 (3), pp. 345–360, DOI: 10.1016/j.ecres.2005.07.005.
- Chiat, S., Armon-Lotem, S., Marinis, T., Poliženská, K., Roy, P. and 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*, Bristol: Multilingual Matters (pp. 56–89).
- Cho, E., Compton, D. L., Fuchs, D., Fuchs, L. S. and Bouton, B. (2014) 'Examining the Predictive Validity of a Dynamic Assessment of Decoding to Forecast Response to Tier 2 Intervention', *Journal of Learning Disabilities*, 47 (5), pp. 409–423, DOI: 10.1177/0022219412466703.
- Cicchetti, D. (1994) 'Guidelines, criteria and rules of thumb for evaluating normed and standardised assessment instruments in psychology', *Psychological Assessment*, 6 (4), pp. 284–290.
- Clay, M. M. (2002, 2013) *An observation survey of early literacy achievement*, Portsmouth NH: Heinemann.
- Cole, P. M., Martin, S. E. and Dennis, T. A. (2004) 'Emotion regulation as a scientific construct: Methodological challenges and directions for child development research', *Child development*, 75 (2), pp. 317–333.
- Cowan, R. and Powell, D. (2014) 'The contributions of domain-general and domain-specific factors to third-grade arithmetic skills and mathematical learning disability', *Journal of Educational Psychology*, 106, pp. 214–229.
- Cowan, R., Donlan, C., Shepherd, D.-L., Cole-Fletcher, R., Saxton, M. and Hurry, J. (2011) 'Basic calculation proficiency and mathematics achievement in elementary school children', *Journal of Educational Psychology*, 103, pp. 786–803.
- Crick, N. R. and Dodge, K. A. (1994) 'A review and reformulation of social information-processing mechanisms in children's social adjustment', *Psychological Bulletin*, 115, pp. 74–101.
- Cryer, D., Harms, T. and Riley, C. (2003) *All about the ECERS–R*, Lewisville NC: Kaplan.
- Darling-Churchill, K. E. and Lippman, L. (2016) 'Early childhood social and emotional development: Advancing the field of measurement', *Journal of Applied Developmental Psychology*, 45, pp. 1–7.
- Demetriou, A., Merrell, C. and Tymms, P. (2017) 'Mapping and predicting literacy and reasoning skills from early to later primary school', *Learning and Individual Differences*, 54, pp. 217–254.
- Denham, S. A., Blair, K. A., Demulder, E., Levitas, J., Sawyer, K., Auerbach-Major, S. and Queenan, P. (2003) 'Preschool emotional competence: Pathway to social competence?', *Child development*, 74 (1), pp. 238–256.
- DfE (2014) *Early Years Foundation Stage Profile handbook*, London: Standards and Testing Agency, http://www.foundationyears.org.uk/files/2014/05/2014_EYFS_handbook.pdf [accessed 19 June 2017].
- DfE (2016) *Phonics screening check and Key stage 1 assessments in England, 2016*, London: Department for Education, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/577806/SFR4_2_Phonics_KS1_2016.pdf [accessed 19 June 2016].

- Dockrell, J. E. and Marshall, C. R. (2015) 'Measurement Issues: Assessing language skills in young children', *Child and Adolescent Mental Health*, 20 (2), pp. 116–125, DOI: 10.1111/camh.12072.
- Dockrell, J. E., Bakopoulou, I., Law, J., Spencer, S. and Lindsay, G. (2015) 'Capturing communication supporting classrooms: The development of a tool and feasibility study', *Child Language Teaching and Therapy*, 31 (3), pp. 271-286, DOI: 10.1177/0265659015572165.
- Dolezal-Sams, J. M., Nordquist, V. M. and Twardosz, S. (2007) 'Family Resources to Support Literacy Interaction: Examples From Families of Children With Disabilities', *Early Education and Development*, 20 (4), pp. 603–630, DOI: 10.1080/10409280802356661.
- Dolezal-Sams, J. M., Nordquist, V. M. and Twardosz, S. (2009) 'Home Environment and Family Resources to Support Literacy Interaction: Examples From Families of Children With Disabilities', *Early Education and Development*, 20 (4), pp. 603–630, DOI: 10.1080/10409280802356661.
- Dowker, A. (2005) *Individual differences in arithmetic: implications for psychology, neuroscience, and education*, Hove: Psychology Press.
- Dowker, A. and Sigley, G. (2010) 'Targeted interventions for children with arithmetical difficulties', in R. Cowan, M. Saxton and A. Tolmie (eds), *Understanding number development and difficulties, British Journal of Educational Psychology Monograph Series II: Psychological Aspects of Education - Current Trends, Volume 7*, Leicester: British Psychological Society (pp. 65–81).
- Duckworth, A. L. and Kern, M. L. (2011) 'A meta-analysis of the convergent validity of self-control measures', *Journal of Research in Personality*, 45 (3), pp. 259–268.
- Ehri, L. (1986) 'Sources and difficulty in learning to spell and read', in M. L. Wolraich and D. Routh (eds), *Advances in developmental and behavioural paediatrics*, Greenwich CT: JAI, Vol 7, pp. 121–195.
- Ehri, L. (1987) 'Learning to Read and Spell Words', *Journal of Literacy Research*, 19, pp. 5–31.
- Ehri, L. (1995) 'Phases of development in learning to read words by sight', *Journal of Research in Reading*, 18 (2), pp. 116–125.
- Ehri, L. C. (1994) 'Development of the ability to read words: Update', in R. Ruddell, M. Ruddell and H. Singer (eds), *Theoretical Models and Processes of Reading* (4th edition), Newark DE: International Reading Association (pp. 323–358).
- Ehri, L. C. and Wilce, L. S. (1985) 'Movement into reading: Is the first stage of printed word learning visual or phonetic?', *Reading Research Quarterly*, 20, pp. 163–177.
- Eisenberg, N., Cumberland, A., Spinrad, T. L., Fabes, R. A., Shepard, S. A., Reiser, M., Murphy, B. C., Losoya, S. H., Guthrie, I. K. (2001) 'The relations of regulation and emotionality to children's externalizing and internalizing problem behavior', *Child Development*, 72 (4), pp. 1,112–34.
- Eisenberg, N., Fabes, R. A. and Spinrad, T. L. (2006) 'Prosocial development', in R. M. Lerner, W. Damon and N. Eisenberg (eds), *Handbook of child psychology*, 6th edn, New York: Wiley (vol. 3, pp. 646–718).
- Elardo, R. and Bradley, R. (1981) 'The Home Observation for Measurement of the Environment (HOME) Scale: A Review of Research', *Developmental Review*, 1 (2), pp. 113–145.
- Farver, J. A. M., Xu, Y. Y., Lonigan, C. J. and Eppe, S. (2013) 'The Home Literacy Environment and Latino Head Start Children's Emergent Literacy Skills', *Developmental Psychology*, 49 (4), pp. 775–791, DOI: 10.1037/a0028766.
- Feigenson, L., Libertus, M. E. and Halberda, J. (2013) 'Links between the intuitive sense of number and formal mathematics ability', *Child Development Perspectives*, 7, pp. 74–79.

- Fekkes, M., Pijpers, F. I., Fredriks, A. M., Vogels, T. and Verloove-Vanhorick, S. P. (2006) 'Do bullied children get ill, or do ill children get bullied? A prospective cohort study on the relationship between bullying and health-related symptoms', *Pediatrics*, 117 (5), pp. 1,568–574.
- Foorman, B. R., Chen, D.-T., Carlson, C., Moats, L., David, J. F. and Jack, M. F. (2003) 'The necessity of the alphabetic principle to phonemic awareness instruction', *Reading and Writing*, 16, pp. 289–324.
- Foster, T. D., Froyen, L. C., Skibbe, L. E., Bowles, R. P. and Decker, K. B. (2016) 'Fathers' and mothers' home learning environments and children's early academic outcomes', *Reading and Writing*, 29 (9), pp. 1,845–863, DOI: 10.1007/s11145-016-9655-7.
- Foulin J. (2005) 'Why is letter-name knowledge such a good predictor of learning to read?', *Reading and Writing*, 18, pp. 129–155, <http://dx.doi.org/10.1007/s11145-004-5892-2>.
- Friberg, J. C. (2010) 'Considerations for test selection: How do validity and reliability impact diagnostic decisions?', *Child Language Teaching and Therapy*, 26 (1), pp. 77–92, DOI: 10.1177/0265659009349972.
- Frith, U. (1985) 'Beneath the surface of developmental dyslexia', in K. Patterson, M. Coltheart and J. Marshall (eds), *Surface Dyslexia*, London: Lawrence Erlbaum Associates Ltd.
- Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B. and Caffrey, E. (2011) 'The Construct and Predictive Validity of a Dynamic Assessment of Young Children Learning to Read: Implications for RTI Frameworks', *Journal of Learning Disabilities*, 44 (4), pp. 339–347, DOI: 10.1177/0022219411407864.
- Fulgini, A. S., Han, W. J. and Brooks-Gunn, J. (2004) 'The Infant-Toddler HOME in the 2nd and 3rd Years of Life', *Parenting: Science and Practice*, 4 (2-3), pp. 139–159, DOI: 10.1207/s15327922par0402and3_3.
- Garon, N., Bryson, S. E. and Smith, I. M. (2008) 'Executive function in preschoolers: a review using an integrative framework', *Psychological Bulletin*, 134 (1), pp. 31–60.
- Geary, D. C. (1994) *Children's mathematical development: Research and practical applications*, Washington DC: American Psychological Association.
- Geeraerts, S. B., Deutz, M. H. F., Deković, M., Bunte, T., Schoemaker, K., Espy, K. A. and Matthys, W. (2015) 'The child behavior checklist dysregulation profile in preschool children: a broad dysregulation syndrome', *Journal of the American Academy of Child and Adolescent Psychiatry*, 54 (7), pp. 595–602.
- Gelman, R. and Gallistel, C. R. (1978) *The child's understanding of number*, Cambridge MA: Harvard University Press.
- Gelman, R. and Meck, E. (1983) 'Preschoolers' counting: Principles before skill', *Cognition*, 13, pp. 343–359.
- Gentry, J. R. (1982) 'An analysis of developmental spelling in GYNS AT WRK', *The Reading Teacher*, 36, pp. 192–200.
- Gilkerson, J. and Richards, J. A. (2008) 'The LENA Developmental Snapshot' (Technical Report No. LTR-07-2), available at http://lena.org/wp-content/uploads/2016/07/LTR-07-2_Snapshot.pdf
- Gilkerson, J., Richards, J., Greenwook, C. and Montgomery, J. (2017) 'Language assessment in a snap: Monitoring progress up to 36 months', *Child Language Teaching and Therapy*, 33 (2), pp. 99–115.
- Gilmore, C. K., Attridge, N. and Inglis, M. (2011) 'Measuring the approximate number system', *Quarterly Journal of Experimental Psychology*, 64, pp. 2,099–109.
- Gini, G. and Pozzoli, T. (2009) 'Association between bullying and psychosomatic problems: A meta-analysis', *Pediatrics*, 123 (3), pp. 1,059–065.

