Measures of Child Development: A review

Helen Bedford
Senior Lecturer Children's Health

Suzanne Walton
Research Associate

Jane Ahn
Research Assistant

Policy Research Unit in the Health of Children, Young People and Families

Centre for Paediatric Epidemiology and Biostatistics
UCL Institute of Child Health

June 2013

Acknowledgements

We would like to thank members of the Policy Research Unit for the health of children, young people and families: Terence Stephenson, Catherine Law, Ruth Gilbert, Russell Viner, Miranda Wolpert, Amanda Edwards, Steve Morris, Helen Roberts, and Catherine Shaw.

The Policy Research Unit in the Health of Children, Young People and Families is funded by the Department of Health Policy Research Programme.

This is an independent report commissioned and funded by the Policy Research Programme in the Department of Health. The views expressed are not necessarily those of the department.

The Centre for Paediatric Epidemiology and Biostatistics was supported in part by the Medical Research Council in its capacity as the MRC Centre of Epidemiology for Child Health. Research at the UCL Institute of Child Health and Great Ormond Street Hospital for Children receives a proportion of the funding from the Department of Health's National Institute for Health Research Biomedical Research Centres funding scheme.

Contents

Executive summary

Purpose of the review

The Department of Health Policy Project Team asked the Policy Research Unit in the Health of Children, Young People and Families (CPRU) to conduct a piece of work as follows.

To identify and review the range of existing standardised instruments/tools that could be used to measure children's developmental progress at age 2–2 1/2 years, to enable a population level outcome measure that meets specific criteria to be selected. This should pay specific attention to tools that could be used as part of the 2-2½ years Healthy Child Programme (HCP) review and, include an analysis of the advantages and disadvantages of the different tools for the purpose of a population level outcome measure. The aspects of children's development to consider are physical, social and emotional, cognitive and speech and language.

Methods

We conducted a review using systematic methods to search the literature for papers citing measures of child development. Other sources were also used including a search of the Internet, gathering review papers and consulting experts. 35 measures were identified for further examination. Only those that covered all the domains of interest were examined in more detail. Finally, the two measures which seemed most suitable for the stated purpose were assessed against pre-determined requirements set out by the Department of Health.

Findings

Thirty five measures met our inclusion criteria, 32 of these were identified through the systematic search, with 13 of these covering all the domains of interest. These included measures completed by parents (n=3), measures completed by health professionals based on the direct observation of a child's skills (n=7), and those involving both parents' report and professionals' observations (n=3). Two parent completed measures, Ages and Stages (ASQ-3) and Parents' Evaluation of Developmental Status (PEDS) meet most of the criteria laid out by DH, and have significant advantages over using a measure that is completed by

professionals. Although they are currently being used in many parts of England in the 2-2½ years HCP review, we could not locate formal peer reviewed evaluations of these measures, and only limited evidence of parental acceptability of PEDS. Importantly neither measure appears to have been validated or standardised for use in the UK.

Conclusions and Recommendations

ASQ and PEDS best satisfy the requirements for a population measure of children's development at 2-2½ years to be incorporated into the HCP review, but both measures require proper evaluation in a representative UK population.

As PEDS is basically a pass/fail screening test it would not be as useful as ASQ in providing a more detailed view of the population year on year. It is suggested that both PEDS and ASQ are tested on different cohorts of children to assess their reliability and acceptability.

A subset of each should also be assessed using an appropriate gold standard test to establish the validity in a representative UK population. Following this, one measure should be chosen for national use as a population measure. By their nature, if the two measures are in use, it will not be possible to aggregate the data from each into one outcome measure to provide informative data on the development of 2 year old children.

Introduction

Purpose of the review

The Department of Health Policy Project Team requested the Children, Young People and Families Policy Research Unit (CPRU) to conduct a piece of work with the following aims and requirements.

To identify and review the range of existing standardised instruments/tools* that could be used to measure children's developmental progress at age 2-2½ years in a way that would enable a population level outcome measure that meets specific criteria to be selected. This should pay specific attention to tools that could be used as part of the 2-2½ years Healthy Child Programme (HCP) review and, include an analysis of the advantages and disadvantages of the different tools for the purpose of a population level outcome measure. The aspects of children's development to consider are physical, social and emotional, cognitive and speech and language.

Scope of the review

In conducting this review, systematic methods were used to search the literature. In addition we gathered papers in which one or more measures had been reviewed, searched websites of relevant professional organisations and academic institutions and contacted experts known to have an interest in this topic. We are confident that, through these sources, we have identified the most relevant measures available. However, in view of the stated purpose of the review and the time available, we were unable to locate and critically appraise all the studies in which the identified measures have been used; in particular we did not access unpublished studies. This may mean that our findings with respect to all that is known about the characteristics and use of the measures may be limited and this should be borne in mind.

Background – Current policy and practice

In recognition of the importance of the early years in laying the foundations for health and wellbeing throughout life, and to ensure equity of health and developmental outcomes for children from all backgrounds, the Government has committed to improving outcomes for young children and families through increased investment in preventive and early intervention services in pregnancy and early years [1, 2].

The Healthy Child Programme (HCP) is a universal early intervention and prevention public health programme offered to all families comprising screening tests, immunisations, developmental reviews and information and guidance to support parenting and healthy choices [3]. As well as providing universal services, it is an important mechanism for identifying families in need of additional support and children who are at risk of poor outcomes. The aim is to offer families the services needed to ensure children and families achieve their potential in terms of health and wellbeing.

Since 2010, a succession of reports focussed on prevention through early intervention [1, 2, 4-8]. Field [5] and Allen [6] called for the creation of an outcome measure of children's health and development between the ages of 2 and 3 years and the Social Mobility Strategy (2011) stated the Government's commitment to *"explore the potential to develop an indicator of young children's health and well-being at age two to three"* [2]. Tickell [7] recommended that early years' practitioners should provide parents and carers with a short summary of their children's communication and language, personal, social and emotional and physical development between 24-36 months and, where possible, this should be shared with health visitors to inform the HV led HCP two year review with an insert in the Personal Child Health Record (red book) documenting this information**.** This recommendation was also commended by the Government in *Supporting Families in the Foundation Year*s [8] in which its vision for services for children, parents and families in the foundation years were described. The Government is exploring the possibilities for bringing this early years' summary together with the HCP two year review into a single integrated review from 2015 when the expanded health visiting service will allow it. To explore this, a

joint working group was established in December 2011 by the Departments of Health and Education with a broad membership representing health and early years' practitioners. The lead author of this report, Helen Bedford, is a member of this group.

Although these reports appear to be making the same recommendation, there are important differences with implications for policy implementation. While Field, Allen and the Government recommend the introduction of a measure to monitor population changes in children's development (although the Government also specifically mentions including a measure of wellbeing), Tickell is referring to an assessment of individual children to determine if they are progressing appropriately, so that action can be triggered if intervention is required.

The Public Health Outcomes Framework for England was published in January 2012 and includes Child Development at age 2-2½ years as an indicator:
"It is intended that this indicator is based on an existing validated measure of an aspect of child development at this age. Which aspect of child development is most meaningful and can be efficiently measured for this purpose is being considered." [9]

It was not our aim to conduct a comprehensive scope of all the relevant policy activity in this area and in view of the current level of interest in early intervention, it is likely we are unaware of all the activity in this area. However, we attempted to contact obvious experts and representatives of organisations with overlapping interests, for example ChiMat. Other on-going relevant research that has come to light while we have been engaged in this review is outlined later in this report.

The Healthy Child Programme two year review

One component of the HCP is a health review at 2-2½ years. This age was selected for the review as it is a key stage for speech and language, social and emotional and cognitive development and allows an assessment to be made of a child's current health status and plans for future health promotion, matching services to need. The specific aim of the two year review is *"to optimise child development and emotional wellbeing and reduce inequalities in outcome"* [3]. The HCP two year review document [10] provided guidance on

the content and process of the review and highlighted the need for further work on the development of quality measures, outcomes and data systems.

As part of their work to test the feasibility of using the 2 year review to collect population outcome data, the Child and Maternal Health Observatory (ChiMat) surveyed progress towards implementing the review. The survey covered 81 PCT areas. 90% reported delivering a 2-2½ years review. Of these, almost 90% were targeting children between 24 and 28 months. The reported percentage of children reviewed in localities was variable, for example in some areas it was 100%, but this partly reflects the differential definition of 'reviewed'; in some areas this simply meant the child had been invited to have a review. A variety of measures and instruments were being used to assess children's development with the Ages and Stages Questionnaire (ASQ) or Ages and Stages Questionnaire – Social and Emotional (ASQ-SE) the most commonly used. Despite the use of 'home-grown' measures being explicitly discouraged [10], these were being used in some areas (personal communication with Helen Duncan, Programme Director ChiMat, January 2012).

Requirements for a population measure of children's development at 2 years of age

In considering which measure(s) would be most suitable for use in the 2 year review, as a population measure, a first requirement is to establish what are we aiming to measure, why and what the results will be used for; these issues have implications for selecting the most appropriate measure.

The measure is intended to measure children's development and the domains of interest, physical, social-emotional, speech and language and cognitive development are defined below. However, measuring child development is fraught with challenges due to its dynamic nature; not only is each developmental domain individually complex they are also inter-related. Children tend to develop in spurts rather than in a linear fashion, developing rapidly yet also slipping in and out of 'normality', particularly at a young age. Because many other factors may affect a child's 'performance' such as hunger, tiredness or being in a

strange place, a single test administered at a particular age provides only a snapshot of their abilities which, if the child was tested a week later, may yield different results. This is particularly true in young children. This was highlighted by Marks *et al* who describe child development as like trying to 'measure a moving target'; to gain a clearer picture of a child's development, they emphasize the value of on-going surveillance rather than a one-off assessment [11].

Measuring child development involves measuring abilities and aptitude and making comparisons with children of the same age. Children's development is usually described in terms of the developmental tasks they can or cannot carry out. Gathering data on children's development at a population level at two years requires an assessment of a large representative sample, a production of scores and collation of all scores to produce data describing the development of the two year old population in an area or for the country as a whole. This information could prove valuable information in determining where more resources are required to reduce inequalities and to provide baseline data to monitor the effects of interventions. However, in view of the lower coverage of universal child health reviews among children living in more deprived areas, a disparity which only increases as children get older [12], considerable thought must be given to the sampling methods to ensure a representative sample.

The practical issues of gathering these data must also be considered. It is intended that this measure should form part of the Healthy Child Programme two year review. Most two year reviews are conducted by members of the health visiting team which includes health visitors, nursery nurses, community children's nurses and others. This skill mix means that any measure administered by professionals should have been validated accordingly, taking account of different levels of skills (inter-rater reliability).

Before conducting the review, its purpose, including that of the specific measure, should be explained to parents. They may be less interested in accepting a measure that is purely being used to measure children's development for population monitoring purposes than if it acts as a guide to their own child's development, which can either be used to reassure them or to plan intervention.

Parents are interested in their children's development and the two year review presents an ideal opportunity to discuss this with parents, indeed parents will want to know how their child 'measures up'. This will be particularly true of a measure applied by parents. It would also be unethical to assess a child, identify a potential problem requiring further investigation or intervention and yet not act on that information. An ideal measure would therefore be one that serves two functions:

i) It can be used to assess the child's development which is fed back to parents and which provides a basis for appropriate health promotion and, if necessary, to offer additional support or refer a child for follow up and support
ii) It can be used to inform a population outcome measure.

In reality, if the sole purpose of any measure introduced is to monitor the population, it may also be used as a means of assessing individual children's development. However, we must be very clear that **this should not be allowed to drift into population screening** without it being assessed against National Screening Committee (NSC) criteria for a good screening test and formally approved by NSC.

Other desirable characteristics of the measure include:

- It should be simple, with a numerical output, a "score" that is characterised by a median and a description of the range. This would allow one to monitor general progress over time, make comparisons between areas as well as observe what happens to the range, so that inequalities are not increased. It would also allow assessments to be made of the impact of interventions. This would be superior to a measure that simply monitors how many children meet a set target year on year.
- There should be up-to-date standardised norms for the population of interest against which to compare scores. In standardisation, the measure is applied to a large group of children for whom it is designed. An individual's or population's scores are then compared to these norms. Ideally, not only should standardisation occur in the country of interest because of differences in population characteristics between countries and different ethnic groups, but it should also be recent. As Johnson and Marlow point out [13], the "Flynn effect" which refers to the upward drift in

standardised scores over time, has not only been found to be of relevance for IQ scores but also for developmental tests. This would result in scores being over estimated if based on older normative data.

- For ease of administration, which may affect acceptability, the ideal instrument should encompass assessment of all the domains of interest. However, each domain would also be scored separately to provide more informative results.
- It should be valid i.e. in this case it should measure the aspects of child development that it is intended to measure. It naturally follows that it should have been validated against a gold standard in the population(s) of interest. Validation is different from standardisation.
- It should be reliable. This relates to the ability of a test to produce stable and consistent results no matter who is performing the test and is of particular relevance with skill mixed teams.
- To ensure acceptability by both parents and health professionals, it should require only one contact.
- It should be quick and easy to administer.
- It should be a positive experience, preferably even fun for the child and so be relevant and appealing.

Acceptability by parents

To monitor population outcomes effectively, high uptake of the measure is required, in this case as it will be incorporated into the review, it will depend on a high uptake of the two year review. Because it will be difficult to separate out the measure from the review as a whole, it is important to ensure the measure is acceptable to parents. The 'right' measure, which is both acceptable to parents and informative, could potentially be pivotal to the success of the two year review as a whole. For example, parents may prefer a measure which they themselves complete as they are more likely to feel a real part of the process. Indeed a measure that involves parents either partly or wholly, but in a meaningful way, is more in keeping with the general ethos of partnership with parents.

Review of Outcome Measures

Aims and Objectives

The aim of the review was to identify existing outcome measures to assess children aged between 2 and 2½ years at a population level in the following domains:

- physical development
- social and emotional development
- cognitive development
- speech and language development

Definitions and considerations

Domains of development

We set out to define these terms before commencing the search, but, in the process of searching, it became clear that there is little consensus about the definition of even these broad aspects of development. In their systematic review of measures of social and emotional skills, Humphrey *et al* highlighted the importance of a clear definition as a basic scientific requirement and yet found this to be lacking [14]. For the purposes of this review we used broad definitions as described by Rydz *et al* in a review of developmental screening [15].

Physical development: Gross and fine motor development – respectively, the control of large groups of muscle involved in walking, sitting or transferring from one position to another and manipulation of objects with the hands in order to eat, draw, play etc.

Social and emotional development: a child's interactions as demonstrated by forming and maintaining relationships and being responsive to others. This also involves personal development – development of self-help skills in activities of daily living, such as feeding, dressing and toileting.

Cognitive development: the ability to problem solve through intuition, perception and verbal and non-verbal reasoning. The ability to retain information learned and understood and to apply it when needed.

Speech and language: articulation, receptive (understanding language) and expressive language skills and the use of non-verbal symbols.

These domains are not discrete, indeed there is considerable overlap particularly between cognition and speech and language development with the latter often used as part of the assessment of cognitive development.

The measures we examined were developed for different purposes and clarity is needed about these before we can determine which is the most appropriate for use as a population measure of children's development at two years. Since there is often a lack of consensus in definitions of some of these purposes, in particular screening, these will be considered next.

Child Health Surveillance

The oversight of the physical, social, and emotional health and development of children. It is initiated by professionals, is synonymous with secondary prevention and includes some screening tests [16].

Developmental delay

Is usually used to mean "the condition in which a child is not developing and/or achieving skills according to the expected time frame" [17]. Unfortunately, most papers are no more definite than this. Should this be statistically defined, i.e. those in the population falling below a certain centile (Limbos and Joyce suggested 10^{th} centile [18]), or should it be absolute, i.e. 'scoring' at a certain level. If the latter, how is that set? Is it those whose function is significantly impaired, or those whose function can be improved by a proven intervention, or those whose function can be improved by a proven intervention *that is readily available.* When assessing eligibility for state services in the USA, individual States have different criteria, but a frequently used cut-off was a 25% delay or 2 standard deviations (SD) below the mean in one or more developmental areas, or a 20% delay or 1.5 SD below the mean in two or more areas [19].

Screening and Screening Tests

An activity which aims to detect defects/disorders using a specific screening test. It may be part of surveillance as defined above. Screening tests are usually applied to whole populations and aim to pick out those individuals who are at greater risk of having a disorder from those who are at lower risk. Tests used in screening are not diagnostic; rather they indicate which individuals are at higher risk of a particular condition and need further, diagnostic testing. Since no screening test is 100% sensitive* or specific†, screening programmes have the potential to do harm by:

i) wrongly labelling individuals who do not have a condition and causing anxiety which may be long lasting or even, through over referral, subjecting them to unnecessary diagnostic tests. Over referral also has implications for costs and for provision of services.
ii) missing affected individuals and thus denying them appropriate interventions.

In the UK, the National Screening Committee (NSC) considers potential screening programmes against criteria based on those developed by Wilson and Jungner [20, 21].

A screening test produces a simple pass/fail result to indicate whether a child is at greater risk of a condition or not. Following a result which indicates a higher risk of the condition, diagnostic testing is needed and if necessary, an intervention with proven effectiveness. Without these, screening has no value. Although some potential screening tests have proven not to satisfy the NSC criteria, e.g. Checklist for Autism in Toddlers (CHAT) and the Edinburgh Postnatal Depression score (EPNDS), they may serve as a useful adjunct to professional judgement.

Because a screening test is purely pass or fail, it would not by itself be a good population measure as it would not enable one to look at the range of abilities, i.e. centiles or quartiles

* Sensitivity refers to the ability of the test to correctly identify people with the condition of interest. This is usually expressed as a percentage or a 0-1.0 scale.

† Specificity is the proportion of people without the condition of interest who are correctly identified as not having the condition.

and how these change over time. A screening test would have been validated on the basis of a fixed cut off, while a continuous measure would allow one to go back and reset the 'pass' level, if appropriate.

Screening for development delay, *"a child who does not meet developmental milestones at the expected age, even after allowing for the range of normality"* [15] has been considered for inclusion as a population screening programme in UK. However, the lack of information about the nature of developmental delay, which is often a complex, imprecise condition, about definitive diagnostic tests and treatment/management options means that whole population screening does not currently meet the UK NSC criteria [22].

It has been suggested that, for developmental screening tests, a sensitivity of 70-80% and a specificity of close to 80% is acceptable [23]. Although lower than acceptable sensitivity rates for many tests in other screening programmes, it is probably not possible to achieve much higher rates that these because of the complex nature of child development [23]. Equally this rate is relatively low for specificity, but the value of using a test with lower specificity was defended in one study of 512 children aged 7 months to 8 years included in a validation study of a number of screening instruments. All were screened with at least two of four measures and also assessed using diagnostic testing. Although 42% had false positive results on one or more screening tests, the authors argued that because these children were also more likely to score lower on other tests which are predictive of poor educational outcome, it was still important to identify them as they might benefit from intervention [24]. However, a test that labels 42% of the population as in need of extra attention raises the question of how that extra attention is defined and whether it is feasible to deliver it.