- Göbel, S., Watson, S. E., Lervåg, A. and Hulme, C. (2014) 'Children's arithmetic development: It is number knowledge, not the approximate number sense that counts', *Psychological Science*, 25, pp. 789–798.
- Goodman, R. (1997) 'The Strengths and Difficulties Questionnaire: A research note', *Journal of Child Psychology and Psychiatry*, 38, pp. 581–586.
- Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S. and Abner, K. (2013) 'An Assessment of the Validity of the ECERS-R With Implications for Measures of Child Care Quality and Relations to Child Development', *Developmental Psychology*, 49 (1), pp. 146–160, DOI: 10.1037/a0027899.
- Gordon, R. A., Hofer, K. G., Fujimoto, K. A., Risk, N., Kaestner, R. and Korenman, S. (2015) 'Identifying High-Quality Preschool Programs: New Evidence on the Validity of the Early Childhood Environment Rating Scale-Revised (ECERS-R) in Relation to School Readiness Goals', *Early Education and Development*, 26 (8), pp. 1,086–110, DOI: 10.1080/10409289.2015.1036348.
- Gough, P. B. (1996) 'How children learn to read and why they fail', *Annals of Dyslexia*, 46, pp. 1–20.
- Gray, S., Plante, E., Vance, R. and Henrichsen, M. (1999) 'The diagnostic accuracy of four vocabulary tests administered to preschool-age children', *Language Speech and Hearing Services in Schools*, 30 (2), pp. 196–206.
- Griffin, S. (2005) 'Fostering the development of whole-number sense: teaching mathematics in the primary grades', in M. S. Donovan and J. D. Bransford (eds), *How students learn: History, mathematics, and science in the classroom*, Washington DC: The National Academies Press (pp. 257–308).
- Grigorenko, E. L. (2009) 'Dynamic Assessment and Response to Intervention: Two Sides of One Coin', *Journal of Learning Disabilities*, 42 (2), pp. 111–132, DOI: 10.1177/0022219408326207.
- Grigorenko, E. L. and Sternberg, R. J. (1999) *Assessing cognitive development in early childhood*, Washington DC: The World Bank.
- Haccoun, R. R. (2001) 'Test review of the Renfrew Bus Story', in B. S. Plake and J. C. Impara (eds), *The 14th Mental Measurements yearbook*, retrieved from <http://marketplace.unl.edu/buros/> [accessed 20 April 2014].
- Halle, T. G. and Darling-Churchill, K. E. (2016) 'Review of measures of social and emotional development', *Journal of Applied Developmental Psychology*, 45, pp. 8–18.
- Harms, T. and Clifford, R. M. (1980) *Early Childhood Environment Rating Scale*, New York: Teachers College Press.
- Harms, T., Clifford, R. M. and Cryer, D. (1998) *Early Childhood Environment Rating Scale—Revised edition*, New York: Teachers College Press.
- Harms, T., Cryer, D. and Clifford, R. M. (2007). *The family child care environment rating scale: Revised edition*. New York: Teachers College Press.
- Hasson, N. and Botting, N. (2010) 'Dynamic assessment of children with language impairments: A pilot study', *Child Language Teaching and Therapy*, 26 (3), pp. 249–272.
- Hasson, N., Camilleri, B., Jones, C., Smith, J. and Dodd, B. (2013) 'Discriminating disorder from difference using dynamic assessment with bilingual children', *Child Language Teaching and Therapy*, 29 (1), pp. 52–70, DOI: 10.1177/0265659012459526.
- Hasson, N., Dodd, B. and Botting, N. (2012) 'Dynamic Assessment of Sentence Structure (DASS): design and evaluation of a novel procedure for the assessment of syntax in children with language impairments', *International Journal of Language and Communication Disorders*, 47 (3), pp. 285–299, DOI: 10.1111/j.1460-6984.2011.00108.x.

- Heckman, J. J., Stixrud, J. and Urzua, S. (2006) 'The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior', *Journal of Labor Economics*, 24 (3), pp. 411–482.
- Henderson, E. (1985) *Teaching spelling*, Boston: Houghton Mifflin.
- Hoff E. (2003) 'The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech', *Child Development*, 74, pp. 1,368–378.
- Hoffman, L. M., Loeb, D. F., Brandel, J. and 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 (6), pp. 1,597–608, DOI: 10.1044/1092-4388(2011/10-0213).
- Holmes, C. J., Kim-Spoon, J. and Deater-Deckard, K. (2016) 'Linking executive function and peer problems from early childhood through middle adolescence', *Journal of Abnormal Child Psychology*, 44 (1), pp. 31–42.
- Hood, M., Conlon, E. and Andrews, G. (2008) 'Preschool home literacy practices and children's literacy development: A longitudinal analysis', *Journal of Educational Psychology*, 100 (2), pp. 252–271, DOI: 10.1037/0022-0663.100.2.252.
- Hurry, J. and Doctor, E. (2007) 'Assessing Literacy in Children and Adolescents', *Child and Adolescent Mental Health*, 12, pp. 38–45.
- Huttenlocher, J., Jordan, N. C. and Levine, S. C. (1994) 'A mental model for early arithmetic', *Journal of Experimental Psychology: General*, 123, pp. 284–296.
- Iuculano, T., Tang, J., Hall, C. W. B. and Butterworth, B. (2008) 'Core information processing deficits in developmental dyscalculia and low numeracy', *Developmental Science*, 11 (5), pp. 669–680.
- Jones, D. E., Greenberg, M. and Crowley, M. (2015) 'Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness', *American Journal of Public Health*, 105, pp. 2,283–290, DOI: 10.2105/AJPH.2015.302630.
- Jones, S. M., Zaslow, M., Darling-Churchill, K. E. and Halle, T. G. (2016) 'Assessing early childhood social and emotional development: Key conceptual and measurement issues', *Journal of Applied Developmental Psychology*, 45, pp. 42–48.
- Jordan, N. C., Glutting, J., Ramineni, C. and Watkins, M. W. (2010) 'Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade', *School Psychology Review*, 39, pp. 181–195.
- Jordan, N. C., Huttenlocher, J. and Levine, S. C. (1994) 'Assessing early arithmetic abilities: Effects of verbal and nonverbal response types on the calculation performance of middle- and low-income children', *Learning and Individual Differences*, 6, pp. 413–432.
- Kalia, V. and Reese, E. (2009) 'Relations Between Indian Children's Home Literacy Environment and Their English Oral Language and Literacy Skills', *Scientific Studies of Reading*, 13 (2), pp. 122-145, DOI: 10.1080/10888430902769517.
- Kim, Y. S., Petscher, Y., Foorman, B.R. and Zhou, C. (2010) 'The contributions of phonological awareness and letter-name knowledge to letter-sound acquisition—A cross-classified multilevel model approach', *Journal of Educational Psychology*, 102, pp. 313–326, <http://dx.doi.org/10.1037/a0018449>.
- Kleemans, T., Peeters, M., Segers, E. and Verhoeven, L. (2012) 'Child and home predictors of early numeracy skills in kindergarten', *Early Childhood Research Quarterly*, 27 (3), pp. 471–477, DOI: 10.1016/j.ecresq.2011.12.004.
- Kline, P (1986) *A Handbook of Test Construction: introduction to psychometric design*, London: Methuen.
- Kluczniok, K., Lehl, S., Kuger, S. and Rossbach, H. G. (2013) 'Quality of the home learning environment during preschool age - Domains and contextual conditions', *European Early*

- Childhood Education Research Journal*, 21 (3), pp. 420–438, DOI: 10.1080/1350293x.2013.814356.
- Kochanska, G., Murray, K. T. and Harlan, E. T. (2000) 'Effortful control in early childhood: continuity and change, antecedents, and implications for social development', *Developmental Psychology*, 36 (2), pp. 220–232.
- Komeili, M. and Marshall, C. R. (2013) 'Sentence repetition as a measure of morphosyntax in monolingual and bilingual children', *Clinical Linguistics and Phonetics*, 27 (2), pp. 152–162, DOI: 10.3109/02699206.2012.751625.
- La Paro, K. M., Pianta, R. C. and Stuhlman, M. (2004) 'The classroom assessment scoring system: Findings from the prekindergarten year', *Elementary School Journal*, 104 (5), pp. 409–426.
- La Paro, K. M., Williamson, A. C. and Hatfield, B. (2014) 'Assessing Quality in Toddler Classrooms Using the CLASS-Toddler and the ITERS-R', *Early Education and Development*, 25 (6), pp. 875–893, DOI: 10.1080/10409289.2014.883586.
- La Paro, K., Hamre, B. K. and Pianta, R. C. (2011) *Classroom Assessment Scoring System Toddler (CLASS)*, Baltimore MD: Brookes.
- Laible, D., McGinley, M., Carlo, G., Augustine, M. and Murphy, T. (2014) 'Does engaging in prosocial behavior make children see the world through rose-colored glasses?', *Developmental Psychology*, 50, pp. 872–880.
- Language and Literacy Research Consortium (2015) 'The Dimensionality of Spanish in Young Spanish-English Dual Language Learners', *Journal of Speech Language and Hearing Research*, 58 (3), pp. 754–766.
- Law, J. and 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 (4), pp. 198–206, DOI: 10.1111/j.1475-3588.2008.00503.x.
- LeFevre, J., Sadesky, G. S. and Bisanz, J. (1996) 'Selection of procedures in mental addition: Reassessing the problem-size effect in adults', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, pp. 216–230.
- Lidz, C. S. and 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), pp. 121–133, DOI: 10.1055/s-0029-1215719
- Lindsay, G. and Desforges, M. (1998) *Baseline Assessment*, London: David Fulton.
- Lindsay, G. and Lewis, A. (2003) 'An evaluation of the use of accredited baseline assessment schemes in England', *British Educational Research Journal*, 29 (2), pp. 149–167, DOI: 10.1080/0141192032000060902.
- Lindsay, G., Dockrell, J. E. and Strand, S. (2007) 'Longitudinal patterns of behaviour problems in children with specific speech and language difficulties: Child and contextual factors', *British Journal of Educational Psychology*, 77, pp. 811–828.
- Lindsay, G., Martineau, E. and Lewis, A. (2004) 'The consistency of baseline assessment schemes as measures of early literacy', *Journal of Research in Reading*, 27 (2), pp. 118–131.
- Linver, M. R., Brooks-Gunn, J. and Cabrera, N. (2004) 'The Home Observation for Measurement of the Environment (HOME) Inventory: The Derivation of Conceptually Designed Subscales', *Parenting: Science and Practice*, 4 (2-3), pp. 99–114, DOI: 10.1207/s15327922par0402and3_1
- Marjanovic Umek, L., Podlesek, A. and Fekonja, U. (2005) 'Assessing the home literacy environment - Relationships to child language comprehension and expression', *European Journal of Psychological Assessment*, 21 (4), pp. 271–281, DOI: 10.1027/1015-5759.21.4.271.

- Maskey, M., Warnell, F., Parr, J. R., Le Couteur, A. and McConachie, H. (2013) 'Emotional and behavioural problems in children with autism spectrum disorder', *Journal of Autism and Developmental Disorders*, 43 (4), pp. 851–859.
- Mazzocco, M. M. M., Feigenson, L. and Halberda, J. (2011) 'Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia)', *Child Development*, 82, pp. 1,224–237.
- McCachlan, C. and Arrow, A. (2014) 'Promoting alphabet knowledge and phonological awareness in low socioeconomic child care settings: a quasi experimental study in five New Zealand centers', *Reading and Writing*, 27, pp. 819–839, doi 10.1007/s11145-013-9467-y.
- McCauley, R. (2001) *Assessment of Language Disorders in Children*, Mahwah NJ: Lawrence Erlbaum.
- McClelland, M. M. and Cameron, C. E. (2012) 'Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures', *Child Development Perspectives*, 6 (2), pp. 136–142.
- Miller, L., Gillam, R. and Peña, E. (2001) *Dynamic Assessment and Intervention: Improving children's narrative abilities*, Austin TX: PRO-ED.
- Miller, P. M. and Wilson, M. (1983) *A Dictionary of Social Science Methods*, New York: Wiley
- Missall, K., Hojnosi, R. L., Caskie, G. I. L. and Repasky, P. (2015) 'Home Numeracy Environments of Preschoolers: Examining Relations Among Mathematical Activities, Parent Mathematical Beliefs, and Early Mathematical Skills', *Early Education and Development*, 26 (3), pp. 356–376, DOI: 10.1080/10409289.2015.968243.
- Mix, K. S., Levine, S. C. and Huttenlocher, J. (1997) 'Numerical abstraction in infants: Another look', *Developmental Psychology*, 33, pp. 423–428.
- Mok, M. M. C. and 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 (2), pp. 203–220.
- Moore, D. S. and Cocas, L. A. (2006) 'Perception precedes computation: Can familiarity preferences explain apparent calculation by human babies?', *Developmental Psychology*, 42, pp. 666–678.
- Moos, R. H. and Moos, R. S. (1986) *Family Environment Scale manual* (2nd edn), Palo Alto CA: Consulting Psychologists Press.
- Mott, F. L. (1993) *Absent fathers and child development: Emotional and cognitive effects at ages five to nine*, Columbus OH: The Ohio State University, Center of Human Resources Research.
- Nantel-Vivier, A., Pihl, R. O., Côté, S. and Tremblay, R. E. (2014) 'Developmental association of prosocial behaviour with aggression, anxiety and depression from infancy to preadolescence', *Journal of Child Psychology and Psychiatry*, 55 (10), pp. 1,135–144.
- Nelson, D. and Crick, N. (1999) 'Rose-colored glasses: Examining the social information processing of prosocial young adolescents', *The Journal of Early Adolescence*, 19, pp. 17–38.
- Neuman, S. B., Koh, S. and Dwyer, J. (2008) 'CHELLO: The Child/Home Environmental Language and Literacy Observation', *Early Childhood Research Quarterly*, 23 (2), pp. 159–172, DOI: 10.1016/j.ecresq.2007.11.001.
- Niklas, F. and Schneider, W. (2014) 'Casting the die before the die is cast: the importance of the home numeracy environment for preschool children', *European Journal of Psychology of Education*, 29 (3), pp. 327–345, DOI: 10.1007/s10212-013-0201-6.
- Norwalk, K., Di Perna, J. and Pui-Wa, L. (2014) 'Confirmatory factor analysis of the Early Arithmetic, Reading, and Learning Indicators (EARLI)', *Journal of School Psychology*, 52 (1), pp. 83–96, DOI: 10.1016/j.jsp.2013.11.006.