Developmental surveillance

This is defined by Drotar *et al* as: *"A flexible, longitudinal, continuous, cumulative process which includes documenting and maintaining the child's developmental history; observing the child's development and identifying potential risk and protective factors for developmental delay" [25].*

International Policy on Developmental Surveillance and Screening in young children

USA

The situation in the USA has been selected for particular attention as over the past decade much work has gone on to encourage routine child health surveillance and periodic screening for developmental delay and, as a result, much attention has been given to identifying and developing suitable measures. This focus on surveillance and screening was prompted by legislation; in the USA the identification of and intervention for disabilities is mandated through the Individuals with Disabilities Education Act (1997).

In 2001, the American Academy of Pediatrics' (AAP) Committee on Children with Disabilities recommended that all infants and young children should be screened for developmental delays [26]. Examples of measures that could be used in screening were listed but little information provided on their characteristics. Five years later, following a survey of AAP members which reported that few paediatricians were using effective means to screen for developmental delay, sub committees of the AAP published further, more detailed guidance, recommending that developmental surveillance and anticipatory guidance (health promotion) should be offered at every well-child visit with screening at the 9, 18 and 30 (or 24) month visits [17].

An algorithm for the provision of developmental surveillance and screening was included together with a review of developmental screening tools. They concluded that no single developmental screening instrument is suitable for all purposes and that child health professionals should choose the tool best fitted to the needs of their populations, practices and their own skill levels. Subsequently, other reviews of available measures have been published [17, 27, 28] and the websites of many US institutions and states contain information with varying levels of detail on measures.

Methods

Summary

We gathered information from a number of sources to identify relevant measures. Key databases were searched to identify papers citing measures, papers reporting reviews of measures were scrutinised, Internet searches were conducted and experts were consulted and only those measures fulfilling specific, pre-determined inclusion criteria were retained. Further information about the characteristics of the retained measures was then gathered from a range of sources again including published papers, review papers, Internet sites and unpublished data provided by experts.

1. Search of key databases

We first conducted a search of key databases using systematic review principles, to identify publications which cited measures used to assess child development.

Defining search terms

To conduct our searches, we identified three broad categories of terms: those related to each of the four domains, those related to measurement, and those related to the population of interest (Table 1). The aim of our search was to identify as many studies as possible that employed any assessment or evaluation tool related to social and emotional, physical and motor, cognitive, and speech and language for children aged 2-2½ years.

For the purposes of this review, outcome measures include any assessment or evaluation tool such as questionnaires, checklists or scales that aim to measure children's development in the domains of interest listed above. The search terms used relating to measurement, as well as those relating to the social and emotional domain, were modelled on Humphrey *et al's* systematic review [14]. For the remaining three domains, preliminary searches were conducted to determine which keywords mapped to the most relevant MeSH terms, and which keywords yielded the most appropriate results.

Table 1: Search terms

Category	Related terms
Measurement	Data collection, assessment, questionnaire, checklist, survey, tool, scale, inventory, diagnosis, test
Development	Development, performance, skills, ability, disability, activity, function
Population of interest	Human, child, infant, preschool, early childhood
Social/emotional	Social, emotional, behaviour, socio - emotional
Cognitive	Cognitive, cognition, learning
Physical/motor	Motor skills, psychomotor, physical,
Speech and language	Speech, language, linguistic, communication

Where possible, searches were conducted using subject headings in combination with keyword terms in order to exclude irrelevant results. For example, searching for combinations of the terms 'language' and 'development' would yield a far greater number of irrelevant hits than if the search was additionally refined by the subject heading 'language development'. While some of the databases did not allow the search to be limited by certain types of parameters (i.e. age group: preschool child, 2-5 years), our search aim was to be inclusive rather than specific, thus our initial searches yielded a large number of results.

Search of key databases

We searched PubMed, ERIC, Web of Knowledge, PsycInfo, and Embase databases for abstracts in English between 01.01.1990 and 31.12.2011. Using the terms defined above, a total of 20620 records were imported into Endnote after duplicates across databases were removed. Table 2, Appendix II shows details of the complete search by database and number of records imported.

2. Filtering

Basic filtering was performed by SW and JA. This involved reading the titles and discarding those not related to children or development. Where the title was insufficient to determine eligibility, the record was kept for the following sorting rounds. In the second round of filtering, abstracts were read and records were excluded if they were for the incorrect age group or did not mention a developmental measure. For the remaining eligible records, papers were retrieved and grouped according to the following categories:

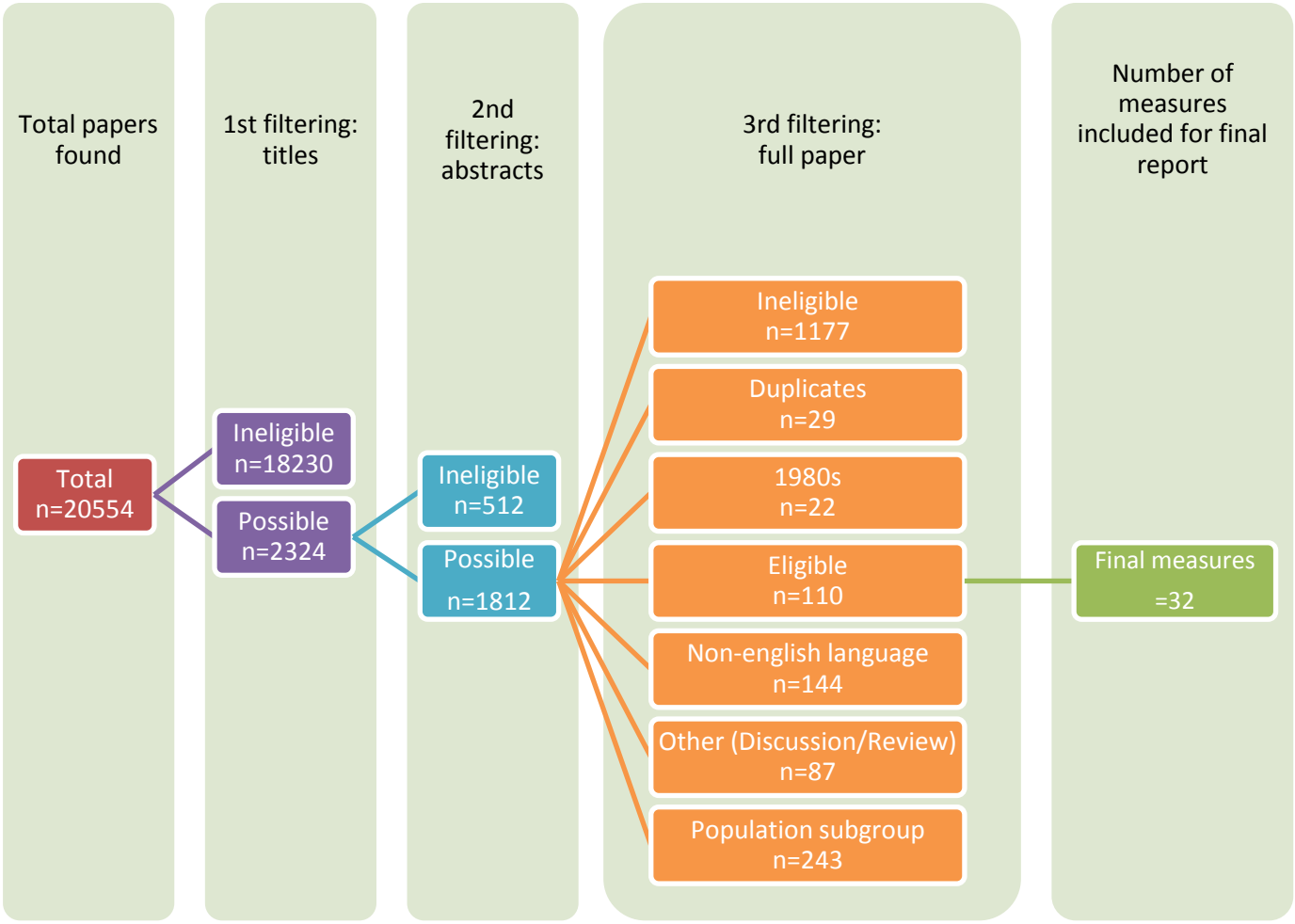
- ineligible
- eligible: 2-year population health measure
- measure used on a population subgroup or as a clinical outcome measure
- measure used in a non-English speaking language situation
- review / other

Papers were deemed eligible if they met the following inclusion criteria:

1. Time period of publication (from 1990-2011)
2. Describes a measurement of development in the relevant domain(s)
3. Describes a measure that can be completed by parents/carers and/or by health professionals
4. The measure has been validated in an appropriate age group (2-2½ years)
5. The measure is available in the English language
6. Requires only one contact

Papers were deemed ineligible if the measure described did not cover a broad range of skills in at least one particular domain, for example it focussed on only one aspect of speech and language development such as expressive communication. Review papers, studies where the developmental assessment or tool was used as an outcome measure in a population subgroup, papers published before 1990, and studies conducted in a non-English speaking country were not included in the main review but were documented. References from these papers were examined for other relevant papers. Additionally, these papers proved useful for further background reading on measures, to determine their diagnostic utility, and to understand how widely the measure is used in other languages and cultural contexts.

Table 3: Filtering stages



3. Search of the Internet

Once the eligible measures were identified from the papers, it was necessary to gather more detailed information about the technical aspects of the measures from a range of sources. These included publishers of the measures via the Internet, published papers and review papers and unpublished information provided by experts and by authors of the measures. Additional searches were also conducted of websites of specialist organisations e.g. National Children's Bureau and professional organisations such as American Academy of Pediatrics to ensure that we had captured all relevant measures. These websites are listed in Appendix I. Many of the websites we identified containing relevant information were discovered by serendipity, and led onto other relevant sources. This was in part because of the many different terms used to describe measures of child development. Most of the information identified from websites tended to be lists or reviews of available measures, developed for use by US paediatricians.

4. Expert Opinion

We contacted individuals known to have expertise in this area to establish whether any similar reviews had been or were being conducted, to gather knowledge of any relevant measures and/or additional information on measures, and to clarify details of on-going research and current policy which might inform the review.

These experts included Professor Clyde Hertzman (Canada), Professor Frank Oberklaid (Australia), Dr Philip Wilson (Glasgow), Professor James Law (Newcastle), Dr. Mike Roblin (PI, Family Nurse Partnership Trial, Cardiff University), Dr David Elliman (National Screening Committee and co-author of Health for All Children), Professor Robert Goodman (author of Strengths and Difficulties Questionnaire), Dr Jane Squires (University of Oregon, Author of Ages and Stages), Professor Frances Glasgoe (Vanderbilt University, author of Parents Evaluation of Developmental Status PEDS), Professor Mitch Blair (Healthy Child Programme Expert Group), Ms Margiad Williams, Bangor University, Professor Lisa Woolfson, (Strathclyde University) Aideen Naughton, (Cardiff University).

5. Construction of Evidence Tables

A table (Table 4 Appendix III) was constructed showing the main features of the measures identified. This table also presents information on a number of measures that did not meet our inclusion criteria. We have included the tools listed in the HCP Two Year Review guidance [10] because we are aware that they are currently being used as part of the 2 year review (personal communication with Helen Duncan, Programme Director ChiMat, January 2012).

Details of measures recorded were:

- Name of measure
- Acronym
- Domains of Development
- Age range
- Administration (by whom)/versions
- Length
- Completion time
- Scales and subscales/areas screened
- Reliability and validity
- Standardisation

- Scoring
- Cost

6. Detailed description of measures

On the basis that acceptability by parents and health professionals, and ease of administration are more likely to be achieved using a single measure embracing a number of domains, rather using a number of measures, we have selected only those that fulfil this requirement. For each of these 13 measures a detailed description of its characteristics follows with consideration of its strengths and limitations as a population measure of children's development at two years. Finally, table 5 (Appendix IV) presents an assessment against the requirements set out by the DH of the two measures considered most suitable for use as population measure of children's development.

DH requirements for the measure:

- It can be updated on a regular basis (e.g. annually) and enables population level child development at age 2-2½ years to be tracked over time.
- It is a valid and reliable measure of the aspects of child development we wish to measure.
- It is applicable to different groups of the population with differing levels of development and needs.
- It has standardised norms for an appropriate population that can be used to benchmark progress in England.
- It can be aggregated at the national and local (local authority) level.
- It is sensitive to changes at a population level.
- It reflects influences on child development during pregnancy and first two years of life as well as being predictive of later life outcomes, especially school readiness.
- It is simple to apply and is acceptable to families and professionals.
- It minimises burdens on professionals and families.
- It can be integrated with existing clinical contacts with all families around this age.

Findings of the Review

No single measure was identified which was specifically developed to be used as a population measure of all four domains of children's development at 2-2½ years. However a number of measures developed for other purposes have subsequently been used in this way.

Table 4 shows the main characteristics of the 35 measures we identified, 32 from the search of papers and 3 through other sources. We specifically included measures described in the 2 year review guidance document [10]. Although seven of these measures did not meet our final inclusion criteria for various reasons, we are aware that some are currently being used as part of HCP two year review (personal communication with Helen Duncan, Programme Director ChiMat, January 2012).

The Measures

In this section we provide detailed descriptions of each of the 13 measures which show most promise for use in the two year review. These are grouped as measures completed by parents: Ages and Stages (ASQ-3), Parents' Evaluation of Developmental Status (PEDS), Parents' Evaluation of Developmental Status – Developmental Milestones (PEDS-DM), by both health professionals and parents through direct assessment of the child and parent report: Bayley, Child Development Inventory (CDI) and Child Development Review (CDR), or by health professionals alone by directly assessing children's skills: Mullen, Battelle, BDI-2, Brigance, Denver –II, Griffiths, Schedule of Growing Skills.

1. Measures completed by parents

Ages and Stages Questionnaires (ASQ-3) (2009) [29]

Purpose: Screening for developmental delay

Age: 1 month to 66 months (5 ½ years).

Format and administration: ASQ-3 is a developmental screening system comprising 21 age specific questionnaires (for 2, 4, 6, 8, 9, 10, 12, 14, 16, 18, 20, 22, 24, 27, 30, 33, 36, 42, 48, 54, and 60 months). The appropriate age (colour-coded) questionnaire can be given to parents in person, mailed or completed online. Each questionnaire has a short demographic section and 30 questions about the child's development divided into five domains with response options of *'yes', 'sometimes' 'not yet'.* Examples of a question in each of these domains are:

Communication:

'Without your showing him, does your child point to the correct picture when you say, "show me the kitty?", or ask, "Where is the dog?" (She needs to identify only one picture correctly.)'

Gross motor:

'Does your child jump with both feet leaving the floor at the same time?'

Fine motor:

'Does your child get a spoon into his mouth right side up so that the food usually doesn't spill?'

Problem solving:

'If your child wants something she cannot reach, does she find a chair or box to stand on to reach it (for example, to get a toy on a counter or to "help" you in the kitchen)?'

Personal-social:

'Does your child copy the activities you do such as wipe up a spill, sweep, shave, or comb hair?'

In addition, open ended questions are included to elicit parents' concerns. In the 24 month questionnaire there are nine additional questions e.g.:

"Do you have any concerns about your child's vision? If yes, explain:"

"Do you have any concerns about your child's behaviour? If yes, explain:"

In the accompanying instructions, parents are advised to try various activities with their child (making it fun for them) and to make sure their child is rested and fed before trying the activities. Parents rate each item as "Yes" the child does the behaviour, "Sometimes," and "Not Yet." If the child is not cooperative they are asked to try again on another occasion. Questions have been phrased at a reading level for 4^{th}-5^{th} US school grade; this is roughly equivalent to a reading age of 9-10 years.

Time required: Approximately 10 to 15 minutes for a parent to complete, 2-3 minutes for professionals to score.

Training and materials: Little training is required for paraprofessionals or office staff to score the questionnaires. A User's Guide and training materials are available. Activity sheets designed to help parents encourage their children's development are included in the User's Guides. The ASQ-3 requires a one-off purchase as the questionnaires, forms, letters, and activity sheets in the user's guides can be reproduced as many times as needed by a single site. Questionnaires are available in English or Spanish. There are online data management systems for single and multisite programmes plus facilities for families to complete questionnaires online.

Scoring: The ASQ-3 results in a score (out of 60) for each area (communication, gross motor, fine motor, problem solving and personal-social) and these are compared to cut-off points on the scoring sheet. Scores beneath the cut-off points indicate a need for further assessment; scores near the cut-off points call for discussion and monitoring; and scores above the cut-off suggest the child is on track developmentally.

Standardisation and psychometrics: The ASQ-3 was standardised on 15,138 children (1,443 aged 24 months) whose parents completed 18,232 questionnaires. Families were educationally and economically diverse, and their ethnicities roughly matched estimates from the 2007 U.S. Census. Sensitivity was .86 and specificity was .85 overall. Figures for sensitivity and specificity at key ages between 24-30 months are given below:

At 24 months: sensitivity 91.2%, specificity 71.9%

At 27 months: sensitivity 77.8%, specificity 86.4%

At 30 months: sensitivity 86.7%, specificity 93.3% [30]

The gold standard test used to assess ASQ was the Battelle Developmental Inventory–II.

The ASQ has also been validated against the Bayley Scales of Infant Development II (BSID-II) and found to have a sensitivity of 100% and specificity of 87% at 24 months for severely delayed status.[31]

No standardised norms for the UK could be located.

Use: The ASQ-3 has been translated and used in a number of European settings (e.g. France [32] Norway [33, 34], Finland, Spain, the Netherlands [35]), Turkey [36] as well as in North America [37, 38], South America, Asia [39, 40] and Australia [41]. However, it has been pointed out that in only a few studies has its psychometric properties been examined in their own cultural setting after translation [35]. In addition to the general paediatric population, it has been used for follow up of children at increased risk for disability such as prematurity (less than 31 weeks gestation)[41], and after various environmental exposures, medical conditions and assisted reproductive technologies. [42-45] Although ASQ is currently used as part of the HCP two year review in some areas of England, and by the Family Nurse Partnership, no information evaluating its use in this setting was found.

Acceptability by parents: ASQ was designed for use with a range of parents and the reading level was kept low, pictures and examples also assist in ensuring clarity. In studies using ASQ, including children at low and high risk of health and development problems, parents indicate that they find the questionnaires easy and quick to complete and that they help them to learn more about their child's growth and development [37, 41, 46]. In a study comparing parent completed ASQ with health professionals' assessments, using the Bayley Scales of Infant Development, low and middle income US parents completed the questionnaire with reasonable accuracy [47].