- Nunes, T., Bryant, P. and Bindman, M. (1997) 'Spelling Acquisition in English', in C. Perfetti, L. Rieben and M. Fayol (eds), *Learning to Spell*, London: Laurence Erlbaum (pp. 151–170).
- Olweus, D. (2013) 'School bullying: Development and some important challenges', *Annual Review of Clinical Psychology*, 9, pp. 751–780.
- Peña, E. D., Reséndiz, M. and Gillam, R. B. (2007) 'The role of clinical judgements of modifiability in the diagnosis of language impairment', *Advances in Speech-Language Pathology*, 9, pp. 332–345, DOI: 10.1080/14417040701413738.
- Peña, E. D., Spaulding, T. J. and Plante, E. (2006) 'The composition of normative groups and diagnostic decision making: Shooting ourselves in the foot', *American Journal of Speech-Language Pathology*, 15 (3), pp. 247–254, DOI: 10.1044/1058-0360(2006/023).
- Perlman, M., Falenchuk, O., Fletcher, B., McMullen, E., Beyene, J. and Shah, P. S. (2016) 'A Systematic Review and Meta-Analysis of a Measure of Staff/Child Interaction Quality (the Classroom Assessment Scoring System) in Early Childhood Education and Care Settings and Child Outcomes', *Plos One*, 11 (12), DOI: 10.1371/journal.pone.0167660.
- Petersen, D. R., Gillam, S. L. and Gillam, R. R. (2008) 'Emerging procedures in narrative assessment - The index of narrative complexity', *Topics in Language Disorders*, 28 (2), pp. 115–130.
- Pianta, R., Hamre, B. and La Paro, K. (2008) *Classroom Assessment Scoring System (CLASS)*, Baltimore MD: Brookes.
- Ponitz, C. C., McClelland, M. M., Matthews, J. S. and Morrison, F. J. (2009) 'A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes', *Developmental Psychology*, 45 (3), pp. 605–619.
- Pullen, P. and Justice, L. (2003) 'Enhancing Phonological Awareness, Print Awareness, and Oral Language Skills in Preschool Children', *Intervention in School and Clinic*, 39, pp. 87–98 [10.1177/10534512030390020401](https://doi.org/10.1177/10534512030390020401).
- Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P. and Dehaene, S. (2009) 'Computer-assisted intervention for children with low numeracy skills', *Cognitive Development*, 24, pp. 450–472.
- Reilly, S., McKean, C. and Levickis, P. (2014). 'Late talking: can it predict later language difficulties?', *Research Snapshot*, (2), Centre for Research Excellence in Child Language, retrieved from https://www.mcri.edu.au/sites/default/files/media/documents/crec_rs2_late-talkers-1_design_v0.1_0.pdf
- Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., ... Wake, M. (2014) 'Specific language impairment: a convenient label for whom?', *International Journal of Language and Communication Disorders*, 49 (4), pp. 416–451, DOI: 10.1111/1460-6984.12102.
- Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., ... Bretherton, L. (2010) 'Predicting Language Outcomes at 4 Years of Age: Findings From Early Language in Victoria Study', *Pediatrics*, 126 (6), E1,530–537, DOI: 10.1542/peds.2010-0254.
- Renfrew C. (1997) *Bus Story Test*, Speechmark.
- Reynolds, C. R. (1981) 'The fallacy of "two years below grade level for age" as a diagnostic for reading disorders', *Journal of School Psychology*, 19 (4), pp. 350–358.
- Riches, N. G., Loucas, T., Baird, G., Charman, T. and 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 (1), pp. 47–60, DOI: 10.3109/13682820802647676.
- Riley, M. S. and Greeno, J. G. (1988) 'Developmental analysis of understanding language about quantities and of solving problems', *Cognition and Instruction*, 5, pp. 49–101.
- Rosenshine, B. and Meister, C. (1994) 'Reciprocal Teaching: A Review of the Research', *Review of Educational Research*, 64 (4), pp. 479–530, <http://www.jstor.org/stable/1170585>.

- Russell, R. L. and Ginsburg, H. P. (1984) 'Cognitive analysis of children's mathematical difficulties', *Cognition and Instruction*, 1, pp. 217–244.
- Rust, J. and Golombok, S. (1999) *Modern Psychometrics: The Science of Psychological Assessment*, Routledge.
- Saxton, M. (2010) *Child Language: Acquisition and Development*, London: Sage Publications.
- Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D. and Foorman, B. R. (2004) 'Kindergarten prediction of reading skills: A longitudinal comparative analysis', *Journal of Educational Psychology*, 96, pp. 265–282.
- Seethaler, P. M., Fuchs, L. S., Fuchs, D. and Compton, D. L. (2012) 'Predicting First Graders' Development of Calculation Versus Word-Problem Performance: The Role of Dynamic Assessment', *Journal of Educational Psychology*, 104 (1), pp. 224–234, DOI: 10.1037/a0024968.
- Semel, E. S., Wiig, E. H. and Secord, W. A. (2004) *Clinical Evaluation of Language Fundamentals-Preschool*, San Antonio TX: The Psychological Corporation.
- Senechal, M. (2006) 'Testing the home literacy model: Parental involvement in kindergarten is differentially related to Grade 4 reading comprehension, fluency, spelling, and reading for pleasure', *Scientific Studies of Reading*, 10, pp. 59–87.
- Siegler, R. S. (1988) 'Individual differences in strategy choices: Good students, not-so-good students, and perfectionists', *Child Development*, 59, pp. 833–851.
- Siegler, R. S. (2009) 'Improving the numerical understanding of children from low-income families', *Child Development Perspectives*, 3, pp. 118–124.
- Siegler, R. S. and Braithwaite, D. W. (2017) 'Numerical development', *Annual Review of Psychology*, 68, pp. 187–213.
- Siraj, I., Kingston, D. and Melhuish, E. (2015) *Assessing Quality in Early Childhood Education and Care: Sustained Shared Thinking and Emotional Well-being (SSTEW) Scale for 2–5-year-olds Provision*, London: Trentham Books.
- Siu, A. L. (2015) 'Screening for speech and language delay in children 5 years old and younger: US preventive services task force recommendation statement', *Pediatrics*, 136 (2), e474–481, DOI: 10.1542/peds.2015–1711.
- Skwarchuk, S. L., Sowinski, C. and LeFevre, J. A. (2014) 'Formal and informal home learning activities in relation to children's early numeracy and literacy skills: The development of a home numeracy model', *Journal of Experimental Child Psychology*, 121, pp. 63–84, DOI: 10.1016/j.jecp.2013.11.006
- Smith, M. W. and Dickinson, D. K. (2002) *Early Language and Literacy Classroom Observation (ELLCO) Toolkit, Research Edition [with] User's Guide*.
- Spaulding, T. J., Hosmer, S. and 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 (5), pp. 453–462, DOI: 10.3109/17549507.2012.762042
- Standards and Testing Agency (2016). *Reception baseline comparability study: Results of the 2015 study*. London: Standards and Testing Agency, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/514581/Reception_baseline_comparability_study.pdf [accessed 24/10/17].
- Stipek, D. & Byler, P. (2004). *The early childhood classroom observation measure*. Early Childhood Research Quarterly, Volume 19, Issue 3, 2004, Pages 375-397.
- Stipek, D. & Byler, P. (2005). *The early childhood classroom observation measure*. Unpublished coding manual. School of Education, Stanford University.

- Starr, A., Libertus, M. E. and Brannon, E. M. (2013) 'Number sense in infancy predicts mathematical abilities in childhood', *Proceedings of the National Academy of Sciences of the U.S.A*, 110 (45), 18,116–120.
- Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A. and Janssens, J. M. (2010) 'Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4- to 12-year-olds: A review', *Clinical Child and Family Psychology Review*, 13, pp. 254–274.
- Swanson, H. L. and Lussier, C. M. (2001) 'A selective synthesis of the experimental literature on dynamic assessment', *Review of Educational Research*, 71 (2), pp. 321–363, DOI: 10.3102/00346543071002321
- Swineford, L. B., Thurm, A., Baird, G., Wetherby, A. M. and Swedo, S. (2014) 'Social (pragmatic) communication disorder: a research review of this new DSM-5 diagnostic category', *Journal of Neurodevelopmental Disorders*, 6, DOI: 10.1186/1866-1955-6-41.
- Sylva, K. (2014) 'The role of families and pre-school in educational disadvantage', *Oxford Review of Education*, 40 (6), pp. 680–695, DOI: 10.1080/03054985.2014.979581.
- 'The Glossary of Educational Reform', <http://edglossary.org/about/>
- Thomason, A. C. and La Paro, K. M. (2009) 'Measuring the Quality of Teacher-Child Interactions in Toddler Child Care', *Early Education and Development*, 20 (2), pp. 285–304, DOI: 10.1080/10409280902773351
- Thompson, R. A. (1994) 'Emotional regulation: A theme in search of definition', in N. A. Fox (ed.), 'The development of emotion regulation: Biological and behavioral considerations', *Monographs of the Society for Research in Child Development*, 59 (2-3, Serial No. 240), pp. 25–52.
- Tiekstra, M., Minnaert, A. and Hessels, M. G. P. (2016) 'A review scrutinising the consequential validity of dynamic assessment', *Educational Psychology*, 36 (1), pp. 112–137, DOI: 10.1080/01443410.2014.915930
- Tindal, G. and Marston, D. (1986) 'Approaches to assessment', in K. Torgesen and B. Y. L. Wong (eds), *Psychology and Educational Perspectives on Learning Difficulties*, New York: Academic Press.
- Totsika, V. and Sylva, K. (2004) 'The Home Observation for Measurement of the Environment Revisited', *Child and Adolescent Mental Health*, 9 (1), pp. 25–35
- Trahan, L., Stuebing, K. K., Hiscock, M. K. and Fletcher, J. M. (2014) 'The Flynn Effect: A Meta-analysis', *Psychological Bulletin*, 140 (5), pp. 1,332–360, <http://doi.org/10.1037/a0037173>
- Treiman, R. and Cassar, M. (1997) 'Spelling Acquisition in English', in C. Perfetti, L. Rieben and M. Fayol (eds), *Learning to Spell*, London: Laurence Erlbaum (pp. 61–80).
- Treiman, R. and Rodriguez, K. (1999) 'Young children use letter names in learning to read words', *Psychological Science*, 10, pp. 334–338.
- Vaillancourt, T., Duku, E., Becker, S., Schmidt, L. A., Nicol, J., Muir, C. and MacMillan, H. (2011) 'Peer victimization, depressive symptoms, and high salivary cortisol predict poorer memory in children', *Brain and Cognition*, 77 (2), pp. 191–199.
- Votruba-Drzal, E. and Miller, P. (2016) 'Reflections on Quality and Dosage of Preschool and Children's Development', *Monographs of the Society for Research in Child Development*, 81 (2), pp. 100–113, DOI: 10.1111/mono.12244
- Wechsler D. (2007) *Wechsler Individual Achievement Test 3rd Edition (WIAT III^{UK})*, London: The Psychological Corp.
- Weinstein, N. and Ryan, R. M. (2010) 'When helping helps: Autonomous motivation for prosocial behavior and its influence on well-being for the helper and recipient', *Journal of Personality and Social Psychology*, 98 (2), pp. 222–244.