Strengths as a population measure

1. ASQ-3 covers the developmental domains of interest although it covers personal-social rather than social-emotional.
2. ASQ-3 has been used as a population measure, and although it is currently being used in the UK as part of the HCP we are not aware of any formal evaluations of its use.
3. ASQ-3 produces scores (out of 60) for each domain and an overall score. This may allow measurement of small changes longitudinally.
4. Its format allows flexibility in administration. For example, it could be incorporated into the two year review in a number of ways: sent to parents in advance of the review, which would allow them to think about their child's development and to gather questions for the later review; adapted for inclusion in the Personal Child Health Record (PCHR), although its length would require a number of pages; for those parents who may have problems with literacy or with language barriers, the individual conducting the review could go through the items with the parent at the time of the review. This would be a useful way of widening access.
5. ASQ allows parents to be active participants in their child's development and encourages enjoyable interaction between parent and child.
6. The results of the ASQ provide a good basis for discussion about the child's current and future development.
7. The authors comment that an important difference between this and other screening tools is that it is designed to show what children can do, not just what they cannot do.
8. Acceptable sensitivity and specificity with figures for these rates among 2 year olds.
9. It has been used among children at high risk of developmental problems.
10. It is quick and easy to complete and to score.
11. Cost efficient as a one-off purchase with questionnaires and other materials being photocopied as required.

Limitations and further questions about ASQ

1. There is a lack of standardised norms for the UK population – this is important as the socio-demographic characteristics of the UK population differ significantly from that of the USA where the measure has been normed.

2. Although ASQ-3 covers all the developmental domains of interest, it focuses on 'personal-social' rather than 'social-emotional', thus issues such as relationships are less well covered. However, ASQ-SE [48], which solely focusses on social and emotional development, could be used in conjunction with ASQ-3. ASQ-SE focuses on a child's social and emotional behaviour in the areas of self-regulation, compliance, communication, adaptive behaviours, autonomy affect and interactions with people.
3. There is a lack of information about acceptability of ASQ-3 among UK (English) parents and health professionals, other than anecdotal reports that 'they like it'.
4. There is a need to evaluate ASQ in the UK (English) population to determine if it can be used with parents with potential language barriers, cultural differences and with literacy problems.
5. Since the 2-2½ year review is currently being conducted at a range of ages between less than 24 months to just less than 36 months (personal communication with Helen Duncan, Programme Director ChiMat, January 2012), different age specific questionnaires would be used. It is not clear whether it is valid to combine the scores from age specific questionnaires into one overall score.
6. ASQ is designed as a system for developmental surveillance and the validity and usefulness of using it as a one-off measure is unclear.
7. Some of the language used in ASQ is 'Americanised'. Parents' understanding of this needs to be assessed and it possibly needs adapted for use in UK.

Parents' Evaluation of Developmental Status (PEDS) (1997) [49]

Purpose: A surveillance tool and screening test to elicit parents' concerns about their child's development and health.

Age: 0 - 8 years of age.

Format and administration: An overall question *'Please list any concerns about your child's learning, development, and behaviour'* is followed by eight short questions to elicit parents' concerns about each developmental domain. It can be conducted as an interview or parents can complete the **"**PEDS Response Form" at home or in a waiting room prior to a consultation. Examples of the questions are:

> *'Do you have any concerns about how your child uses his or her arms and legs?*
>
> *Circle one: No Yes A little COMMENTS:'*

'Do you have any concerns about how your child behaves?

Circle one: No Yes A little COMMENTS:'

Time required: Completion of the response form takes about five minutes with a further two minutes to score.

Scoring: An eight page booklet, *"PEDS Brief Administration and Scoring Guide"* is required to score the "PEDS Response Form". The "PEDS Score Form" has columns for each age range (which allows children's progress to be tracked over time) and identifies which concerns predict problems. On the reverse of the Score Form is the "**PEDS Interpretation Form"** which includes an algorithm for deciding whether to refer, screen further, watch carefully, counsel parents, or simply reassure them.

Training and materials: Minimal training is needed for users and various training materials are available. PEDS Forms are available in English, Spanish, and Vietnamese, with licensed translations in several other languages and the Form is written at the equivalent of 4^{th}-5^{th} US school grade, roughly equivalent to a reading age of 9-10 years. PEDS is also available electronically and provides automated scoring and other resources. The sole UK supplier of PEDS materials has adapted the PEDS response form as a page for inclusion in the personal child health record.

Standardisation and psychometrics: Validation studies in 1997 included 771 children across the US in various settings. PEDS has a sensitivity of 74% to 80% and a specificity of 70% to 80% among 0-8 year olds. For 1 to 3 year olds, sensitivity is 79% and specificity is 79% [50]. It was standardized on 2823 families in the USA from various backgrounds, including different levels of socioeconomic status and varying ethnicity.

Many further studies have been carried out to determine the validity of PEDS. A review by Halle *et al* [28] reported that PEDS has been compared with 14 other developmental assessments including the Child Development Inventory, Bayley Scales of Infant Development, Brigance Screens (short screening test), and Batelle Developmental Inventory Screening Test and been found to compare well.

Use: PEDS has been used in population based surveys, including the US national survey of early childhood of 2068 parents of young children [51] [52].The versions used have traditionally eliminated all open-ended questions and included several other items to

encourage parents to indicate whether they have concerns about their child's vision, hearing, health, and global developmental status. When used in such a way PEDS cannot be used to guide clinical care because it does not elicit parents' specific concerns. As a solution to some of the challenges, an official electronic survey version of PEDS now exists with automated scoring. Children are categorised as high risk, moderate risk, low risk but concerned, or low risk/not concerned. The Survey PEDS also provides options to type in parents' exact comments which enables use of the Survey PEDS in both population research and clinical care.

The Centre for Community Child Health, Royal Children's Hospital, Melbourne, has used PEDS with *"hundreds of children and families across a variety of community-based settings".* The language of the PEDS was changed to conform to Australian language usage. PEDS has been successfully used within Australian day care, [53] in which 98% of 233 parents found the questionnaire easy to complete and 89% felt the tool would be helpful or very helpful to health professionals.

Four out of the ten PEDS questions have been used in conjunction with elements from the MacArthur Communication Development Inventory-UK Short Form (MCDI-UKSF) to develop the Sure Start Language Measure (SSLM)[54].

Although reference is made to PEDS having been *"validated on thousands of children in America, Australia, Great Britain and elsewhere, at pediatric offices, outpatient clinics, day care centers, and schools"* and being used in *"Great Britain's Sure Start program and in trials for the UK National Health Service"* [55], only one published study conducted in UK evaluating the use of PEDS was identified [56]. This describes a pilot study in which 100 two year old children living in the Milton Keynes Sure Start programme area were invited for a review of their health and development. The aims of the project were to determine whether PEDS was parent friendly and time efficient and its effectiveness in identifying physical, behavioural and social developmental issues. 76 parents and children attended the review, with 36% of these parents completing a questionnaire on their views. Most felt fully involved in their child's review but some expressed surprise that the focus had been on their concerns and not on the positive aspects of their child's development. No information was available about the predictive value of PEDS. Subsequently, following contact with the author of the paper, it was established that PEDS has continued to be used in Milton Keynes

though no further data are available on validity. Indeed there is a local review in place to determine which instrument should be used in the future in the two year review (personal communication with Practice Development Lead (Health Visiting), Milton Keynes PCT).

A small study was also conducted in 2008 in four clinic bases in two English boroughs. The aim was to test the feasibility of using PEDS to gather information about parents' views of their child's development at 2 years to facilitate identification of children needing further assessment or intervention. 100 PEDS forms were sent to parents at each clinic base. Parents could return the form by post or use an online version of PEDS to reply. Response rates were poor even in the two areas with a high socio-economic profile (30% and 20%). The authors concluded that PEDS might best be used in early childcare setting and with health professionals rather than by post. Some parents commented that they lacked knowledge of developmental norms and so felt unable to assess their child's development (personal communication with Dr Mitch Blair).

Strengths of PEDS as a population measure

1. PEDS covers the developmental domains of interest.
2. PEDS or parts of PEDS have been used in population surveys although we could not ascertain any detailed published studies of its use in the UK setting.
3. PEDS encourages parents' involvement in assessing their child's development which is in keeping with the general philosophy of the HCP.
4. Its format allows flexibility in administration. For example, it could be incorporated into the two year review in a number of ways: sent in advance to parents for completion before the review; included as a page in the PCHR; completed by parents online; for parents with limited literacy or language barriers, completed at the time of the review with the reviewer. However, the results from one small unpublished UK study suggest that requesting parents to return the form prior to the review may not be the best approach.
5. PEDS is quick and easy to complete and is written at the equivalent of 5-7 year old reading level.
6. It is acceptable to parents and professionals and requires minimal training.
7. Published rates of sensitivity and specificity are acceptable.

8. It has been used among children at risk of developmental problems.
9. Compared with alternatives, the costs are relatively low.

Limitations and Further Questions about PEDS

1. Results of PEDS are produced in five outcome categories. This may make changes in proportions of children in each category and small changes over time difficult to interpret. There may also be limited discrimination between children.
2. As it is not a continuous variable, it is less useful as a population monitoring measure.
3. As PEDS offers no specific opportunity to try things out with the child, it may be a more subjective measure. However, parental concerns have been found to accord well with a child having difficulties.
4. It is unclear whether comparisons with PEDS between very different areas (small areas) e.g. in terms of socio-economic status are valid. For example, parents tend to compare their children's development with others in their social circle. It is not clear to what extent parents' judgements about their child's development, and their impressions of normality and in this case whether 'they have concerns' differ according to the socio - demographic characteristics of where they live.
5. Similarly, asking parents whether they have concerns about aspects of their child's development assumes a level of knowledge about what is normal at a particular age. In US, where much of the work on PEDS has been conducted, and where it has been used as part of the child health surveillance and screening programme, parents are also offered anticipatory guidance. This may mean they are better equipped to judge whether or not their child has a developmental issue.

Parents' Evaluation of Developmental Status – Developmental Milestones (2008) [57] [58] [28]

Purpose: Brief Screening tool for developmental delay, to replace use of informal developmental milestone checklists. It is intended for longitudinal monitoring of developmental progress and whilst it can be used as a stand-alone test, its authors recommend that it is administered alongside PEDS to give a more complete picture of development, as it measures actual skill levels.

Age: From birth to 7 years, 11 months

Format and administration: The test was designed for parent completion but can also be directly administered. It consists of a book of laminated forms, one for each age range that parents complete with a dry erase marker. It is written at a 5-7 year old reading level and parents indicate their answer from a multiple choice format. There are between 6 and 8 items per test depending on the age of the child. Each item covers a different domain and acts as a screen for that domain. Domains include fine motor, gross motor, expressive language, receptive language, self-help, social-emotional, and for older children (4 to 8 years) reading and arithmetic. After completion of the questions, parents are encouraged to read a short story, (which is presented in the book on the opposite page), to their child. The stories focus on child development and positive parenting practices.

PEDS-DM consists of 20 age groupings: 1-month to 3-month intervals in the first and second years of life, 4-month to 6-month intervals up to 5.5 years of age, and in half yearly intervals up to the age of 8 years. In the range of 2-2½ years there are measures for 23 to 25 months, 26 to 28 months and 29 to 33 months.

At 29 to 33 months questions include:

"Can your child scribble with a crayon or marker without going off the page much?"

Choice of answers: *"No", "A little", "Yes"*

"*When your child talks, how many words does he or she usually use at a time?"*

Choice of answers: *"None", "1", "2 or more"*

Time required: the test takes up to 5 minutes to administer and 1 minute to score

Training and materials: No specific training or qualifications are required to use the test. Various on-line training materials (videos and PowerPoint presentations) are available. *"The accompanying professional manual contains a list of items in developmental order by domain so that clinicians can probe the extent of weaknesses or strengths and check the reliability of parents' answers (or administer the measure directly to children)."* The PEDS:DM Family Book contains supplementary measures helpful in screening and surveillance.

Scoring: A single scoring template is placed over the parent's responses to score answers. These are then transferred to a longitudinal developmental chart, which the authors refer to as a 'growth chart'. Failure on any item suggests probable difficulties in that domain and performance below the 16th percentile.

Standardisation and psychometrics: The items for PEDS-DM were selected from the Brigance Inventory of Early Development-II (IED-II), created in 2004, and the Brigance Comprehensive Inventory of Basic Skills-Revised (CIBS-R), developed in 1999. Thus, the norms for the PEDS-DM are based on the norms for these two other tools. The norming sample for PEDS-DM included 1619 children included in the norming studies for these other two tools. The sample was considered representative of the 2006 US population. To examine the relationships between the PEDS-DM and other developmental screeners, children were screened using the PEDS-DM and either the IED-II or CIBS-R. Sensitivity was 70% or greater (average 83%) and specificity 77% to 93% (average 84%) across ages and developmental domains. For the ages 23 to 33 months, sensitivity ranged between 80% and 93% and specificity 82% to 93%. However because the items on the PEDS-DM are taken from IED-II and the CIBS-R, there is an inherent overlap between the tools. Therefore if the PEDS-DM was validated against other tools, its psychometric properties may be less favourable.

Use: An "Australian - English" version of PEDS-DM is available but no references were found for a UK version.

There is an assessment level version of PEDS-DM for use in NICU and early intervention programmes where more detailed test results and follow-up measurements are required. This has more items and gives age-equivalent and percentage of delay scores, no further information could be located on this version.

Acceptability by parents: No specific information could be found.

Strengths of PEDS-DM

1. It covers the developmental domains of interest.
2. Appears to have good psychometric qualities but validation with other (non-related) tools would be important.
3. It allows an opportunity to assess the child's skills and thus could overcome one of the potential limitations of PEDS.
4. Its format allows flexibility in administration. For example, it could be incorporated into the two year review in a number of ways: the parents could be asked to complete it in advance of the review or for parents with limited literacy or language barriers, completed at the time of the review with the reviewer.

5. Relatively inexpensive as main materials can be reused.
6. Incorporates materials which encourage parental learning.

Limitations and Further questions about PEDS-DM

1. PED-DM has to be repeated to give longitudinal data, although the authors say it can be used for a stand-alone test. Its value as a population measure is less clear.
2. As it is better used with PEDS, it adds another dimension to using PEDS alone. The package may be less easy to explain to parents and less quick and easy to administer. There is a need to properly evaluate its use in the UK.
3. PEDS-DM has not been validated for use in the UK nor are there any UK norms.
4. Relatively new, so less experience with its use.
5. Its acceptability in a UK population needs to be assessed.

2. Measures completed by professionals with varying involvement of parents

Child Development Inventory (CDI) (1992) [59]

Purpose: Screening and assessment of children where there are concerns about development. However, it has subsequently been deemed as too long for screening apparently 'normal' children.

Age: 15 months to 6 years (and for older children who are judged to be functioning in this range)

Format and administration: The CDI is an assessment tool used by professionals when there are concerns about a child's development. The CDI consists of a 300 item booklet and an answer sheet for parents to complete. There are 270 statements relating to developmental skills of young children that are observable by parents in everyday situations. These items measure the child's present development in eight areas: social, self-help, gross motor, fine motor, expressive language, language comprehension, letters, and numbers. It also includes a General Development Scale and 30 items to identify parent's concerns about their child's health and growth, vision and hearing, development and behaviour.

Time required: 30 to 50 minutes to complete, and 10 minutes to score

Scoring: Parents are asked to "Answer YES or NO to each statement in the booklet to report what you have seen your child doing." Scoring is done by counting the number of "Yes"

responses for each scale. The scores are recorded on the CDI profile which is used to make comparisons to norms for a child of that age. For the CDI, a child's development is considered to be within the normal range if their scores on the developmental scales are at or above the mean scores for children who are 30% younger (this is equivalent to -2 S.D. below the mean). Problem items are recorded at the bottom of the CDI profile. When interpreting the CDI, children's strengths as well as problems should be identified.

Training and materials: (No information found)

Standardisation and psychometrics: The CDI was developed as a result of 30 years research and clinical experience with the Minnesota Child Development Inventory (MCDI). The CDI was standardised on a sample of 568 children, aged one year to six years three months, from Minnesota[60]. The date of this standardisation study is not reported in the paper, but it was conducted no later than 1992. The standardisation sample was 95% white and the authors say [61] "*The CDI norms established on this normative sample should not be generalized to groups of children who are significantly different from the norm group. It is best to develop local norms for particular communities or school systems."* They go on to say *"The inventory format may be inappropriate for parents of some racial and cultural groups and for parents with less than a high school education."*

The CDI manual says "*validity of the CDI was determined in three ways: first, by examining results for norm group children at younger and older ages; second, by comparing CDI results to psychological test results; and third, by looking at CDI results for children with developmental and other problems."* Sample sizes for these studies were small and did not result in figures for sensitivity and specificity.

Use: The CDI has been used to follow-up high risk children [62].

Glascoe [63] reviewed a number of tests relying on parent information and found the Child Development Inventories to be amongst the best performing tests with a sensitivity approximating 80% and sensitivity approaching 90%.

The CDI has been validated in France with a community sample of 1278 children aged 15 to 72 months. Sensitivity was 84% and specificity was 92% [64].

The sensitivity and specificity of the CDI is said to be lower for younger children. Rydz *et al* [65] carried out a study in Quebec with children recruited at the age of 18 months. Of 152 parents sent the CDI, 114 (75%) completed and returned it. The CDI had poor sensitivity (0.50) but good specificity (0.86). Parents were also asked, "Did you find this questionnaire

easy to complete?" Out of 112, 54 (48%) said it was "very easy" and 49 (44%) said it was "easy".

Strengths of the Child Development Inventory (CDI) as a population measure

1. It covers the developmental domains of interest
2. Relatively low costs involved
3. Parents find it easy to complete.
4. The measure has been shown to have good specificity among 18 month old children in one study.
5. It has been used among children at high risk for developmental problems.

Limitations of the Child Development Inventory (CDI) as a population measure

1. Its original purpose is for use among children where there are concerns about development rather than as an assessment for developmental delay per se.
2. The original standardisation was conducted over 30 years ago and in a largely white US sample.
3. There are no UK norms for this measure, although it could be re normed for a UK population the issue is whether it would require re-validation. Ideally it would include evaluation of the measure against a gold standard.
4. The measure was shown to have poor sensitivity among younger children in one study.
5. Although one study reported that parents found it easy to complete the authors originally stated that its format may not be appropriate for parents with less than a high school education.

Child Development Review (CDR) (1990) [66] [67]

The CDR is used for brief screening to help identify children with health problems, developmental delays and behaviour problems.

Age: 18 months to 5 years

Format and administration: The Child Development Review consists of two sections - a parent questionnaire (CDR-PQ) and a child development chart. The Parent Questionnaire provides information about toddlers' and pre-schoolers' health, development and

adjustment, as well as enquiring about the parent's functioning. Parents complete the front of the form which has six open ended questions and a checklist of 25 possible problems. The problems list covers health, growth, hearing, vision, habits (eating, elimination, sleeping), aches and pains, energy, motor symptoms, language symptoms, behaviour and emotional problems. Parents' responses are classified as "no problem", "a possible problem" or "possible major problem". One question relates directly to the parents "How are you doing as a parent and otherwise, at this time?"