- Weisleder, A. and Fernald, A. (2013) 'Talking to children matters', *Psychological Science*, 24, pp. 143–152.
- Wichstrøm, L., Belsky, J. and Steinsbekk, S. (2017, online publication) 'Homotypic and heterotypic continuity of symptoms of psychiatric disorders from age 4 to 10 years: a dynamic panel model', *Journal of Child Psychology and Psychiatry*, doi:10.1111/jcpp.12754.
- Wolke, D., Woods, S., Bloomfield, L. and Karstadt, L. (2000) 'The association between direct and relational bullying and behaviour problems among primary school children', *Journal of Child Psychology and Psychiatry*, 41 (8), pp. 989–1,002.
- Wynn, K. (1992) 'Addition and subtraction by human infants', *Nature*, 358, pp. 749–750.
- Zhou, Q., Chen, S. H. and Main, A. (2012) 'Commonalities and differences in the research on children's effortful control and executive function: A call for an integrated model of self-regulation', *Child Development Perspectives*, 6 (2), pp. 112–121.

Appendix A: Assessments identified in search

Name of Test	Author/s	Publisher
(ABAS-3) Adaptive Behaviour Assessment System, third edition	P. Harrison and T. Oakland	Hogrefe
Action Picture Test revised	C. Renfrew	Speechmark publishing (Routledge)
(ASBI) Adaptive Social Behaviour Inventory		
ASPECTS		CEM Center for Evaluation and Monitoring
(ASQ-3) Age and Stages Questionnaire, 3rd edition	J. Squire and D. Bricker	Paul H. Brookes Publishing Co. Inc.
(BAS III-Early years) British ability Scales	C. D. Elliot, P. Smith and K. McCulloch	GL-Assessment
(BASC-3) Behaviour Assessment System for Children, 3rd edition	C. R. Reynolds and R. W. Kamphaus	Pearson
Baseline and Baseline Progress		GL-Assessment
(Bayley-III) Bayley Scales of Infant and Toddler Development, third edition	N. Bayley	Pearson
(BBCS-3) Bracken Basic Concept Scale	B. A. Bracken	The Psychology Corporation
(BDI-2 Screening Test) Battelle Developmental Inventory Test, second edition	J. Newborg	Riverside Publishing
(BERS-2) Behavioural and Emotional Rating Scale, second edition	M. H. Epstein and J. M Sharma	Houghton Mifflin Harcourt
(BESA) Bilingual English Spanish Assessment	E. Peña, V. F. Gutierrez-Ciellen, A. Iglesias, B. A. Goldstein, L. M. Bedore	AR-Clinical Publications
(BESSI) Brief Early Skills and Support Index	C. Hughes, I. Daly, S. Foley, N. White and R. T. Devine	Centre for Family Research, University of Cambridge
(Boehm-3) Boehm Test Preschool, third edition	A. E. Boehm	Pearson
Boxall Profile	M. Boxall	The Nurture Group Network
(BPVS-3) British Picture Vocabulary Scale, third edition	L. M. Dunn, D. M. Dunn, B. Styles and J. Sewell	GL-Assessment
Brigance Diagnostic Inventory of Early Development II		Curriculum Associates, Inc.
Bus Story	C. Renfrew	The Centerville School Delaware
(CASL-2) Comprehensive Assessment of Spoken Language, second edition	E. Carrow-Woolfolk	Pearson

(CATS) Comprehensive Affect Testing System	S. G. Weiner, A. Gregory, K. B. Froming, C. M. Levy and P. Ekman	Not published
(CBCL) Child Behaviour Checklist	T. M. Achenbach and L. A. Rescorla	Achenbach System of Empirically Based Assessment (ASEBA)
(CBQ) Child Behaviour Questionnaire	S. P. Putnam and M. K. Rothbart	University of Oregon
(CCC-2) Children's Communication Checklist	D. Bishop	Pearson
(CDI) Child Development Inventory	H. Ireton	Ellsworth and Vandermeer Press LLC
(CELDT) California English Language Development Test		Not published. Available at the California Department for Education
(CELF-Preschool 2) Clinical Evaluation of Language Fundamentals, second edition	E. H. Wiig, W. Secord and E. Semel	Pearson
Child Math Assessment	A. Klein, A. and P. Starkey	Not published
CIRCLE (formerly C-PALLS+STEM: Phonological Awareness, Language and Literacy System + Science, Technology, Engineering and Math)	S. H. Landry, M. Assel, J. Williams, T. A. Zucker, P. R. Swank and S. Gunnewig	University of Texas, Health Science Center
(CISU) Conceptual Interview on Scientific Understanding	Spycher, P.	Not published
(CSBI) Child Sexual Behaviour Inventory	W. N. Friedrich	Hogrefe
(DANVA-2) Diagnostic Analysis of Nonverbal Accuracy, second edition	S. Nowicki and M. Duke, M.	Dyssemia Inc.
(DAYC-2) Developmental Assessment of Young Children	J. K. Voress and T. Maddox	Pro-Ed.
(DCCS) Dimensional Change Card Sort	P. D. Zelazo	Not published
(DCS) Developing Skills Checklist		CTB Macmillan/MacGraw-Hill
(DEAP) Diagnostic Test of Articulation and Phonology	B. Dood, Z. Hua, S. Crosbie, A. Holm and A. Izanne	Pearson
(DECA-P2) Devereux Early Childhood Assessment, second edition	P. LeBuffe and J. Naglieri	Kaplan Early Learning Company
(DELV) Diagnostic Evaluation of Language Acquisition	H. N. Seymour, W. R. Thomas, J. deVilliers and A. deVilliers	Pearson
(DEST-2) Dyslexia Early Screening Test, second edition	R. Nicolson and A. Fawcett	Pearson
Developmental Snapshot	J. Richards, C. R. Greenwood and J. K.	LENA Research Foundation

	Montgomery	
Diagnostic Test of Word Reading Performance	H. N. Seymour, W. R. Thomas, J. deVilliers and A. deVilliers	Pearson
(DIAL-4) Developmental Indicators for the Assessment of Learning	C. Mardell and D. S. Goldenberg	Pearson
(DIBELS-6) Dynamic Indicators of Basic Early Literacy skills, 6th edition	R. H. Good and R. A. Kaminsky	http://dibels.uoregon.edu
(EARLI) The Early Arithmetic, Reading, and Learning Indicators	K. E. DiPerna, J. C. Morgan and P. Lei	Not published
Early Functional Communication Profile	S. L. Jensen	Linguisticsystems
(ECADES) Early Childhood Attention Deficit Disorders Evaluation Scale	S. McCartney	Hawthorne-ed
Early Years Foundation Stage Profile		DfE
(ECBI) Eyberg Child Behaviour Inventory	S. Eyberg	PARINC
(ECI-5) Early Childhood Inventory, 5th edition	J. Sprafkin and K. D. Gadow	Checkmate Plus, LTD.
(EExBA-R) Early Excellence Baseline Assessment	S. Riley-Ayers, J. Stevenson-Garcia, E. C. Frede and K. Brennehan	Not published; available at http://datarecognitioncorp.com
Effortful Control Battery	G. Kochanska	Not published
(ELP-MAT) Early Literacy Progress Monitoring Assessment Tool	J. Kaderavek	Not published
(ELS) Early Learning Scale Preschool Assessment	S. Riley-Ayers, J. Stevenson-Garcia, E. C. Frede and K. Brennehan	Not published; available at http://datarecognitioncorp.com
(ELSA) Early Literacy Skills Assessment	A. DeBruin-Parecki	High/Scope Educational Reading Foundation
(EMDA) Early Maths Diagnostic Assessment	L. Fuchs, D. Fuchs, S. Eaton and C. Hamlett	Pearson
Emerging Literacy Screening	L. H. Paulson, I. Noble, S. Jepson and R. A. van den Pol	Not published
(ENT-R) Early Numeracy Test	J. E. van Luit and B. A. Van de Rijt	Not published
(EOWPVT-4) Expressive One-Word Picture Vocabulary Test, 4th edition	N. A. Martin and R. Brownell	Pearson
(ERB) Early Repetition Battery	B. Seeff-Gabriel, S. Chiat and P. Roy	Pearson
(ERDA-2) Early Reading Diagnostic Assessment, second edition	J. L. Wiederholt and B. R. Bryant	Pearson

(ESI-R) Early Screening Inventory-revised	S. J. Meisels, D. B. Marsden, M. S. Wiske and L. W. Henderson	Pearson.
(ESVA) Emergent Science Vocabulary Assessment	Spycher, P.	Not published
(EVT-2) Expressive Vocabulary Test, second edition	K. T. Williams	Pearson
FirstSTEP Screening Test for Evaluating Preschoolers	L. Miller	Not published
(Fluharty-2) Fluharty Preschool Speech and Language Screening Test, second edition	N. B. Fluharty	Pearson
(GAPS) Grammar and Phonology Screening test	H. van der Lely, H. Gardner, K. Froud and A. McClelland	Pearson
(GARS-2) Gilliam Autism Rating Scale, second edition	J. E. Gillian	Pearson
(GFTA-3) Goldman-Fristoe Test of Articulation, third edition	R. Goldman and M. Fristoe	Pearson
(GMDS-ER) Griffiths Mental Development Scales, extended, revised	L. B. Faragher, A. A. Barnard, N. Knosen, N. Kotras, L. E. Burns and D. Challis	Hogrefe
(GRTR-R) Get Ready to Read!–revised	J. Grover Whitehurst and C. Lonigan	Pearson
(HBQ) MacArthur Health and Behaviour Questionnaire	J. M. Armstrong and L. H. Goldstein	John D. and Catherine T. MacArthur Foundation Research Network on Psychopathology and Development (available only as questionnaire)
Hodder Oral Reading Test		
(HTKS)Head-Toes-Knees-Shoulders	C. C. Ponitz, M. M. McClelland, A. M. Jewkes, C. M. Connor and F. J. Morrison	Not published
(IGDIs) Individual Growth and Development Indicators of Early Numeracy	R. Hojnoski and R. Floyd	Early Learning Lab, Inc.
InCAS (Interactive Computerised Assessment System)-PIPS		CEM (Center for Evaluation and Monitoring)
(KABC-II) Kaufman Assessment Battery for Children, second edition	A. S. Kaufman and N. L. Kaufman	Pearson
KeyMaths-3	A. J. Connolly	Pearson
(KLPA-2) Khan Lewis Phonological Assessment, second edition	L. Khan and N. Lewis	Pearson

(LAP-D Screens) Learning Accomplishment Profile-Diagnostic	B. J. Hardin, E. S. Peisner-Feinberg and S. W. Weeks	Kaplan Early Learning Company
MacArthur-Bates Communicative Development Inventories - Words and Sentences	L. Fenson, V. A. Marchman, D. J. Thal, P. S. Dale, J. S. Reznick, J. S. and E. Bates	Brookes Publishing Co.
(MAIN) The Multilingual Assessment Instrument for Narratives	N. Gagarina, D. Klop, S. Kunnari, K. Tantele, T. Välimaa, I. Balciuniene, U. Bohnacker and J. Walters	Not published
(MaLT-CAP) Mathematics Assessment for Learning and teaching - Computer Adaptive Testing	University of Manchester	Hodder Education
Quantitative Reasoning Assessment	T. Nunes and P. Bryant	
(MSCA) McCarthy Scales of Children's Abilities	D. McCarthy	The Psychological Corporation
Mullen Scales of Early Learning	E. M. Mullen	Not published
(NAP) Narrative Assessment Protocol	K. Pence, L. Justice and C. Gosse	Not published
NFER Baseline Assessment		NFER: Evidence for Excellence in Education
(NRDLS-4) New Reynell Developmental Language Scales, 4th edition	S. Edwards, C. Letts and I. Sinka	GL-Assessment
(NSS) Number Sense Screener	N. C. Jordan, J. J. Glutting and N. Dyson	Brookes Publishing Co.
Number Knowledge Test	S. Griffin	McGraw-Hill/SRA
Observation Survey of Early Literacy Achievement	M. Clay	Heineman
(OWLS-2) Oral and Written Language Scales, second edition	Carrow-Woolfolk, E.	Pearson
(PALS-PreK) Phonological Awareness Literacy Screening	M. Invernizzi, A. Sullivan, J. Meier and L. Swank	Virginia Department of Education; University of Virginia
Parents' Evaluation of Developmental Status–Developmental Milestones	N. S. Robertshaw	Available at www.pedtest.com
(PES-2) Preschool Evaluation Scale	S. McCarney and T. J. Arthaud	Hawthorne ed.
(PhAB-2) Phonological Assessment Battery, second edition	S. Gibbs and S. Dodman	GL-Assessment
(PIPA) Pre-reading Inventory of Phonological Awareness	B. Dodd, S. Crosbie, B. McIntosh, B., T. Teitzel and A. Ozanne	Pearson
(PIPPS) Penn Interactive Peer Play Scale	Fantuzzo and McWayne	Not published

(PiRA) Progress in Reading Assessment		
(PTM) Progress Test in Maths		GL Assessment
(PKBS-2) Preschool and kindergarten Behaviour Scales, second edition	K. W. Merrell	HMH
(PLAI-2) Preschool Language Assessment Instrument, second edition	B. Dodd, S. Crosbie, B. McIntosh, T. Teitzel and A. Ozanne	Pearson
(PLS-5 UK) Preschool Language Scale, 5th edition	I. L. Zimmerman R. E. Pond and V. G. Steiner	Pearson
(PNI) Preschool Numeracy Indicators	R. G. Floyd, R. L. Hojnoski and J. M. Key	Not published
(Pre-CTOPPP) Preschool Comprehensive Test of Phonological and Print Processing	C. J. Lonigan, R. K. Wagner, J. K. Torgesen and C. A. Rashotte	Florida State University
(PSA) Preschool Science Assessment	D. B. Greenfield, X. Dominguez, A. Greenberg, J Fuccillo, M. F. Maier and R. Penfield,	Not published
(PSRA) Preschool Self-Regulation Assessment	R. Smith-Donald, C. Cybele Raver, T. Hayes and B. Richardson	Downloadable from CSRP Chicago School Readiness Project
(PUMA) Progress in Understanding Mathematics Assessment	C. Cooke and C. McCarty	Hodder Education
(PWPA) Preschool Word and Print Awareness	L. M. Justice, R. P. Bowles and L. E. Skibbe	Not published
(QELI) Qualls Early Learning Inventory	A. L. Qualls, H. D. Hoover, S. B. Dunbar and D. A. Frisbie	Houghton Mifflin Harcourt
(RDC-PA) Research Diagnostic Criteria-Preschool Age	M. Scheeringa	Not published
Receptive grammar task in Comprehension of Third Person Singular /s/ in AAE-Speaking Children	V. Johnson	Not published
(REEL-3) Receptive Expressive Emergent Language, third edition	K. Bzoch, R. League and V. Brown	Pro-Ed
Sandwell Early Numeracy Test, revised	C. Arnold, P. Bowden, M. Talents and R. Waldon	Imaginative Minds
(SAT 10) Stanford Achievement Test		Center for Advanced Study of Teaching and Learning, University of Virginia
Schedule of Growing Skills	M. Bellman, S. Lingam and A. Aukett	GL-Assessment
(SCQ) Social Communication Questionnaire	M. Rutter and A. Bailey	Hogrefe
(SDQ) Strengths and Difficulties	R. N. Goodman	DAWBA

Questionnaire		
(SELF) vocabulary measure	L. Santiago	Not published
(SESBI-R) Sutter-Eyberg Student Behaviour Inventory, revised	S. Eyberg	PARINC
(SNAP) Strong Narrative Assessment Procedure	C. Strong	Not published
(SSBS-2) School Social Behaviour Scales, second edition	K. W. Merren and P. Caldarella	Brookes Publishing Co.
(SSIS) Replaced by Social Skills Improvement System	F. M. Gresham and S. N. Elliott	Pearson
(TEAM) Tools for Early Assessment in Mathematics		McGraw-Hill Education
(TELD-3) Test of Early Language Development, third edition	W. P. Hresko, D. K. Reid and D. D. Hammill	Pro-Ed
(TEMA-3) Test of Early Mathematics Ability, third edition	H. P. Ginsburg and A. J. Baroody	WPS
(TERA-3) Test of Early Reading Ability, 3rd edition	K. Reid, P. Hresko and D. Hamill	WPS
(TEWL-3) Test of Early Writing, third edition	W. P. Hresko and S. R. Herron	WPSPublish
The Preschool Numeracy Scales.	D. J. Purpura and C. J Lonigan	Not published. See <i>School Psychology Review</i> (2015), 44 (1), pp. 41–59.
(TNR) The Test of Narrative Retell	T. D. Spencer and D. B. Petersen	Not published
(TOLD-4) Development Primary, 4th edition	P. L. Newcomer and D. D. Hammill	Pro-Ed
(TOPEL) Test of Preschool Early Literacy	C. Lonigan, R. Wagner, J. Torgesen and C. Rashotte	Pro-Ed
(TOPL) The Test of Pragmatic Language	D. Phelps-Terasaki and T. Phelps-Gunn	TEA
(TOWK) Test of Word Knowledge	E. Wiig and W. Secord	Pearson
(TROG-2) Test of reception of grammar, second edition	C. Bishop	Pearson
(TWF-3) Test of word finding, third edition	D. J. German	Pro-Ed
(Vineland-2) Vineland Adaptive Behaviour Scales	S. S. Sparrow and D. V Cicchetti	Pearson
Vocabulary task in Learning New words from Storybooks	L. Justice	Not published
(Wellcomm: The Complete Speech Language Tool Kit	A Hurd and D. McQueen	GL-Assessment
(WIAT-3UK) Weschler Individual Achievement Test	D. Wechsler	Pearson
Woodcock Reading Mastery Test		