The reverse side, for professionals to complete, contains the child development chart which covers social, self-help, gross motor skills, fine motor skills and language for the first 5 years. The professional can use the chart as an observational guide, a parent interview guide or a parent hand-out.

Typically the parent questionnaire and/or the development chart are used for brief screening purposes. Physicians may be more inclined to use the development chart whereas teachers may be more likely to use the parent questionnaire. However, use of both tools together gives a more complete picture, and can be used to conduct a comprehensive review.

Time required: 5 minutes to administer and 5 minutes to score

Scoring: The parent's responses to the six questions are marked with one of the following symbols: "OK" (No problems or doing well); "?" (Possible Problem – ask for more information); or "P" (Possible Major Problem – ask for more information and consider referral). The development chart results are compared to age norms with a cut off performing at a level below that equivalent to 70% of the child's age and classified as "typical" for age in all areas, or as "borderline" or "delayed" in one or more areas of development.

Training and materials: The Instruction manual gives information on administration of the test. Additional training materials do not appear to be available or necessary.

Standardisation and psychometrics: The tool was validated on 220 predominantly white (95%) children aged 3 and 4 years from Minnesota. Sensitivity was .68 and specificity .88.

Use: The CDR has been used in a variety of educational and health care settings. The authors say that "Using this Chart in other communities and with children of diverse cultural backgrounds should be preceded by a careful review of its contents." This would be a major limitation if applying to a UK population of different ethnicities.

Strengths of the Child Development Review as a population measure

1. It covers the developmental domains of interest
2. Parents are involved in the process.
3. It elicits parents' concerns as well as using more objective measures of the child's skills.
4. It is quick and easy to administer.
5. It is relatively inexpensive.

Limitations and further questions about the Child Development Review as a population measure

1. The measure has not been standardised among a UK population.
2. The authors advise caution using this measure among children with diverse cultural backgrounds.

Bayley Scales of Infant and Toddler Development, Third edition (2005) [68] [69]

Purpose: To examine all the facets of a young child's development

Age: 1 to 42 months

Format and administration: Bayley-III covers five developmental domains. Cognitive, motor and language are administered with the child; interaction, social-emotional and adaptive behaviour are administered with parent questionnaires. The test is suited to administration in multidisciplinary teams of professionals. Domain subtests can be administered individually. For the cognitive, language and motor scales, items are administered in such a way as to establish basal and ceiling levels of performance.

Time required: 90 minutes to administer the test to children aged 13 months and over

Scoring: Scoring of Bayley-III has been simplified from previous versions. Scoring for every item is either 1 (credit) or 0 (no credit). Scores available include raw scores, scaled scores, composite scores, percentile ranks and confidence intervals. Normative scores are available.

Training and materials: Although the publishers of the test describe Bayley-III as being easy to use, they also say the users of the test are likely to have at least a Master's degree. In order to administer the test, analyse and interpret results, qualified personnel are likely to need formal training in the use of assessment tools, mental health and /or educational training specific to working with parents and assessing young children and training in infant

and child development. Training materials are available and include manuals and DVD resources. The administration manual provides clear guidelines. Scoring software is available. The stimulus materials are attractive to infants and toddlers.

Standardisation and psychometrics: The standardisation sample for the cognitive, language and motor scales, was based on the 2000 US census, and included 1700 children between the ages of one to 42 months, broken down into 17 separate age groups with 100 children in each group. The standardisation sample for the social-emotional scale was based on 456 children and the adaptive behaviour scale was based on 1,350 children. The Psychometrics Centre at the University of Cambridge has carried out work to establish the validity of the norms for use in the UK[70]. A Bayley-III UK and Ireland supplement report gives the results of a UK validation study based on 221 children 12 to 24 months[71].

Use: The Bayley scales are described as being the most widely used developmental assessment scheme [72]. Bayley-III and previous versions have been used with children with a wide range of clinical conditions including prematurity such as the EPICure study, small for gestational age, Downs syndrome, pervasive developmental disorder, asphyxia, cerebral palsy and language impairment. The Bayley scales have been used in many different countries and tend to be used as a standard against which other tests are compared. However, Bayley-III has not been as widely adopted as previous versions [31] and its scores tend to be higher than previous versions. Anderson *et al* found that Bayley-III underestimates developmental delay in 2 year old Australian children [73] and this finding has been more recently confirmed in a UK study [72].

Strengths of Bayley-III as a population measure

1. Bayley-III covers the developmental domains of interest
2. The measure has been validated for use in the UK.
3. Parents are involved in the assessment although to a lesser extent than other measures.
4. The materials used are attractive for children.
5. Bayley-III has been used with children with a wide range of clinical conditions and developmental disorders.
6. It is often used as a gold standard against which other tests are compared.

Limitations and further questions about Bayley-III as a population measure

1. High levels of training are required to administer the measure and to analyse and interpret results.
2. Since the test takes 90 minutes to administer, this is a considerable limitation when assessing children who may have difficulties with attention or other developmental problems as well as many normal two year olds. It is also a limitation for a population measure in terms of professionals' time.
3. Studies have reported that Bayley-III underestimates developmental delay and thus more work is required on its scoring format.
4. Costly for both training and for the materials required for the assessment.
5. Its widest use is in a research setting or for follow up of children at high risk of developmental problems.
6. Rather than being used as a population measure it is more frequently used as a gold standard against which to validate other tests.

3. Measures completed by professionals

Mullen Scales of Early Learning (MSEL) (1995) [74] [75] [76]

Age: For use with children from birth to 68 months.

Format and administration: The assessment is based on the child's responses to activities prepared by the examiner. The test provides complete information on a child's cognitive and motor ability through the use of five scales: Gross Motor and four "cognitive scales": Visual Reception, Fine Motor, Expressive Language, and Receptive Language

Time required: It takes around 30 minutes to administer the test to a three year old child.

Scoring: Instructions for scoring are in the administration book and scoring is done on a record form. Each scale produces a raw score which can be compared against age equivalents and the "cognitive" scores can be summarised into an Early Learning Composite (ELC) score. The scores can be used to obtain the child's percentile rank and age equivalent score.

Training and materials: The test is administered and scored by "highly trained" professionals with training or practical experience in the clinical assessment of infants and young children. Items can either be scored by hand or computerised software is available.

Standardisation and psychometrics: The normative sample was based on a US sample of 1849 children, which excluded those with known physical or mental disabilities. Data were collected between 1981 and 1989. A review of cognitive tests by Bradley – Johnson [77] found limited evidence for the concurrent, content and construct validity of the test.

Use: According to Bishop et al[78] the MSEL is now commonly used as a measure of cognitive and / or language skills in research protocols and is less commonly mentioned in the general child assessment literature. It is used in research, clinical evaluations and longitudinal investigation of children with autistic spectrum disorders.

Strengths of Mullen Scales of Early Learning as a population measure

1. Covers the developmental domains of interest.
2. Relatively easy to score.

Limitations and further questions about Mullen Scales of Early Learning as a population measure

1. Professionals need to be highly trained and be experienced in assessing young children.
2. No evidence found for use as a population measure.
3. The normative data, standardised 23-30 years ago, are now out of date in comparison with other tests.
4. The normative sample on which standardisation was based excluded children with known disabilities, it would be important to include these children if the measure is to be used for population monitoring.
5. There are no UK norms for this measure.
6. Used in research protocols rather than in general child developmental assessment.
7. No evidence regarding its acceptability by parents.
8. Parents are not involved in the assessment process.
9. Takes 30 minutes to administer.
10. It is relatively costly.

Battelle Developmental Inventory, Second Edition (BDI-2) (2004) [79, 80]

Age: The BDI-2 is a comprehensive test used by professionals to assess the development of children from birth to seven years and eleven months of age.

Format and administration: It was primarily designed for use by preschool, kindergarten, and primary school teachers and covers personal-social, adaptive, motor, communication and cognitive domains. There are five item test books which provide specific instructions for the examiner. Many items give the examiner a choice in how to administer each component: direct assessment (using toys, games and tasks), observation (ideally over a few weeks) or via parent report.

Time required: It is comprehensive with 450 items and so typically takes 1 ½ hours to administer.

Scoring: Scoring of the BDI-2 is considered straightforward. The examiner generally scores 2, 1, or 0 on each of the items. If the child is able to demonstrate each skill on a regular basis, they are given a score of 2; if the skill is emerging, the child is given a score of 1 and if the child is unable to demonstrate the skill, they are given a score of 0. At the end of each subdomain, the examiner totals the scores the child received. Norm referenced scores are provided at the subdomain level. The subdomain scores combine to form the five domain scores (representing the child's overall abilities in each of these areas) and the overall Developmental Quotient, which is a summary of the child's general level of development. Percentiles and confidence intervals are also provided for domain scores.

Training and materials: Materials used are child friendly. The test can be administered by a team of professionals or an individual. These may include early childhood teachers, early interventionists, psychologists, physiotherapists, speech and language therapists and other health professionals. Training is required for administration of the test. An interpretation guide is included in the Examiners Manual. A web based scoring software programme is available. This also allows reports to be generated in a range of formats.

Standardisation and psychometrics: Normative data were gathered from 2,500 children (closely resembling the 2000 US census) between the ages of birth to 7 years 11 months. Reliability data are strong, and validity data indicate moderate correlations with other established tests[80].

Use: The Examiner's manual says that caution should be exercised when interpreting the results if the child is not familiar with the culture of the United States or with specific regional cultures or if the child's first language is not English. The BDI-2 has been used among children with autism, developmental delays, motor delays, speech and language delays and prematurity. Administration adaptations for children with disabilities are

provided but implications for scoring are unclear. These are major limitations if it were to be considered for a UK population.

Strengths of Battelle Developmental Inventory, Second Edition as a population measure

1. It covers all the developmental domains of interest.
2. Adaptions are available for children with disabilities.

Limitations and further questions about of Battelle Developmental Inventory, Second Edition as a population measure

1. Training is required in order to administer the measure.
2. The measures takes 1 ½ hours to administer which is too long for children of this age and may be a particular issue among children with attention difficulties.
3. The measure is not standardised for use in the UK; caution is required if it is used in non-US populations.
4. It does not involve parents in the process.
5. Evidence is lacking about acceptability by parents.
6. It is relatively costly.

The BDI-2 Screening Test (2004) [81]

See information on BDI-2. Presented below are the main features and changes with the screening test.

Format and administration: Like the BDI (from which the items were extracted), it has subtests for fine and gross motor, adaptive, personal-social, receptive and expressive language, and cognitive skills. It is administered in the same manner as the full BDI-2.

Time required: The BDI-2 consists of 96 items and can be administered in 10 to 30 minutes depending on the age of the child.

Scoring: Similar to the BDI-2 but cut-off scores are provided to aid in identification of children who may need additional follow-up.

Standardisation and psychometrics: Papers suggest there are some concerns over the psychometric properties of the BDI-2 screening test in that it was not part of the standardised sample, so its reliability and validity data were extracted from data for the full test [80].

Use: The BDI-2 Screening test was designed as a method for determining whether a child needs further and more in-depth evaluation.

Strengths of BDI-2 Screening Test as a population measure

1. It covers the developmental domains of interest

Limitations and further questions about BDI-2 Screening Test as a population measure

1. Training is required in order to administer the measure.

2. The measure itself has not been standardised and relies on the reliability and validity data of the full BDI test.

3. Evidence is lacking about acceptability by parents.

Brigance Early Childhood Screens - BRIGANCE Early Childhood II 0–35 Months Screening Kit [82]

Age: 0 to 35 months

Format and administration: The Brigance Early Childhood-II 0 – 35 months Screening Kit is a reconfiguration of the earlier Infant and Toddler Screen-II (for birth to 23 months) and the Early Preschool Screen-II (for 2 year olds and 2 ½ year olds) and so has four separate data sheets for assessment of infants, toddlers, 2 year olds and 2 ½ year olds. The screens, widely used in educational settings, enable assessment of language, motor, self-help, social-emotional and cognitive skills. Flexible administration allows for assessment through parent/caregiver interview, child performance, and observation of a child in the natural setting.

Time required: 10 to 15 minutes per child.

Scoring: In addition to identifying potential learning delays the screens also identify children who may have academic giftedness. Age appropriate cut-offs allow children to be grouped: (1) Children who should be evaluated for special education services due to a high probability of developmental delays or difficulties (2) Children who should be evaluated to determine whether they are gifted or academically talented and (3) Children who are performing adequately for their age or grade placement.

Training and materials: Use of the screens requires no specialised training; the examiners' pages of the screen provide clear instructions on how to administer the test and score

responses. Free electronic training modules are available. Additional resources are available for parents and teachers such as take home activity books.

Standardisation and psychometrics: Sensitivity 82%, specificity 84%. Also identifies 86% of children over the age of 2 years with potential academic giftedness. Validity was determined through comparison with a battery of age-appropriate developmental assessment tools such as the Bayley Scales of Infant Development–II (BSID-II) [83]. The original study in which the extension of the Brigance screens (original version) was extended to children ages 0 to 2 years, found the screen maintained its sensitivity (76% to 77%) and specificity (85% to 86%) [84].

Use: The screens have been used widely in educational settings in the United States but are less commonly used by health professionals [83] [85].

Strengths of the Brigance Early Childhood Screens as population measure

1. It covers the developmental domains of interest.
2. It is relatively quick and easy to use.
3. It has acceptable psychometric properties
4. Flexibility of administration is possible.

Limitations and further questions about Brigance Early Childhood Screens as population measure

1. It appears to be more focussed on academic performance.
2. It has been used mostly in educational rather than health settings.
3. It has not been standardised for use in the UK.
4. There is no evidence of its use as a population measure.

Denver II (1990) [86]

Age: Denver II is a screening tool for children from birth to six years to detect developmental delay.

Format and administration: It was devised to give a brief overview of the child's development to identify those who are not performing as well as other children of the same age. It includes personal-social, fine motor adaptive, language and gross motor items.

Denver II can be conducted by Professionals or para-professionals. The test utilises both parent observation and direct observation.

Time required: It takes around 20 to 30 minutes to complete the 125 items. The total number of items administered will vary with the child's age and ability.

Scoring: The child's responses are recorded as Pass or Fail on the score sheets. These are then examined to see if they fall into or outside the normal expected range of success on that item for the child's age. The child is either classified as normal range, suspect, or delayed. Results are presented like a growth curve, with a display of norms over time.

Training and materials: In order to administer and interpret the Denver II, training (e.g. two days) from a master instructor is recommended. Training manuals and DVDs are available. "Anyone who works well with children and meticulously follows directions for administration can be a screener."

Standardisation and psychometrics: This test differs in that its authors have made no attempt to measure the validity of the tool in the conventional way to estimate its sensitivity and specificity. Instead, they present norms based on representative population data (based on 1980 US census)[87]. Others have criticised this approach and question whether it is appropriate to generalise to different and more heterogeneous populations. Studies have demonstrated that Denver II has good sensitivity but an unacceptably low specificity: 43% in one study [88] and 26% in another[83].

Use: The Denver II is widely used especially in clinical settings and as the gold standard against which other measures are compared.

Strengths of Denver II as a population measure

1. It covers the developmental domains of interest.
2. It is well known and widely used.
3. Denver II is reported to have good sensitivity.

Limitations and further questions about Denver II as a population measure

1. Not only is training required, but the measure requires meticulous administration making it less appropriate for use by skill mix teams.
2. It has poor specificity.

3. The measure was standardised in 1980 making the norms outdated, and in a US population which may be not applicable to a contemporary UK population.
4. There is no evidence for the use of Denver II as a population measure.

Griffiths Mental Development Scales-Extended revised (GMDS-ER) (2006) [89]

Purpose: To measure the rate of development of young children

Age: Two to eight years

Format and Administration: The Griffiths scales were originally developed in the 1960s and designed to measure children aged 0-2 years. These were later extended to cover birth to 8 years. The third most current edition was published in 2006. A kit of standard equipment is needed to administer the Griffiths scales. This consists of 39 pieces of equipment such as building blocks; a drawing book and record form are also supplied. Griffiths scales comprise six sub scales: Locomotor, personal social, language (receptive and expressive), hand and eye coordination, performance, practical reasoning. The latter is only used in older children.

Time required: 50-60 minutes.

Scoring: Individual items are scored and written into a record book. The items are colour coded to identify which items are similar. Raw scores are computed for each individual sub-scale and can be converted to four types of standard scores: percentiles, z scores, age equivalents, general quotient.

Training and materials: The scales are only supplied to paediatricians and health professionals who have successfully completed a five day intensive training course accredited by the Association for Research in Infant and Child Development (ARICD).

Standardisation and psychometrics:

The measure was normed on a national representative sample of children in UK between 2-8 years of age. This sample was stratified according to geographic region and proportionate to the population ratios obtained in 1997 by ONS for children of the same age. Coefficients were calculated for each of the sub scales using all the items in the scales. The publishers state that with the exception of Scale E (performance) in children with chronological age less than 48 months the coefficients *'all comfortably exceed the minimum acceptable value of 0.70'*

Use: The Scales are widely used for both clinical and research purposes. Clinical use of the Scales is restricted to psychologists and developmental paediatricians. Training courses are

organised and run throughout the UK and in many overseas countries by the Association for Research in Infant and Child Development (ARICD).

Strengths of GMDS-ER as a population measure

1. It covers the developmental domains of interest, although personal social rather than social-emotional.
2. It is widely used.
3. It appears to have acceptable sensitivity and specificity.
4. It has recently been standardised on a representative UK population.

Limitations and further questions about Griffiths as a population measure

1. Intensive training is required. The ARICD website states that access to the scales for clinical use is restricted to developmental paediatricians and psychologists.
2. No evidence for use as a population measure.
3. No evidence about acceptability by parents.
4. Lengthy to administer.
5. Little published evidence on validity.

Schedule of Growing Skills –II (SGS-II) 1996 [90]

Purpose: To establish children's developmental level

Age: For use in children 0-5 years.

Format and administration: Originally developed to be used in the British National Childhood Encephalopathy Study (NCES) (1976-1979) investigating the cause and outcome of serious neurological illness in young children [91]. The particular focus of this study was the potential role of immunisation in the aetiology of neurological illness. When no suitable measure could be found for the study, SGS was developed based on Mary Sheridan's STYCAR sequences, and originally designed for use in children aged 2-36 months. It examines nine key areas, passive posture, active posture, locomotor, manipulative, visual, hearing and language, speech and language, interactive social and self-care social.

Scoring: The score for the highest item for each subscale is transferred to the SGS II profile form. The child's chronological age is added to this form and, if the child performs within one age band of their chronological age, they are considered to be developing normally. If

their performance is two or more age bands below their chronological age, they are considered to require further assessment.

Training and materials: A short training course, of a half or full day, is required.

Standardisation and psychometrics: Validity of the original tool was assessed by comparison with the Griffiths test with the NCES tool showing significant correlations: sensitivity levels ranged from .44 to .82 and specificity from .94 to 1.0. Subsequently modifications were made to the test, including renaming it the Schedule of Growing Skills (SGS), and as validity and reliability had been assessed for the 2-36 months age range, further validation was only performed for the three to five year age range. This was again compared with Griffiths but no estimates of sensitivity and specificity were made. In 1996 the measure was revised (SGS-II) and standardised among 348 children for use in the UK. It was also compared with Denver. Details of this standardisation are only available in the reference manual and have not been published in a peer review journal. The publishers of SGS-II state that the technical manual contains 14 case studies highlighting the concurrent validity of SGS-II, and 9 case studies examining construct validity, but these could not be located in peer review journals. The publishers of SGS-II report that a new edition is being developed to fit in specifically with the HCP, the estimated publication date is summer 2014.

One of the potential limitations identified with SGS-II, is that the breadth of age band widens with age, such that at 18 months the developmental windows are 6 months wide and by 36 months, 12 months wide. This could result in difficulties assessing children over time and in making comparisons between children. Furthermore, SGS-II was assessed against Denver which is known not to have robust sensitivity and specificity. In recognition of this, pilot work has been conducted by Williams *et al* at Bangor University to develop a new scoring method. The authors reported that the new scoring method demonstrated better criterion – related validity with higher sensitivity [92]. There are no data available on acceptability by parents or professionals other than anecdotal reports by the publishers that both groups like the measure.

Use: The UK suppliers of SGS report that it is widely used both in the UK and internationally. It is being used in the Welsh Flying Start programme to assess children as they enter the programme at 2 years and leave it at 3 years. It is used in some parts of England in the HCP as a second tier assessment for children at two years of age.

Strengths of SGS as a population measure

1. It covers the developmental domains of interest.
2. It is reported by the publishers to be widely used.
3. The original estimates for specificity are good.
4. Completion time is relatively short.

Limitations and further questions about SGS as a population measure

1. The original estimates of sensitivity range from poor to good, depending on the domain being assessed. There are no recent estimates of the sensitivity and specificity of this measure.
2. The original validation and reliability assessments of SGS were carried out over 30 years ago and are now outdated.
3. Although SGS is being used in the Flying Start programme, there is no published evidence of its use in this way.
4. SGS does not actively involve parents in the assessment.
5. There is no information available on acceptability by parents or professionals.
6. Although reported to be widely used, little information about SGS-II could be located in peer reviewed journals; we did not identify this measure from the initial search of papers.

Assessment against DH requirements for a population measure of children's development

Table 5 (appendix IV) shows an assessment of the two measures (Ages and Stages and PEDS) which, on the basis of our detailed descriptions, are the most suitable measures to be incorporated into the HCP two year review as a population measure of children's development. This assessment was made against Department of Health's stated requirements for such a measure. They are completed by parents, which not only saves professionals' time but also reflects the ethos of partnership with parents. Although the purpose of this measure is primarily to inform a population measure of children's development, these two measures were originally developed as a means of assessing individual children's development and their characteristics reflect this.

Other instruments not meeting inclusion criteria

Although not meeting our inclusion criteria, we have included information about the Early Development Instrument (EDI) in our report. The EDI focusses on children in their first school year but it is of interest because it was specifically developed as a population measure of children's development. Work is currently on-going in Canada to develop a similar measure for use among 18 month old children.

The Early Development Instrument (EDI) [93]

The EDI originally developed in Canada is a population-based measure of children's development at age 4-5 years. It is a teacher-completed checklist of 104 items, completed half way through the first kindergarten / school year. Data are aggregated at a group level such as school, neighbourhood, region or country. It is not reported at an individual level or class level and is not used as a diagnostic tool for individual children. It provides assessment over five developmental domains: physical health and wellbeing; social competence; emotional maturity; language and cognitive development; communication skills and general knowledge. Each child's EDI takes about 20 to 30 minutes to complete and is geographically coded according to home address and then presented using Geographic Information System technologies. Maps show the percentage of children vulnerable in each developmental domain by geographical region. This allows stakeholders to identify areas of greatest need, make comparisons with socio-economic indicators to understand reasons for observed patterns, identify gaps in services, and over time to observe the effects of interventions and changes in policy.

The instrument has been adapted for use in Australia (AEDI)[94] where documented outcomes included: increased community awareness of the importance of early childhood development; increased collaborative working between stakeholders; better informed planning processes; and strengthened grant applications for funding [95].

In British Colombia, the EDI data has been linked at an individual level with the Ministry of Education's Foundation Skills Assessment (FSA), a Grade 4 measure of numeracy, reading comprehension, and writing skills to provide a Community Index of Child Development

(CICD) for each geographic area. This provides a means of summarising children's longitudinal development which in turn allows additional uses of the data [96, 97].

Work is being conducted in Scotland to develop the EDI for use among primary school aged children (see page 56) and a pan-Canadian group are developing a population health measurement tool for use at 18 months of age.

Other current relevant research and considerations

- In 2010, the Scottish Chief Medical Officer proposed the re-introduction of a universal 24-30 month child health assessment focussing on child development, parenting and health promotion: the 'Ready to Learn' contact. The 30 month visit involves the use of the Strengths and Difficulties Questionnaire (SDQ) and the Law-Miniscalco two-item language screen and the Sure Start Language Measures. Results of the pilot evaluation in Glasgow show high uptake (90%). A significant group of children with previously unsuspected developmental problems were identified. More detailed results are awaited.
- The National Screening Committee has commissioned an update of a review of screening for speech and language delay. At the time of writing a consultation document is to be published imminently.
- Results are awaited from the RCT of the Family Nurse Partnership programme in which children's outcomes are being assessed at two years. It is unclear how this is being assessed but this study should provide much valuable information on method of assessment as well as its acceptability for parents and professionals.
- The Child Health Sub-Group of the National Screening Committee reviewed the evidence on screening for autism in young children in and decided that the introduction of screening could not be recommended to the UK NSC. This policy will be reviewed again in 2015/16.
- A Systematic Review has been completed exploring risk factors for emotional abuse and neglect in the preschool child, particularly aspects of child/carer interaction and for tools with which to measure parent/child interaction. A paper has been submitted for publication. This information could also inform an appropriate measure to be incorporated into the HCP (Cardiff University).
- National Institute for Health and Clinical Excellence, Public Health Intervention Advisory Committee issued guidance on supporting the social and emotional wellbeing of vulnerable children aged under 5 years through home visiting, childcare and early education. Professor Adrian Angold (Duke University) has been conducting work to validate the Strengths and Difficulties Questionnaire (SDQ) for use in 2 year old children. This work has not yet been published and information about the validity of the measure

in this age group is not yet readily accessible. This measure is being used in the 'Ready to Learn' contact in Scotland about which detailed results are awaited.

- In collaboration with East Lothian Local Authority and McMaster University, Canada, (lead, Rosemary Geddes) and the Head of the School of Psychological Sciences and Health at Strathclyde University, Professor Lisa Woolfson has piloted a project to test the feasibility in Scotland of using the Early Development Instrument (EDI). This would be completed triennially for all children in Primary 1 (P1), four months after school entry, to be a population-level tool to monitor the global developmental status and "school readiness" of each birth cohort aged 5 years. Phase 1 of the project (2011) showed that the majority of teachers found the Canadian-designed EDI to be acceptable and feasible, and only minor adjustments to terminology were required to adapt the EDI for the Scottish context. During phase 2 (starting in January 2012), all 1200 P1 children in East Lothian were assessed by their 70 P1 teachers. Although not within the scope of this review, it has been suggested that a measure of parent/carer-child interaction should be included in the HCP 2 year review. Poor parent/carer-child interaction is a predictor for poor child developmental outcomes. A review of the evidence around identification of features in the child and in the parent/carer-child interaction has recently been conducted and has been submitted for publication (personal communication with Aideen Naughton, Cardiff University).

Summary of Findings

- The review of papers identified 32 measures of various aspects of child development which met our inclusion criteria.
- A further 3 measures were identified through other means.
- Of the 35 measures, 13 covered all the developmental domains of interest.
- These 13 measures included those completed by parents (n=3); measures completed by professionals with parental input (n=3); and measures completed by professionals based on direct observation of children's skills (n=7).
- The 13 measures which met our criteria were described in detail to elicit further information about their characteristics, format of administration, time taken to administer, as well as estimates of sensitivity and specificity, information about acceptability and use as a population measure.
- Two measures (ASQ-3 and PEDS) emerged as the most suitable to be included in the 2 HCP 2 year review as a population measure of children's development.
- Although these measures are currently in use in the HCP 2 year review, there is a lack of formal evaluation of their validity and acceptability in the UK context.

Discussion

In this review, we identified tools to measure children's development from a variety of sources: via a search of the literature, Internet searches, contact with publishers of tools, other papers and word of mouth. We identified 13 measures which cover all the developmental domains of interest for use with children aged 2-2½ years.

Although we could identify no single measure developed specifically for use as a population measure of all domains of children's development at 2 years of age, a number originally developed to screen for development delay have been used at a population level (ASQ and PEDS).

There is a range of single domain measures available, but administration of the outcome measure would be complicated if it comprised a combination of a number of individual measures. In addition, as the possible combinations of single domain measures are numerous, evaluating the most appropriate would be complex. Since acceptability and ease of administration are key considerations in the choice of a suitable measure, it is preferable to use a single measure in which all the developmental domains are assessed.

Above all other considerations, there is an ethical imperative to ensure that the measure selected is appropriate for the intended purpose and is both well validated and reliable. It is unethical to use home grown or poorly validated measures which potentially do more harm than good by either wrongly labelling children who have no developmental problems or, in missing children who do, denying them the necessary support and interventions. This is not only because of the possible impact on families and children, but also the resulting inappropriate use of services at great cost to the NHS.

The identified measures include well established and widely used measures for assessing children, administered by professionals. In some cases these require a high level of training. This immediately reduces their usefulness as a measure that can be easily incorporated into the HCP 2 year review; the review is often conducted by members of the health visiting team, who have varying levels of skills, the most suitable measure therefore will of necessity

not require a high level of skill. Furthermore, many of the identified measures are lengthy, requiring considerable time to administer. To reduce the burden on parents, children and health professionals and to maximise uptake of the review, the ideal measure must be quick and easy to administer.

Although the purpose is to have a tool to measure children's development at 2-2½ years at a population level, we consider that this would be difficult to implement without the full cooperation of parents. They are more likely to accept a measure which also provides an opportunity for their child's development to be assessed and involves them in the process.

Few of the measures have been standardised in a UK population, an important requirement for a population measure since population norms are needed against which to monitor change. However, gathering these data is a relatively straightforward process and so does not pose a serious limitation to the use of a measure which in other respects is suitable.

Measures completed by parents have been shown to be as accurate as those administered by professionals in identifying children with developmental problems [63, 98]. They have the additional advantage of involving parents wholly in the process, using them as experts in their child's development. If the parents have had the opportunity to complete the measure in advance of the review, it may allow the review as a whole to be more focussed and ultimately more useful for the parent and the child. Given the difficulties of assessing such young children accurately, a measure that the parent completes is also more likely to provide a better overall assessment than a one-off test which may be subject to too many external factors such as how the child feels on the day of assessment. This is also in keeping with the ethos of working in partnership with parents; they can best describe what their child can do, even if they cannot always interpret what it means.

Bearing these issues in mind, the two measures that fulfil most of these criteria are ASQ and PEDS. However, they are not without some limitations. Although reported to be widely used in the UK, and to be well liked, such reports are largely anecdotal and there is a lack of well conducted evaluation studies and of UK norms. Most use of these two measures has been in the USA, where the surveillance and screening programme offers many more well-child

contacts along with anticipatory guidance. It is not clear whether UK parents with fewer well-child contacts, are as well-equipped as their US counterparts to assess their child's development, and in particular to recognise whether or not they should have concerns which PEDS requires. Despite evidence of the validity of PEDS in identifying children with developmental delay, there is some evidence to suggest that ASQ, which requires an objective assessment of a child's skills, may be superior [18]. Furthermore, ASQ-3 includes a domain focussing on personal-social rather than social-emotional development. Social-emotional development is covered in questionnaires specifically developed for this purpose (ASQ-SE)[48] . To use both questionnaires as a population measure would complicate the process, making it more time consuming, and it is not clear whether using ASQ-SE in addition to ASQ-3 would identify children as having social-emotional problems that would be missed using ASQ-3 alone. This requires further investigation. Other questions regarding ASQ-3 and PEDS also need to be addressed. They include the validity of using ASQ as a one-off measure, the effect of combining scores from different ASQ age specific questionnaires into one overall score; whether PEDS scores can be used to detect small changes at a population level and finally the acceptability of both measures to UK parents.

Recommendations

ASQ and PEDS best satisfy the requirements for a population measure of children's development but both measures require proper evaluation in a representative UK population.

It is suggested that both PEDS and ASQ are tested on different cohorts of children to assess their reliability and acceptability.

A subset of each should also be assessed using an appropriate gold standard test to establish the validity of the measures in a representative UK population.

NB. Since making these recommendations it has come to light that no district in England is currently using PEDS as part of the two year review. The only district that was, has discontinued its use in favour of ASQ.

Appendix I

Website findings

Bright Futures (American Academy of Pediatrics)
http://brightfutures.aap.org/goals.html

The Longitudinal Study of Australian Children Outcome Indices, Waves 2 and 3
http://www.aifs.gov.au/institute/pubs/rp50/rp50c.html

Royal Children's Hospital Melbourne
Parents' Evaluation of Developmental Status (PEDS)
Authorised Australian Version of Parents' Evaluation of Developmental Status (PEDS).
http://www.rch.org.au/ccch/resources.cfm?doc_id=10963

First five association of California
http://www.first5ecmh.org/

Canadian Pediatric Society
http://www.cps.ca/

CHADIS
CHADIS is online system that delivers questionnaires that help Pediatricians review the health and development of children.
http://www.chadis.com/families/about_chadis.html

Canada Offord Centre for Child Studies, McMaster University
http://www.offordcentre.com/

Child Trends
US based Independent research and policy center focused exclusively on improving outcomes for children
http://www.childtrends.org/_listAllPubs.cfm?LID=73039143-C617-411A-A4A1D55FAEC978CF

National Children's Bureau
http://www.ncb.org.uk/

Children's Hospital Boston
Developmental screening Toolkit for Primary Care providers
http://www.developmentalscreening.org/about.htm

Center on the Developing Child, Harvard University
http://developingchild.harvard.edu/

First Signs
https://www.firstsigns.org/

Minnesota Department of Health
http://www.health.state.mn.us/divs/fh/mch/devscrn/index.html

Pennsylvania's Departments of Education and Public Welfare
http://www.pakeys.org/docs/EarlyChildhoodAssessment.pdf

The National Academies (USA)
http://www.bocyf.org/head_start_brief.pdf

Washington State
A Guide to Assessment in Early Childhood
http://www.k12.wa.us/EarlyLearning/pubdocs/assessment_print.pdf

Appendix II

Table 3: Complete search by database and number of records imported

Database	Search terms	Imported to Endnote
PubMed	**social emotional (5805)** ("Data Collection"[Mesh] OR "Outcome and Process Assessment (Health Care)"[Mesh] OR "Diagnostic Techniques and Procedures"[Mesh]) *AND ("humans"[MeSH Terms] AND English[lang] AND ("infant"[MeSH Terms:noexp] OR "child, preschool"[MeSH Terms]) AND ("1980"[PDAT] : "2011/10/31"[PDAT]))* **AND ("Psychological Tests"[Mesh] OR "Psychiatric Status Rating Scales"[Mesh] OR "Emotional Intelligence"[Mesh] OR "Social Behavior"[Mesh] OR "Social Behavior Disorders"[Mesh])** **physical/motor (2665)** ("Data Collection"[Mesh] OR "Outcome and Process Assessment (Health Care)"[Mesh] OR "Diagnostic Techniques and Procedures"[Mesh]) *AND ("humans"[MeSH Terms] AND English[lang] AND ("infant"[MeSH Terms:noexp] OR "child, preschool"[MeSH Terms]) AND ("1980"[PDAT] : "2011/10/31"[PDAT]))* **AND ("Motor Activity" OR "Motor Skills" OR "Motor Skills Disorders" OR "psychomotor performance")**	9647

	speech and language (3638) ("Data Collection"[Mesh] OR "Outcome and Process Assessment (Health Care)"[Mesh] OR "Diagnostic Techniques and Procedures"[Mesh]) *AND ("humans"[MeSH Terms] AND English[lang] AND ("infant"[MeSH Terms:noexp] OR "child, preschool"[MeSH Terms]) AND ("1980"[PDAT] : "2011/10/31"[PDAT]))* **AND ("Language Tests"[Mesh] OR "Language Disorders"[Mesh] OR "Language Development"[Mesh] OR "speech"[Mesh])** **cognition (2921)** ("Data Collection"[Mesh] OR "Outcome and Process Assessment (Health Care)"[Mesh] OR "Diagnostic Techniques and Procedures"[Mesh]) *AND ("humans"[MeSH Terms] AND English[lang] AND ("infant"[MeSH Terms:noexp] OR "child, preschool"[MeSH Terms]) AND ("1980"[PDAT] : "2011/10/31"[PDAT]))* **AND ("Cognition"[Mesh] OR "Cognition Disorders"[Mesh] OR "Learning Disorders"[Mesh])**	
Web of Knowledge	#4 OR #3 OR #2 OR #1 Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=1980-2011 Lemmatization=On #4 ((((TS=(early childhood) OR TS=(preschool)) AND (TS=(assessment) OR TS=(measurement) OR TS=(tool) OR TS=(scale) OR TS=(questionnaire) OR TS=(survey)) AND (TS=(psycholog*) OR TS=(emotion*) OR TS=(social	5138

	behaviour))))) AND Language=(English) Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=1980-2011 Lemmatization=On #3 ((((TS=(early childhood) OR TS=(preschool)) AND (TS=(assessment) OR TS=(measurement) OR TS=(tool) OR TS=(scale) OR TS=(questionnaire) OR TS=(survey)) AND (TS=(motor activity) OR TS=(motor skill) OR TS=(psychomotor performance) OR TS=(physical development) OR TS=(motor development))))) AND Language=(English) Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=1980-2011 Lemmatization=On #2 (((TS=(early childhood) OR TS=(preschool)) AND (TS=(assessment) OR TS=(measurement) OR TS=(tool) OR TS=(scale) OR TS=(questionnaire) OR TS=(survey)) AND (TS=(language test) OR TS=(language development) OR TS=(language disorder) OR TS=(speech) OR TS=(linguistic)))) AND Language=(English) Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=1980-2011 Lemmatization=On #1 ((((TS=(early childhood) OR TS=(preschool)) AND (TS=(assessment) OR TS=(measurement) OR TS=(tool) OR TS=(scale) OR TS=(questionnaire) OR TS=(survey)) AND (TS=(cognition) OR TS=(cogni* disorder) OR TS=(learning) OR TS=(cogni* development) OR TS=(learning disorder))))) AND Language=(English)	

	Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=1980-2011 Lemmatization=On	
Embase	(human and english language and yr="1980 -Current" and preschool child <1 to 6 years>) AND ("diagnosis, measurement and analysis" or checklist or clinical assessment tool or rating scale or questionnaire or data collection method or clinical assessment or clinical assessment tool or functional assessment or needs assessment or outcome assessment or "named inventories, questionnaires and rating scales") AND ((cognition or cognitive development or learning disorder) OR (psychologic assessment or psychologic test or social behavior or emotional intelligence or emotional disorder or emotional stability) OR ("speech and language assessment" or language development or language test or language disability or speech development or language ability or "speech and language") OR (motor activity or physical development or motor performance or physical performance))	941
PsychInfo	(human and english language and yr="1980 -Current" and preschool child <1 to 6 years>) AND ("diagnosis, measurement and analysis" or checklist or clinical assessment tool or rating scale or questionnaire or data collection method or clinical assessment or clinical assessment tool or functional assessment or needs	8

	assessment or outcome assessment or "named inventories, questionnaires and rating scales") AND ((cognition or cognitive development or learning disorder) OR (psychologic assessment or psychologic test or social behavior or emotional intelligence or emotional disorder or emotional stability) OR ("speech and language assessment" or language development or language test or language disability or speech development or language ability or "speech and language") OR (motor activity or physical development or motor performance or physical performance))	
ERIC	Limit 19800101 Titles and abstracts (preschool OR child OR infant) AND (checklist OR "clinical assessment tool" OR "rating scale" OR questionnaire OR "data collection method" OR "clinical assessment" OR "clinical assessment tool" OR "functional assessment" OR "needs assessment" OR "outcome assessment" OR inventory OR "rating scales") AND (cognition OR cognitive development OR learning disorder) OR (speech assessment OR language assessment OR language development OR language test OR language disability OR speech development OR language ability) OR (motor activity OR physical development OR motor performance OR physical performance) OR (social behavior OR emotional intelligence OR emotional disorder OR emotional stability OR psychologic assessment OR psychologic test)	2483

NB: The results from the database searches were imported into Endnote in the order they are listed, thus the number of records imported excludes duplicate results.