Woodcock-Johnson Test of Achievement-IV	R. W. Woodcock, F. A. Shrank, K. S. McGrew and N. Mather	HMH Assessments
Word Finding Vocabulary Test	C. Renfrew and P. Mitchell	Speechmark (Routledge)
(WPPSI-IVUK) Weschsler Preschool and Primary Scale of Intelligence, 4th U.K. edition	D. Wechsler	Pearson
(WRAT-4) Wide Range Achievement Test	G. J. Robertson and G. S Wilkinson	Pearson
(YARC-Early Reading) York Assessment of Reading for Comprehension	M. J. Snowling, S. E. Stothard, P. Clarke, C. Bowyer-Crane, A. Harrington, E. Truelove and K. Nation	GL-Assessment

Appendix B: Assessments by domain

Domain 1: Omnibus

Name of Test	Author/s	Publisher
ASPECTS		CEM (Center for Evaluation and Monitoring)
(BAS III-Early years) British Ability Scales	C. D. Elliot, P. Smith and K. McCulloch	GL-Assessment
Baseline and Baseline Progress		GL-Assessment
(Bayley-III) Bayley Scales of Infant and Toddler Development, third edition	N. Bayley	Pearson
(BBCS-3) Bracken Basic Concept Scale	B. A. Bracken	The Psychology Corporation
(BDI-2 Screening Test) Battelle Developmental Inventory Test, second edition	J. Newborg	Riverside Publishing
(BESSI) Brief Early Skills and Support Index	C. Hughes, I. Daly, S. Foley, N. White and R. T. Devine	Centre for Family Research, University of Cambridge
(Boehm-3) Boehm Test Preschool, third edition	A. E. Boehm	Pearson
Brigance Diagnostic Inventory of Early Development II		Curriculum Associates Inc.
(CDI) Child Development Inventory	H. Ireton	Ellsworth and Vandermeer Press LLC
CIRCLE (formerly C-PALLS+STEM: Phonological Awareness, Language and Literacy System + Science, Technology, Engineering and Math)	S. H. Landry, M. Assel, J. Williams, T. Zucker, T. A., Swank, P. R. and S. Gunnewig	University of Texas, Health Science Center
(DAYC-2) Developmental Assessment of Young Children	J. K. Voress and T. Maddox	Pro-Ed
(DCS) Developing Skills Checklist		CTB Macmillan/MacGraw-Hill
(DIAL-4) Developmental Indicators for the Assessment of Learning	C. Mardell and D. S. Goldenberg	Pearson
(EARLI) The Early Arithmetic, Reading, and Learning Indicators	K. E. DiPerna, J. C. Morgan and P. Lei	Not published
Early Functional Communication Profile	S. L. Jensen	Linguisticsystems
(EExBA-R) Early Excellence Baseline Assessment	S. Riley-Ayers, J. Stevenson-Garcia, E. C. Frede and K. Brenneman	Not published; available at http://datarecognitioncorp.com
(ELS) Early Learning Scale Preschool	S. Riley-Ayers, J.	Not published; available at

Assessment	Stevenson-Garcia, E. C. Frede and K. Brenneman	http://datarecognitioncorp.com
(ESI-R) Early Screening Inventory-Revised	S. J. Meisels, D. B. Marsden, M. S. Wiske and L. W. Henderson	Pearson
Early Years Foundation Stage Profile		DfE
(GMDS-ER) Griffiths Mental Development Scales, extended, revised	L. B. Faragher, A. A. Barnard, N. Knosen, N. Kotras, L. E. Burns and D. Challis	Hogrefe
(IGDIs) Individual Growth and Development Indicators of Early Numeracy	R. Hojnoski and R. Floyd	Early Learning Lab Inc.
(InCAS) (Interactive Computerised Assessment System)-PIPS		CEM (Center for Evaluation and Monitoring)
(LAP-D Screens) Learning Accomplishment Profile-Diagnostic	B. J. Hardin, E. S. Peisner-Feinberg and S. W. Weeks	Kaplan Early Learning Company
(MSCA) McCarthy Scales of Children's Abilities	D. McCarthy	The Psychological Corporation
Mullen Scales of Early Learning	E. M. Mullen	Not published
NFER Baseline Assessment		NFER: Evidence for Excellence in Education
Parents' Evaluation of Developmental Status–Developmental Milestones	N. S. Robertshaw	Available at www.pedtest.com
(PES-2) Preschool Evaluation Scale	S. McCarney and T. J. Arthaud	Hawthorne ed.
(QELI) Qualls Early Learning Inventory	A. L. Qualls, H. D. Hoover, S. B. Dunbar and D. A. Frisbie	Houghton Mifflin Harcourt
Schedule of Growing Skills	M. Bellman, S. Lingam and A. Aukett	GL-Assessment
(Vineland-2) Vineland Adaptive Behaviour Scales	S. S. Sparrow and D. V Cicchetti	Pearson
(WIAT-3UK) Weschler Individual Achievement Test	D. Wechsler	Pearson
Woodcock-Johnson Test of Achievement-IV	R. W. Woodcock, F. A. Shrank, K. S. McGrew and N. Mather	HMH Assessments
(WPPSI-IVUK) Weschsler Preschool and Primary Scale of Intelligence, 4th U.K. edition	D. Wechsler	Pearson
(WRAT-4) Wide Range Achievement Test	G. J. Robertson and G. S Wilkinson	Pearson

Domain 2: Language

Name of Test	Author/s	Publisher
Action Picture Test revised	C. Renfrew	Speechmark (Routledge)
(BESA) Bilingual English Spanish Assessment	E. Peña, V. F. Gutierrez-Ciellen, A. Iglesias, B. A. Goldstein and L. M. Bedore	AR-Clinical Publications
(BPVS-3) British Picture Vocabulary Scale, third edition	L. M. Dunn, D. M. Dunn, B. Styles and J. Sewell	GL-Assessment
Bus Story	C. Renfrew	The Centerville School Delaware
(CASL-2) Comprehensive Assessment of Spoken Language, second edition	E. Carrow-Woolfolk	Pearson
(CCC-2)Children's Communication Checklist	D. Bishop	Pearson
(CELF-Preschool 2) Clinical Evaluation of Language Fundamentals, second edition	E. H. Wiig, W. Secord and E. Semel	Pearson
(CISU) Conceptual Interview on Scientific Understanding	Spycher, P.	Not published
(DEAP) Diagnostic Test of Articulation and Phonology	B. Dood, Z. Hua, S. Crosbie, A. Holm and A. Izzanne	Pearson
(DELV) Diagnostic Evaluation of Language Acquisition	H. N. Seymour, W. R. Thomas, J. deVilliers and A. deVilliers	Pearson
Developmental Snapshot	J. Richards, C. R. Greenwood and J. K. Montgomery	LENA Research Foundation
(EOWPVT-4) Expressive One-Word Picture Vocabulary Test, 4th edition	N. A. Martin and R. Brownell	Pearson
(ERB) Early Repetition Battery	B. Seeff-Gabriel, S. Chiat and P. Roy	Pearson
(ESVA) Emergent Science Vocabulary Assessment	P. Spycher	Not published
(EVT-2) Expressive Vocabulary Test, second edition	K. T. Williams	Pearson
(FirstSTEP) Screening Test for Evaluating Preschoolers	L. Miller	Not published
(Fluharty-2) Fluharty Preschool Speech and Language Screening Test, second edition	N. B. Fluharty	Pearson
GAPS	H. van der Lely, H. Gardner, K. Froud and A. McClelland	Pearson

(GRTR-R) Get Ready to Read!– revised	J. Grover Whitehurst and C. Lonigan	Pearson
(GFTA-3) Goldman-Fristoe Test of Articulation, third edition	R. Goldman and M. Fristoe	Pearson
(KLPA-2) Khan Lewis Phonological Assessment, second edition	L. Khan and N. Lewis	Pearson
MacArthur-Bates Communicative Development Inventories - Words and Sentences	L. Fenson, V. A. Marchman, D. J. Thal, P. S. Dale, J. S. Reznick and E. Bates	Brookes Publishing Co.
(MAIN) The Multilingual Assessment Instrument for Narratives	N. Gagarina, D. Klop, S. Kunnari, K. Tantele, T. Välimaa, I. Balciuniene, U. Bohnacker and J. Walters	Not published
(NAP) Narrative Assessment Protocol	K. Pence, L. Justice and C. Gosse	Not published
(NRDLS-4) New Reynell Developmental Language Scales, 4th edition	S. Edwards, C. Letts and I. Sinka	GL-Assessment
(OWLS-2) Oral and Written Language Scales, second edition	Carrow-Woolfolk, E.	Pearson
(PLAI-2) Preschool Language Assessment Instrument, second edition	B. Dodd, S. Crosbie, B. McIntosh, T. Teitzel and A. Ozanne	Pearson
(PLS-5 UK) Preschool Language Scale, 5th edition	I. L. Zimmerman R. E. Pond and V. G. Steiner	Pearson
(PSA) Preschool Science Assessment	D. B. Greenfield, X. Dominguez, A. Greenberg, J. Fuccillo, M. F. Maier and R. Penfield	Not published
Receptive grammar task in Comprehension of Third Person Singular /s/ in AAE-Speaking Children	V. Johnson	Not published
(REEL-3) Receptive Expressive Emergent Language, third edition	K. Bzoch, R. League and V. Brown	Pro-Ed
SELF vocabulary measure	L. Santiago	Not published
(SNAP) Strong Narrative Assessment Procedure	C. Strong	Not published
(TELD-3) Test of Early Language Development, third edition	W. P. Hresko, D. K. Reid and D. D. Hammill	Pro-Ed
(TNR) The Test of Narrative Retell	T. D. Spencer and D. B. Petersen	Not published
(TOLD-4) Test of Oral Language Development Primary, 4th edition	P. L. Newcomer and D. D. Hammill	Pro-Ed