Appendix III

Table 4: Main characteristics of measures identified in review

Measure	Acronym	Domain(s)	Age	Administration / Versions	Length	Completion time (minutes)	Scales and subscales / Areas screened	Reliability & Validity	Standardized	Scoring [75, 99] [100]	Costs (examples of costs - March 2012)
GENERAL DEVELOPMENT											
Ages and Stages Questionnaire [101]	ASQ-3	General development	1 to 66 months (5 ½ years)	Parental questionnaire (21 questionnaires dependant on child's age)	~30 items	10 to 15 minutes (1 to 3 minutes for professional to score)	Communication, gross motor, fine motor, problem solving, and personal-social	Concurrent validity 86% overall agreement. Sensitivity 86%; Specificity 85%. Validity .82 to .88, test-retest reliability is .91, and inter-rater reliability is .92 [102]	normative samples of more than 18,000	Sum for each developmental area. 2 SD below the mean cut-off score is used	Starter kit with English questionnaires $275 [103] (Once purchased a site can photocopy materials as required) Users guides, training DVDs, online options

Measure	Acronym	Domain(s)	Age	Administration / Versions	Length	Completion time (minutes)	Scales and subscales / Areas screened	Reliability & Validity	Standardized	Scoring [75, 99] [100]	Costs (examples of costs - March 2012)
Battelle Developmental Inventory [81]	BDI-2 or BDI-2 screening test	Personal – social, adaptive, motor, communication, cognitive	Birth to 8 years	Professionals – direct assessment / observation and parental interview	450 test items (BDI)	1 to 2 hours (complete BDI), 10 to 30 minutes (Screening Test)	Personal – social, adaptive, motor, communication, cognitive	BDI-2 total score reliability .98 to .99. Test-retest reliability generally above .80 [80]	over 2,500 children between the ages of birth to 7 years 11 months (USA)	Scores for subdomains, domains and an overall Development Quotient. Norm references scores, percentiles and confidence intervals available.	BDI-2 Complete kit $1825 BDI-2 complete screener kit $445
Bayley Scales of Infant and Toddler Development, Third edition [68] [69]	BSID-III	Adaptive behaviour, cognitive, language, motor, social-emotional	1 to 42 months	Experienced practitioners	91 items for cognitive scale; 49 items receptive communication ; 48 items expressive communication; 66 items fine motor; 72 items	30 to 90 minutes	Three scales administered with child interaction – cognitive, motor, language. Two scales conducted with parent questionnaires – social-emotional, adaptive behaviour.	Scale composite average reli- ability coefficients ranged from .91 to .93. Across all ages, average stability coefficients were .80 or higher.	1,700 children age 1 to 42 months, stratified according to age, based on the 2000 U.S. Census	raw scores, scaled scores (ranging from 1 to 19), composite scores, and percentile ranks	DVD training resources; computerised or manual scoring. Complete kit $978.45 2 days UK training workshops available [104]

Measure	Acronym	Domain(s)	Age	Administration / Versions	Length	Completion time (minutes)	Scales and subscales / Areas screened	Reliability & Validity	Standardized	Scoring [75, 99] [100]	Costs (examples of costs - March 2012)
					gross motor. 35 items social-emotional;						
Bayley Scales of Infant and Toddler Development, Third edition – Screening test [105]	Bayley-III screening test	Cognitive, language and motor	1 to 42 months	Combination of incidental observation and direct administration		15 to 25 minutes					Bayley-III Screening Test Kit $215.35
Brigance Early Childhood Screens [82] [106] (Incorporates Brigance Infant and Toddler Screen-II and Early Preschool Screen-II)		language, motor, self-help, social-emotional and cognitive skills	0 to 35 months	Administered by paraprofessionals		10 to 15 minutes		Sensitivity 82%. Specificity 84%. Identifies 86% of children > 2 years with potential academic giftedness. Internal consistency of 0.84 to 0.99; test-retest	1,366 children from across the United States. Representative of US population.		Kits $299 2 and 2 ½ year data sheets available $59 (for 60) Free e-training modules.

Measure	Acronym	Domain(s)	Age	Administration / Versions	Length	Completion time (minutes)	Scales and subscales / Areas screened	Reliability & Validity	Standardized	Scoring [75, 99] [100]	Costs (examples of costs - March 2012)
								reliability 0.84 to 0.99; inter-rater reliability 0.90 to 0.99; concurrent validity 0.66 to 0.97			
Child Development Inventory [107] [108] [61] (This replaced Minnesota Child Development Inventory - MCDI)	CDI	social, self-help, gross motor, fine motor, expressive language, language comprehension, letters, and numbers	15 months to 6 years	Booklet and answer sheet for parents to complete	300 items "too long for groups of presumably 'normal' children"		Social, Self Help, Gross Motor, Fine Motor, Expressive Language, Language Comprehension, Letters, Numbers, and General Development		568 children aged 1 to 6 years, in Minnesota	a single cut-off tied to 1.5 standard deviations. T-scores may be calculated from this information.	Starter set $85 to $150.
(Capute Scales): Cognitive Adaptive Test / Clinical Linguistic and Auditory Milestone Scale [109]	CAT /CLAMS	CAT: visual-motor functioning; CLAMS: expressive and receptive language	1 to 36 months	(For use in clinical settings)	100 items (58 + 42)	6 to 20 minutes		Interrater reliability 80%	1055 US children of 0 to 3 years (1999-2001)	Raw scores, subscale and total scores. Developmental quotient.	Manual available. Complete system $375

Measure	Acronym	Domain(s)	Age	Administration / Versions	Length	Completion time (minutes)	Scales and subscales / Areas screened	Reliability & Validity	Standardized	Scoring [75, 99] [100]	Costs (examples of costs - March 2012)
[110] [111] [112]		development									
Child Development Review [66] [67] [113] (Developed from Child Development Inventories)	CDR Consists of parent question naire (CDR-PQ) and Child Develop ment Chart	Health, behaviour, development social, self-help, gross motor skills, fine motor skills and language	18 months to kinderg arten Covers first 5 years	Parent completion of open ended questions and problem checklist Professional observation or parent interview	6 question s and 26 item problem checklist	5 minutes to administer (and 5 minutes to score)		Sensitivity .68 Specificity .88	220 predominantly (95%) white children, aged 3 and 4 years, from Minnesota.	“no problem”, “a possible problem” or “possible major problem” "typical" for age in all areas, or as "borderline" or "delayed" in one or more areas of development	Manual $40 $45 for 75 questionnair es / charts
Denver II [86] [87] (Derived from: Denver Developmental Screening Test (DDST)	Denver II	Personal social, Fine motor adaptive, language and gross motor	Birth to 6 years	Professional, para-professionals: direct assessment and parent report	125 items	20 to 30 minutes		The authors have made no attempt to measure sensitivity and specificity of the DENVER II (instead “norms based on representative population”).	Standardised in 1988 /89 on 2096 children from Colorado (based on 1980 US census population).	Approximates a growth curve in its display of norms over time. Overall categories: Normal, Abnormal,	Complete package $140 (100 forms). Online version $49.99 per month. Training materials available

Measure	Acronym	Domain(s)	Age	Administration / Versions	Length	Completion time (minutes)	Scales and subscales / Areas screened	Reliability & Validity	Standardized	Scoring [75, 99] [100]	Costs (examples of costs - March 2012)
								Average inter-rater and test-retest reliabilities were 0.99 and 0.90		Questionable, and Untestable	
Griffiths[89] Mental Development Scales-Extended Revised	GMDS-ER 2-8	Locomotor, Expressive and receptive language, Personal-social, Hand and eye coordination, Performance, Practical reasoning	2-8 years	Professional, direct assessment		50-60 mins		*'With the exception of the scale assessing performance, the co-efficients all comfortably exceed the minimum acceptable value of 0.70'*	Norms are based on a sample of 1026 UK children between 2-8 years of age.	Raw scores computed to each sub scale. Converted into four types of standard score: percentiles, Z-Scores, age equivalents or general Quotient	Griffiths 2-8 years comprehensive starter kit available from HOGREFE (£754). The cost of 5 day training is being ascertained
The Motor and Social Development Scale [114] [115]	MSD	Motor, social, cognitive	Birth to 3 years	(previously used as a component within larger health surveys)	15 out of 48 items dependant on child's age			"The MSD tends to 'top out' for three-year-olds and does not provide a sensitive ceiling for	On children from USA participating in 1981 National Health Interview Survey		

Measure	Acronym	Domain(s)	Age	Administration / Versions	Length	Completion time (minutes)	Scales and subscales / Areas screened	Reliability & Validity	Standardized	Scoring [75, 99] [100]	Costs (examples of costs - March 2012)
								these older children."			
Mullen Scales of Early Learning [74, 76]	MSEL	Gross Motor, Visual Reception, Fine Motor, Expressive Language, and Receptive Language	Birth to 68 months	Direct assessment by professional		25 to 35 minutes (at 3 years)	5 scales "Pinpoints strengths and weaknesses"	Reliability: High (.65 or higher) Concurrent Validity: .50 or higher Internal consistency reliability .91 for the composite. Test-retest reliability .71 to .96. Inter-rater reliability .91 to .99.	Sample included 1,849 children, representative of the U.S. Population. (1981-1989)	T score; confi dence intervals, percentile rank, age equivalent, developmenta l stage, descriptive category, profile analysis; an early learning composite	Hand or computer scoring. Complete kit with computer scoring $929.75. Manual ($93.20) Training video ($152.75)
Parents' Evaluation of Developmental Status [49]	PEDS	General development	Birth to 8 years	Parental questionnaire or interview	10 questions	5 minutes for parents to complete (and 1-2 minutes for	PEDS provides evidence on when to refer, when to give parents advice, when to watch carefully, and when to look	sensitivity 74% to 80% specificity 70% to 80%	Standardized on 2823 families from USA, from various backgrounds, including levels of SE	responses grouped into low, medium or high risk for developmenta l and behavioural /mental	Available on-line, requires minimal training, various training materials

Measure	Acronym	Domain(s)	Age	Administration / Versions	Length	Completion time (minutes)	Scales and subscales / Areas screened	Reliability & Validity	Standardized	Scoring [75, 99] [100]	Costs (examples of costs - March 2012)
						clinician to score)	further at development		status and varying ethnicity	health problems. A longitudinal score available	available [116] Complete set $36 (50 forms), Manual $79.95 [117]
Parents' Evaluation of Developmental Status – Developmental Milestones[57] [58]	PEDS-DM	fine motor, gross motor, expressive language, receptive language, self-help, social-emotional, and for older children (4 to 8 years) reading and math	Birth to 8 years	Parent report (but can also be administered directly to children). Items answered in a multiple choice format.	6-8 items (one for each domain) per age / encounter.	Less than 5 minutes Easy and quick to score (takes about one minute)		Sensitivity 70% or greater (average 83%) and specificity 77% to 93% (average 84%) across ages and developmental domains. Test–retest reliability, .98 to .99 Interrater reliability .82 to .96	Standardized and validated on 1619 children around the US. (Population similar to US population in 2006)	cut-offs tied to the 16th percentile and below. The PEDS-DM screener describes milestones in each domain as "met" or "unmet."	Starter kit $275 Training materials available. Online versions available.
Schedule of Growing Skills [90] (Second edition produced 1996 – further information not readily	SGS	Passive Posture, Active Posture, Locomotor, Manipulative, Visual, Hearing and Language,	0 to 5 years	Suitable for Educational Psychologists, SENCOs, Nursery Teachers, Paediatricians and Health Visitors		~ 20 minutes for 2 year olds	Scores are given in each of the 9 domains.	Sensitivity .44-.82 and specificity from .94-1.0. These estimates are based on the original test developed in 1970s. No further	Standardised nationally in England among 348 children.		Complete kits £190 + VAT (only available to qualified, registered test users)

Measure	Acronym	Domain(s)	Age	Administration / Versions	Length	Completion time (minutes)	Scales and subscales / Areas screened	Reliability & Validity	Standardized	Scoring [75, 99] [100]	Costs (examples of costs - March 2012)
accessible.)		Speech and Language, Interactive Social, Self-Care Social						validation has been performed following a second revision.			Onsite training courses – half day and full day. Manual and training DVD available

LANGUAGE											
Early Language Milestone Scale [118] [119]	ELM Scale-2	Speech and language development	Birth to 36 months	healthcare personnel; early childhood providers; and other early childhood specialists	43 items	1 to 10 minutes	Auditory Expressive (further subdivided into Content & Intelligibility), Auditory Receptive, and Visual				Complete kit $192
Language Development Survey [120] [121] [122] [123]	LDS	Language	18-35 months	Parent-completed report	310 words	10 minutes	14 semantic categories	test-retest reliability (.97-.99); Cronbach's alpha internal consistency (.99)	In the 1999-2000 National Survey of Children, Youths, and Adults normative data for the LDS were obtained for 278 children 18 to 35 months.		50 forms for $25
MacArthur-Bates Communicative Development Inventories - Words and Sentences (Toddler form) [124]	CDI-WS	Language and communication skills	16 to 30 months	Parents / caregivers complete (professionals score)	Part I (Words Children Use) contains 685 items. Part II (Sentences and	20–40 minutes for Parents or caregivers to complete and 10–15 minutes for	Part 1 – two subtests Part 11 – six subtests	Internal consistency reliability: .86 for the Words Produced scores and .95 for the Complexity scores. Test-	Updated norming sample 2007 – 2550 children in USA (not nationally representative).	subtest raw scores.	Users guide & technical manual ($59.95), CDI: Words and Sentences (package of 25) $25 Numerous

[125] [126] [127]					Grammar) contains 113 items	professionals to score		retest reliability: .95. Concurrent validity .40 to .88			languages available [3]
New Reynell Developmental Language Scales [128]	NRDLS	Language	3 years to 7 years 6 months (can be used from age 2 years)	mix of play-based activities			Comprehension Scale & Production Scale		Newly standardised on more than 1,200 children in the UK. Provides data from typically developing children between the ages of 2:0 and 7:6		Available from representatives in the UK. Multilingual toolkit available
Preschool Language Scale, 5th edition [129]	PLS-5	Language	Birth to 7.11	interactive, play-based assessment with two page Home Communication questionnaire		45 to 60 minutes	Auditory Comprehension, Expressive Communication	Sensitivity for the Total Language score is .83; specificity is .80	1,400 children participated in the standardization normative sample, collected in more than 45 states in the United States	Raw scores, standard scores, percentile ranks, and age equivalents (6m intervals). A Total Language Score, Auditory Comprehension and	Complete kit $339

										Expressive Communication scores.	
Receptive and Expressive One-Word Picture Vocabulary Tests, Fourth edition [130] [100]	EOWPVT-4, ROWPVT-4	Receptive and Expressive vocabulary	2 to 80+ years	Individually administered tests. Assessors need to be professionally trained.	Each test has 190 items with age-related starting points and ceilings so only a subset used.	15 to 20 minutes per test to administer plus up to 5 minutes to score	The ROWPVT-4 tests an individual's ability to match a spoken word with an image of an object, action, or concept. The EOWPVT-4 tests an individual's ability to name, with one word, objects, actions, and concepts when presented with colour illustrations.	EOWPVT-4: Internal consistency .94 to .95 (for 2 to 5 year olds). Test-retest reliability .98 Inter-rater reliability .95 (for 3 to 17 year olds)	Test co-normed. 2010 normative sample. Based on representative US population.	Raw scores are reported as standard scores, percentile ranks, and (if necessary) age equivalents.	Complete kit is $175 for each test
Sequenced Inventory of Communication Development – Revised [131] [132]	SCID-R	Receptive and expressive language	4 to 48 months	Individually administered by trained professional		30 to 75 minutes	Receptive: discrimination, awareness, and understanding. Expressive: three types of behaviour (imitating, initiating, and responding), and linguistic behaviour (Verbal Output, and Articulation).	Test-retest .88 to .98 Concurrent validity .74 to .95	Norms based on sample of 252 white children aged 4 months to 4 years.		Kit $450 Instruction manual $37 Test manual $37 Test forms (25) $57

PHYSICAL											
Peabody Development Motor Scales [133] [134] [135]	PDMS-2	gross and fine motor skills	Birth to 5 years	Individually administered motor performance tasks	Total of 249 items – depends on child's age	45 to 60 minutes	six subtests: Reflexes, Stationary (body control and equilibrium), Locomotion, Object Manipulation, Grasping, and Visual-Motor Integration. 3 Composite scores: Fine Motor Quotient; Gross Motor Quotient and Total Motor Quotient		Based on a nationally representative age-stratified sample of 2,000+ children from 46 USA states (1997-1998)	Raw scores, percentiles, age equivalents, and standard scores.	Test Kit $430+ or £430

COGNITIVE											
Cognitive Abilities Scale [136] [137]	CAS-2	Cognitive development	3 months to 3 years	"playful" professional interaction / observation	Preschool form: 88 items	20 to 30 minutes	5 sections: Oral Language, Reading, Math, Writing, Enabling Behaviours	(At age 3 years) interscorer reliability .99 test-retest reliability .94	1,106 children from 27 US states		Complete kit $521. Examiners manual available
Parent Report of Children's Abilities [138] [139] (The revised version PARCA-R is for 2 year old children born very preterm) [140] [141]	PARCA	Non-verbal cognitive	2 years	(1) Parent report component and (2) parent administered component	(1) 26 questions (2) 24 items	Approx. 60 minutes	(1) Quantitative skills, special abilities, symbolic play, planning and organising, adaptive behaviours and memory (2) design drawing, match-to-sample, block building, imitative action	107 children (twins or triplets) in UK. 96% of mothers Caucasian, predominantly from middle to upper-middle class. Internal consistency .74 and .83 for the two parts. Correlation between total PARCA score and MDI of BSID-II was .52 Correlation of .51 with Bayley scale (non-language). Adding parent reports of		(1) Parents asked (yes/no) whether seen child perform activity – "Yes"=score of 1. "No"/ "Don't know" = score of 0. Total score derived. (2) Total score	(Test not found as a commercially available product)

								language development significantly improved the predication of MDI.			

SOCIAL EMOTIONAL											
Ages and Stages – Social Emotional [142] [143] [144]	ASQ - SE	Social emotional screening	3 to 66 months	Parental questionnaire (8 questionnaires dependant on child's age)	~30 items	10 to 15 minutes (1 to 3 minutes for professional to score)	Self-regulation, compliance, communication, adaptive functioning, autonomy, affect, and interaction with people	Internal consistency 67-91%; Test-retest reliability 94%; Concurrent validity 93%; Sensitivity 78%; Specificity 94% [145] [146]	Normative studies included 3,014 preschool-age (6 to 60 months) children and their families (USA)	Responses converted to a numerical value. These are totalled and compared with the empirically derived cut-off score	Starter kit with English questionnaires $225 [147] Training DVD $49.95
Brief Infant Toddler Social Emotional Assessment [148] [149]	BITSEA	Social and emotional behaviour	12 to 36 months	Parent form, Child care provider form	42 item parent form	7 to 10 minutes	Problems, Competence, combined scales	Reliability high (.80 or higher). Sensitivity and Specificity both high (.80 or higher) [75]	Parent form - national sample of 600 children from 42 states (USA, 2002)	Total Problem and Competence scores. Cut points for age bands for boys and girls	BITSEA Kit $110.25 (manual +25 parent forms + 25 Childcare provider forms)
The Infant-	ITSEA	4 Domains:	12 to	Parent Form	166 item	25 to 30	17 subscales	Test–retest	National	Domain scores,	Complete kit

Toddler Social and Emotional Assessment [150]		Externalizing, Internalizing , Dysregulation and Competence	36 months	and Child Care Provider Form		minutes as questionnaire 35 to 45 minutes as an interview	address four domains. Plus Index Scores: Maladaptive Cluster; Social Relatedness Cluster; Atypical Cluster “gives in-depth analysis to guide intervention planning”	coefficients for domains ranged from 0.82 to 0.90 and from 0.69 to 0.85 for scales. [151] Reliability: High (.80 or higher) Concurrent Validity: adequate (.50 to .69) [75]	sample of 600 US children. Clinical groups included language delayed, premature, and other diagnosed disorders	subscale scores, and item cluster scores. T-scores and percentile ranks divided by 6-month age bands and gender	$230

ADDITIONAL TESTS LISTED IN 'HEALTHY CHILD PROGRAMME, THE TWO YEAR REVIEW' [10]											
Achenbach Child Behavior Check List [152] [153] [154]	CBCL/1½ -5 and C-TRF	Used to detect behavioural and emotional problems	1 ½ to 5 years	Parents forms (CBCL/1½ - 5) or Teachers forms (C-TRF). Restricted for use by suitably qualified professionals (training in standardised assessment to Master's degree or higher)	99 problem items plus other items e.g. concerns	10 to 20 minutes (Questions answered based on the preceding two months)	7 Syndrome Scales: e.g. Emotionally Reactive; Anxious/Depressed; Aggressive Behaviour. 5 DSM-Oriented Scales: e.g. Affective Problems; Attention Deficit/Hyperactivity Problems;	Mean test-retest reliability .85 for CBCL and .81 for C-TRF. Inter-parent agreement (CBCL) .61 Concurrent validity: The CBCL correctly classified 84% of a sample of children and the CTR correctly classified 74% of the children. [100]	Originally normed on 1728 US children. Multicultural norms available	Item scoring: 0 = not true, 1 = somewhat / sometimes true, or 2 = very true or often true of the child. Internalizing, Externalizing, and Total Problems scales and a Stress Problems scale. T scores available.	Available in >85 languages. Hand-scoring starter kit £140 Computer-scoring starter kit £210 ADM computer software £170 [155]
First Words and First sentence Tests [156]	FWT/FST	Receptive and expressive language and first word combinations	18 to 36 months	Parent checklist and pictures		10 to 15 minutes			In UK		NOT CURRENTLY AVAILABLE - OUT OF PRINT SINCE DEC 2001
The Home Observation for Measurement of the	HOME (IT HOME)	quality and quantity of stimulation and support available to a child in the	Birth to 3 years	Home visit with semi-structured observation and parent interview	45 items	45 to 90 minutes (when the	six subscales: 1) Parental Responsivity, 2) Acceptance of Child, 3)	Inter-observer agreement for each measure is 80% or higher.		Yes/No format is used in scoring items.	Comprehensive manual $50. Forms $15

Environment (HOME) Inventory - Infant / Toddler HOME [157] [158]		home environment		Requires experienced interviewers and special training. There is no standard question format or standardised procedure for administration		child is awake and can be observed interacting with the mother or primary caretaker)	Organization of the Environment, 4) Learning Materials, 5) Parental Involvement, and 6) Variety in Experience.	Internal consistency .89 for total HOME and averaged .70 for the 6 subscales		Higher total HOME scores indicate more enriched home environment. Scores in the lowest fourth of the score range indicate an environment that may pose a risk to the child's development.	(for 50)
The Modified Checklist for Autism in Toddlers [159] [160, 161]	M-CHAT	Screening tool for Autism Spectrum Disorders (ASD)	16 to 30 months	Parental questionnaire followed by M-CHAT interview if positive score (~10% of community sample). Interview reduces number of false positives. Should be administered by trained health professional.	23 items in question naire.	"few minutes" to complete. Can be scored in less than 2 minutes. (+) 5 to 15 minutes for interview.	Includes reciprocal social interaction, language and communication, and repetitive, stereotyped patterns of behaviour	Originally validated (2001) on 1293 children. Set up to have maximum sensitivity, so has high false positive rate. Based on scores alone for children 2 to 3 years referred for		Yes / No responses. Children who fail 3 or more items total or 2 or more critical items should be referred for diagnostic evaluation.	Free online access Available in many languages

								assessment of suspected Autism: sensitivity .77 to .92 and specificity .27 to .43 [162] Positive predictive value of M-CHAT plus interview .57 [161]			
Social and Communication Questionnaire [163, 164] (Previously known as the Autism Screening Questionnaire - ASQ)	SCQ	Screen for autism spectrum disorders (not suitable for diagnosis) (SCQ can also be used to compare levels of ASD symptomatology across various groups)	**Over 4 years**	Parent completed questionnaire. The *Lifetime Form* focuses on the child's entire developmental history. The *Current Form* looks at the child's behaviour over the most recent 3-month period.	40 items	Less than 10 minutes		In a UK sample of children 9 to 13 years with special educational needs: Sensitivity 0.86 and specificity 0.78 [165] Similar findings found with 9 and 10 year olds. [166] Sensitivity and specificity lower with younger children.[167]		Yes / No responses. Each item then scored 0 or 1 (with 1 indicating a symptom of Autism). Total score (range 0 to 39) with cut-off points.	Kit $115 (20 forms and manual) or £108
Strengths and Difficulties Questionnaire	SDQ	Psychological attributes	**3 to 16 year olds**	Parents or nursery teachers	25 items	10 minutes	5 scales: (1) emotional symptoms (2)	In a British Study [170] multi-	British sample included 10,438	Sub-scale scores and total	Free to download in numerous

[168]			(one version for 3 and 4 year olds)				conduct problems (3) hyperactivity / inattention (4) peer relationship problems (5) prosocial behaviour "May be better to use the 3 subscale version for general population samples" [169]	informant SDQs (parents, teachers, older children) identified individuals (5 to 15 years) with a psychiatric diagnosis with a specificity of 94.6% and a sensitivity of 63.3%	individuals aged between 5 and 15 years. Norms available for other countries with different age ranges.	difficulties score	languages [171]
Sure Start Language Measure [172, 173]	SSLM	Language (Used to monitor language performance of 2 years old children in Sure Start Programmes 2001 to 2006).	23 to 27 months	Face-to-face interview with child's main caregiver. Designed for use by range of workers, but largely used by speech and language therapists.	50 (or 100) word vocabulary checklist (from MCDI-UKSF)); and 4 questions about parental concerns (from PEDS)	5 to 15 minutes (including explanation of study and signing consent form)	Word count, use of words, parental concerns about general and language development		Standardized on 1290 English children aged 16 to 30 months with a broad range of ethnic backgrounds (2004/2005). [174]	Word count scores; word use scores; parental concerns – general and language. Composite measure "PQRS" for comparisons across populations.	

Appendix IV

Table 5: ASQ-3 and PEDS assessed against Department of Health requirements for a Population measure of Child Development in 2 year old children

DH Requirements	ASQ-3	PEDS
Updatable annually and enables population level child development at age 2-2½ years to be tracked over time	✓	✓
Valid and reliable measure of the aspect of child development of interest	✓/? May need addition of ASQ-SE	✓
Applicable to different groups with differing levels of development and need.	✓	✓
It has standardised norms for 2 year old children in England that can be used to benchmark progress	X	?
It can be compiled at local authority and at national level	✓	✓
It is sensitive to population level changes	✓	X
It reflects influences on child development during pregnancy and the first two years of live as well as being predictive of later outcomes, especially school readiness.	?	?
It is simple to apply and acceptable to families and professionals	✓	✓
Minimises burdens on families and professionals	✓	✓
It can be integrated with existing contacts with all families around this age	✓	✓

References

1. Secretary of State for Health, *Healthy Lives, Healthy People: Our Strategy for Public Health in England*, 2010, The Stationery Office.

2. HM Government. *Opening Doors, Breaking Barriers: A Strategy for Social Mobility*. 2011; Available from: https://update.cabinetoffice.gov.uk/resource-library/opening-doors-breaking-barriers-strategy-social-mobility.

3. Department of Health, *Healthy Child Programme, Pregnancy and the First Five Years of Life.* http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_107563
2009.

4. Marmot, M., et al., *Fair society, healthy lives: Strategic review of health inequalities in England post-2010*, 2010.

5. Field, F., *The Foundation Years: preventing poor children becoming poor adults*2010: The Stationery Office/Tso.

6. Allen, G., *Early intervention: the next steps.* London: HM Government, 2011.

7. Tickell, C., *The Early Years: Foundations for life, health and learning—An Independent Report on the Early Years Foundation Stage to Her Majesty's Government*, 2011. p. http://media.education.gov.uk/MediaFiles/B/1/5/%7BB15EFF0D-A4DF-4294-93A1-1E1B88C13F68%7DTickell%20review.pdf.

8. Department of Education and Department of Health, *Supporting Families in the Foundation Years.* http://www.education.gov.uk/home/childrenandyoungpeople/earlylearningandchildcare/early/a00192398/supporting-families-in-the-foundation-years
July 2011.

9. Department of Health, *Improving outcomes and supporting transparency. Part 1: A public health outcomes framework for England, 2013-2016*, Department of Health, Editor 2012, Crown copyright. .

10. Department of Health, *Healthy Child Programme: The two year review.* http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_108329.pdf, 2009.

11. Marks, K., et al., *The thorny nature of predictive validity studies on screening tests for developmental-behavioral problems.* Pediatrics, 2008. **122**(4): p. 866-868.

12. Wood, R., et al., *Trends in the coverage of 'universal' child health reviews: observational study using routinely available data.* BMJ Open, 2012. **2**(2): p. e000759.

13. Johnson, S. and N. Marlow, *Developmental screen or developmental testing?* Early Human Development, 2006. **82**(3): p. 173-83.

14. Humphrey, N., et al., *Measures of Social and Emotional Skills for Children and Young People A Systematic Review.* Educational and Psychological Measurement, 2011. **71**(4): p. 617-637.

15. Rydz, D., et al., *Developmental screening.* J Child Neurol, 2005. **20**(1): p. 4-21.

16. Hall, D.M.B. and D. Elliman, *Health for all children*. 4th Edition, revised ed2006: Oxford University Press, 4th edition revised.

17. Council on Children with Disabilities, Section on Developmental Behavioral Pediatrics, and Bright Futures Steering Committee and Home Initiatives for Children with Special Needs Project Advisory Committee, *Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening.* Pediatrics, 2006. **118**(1): p. 405-20.

18. Limbos, M.M. and D.P. Joyce, *Comparison of the ASQ and PEDS in screening for developmental delay in children presenting for primary care.* J Dev Behav Pediatr, 2011. **32**(7): p. 499-511.

19. Danaher J. *Eligibility policies and practices for young children under Part B of IDEA*. 2011 [cited 2012 04]; Available from: http://www.nectac.org/~pdfs/pubs/nnotes27.pdf.

20. Wilson, J. and G. Jungner, *Principles and practice of screening for disease. 1968.* Geneva: WHO, 2009: p. 15-39.

21. UK National Screening Committee. *Criteria for appraising the viability, effectiveness and appropriateness of a screening programme*. 2012, [cited 2012 4th April]; Available from: http://www.screening.nhs.uk/criteria.

22. UK National Screening Committee. *UK NSC Policy database*. 2012 [cited 2012 4/04]; Available from: http://www.screening.nhs.uk/policydb.php?sortby=Policy_short_name&direction=DESC.

23. Hamilton, S., *Screening for developmental delay: Reliable, easy-to-use tools.* The Journal of Family Practice; The Journal of Family Practice, 2006.

24. Glascoe, F.P., *Are overreferrals on developmental screening tests really a problem?* Arch Pediatr Adolesc Med, 2001. **155**(1): p. 54-9.

25. Drotar, D., T. Stancin, and P. Dworkin. *Pediatric developmental screening: understanding and selecting screening instruments*. New York, NY: Commonwealth Fund 2008; Available from: http://www.commonwealthfund.org/Publications/Fund-Manuals/2008/Feb/Pediatric-Developmental-Screening--Understanding-and-Selecting-Screening-Instruments.aspx.

26. Committee on Children with Disabilities, *Developmental surveillance and screening of infants and young children.* Pediatrics, 2001. **108**(1): p. 192-6.

27. Grant R, G.D., Brito A,. *Developmental and Social-Emotional Screening Instruments for Use in Pediatric Primary Care in Infants and Young Children*. Children's Health Fund New York, NY; Available from: http://www.childrenshealthfund.org/sites/default/files/dev-and-mental-health-primary-care-screening-tools.pdf.

28. Halle, T., et al., *Understanding and Choosing Assessments and Developmental Screeners for Young Children Ages 3-5: Profiles of Selected Measures* 2011, Child Trends.

29. Squires, J., et al., *ASQ-3 User's Guide*. Third ed2009, Baltimore: Brookes.

30. Squires, J. and D. Bricker. *Concurrent Validity for ASQ-3 Questionnaire Intervals*. Available from: http://agesandstages.com/pdfs/asq3_concurrent_validity.pdf.

31. Gollenberg, A.L., et al., *Concurrent validity of the parent-completed Ages and Stages Questionnaires, 2nd Ed. with the Bayley Scales of Infant Development II in a low-risk sample.* Child: Care, Health & Development, 2010. **36**(4): p. 485-90.

32. Troude, P., et al., *Ages and Stages Questionnaires: feasibility of postal surveys for child follow-up.* Early Human Development, 2011. **87**(10): p. 671-6.

33. Richter, J. and H. Janson, *A validation study of the Norwegian version of the Ages and Stages Questionnaires.* Acta Paediatrica, 2007. **96**(5): p. 748-52.

34. Janson, H. and J. Squires, *Parent-completed developmental screening in a Norwegian population sample: a comparison with US normative data.* Acta Paediatrica, 2004. **93**(11): p. 1525-9.

35. Kerstjens, J.M., et al., *Support for the global feasibility of the Ages and Stages Questionnaire as developmental screener.* Early human development, 2009. **85**(7): p. 443-447.

36. Kapci, E.G., S. Kucuker, and R.I. Uslu, *How Applicable Are Ages and Stages Questionnaires for Use With Turkish Children?* Topics in Early Childhood Special Education, 2010. **30**(3): p. 176-188.

37. Elbers, J., et al., *The Ages and Stages Questionnaires: feasibility of use as a screening tool for children in Canada.* Can J Rural Med, 2008. **13**(1): p. 9-14.

38. Radecki, L., et al., *Trends in the use of standardized tools for developmental screening in early childhood: 2002-2009.* Pediatrics, 2011. **128**(1): p. 14-9.

39. Bian, X.-y., et al., *[Studies of the norm and psychometric properties of Ages and Stages Questionnaires in Shanghai children].* Zhonghua Erke Zazhi, 2010. **48**(7): p. 492-6.

40. Yao, G.-y., et al., *[Cutoff scores of the Ages and Stages Questionnaire-Chinese for screening infants and toddlers].* Zhonghua Erke Zazhi, 2010. **48**(11): p. 824-8.

41. Skellern, C.Y., Y. Rogers, and M.J. O'Callaghan, *A parent-completed developmental questionnaire: follow up of ex-premature infants.* J Paediatr Child Health, 2001. **37**(2): p. 125-9.

42. Lindsay, N.M., et al., *Use of the Ages and Stages Questionnaire to predict outcome after hypoxic-ischaemic encephalopathy in the neonate.* J Paediatr Child Health, 2008. **44**(10): p. 590-5.

43. Chiu, S.-H. and M.A. DiMarco, *A pilot study comparing two developmental screening tools for use with homeless children.* Journal of Pediatric Health Care, 2010. **24**(2): p. 73-80.

44. Yu, L.-M., et al., *Evaluation of the Ages and Stages Questionnaires in identifying children with neurosensory disability in the Magpie Trial follow-up study.* Acta Paediatrica, 2007. **96**(12): p. 1803-8.

45. Squires, J., A. Carter, and P. Kaplan, *Developmental monitoring of children conceived by intracytoplasmic sperm injection and in vitro fertilization.* Fertil Steril, 2003. **79**(2): p. 453-4.

46. Rydz, D., et al., *Screening for developmental delay in the setting of a community pediatric clinic: a prospective assessment of parent-report questionnaires.* Pediatrics, 2006. **118**(4): p. e1178-86.