(TOPL) The Test of Pragmatic Language	D. Phelps-Terasaki and T. Phelps-Gunn	TEA
(TROG-2) Test of Reception of Grammar, second edition	C. Bishop	Pearson
(TWF-3) Test of word finding, third edition	D. J. German	Pro-Ed
Vocabulary task in Learning New Words from Storybooks	L. Justice	Not published
Wellcomm: The Complete Speech Language Tool Kit	A Hurd, D. McQueen,	GL-Assessment
Word Finding Vocabulary Test	C. Renfrew and P. Mitchell	Speechmark (Routledge)

Domain 3: Literacy

Name of Test	Author/s	Publisher
(CELDT) California English Language Development Test		Not published. Available at the California Department for Education
(DEST-2) Dyslexia Early Screening Test, second edition	R. Nicolson and A. Fawcett	Pearson
Diagnostic Test of Word Reading Performance	H. N. Seymour, W. R. Thomas, J. deVilliers and A. deVilliers	Pearson
(DIBELS-6) Dynamic Indicators of Basic Early Literacy skills, 6th edition	R. H. Good and R. A. Kaminsky	http://dibels.uoregon.edu
(ELP-MAT) Early Literacy Progress Monitoring Assessment Tool	J. Kaderavek	Not published
(ELSA) Early Literacy Skills Assessment	A. DeBruin-Parecki	High/Scope Educational Reading Foundation
Emerging Literacy Screening	L. H. Paulson, I. Noble, S. Jepson and R. A. van den Pol	Not published
(ERDA-2) Early Reading Diagnostic Assessment, second edition	J. L. Wiederholt and B. R. Bryant.	Pearson
Hodder Oral Reading Test		
Observation Survey Of Early Literacy Achievement	M. Clay	Heineman
(PALS-PreK) Phonological Awareness Literacy Screening	M. Invernizzi, A. Sullivan, J. Meier and L. Swank	Virginia Department of Education; University of Virginia
(PhAB-2) Phonological Assessment Battery, second edition	S. Gibbs and S. Dodman	GL-Assessment
(PIPA) Pre-reading Inventory of	B. Dodd, S. Crosbie, B.	Pearson

Phonological Awareness	McIntosh, B. T. Teitzel and A. Ozanne	
(PiRA) Progress in Reading Assessment		
(Pre-CTOPPP) Preschool Comprehensive Test of Phonological and Print Processing	C. J. Lonigan, R. K. Wagner, J. K. Torgesen and C. A. Rashotte	Florida State University
(PWPA) Preschool Word and Print Awareness	L. M. Justice, R. P. Bowles and L. E. Skibbe	Not published
(SAT 10) Stanford Achievement Test		Center for Advanced Study of Teaching and Learning, University of Virginia
(TERA-3) Test of Early Reading Ability, third edition	K. Reid, P. Hresko and D. Hamill	WPS
(TEWL-3) Test of Early Writing, third edition	W. P. Hresko and S. R. Herron	WPS
(TOPEL) Test of Preschool Early Literacy	C. Lonigan, R. Wagner, J. Torgesen and C. Rashotte	Pro-Ed
(TOWK) Test of Word Knowledge	E. Wiig and W. Secord	Pearson
Woodcock reading Mastery Test		
(YARC-Early Reading) York Assessment of reading for Comprehension	M. J. Snowling, S. E. Stothard, P. Clarke, C. Bowyer-Crane, A. Harrington, E. Truelove and K. Nation	GL-Assessment

Domain 4: Social-emotional

Name of Test	Author/s	Publisher
(ABAS-3) Adaptive Behaviour Assessment System, third edition	P. Harrison and T. Oakland	Hogrefe
(ASBI) Adaptive Social Behaviour Inventory		
(ASQ-3) Age and Stages Questionnaire, third edition	J. Squire and D. Bricker	Paul H. Brookes Publishing Co. Inc.
(BASC-3) Behaviour Assessment System for Children, third edition	C. R. Reynolds and R. W. Kamphaus	Pearson
(BERS-2) Behavioural and Emotional Rating Scale, second edition	M. H. Epstein and J. M. Sharma	Houghton Mifflin Harcourt
Boxall Profile	M. Boxall	The Nurture Group Network
(CATS) Comprehensive Affect Testing System	S. G. Weiner, A. Gregory, K. B. Froming, C. M. Levy and P.	Not published

	Ekman	
(CBCL) Child Behaviour Checklist	T. M. Achenbach and L. A. Rescorla	Achenbach System of Empirically Based Assessment (ASEBA)
(CBQ) Child Behaviour Questionnaire	S. P. Putnam and M. K. Rothbart	University of Oregon
(CSBI) Child Sexual Behaviour Inventory	W. N. Friedrich	Hogrefe
(DANVA-2) Diagnostic Analysis of Nonverbal Accuracy, second edition	S. Nowicki and M. Duke, M.	Dyssemia Inc.
(DCCS) Dimensional Change Card Sort	P. D. Zelazo	Not published
(DECA-P2) Devereux Early Childhood Assessment, second edition	P. LeBuffe and J. Naglieri	Kaplan Early Learning Company
(ECADES) Early Childhood Attention Deficit Disorders Evaluation Scale	S. McCartney	Hawthorne-ed
(ECBI) Eyberg Child Behaviour Inventory	S. Eyberg	PARINC
(ECI-5) Early Childhood Inventory, 5th edition	J. Sprafkin and K. D. Gadow	Checkmate Plus LTD.
Effortful Control Battery	G. Kochanska	Not published
(GARS-2) Gilliam Autism Rating Scale, second edition	J. E. Gillian	Pearson
(HBQ) MacArthur Health and Behaviour Questionnaire	J. M. Armstrong and L. H. Goldstein	John D. and Catherine T. MacArthur Foundation Research Network on Psychopathology and Development (available only as questionnaire)
(HTKS) Head-Toes-Knees-Shoulders	C. C. Ponitz, M. M. McClelland, A. M. Jewkes, C. M. Connor and F. J. Morrison	Not published
(PIPPS) Penn Interactive Peer Play Scale	Fantuzzo and McWayne	Not published
(PKBS-2) Preschool and Kindergarten Behaviour Scales, second edition	K. W. Merrell	HMH
(PSRA) Preschool Self-Regulation Assessment	R. Smith-Donald, C. Cybele Raver, T. Hayes and B. Richardson	Downloadable from CSRP Chicago School Readiness Project
(RDC-PA) Research Diagnostic Criteria-Preschool Age	M. Scheeringa	Not published
(SCQ) Social Communication Questionnaire	M. Rutter and A. Bailey	Hogrefe
(SDQ) Strengths and Difficulties Questionnaire	R. N. Goodman	DAWBA

(SESBI-R) Sutter-Eyberg Student Behaviour Inventory, revised	S. Eyberg	PARINC
(SSBS-2) School Social Behaviour Scales, second edition	K. W. Merren and P. Caldarella	Brookes Publishing Co.
(SSIS) Replaced by Social Skills Improvement System	F. M. Gresham and S. N. Elliott	Pearson

Domain 5: Numeracy

Name of Test	Author/s	Publisher
Child Math Assessment	A. Klein and P. Starkey	Not published
(EMDA) Early Maths Diagnostic Assessment	L. Fuchs, D. Fuchs, S. Eaton and C. Hamlett	Pearson
(ENT-R) Early Numeracy Test	J. E. van Luit and B. A. Van de Rijt	Not published
(KABC-II) Kaufman Assessment Battery for Children, second edition	A. S. Kaufman and N. L. Kaufman	Pearson
KeyMaths-3	A. J. Connolly	Pearson
(MaLT-CAP) Mathematics Assessment for Learning and teaching - Computer Adaptive Testing	University of Manchester	Hodder Education
(NSS) Number Sense Screener	N. C. Jordan, J. J. Glutting and N. Dyson	Brookes Publishing Co.
Number Knowledge Test	S. Griffin	McGraw-Hill/SRA
(PNI) Preschool Numeracy Indicators	R. G. Floyd, R. L. Hojnoski and J. M. Key	Not published
Progress Test in Maths		GL-Assessment
(PUMA) Progress in Understanding Mathematics Assessment	C. Cooke and C. McCarty	Hodder Education
Quantitative Reasoning Assessment	T. Nunes and P. Bryant	
Sandwell Early Numeracy Test, revised	C. Arnold, P. Bowden, M. Talents and R. Waldon	Imaginative Minds
(TEAM) Tools for Early Assessment in Mathematics		McGraw-Hill Education
(TEMA-3) Test of Early Mathematics Ability, third edition	H. P. Ginsburg and A. J. Baroody	WPS
The Preschool Numeracy Scales	D. J. Purpura and C. J. Lonigan	Not published; see <i>School Psychology Review</i> (2015), 44 (1), pp. 41–59.