47. Squires, J.K., et al., *Parent-completed developmental questionnaires: Effectiveness with low and middle income parents.* Early Childhood Research Quarterly, 1998. **13**(2): p. 345-354.

48. Squires, J., et al., *Identification of social-emotional problems in young children using a parent-completed screening measure.* Early Childhood Research Quarterly, 2001. **16**: p. 405-419.

49. *Welcome to PEDStest.com, Tools for Developmental-Behavioral Screening and Surveillance*. Available from: http://www.pedstest.com/Home.aspx.

50. *What is the science behind screening and how can we recognize well-constructed, accurate tools?* . [cited 2012 04]; Available from: http://www.pedstest.com/Portals/0/downloads/science.pdf.

51. Halfon, N., et al., *Assessing development in the pediatric office.* Pediatrics, 2004. **113**(Supplement 5): p. 1926-1933.

52. *Welcome to PEDStest.com Use of "Parents' Evaluation of Developmental Status (PEDS) in Population Surveys* Available from: http://www.pedstest.com/Research/tabid/91/ID/27/Use-of-Parents-Evaluation-of-Developmental-Status-PEDS-in-Population-Surveys.aspx.

53. Coghlan, D., J.S.H. Kiing, and M. Wake, *Parents' Evaluation of Developmental Status in the Australian day-care setting: developmental concerns of parents and carers.* Journal of Paediatrics & Child Health, 2003. **39**(1): p. 49-54.

54. Harris, F., *The first implementation of the Sure Start Language Measure.* DfES publications, 2002.

55. PEDSTest.com. *Welcome to PEDStest.com, Tools for Developmental-Behavioral Screening and Surveillance*. Available from: http://www.pedstest.com/Home.aspx.

56. Davies, S. and H. Feeney, *A pilot of the Parents' Evaluation of Developmental Status tool.* Community Pract, 2009. **82**(7): p. 29-31.

57. PEDSTest.com. *Parents' Evaluation of Developmental Status -- Developmental Milestones (PEDS:DM)©*. Available from: http://www.pedstest.com/AboutOurTools/LearnAboutPEDSDM.aspx.

58. Brothers, K.B., F.P. Glascoe, and N.S. Robertshaw, *PEDS: Developmental Milestones—An Accurate Brief Tool for Surveillance and Screening.* Clinical Pediatrics, 2008. **47**(3): p. 271-279.

59. Ireton, H. *Child Development Inventory Manual*. 1992; Available from: http://www.childdevrev.com/page11/page59/files/cdi-manual.pdf.

60. Ireton H and Glascoe FP, *Assessing children's development using parents' reports. The Child Development Inventory.* Clin Pediatr (Phila). 1995. **34**(5): p. 248-55.

61. Ireton Harold. *Child Development Inventory Manual*. 1992; Available from: http://www.childdevrev.com/page11/page59/files/cdi-manual.pdf.

62. Doig, K.B., et al., *The Child Development Inventory: A developmental outcome measure for follow-up of the high-risk infant.* The Journal of Pediatrics, 1999. **135**(3): p. 358-362.

63. Glascoe, F.P. and P.H. Dworkin, *The Role of Parents in the Detection of Developmental and Behavioral Problems.* Pediatrics, 1995. **95**(6): p. 829-836.

64. Duyme, M., et al., *French Norms and Validation of the Child Development Inventory (CDI): Inventaire du Développement de l'Enfant (IDE).* Clinical Pediatrics, 2011. **50**(7): p. 636-647.

65. Rydz D, et al., *Screening for developmental delay in the setting of a community pediatric clinic: a prospective assessment of parent-report questionnaires.* Pediatrics., 2006. **118**(4): p. e1178-86.

66. Child Development Review. *CDR - Parent Questionnaire - for Toddlers and Preschoolers*. Available from: http://www.childdevrev.com/page4/page10/cdrhealthcare.html.

67. Ireton, H. and H. Vader. *Child Development Review Instructions Manual*. 2004 [cited 2012 04]; Available from: http://www.childdevrev.com/page4/page58/files/cdr_manual.pdf.

68. Pearson. *Bayley Scales of Infant and Toddler Development®, Third Edition (Bayley-III®)*. Available from: http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8027-23X.

69. Albers, C.A. and A.J. Grieve, *Test Review: Bayley, N. (2006). Bayley Scales of Infant and Toddler Development– Third Edition. San Antonio, TX: Harcourt Assessment.* Journal of Psychoeducational Assessment, 2007. **25**(2): p. 180-190.

70. Centre, U.o.C.T.P. *Bayley Scales of Infant Development, 3rd Edition*. [cited 2012 04]; Available from: http://www.psychometrics.cam.ac.uk/page/92/bayley-scales.htm.

71. Pearson. *Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III), Bayley-III UK Validation* [cited 2012 04]; Available from: http://www.pearsonclinical.co.uk/Psychology/ChildCognitionNeuropsychologyandLanguage/ChildGeneralAbilities/BayleyScalesofInfantandToddlerDevelopmentThirdEdition%28Bayley-III%29/ForThisProduct/Bayley-IIIUKValidation.aspx.

72. Moore, T., et al., *Relationship between test scores using the second and third editions of the Bayley Scales in extremely preterm children.* J Pediatr, 2012. **160**(4): p. 553-8.

73. Anderson, P.J., et al., *Underestimation of Developmental Delay by the New Bayley-III Scale.* Arch Pediatr Adolesc Med, 2010. **164**(4): p. 352-356.

74. Inc, P.E. *Mullen Scales of Early Learning*. Available from: http://psychcorp.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=PAa11150.

75. Washington State Office of Superintendent of Public Instruction. *A Guide to Assessment in Early Childhood; Infancy to Age Eight*. 2008 [cited 2012 04]; Available from: http://www.k12.wa.us/EarlyLearning/pubdocs/assessment_print.pdf.

76. U.S. Department of Health and Human Services; Administration for Children and Families. *Mullen Scales of Early Learning, AGS Edition, 1995* [cited 2012 04]; Available from: http://www.acf.hhs.gov/programs/opre/ehs/perf_measures/reports/resources_measuring/res_meas_cdibb.html.

77. Bradley-Johnson, S., *Cognitive Assessment for the Youngest Children: A Critical Review of Tests.* Journal of Psychoeducational Assessment, 2001. **19**(1): p. 19-44.

78. Bishop, S.L., et al., *Convergent validity of the Mullen Scales of Early Learning and the differential ability scales in children with autism spectrum disorders.* American Journal on Intellectual & Developmental Disabilities, 2011. **116**(5): p. 331-43.

79. Education, N. *Battelle Developmental Inventory, Second Edition*. Available from: http://www.assess.nelson.com/test-ind/bdi.html.

80. Bliss, S.L., *Test Reviews: Newborg, J. (2005). Battelle Developmental Inventory–Second Edition. Itasca, IL: Riverside.* Journal of Psychoeducational Assessment, 2007. **25**(4): p. 409-415.

81. Advantage, N.E.C.s.L. *Battelle Developmental Inventory, Second Edition*. Available from: http://www.assess.nelson.com/test-ind/bdi.html.

82. Curriculum Associates. *BRIGANCE Early Childhood Screens*. [cited 2012 04]; Available from: http://www.curriculumassociates.com/products/detail.aspx?title=BrigEC-Screens&topic=CEC0.

83. Hamilton, S., *Screening for developmental delay: reliable, easy-to-use tools.* Journal of Family Practice, 2006. **55**(5): p. 415-22.

84. Glascoe, F., *The Brigance Infant and Toddler Screen: Standardization and Validation.* Journal of Developmental & Behavioral Pediatrics, 2002. **23**(3): p. 145-150.

85. Lee LL and H. SR., *Psychometric properties and standardization samples of four screening tests for infants and young children: a review* Pediatr Phys Ther., 2005. **17**(2): p. 140-7.

86. Denver Developmental Materials Inc. *DENVER II* [cited 2012 04]; Available from: http://www.denverii.com/home.html.

87. Frankenburg, W.K., et al., *The Denver II: a major revision and restandardization of the Denver Developmental Screening Test.* Pediatrics, 1992. **89**(1): p. 91-7.

88. Glascoe, F.P., et al., *Accuracy of the Denver-II in developmental screening.* Pediatrics, 1992. **89**(6 Pt 2): p. 1221-5.

89. Luiz D, et al. *GMDS-ER: Griffiths Mental Development Scales-Extended Revised: Two to Eight years.* Administration and analysis manual 2006; Available from: http://www.hogrefe.co.uk/clinical-and-educational/child-development/griffiths-mental-development-scales-extended-revised-2-to-8-years-gmds-er-2-8.html.

90. GL assessment. *Schedule of Growing Skills*. Available from: http://shop.gl-assessment.co.uk/home.php?cat=360.

91. Bellman, M., et al., *A developmental test based on the STYCAR sequences used in the national childhood encephalopathy study.* Child: care, health and development, 1985. **11**(5): p. 309-323.

92. Williams, M.E., et al., *Schedule of Growing Skills II: Pilot Study of an Alternative Scoring Method.* Psychology, 2013. **4**(3): p. 143-152.

93. Janus, M., et al. *The Early Development Instrument: A population-based measure for communities. A handbook on development, properties and use. Offord Centre for Child Studies, Hamilton, Ontario, Canada*
2007; Available from: http://www.rch.org.au/emplibrary/australianedi/2007_12_FINALEDIHANDBOOK.pdf .

94. Goldfeld, S., et al., *The process and policy challenges of adapting and implementing the Early Development Instrument in Australia.* Early Education and Development, 2009. **20**(6): p. 978-991.

95. Centre for Community Child Health. *Australian Early development Index: Building better Communities for Children. Final Evaluation Report.* 2007; Available from: http://www.rch.org.au/ccch/research.cfm?doc_id=10617.

96. Lloyd J, H.C., *From Kindergarten readiness to fourth-grade assessment: Longitudinal analysis with linked population data. .* Social Science & Medicine 2009. **68**: p. 111-123.

97. Janus, M.J. and E. Duku, *The school entry gap: Socioeconomic, family, and health factors associated with children's school readiness to learn.* Early Education and Development, 2007. **18**(3): p. 375-403.

98. Squires, J., D. Bricker, and L. Potter, *Revision of a parent-completed development screening tool: Ages and Stages Questionnaires.* J Pediatr Psychol, 1997. **22**(3): p. 313-28.

99. Ringwalt, S., *Developmental Screening and Assessment Instruments with an Emphasis on Social and Emotional Development for Young Children Ages Birth through Five*

2008, Chapel Hill: The University of North Carolina, FPG Child Development Institute.

100. Administration for Children & Families. *Child Development Instruments*. Available from: http://www.acf.hhs.gov/programs/opre/ehs/perf_measures/reports/resources_measuring/res_meas_cdi.html.

101. Brookes Publishing Co. *ASQ. Ages and Stages Questionnaires - What is ASQ-3?* ; Available from: http://agesandstages.com/asq-products/asq-3/.

102. Squires, J., et al., *Psychometric Studies of ASQ, Third Edition. Excerpted from ASQ-3 User's Guide.*, 2009.

103. Brookes. *Ages & Stages Questionnaires®, Third Edition (ASQ-3™)*. Available from: http://www.brookespublishing.com/store/books/squires-asq/index.htm#asqkit.

104. Pearson. *Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III) - Training*. Available from:

http://www.pearsonclinical.co.uk/Psychology/ChildCognitionNeuropsychologyandLanguage/ChildGeneralAbilities/BayleyScalesofInfantandToddlerDevelopmentThirdEdition%28Bayley-III%29/ForThisProduct/Training.aspx.

105. Inc, P.E. *Bayley Scales of Infant and Toddler Development®, Third Edition (Bayley-III®) Screening Test*. Available from: http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8027-256.

106. Brigance. *Brigance Early Childhood Screens. Standardization & Validation Research Highlights.* . Available from: http://www.casamples.com/downloads/Brig-EC-research.pdf.

107. Review, C.D. *Child Development Inventory*. Available from: http://www.childdevrev.com/page15/page17/cdi.html.

108. Review, C.D. *The Child Development Inventory (CDI)*. Available from: http://www.childdevrev.com/page11/page24/cdiresearch.html.

109. Nosarti, C., et al., *Impaired executive functioning in young adults born very preterm.* Journal of the International Neuropsychological Society, 2007. **13**(4): p. 571-81.

110. Visintainer, P.F., et al., *Standardization of the Capute Scales: Methods and Results.* Journal of Child Neurology, 2004. **19**(12): p. 967-972.

111. Brookes. *The Capute Scales - Cognitive Adaptive Test and Clinical Linguistic and Auditory Milestone Scale (CAT/CLAMS)*. Available from: http://www.brookespublishing.com/store/books/accardo-8167/index.htm.

112. O'Connor Leppert, M.L., et al., *The Capute Scales: CAT/CLAMS—A pediatric assessment tool for the early detection of mental retardation and communicative disorders.* Mental retardation and developmental disabilities research reviews, 1998. **4**(1): p. 14-19.

113. Child Development Review. Available from: http://www.childdevrev.com/page11/page43/idicdrresearch.html.

114. *Motor and Social Development (MSD),*. Available from: http://www.nlsinfo.org/childya/nlsdocs/guide/assessments/MotorSocialDevelopment.htm.

115. Baker, P., et al. *NLSY Child Handbook. Revised Edition. A Guide to the 1986-1990 National Longitudinal Survey of Youth Child Data.* . 1993; Available from: http://www.nlsinfo.org/childya/nlsdocs/guide/previousrounds/ChildHandbook1990.pdf.

116. PEDSTest.com, *Slide shows, Videos, Audience Handouts and Other Training Materials.*

117. PEDSTest.com. *PEDS Products*. Available from: http://www.pedstestshop.com/home.php?cat=2.

118. pro.ed - An International Publisher. *Early Language Milestone Scale (ELM Scale-2)*. Available from: http://www.proedinc.com/customer/ProductView.aspx?ID=784&sSearchWord=james%2520coplan.

119. Coplan, J. and J.R. Gleason, *Test-retest and interobserver reliability of the Early Language Milestone Scale, second edition.* Journal of Pediatric Health Care, 1993. **7**(5): p. 212-9.

120. ASEBA - Achenbach System of Empirically Based Assessment. *The Language Development Survey (LDS)*. Available from: http://www.aseba.org/research/language.html.

121. Rescorla, L. and T.M. Achenbach, *Use of the language development survey (LDS) in a national probability sample of children 18 to 35 months old.* Journal of Speech Language & Hearing Research, 2002. **45**(4): p. 733-43.

122. Crandell, L.E. and R.P. Hobson, *Individual differences in young children's IQ: a social-developmental perspective.* Journal of Child Psychology & Psychiatry & Allied Disciplines, 1999. **40**(3): p. 455-64.

123. Rescorla, L., *The Language Development Survey: a screening tool for delayed language in toddlers.* J Speech Hear Disord, 1989. **54**(4): p. 587-99.

124. MacArthur Bates Communicatitive Development Inventories. *Welcome to the MacArthur-Bates Communicative Development Inventories*. Available from: http://www.sci.sdsu.edu/cdi/.

125. Brookes. *MacArthur-Bates Communicative Development Inventories (CDIs), Second Edition*. Available from: http://www.brookespublishing.com/store/books/fenson-cdi/index.htm.

126. U.S. Department of Health and Human Services - Administration for Children & Families. *MacArthur-Bates Communicative Development Inventories (CDIs) — Second Edition, 2007* Available from: http://www.acf.hhs.gov/programs/opre/ehs/perf_measures/reports/resources_measuring/res_meas_cdiaa.html#foot1.b.

127. Law, J. and P. Roy, *Parental Report of Infant Language Skills: A Review of the Development and Application of the Communicative Development Inventories.* Child and Adolescent Mental Health, 2008. **13**(4): p. 198-206.

128. GL assessment. *New Reynell Developmental Language Scales*. Available from: http://reynell.gl-assessment.co.uk/?gclid=CJeDlNz8xa4CFUEMfAodHHHwXA.

129. Pearson, *Preschool Language Scales, Fifth Edition (PLS™-5).*

130. Ann Arbor Publishers. *EOWPVT-4 - Expressive One-Word Picture Vocabulary Test-4*. Available from: http://www.annarbor.co.uk/index.php?main_page=index&cPath=251_363.

131. WPS Western Psychological Services. *Sequenced Inventory of Communication Development, Revised (SICD-R)*. Available from: http://portal.wpspublish.com/portal/page?_pageid=53,70239&_dad=portal&_schema=PORTAL.

132. Child Trends. *Early Childhood Measures Profile*. Available from: http://www.childtrends.org/files/child_trends-2004_09_01_FR_ECMeasures.pdf.

133. WPS Western Psychological Services. *Peabody Developmental Motor Scales, Second Edition (PDMS-2)*. Available from: http://portal.wpspublish.com/portal/page?_pageid=53,69919&_dad=portal&_schema=PORTAL.

134. Pearson. *Peabody Developmental Motor Scales, Second Edition (PDMS-2)* Available from: http://www.pearsonclinical.co.uk/AlliedHealth/PaediatricAssessments/Motor/PeabodyDevelopmentalMotorScalesSecondEdition%28PDMS-2%29/PeabodyDevelopmentalMotorScalesSecondEdition%28PDMS-2%29.aspx.

135. pro.ed - An International Publisher. *PDMS-2: Peabody Developmental Motor Scales – Second Edition*. Available from: http://www.proedinc.com/customer/productView.aspx?ID=1783.

136. pro.ed - An International Publisher. *CAS-2: Cognitive Abilities Scale - Second Edition*. Available from: http://www.proedinc.com/customer/productView.aspx?ID=686.

137. Swanson, J.R., et al., *The Cognitive Abilities Scale—Second Edition Preschool Form.* Journal of Psychoeducational Assessment, 2009. **27**(1): p. 46-56.

138. Saudino, K.J., et al., *The validity of parent-based assessment of the cognitive abilities of 2-year-olds.* British Journal of Developmental Psychology, 1998. **16**(3): p. 349-362.

139. Plomin, R., et al., *Associations between behaviour problems and verbal and nonverbal cognitive abilities and disabilities in early childhood.* Journal of Child Psychology & Psychiatry & Allied Disciplines, 2002. **43**(5): p. 619-33.

140. Johnson, S., et al., *Validation of a parent report measure of cognitive development in very preterm infants.* Dev Med Child Neurol, 2004. **46**(6): p. 389-97.

141. Johnson, S., et al., *Developmental assessment of preterm infants at 2 years: validity of parent reports.* Developmental Medicine & Child Neurology, 2008. **50**(1): p. 58-62.

142. Brookes Publishing Co. *ASQ; Ages and Stages Questionnaires - ASQ:SE*. Available from: http://agesandstages.com/asq-products/asqse/.

143. Squires, J., D. Brickler, and E. Twombly, *The ASQ:SE User's Guide for the Ages & Stages Questionnaires: Social-Emotional*, 2003.

144. Squires, J., D. Bricker, and E. Twombly, *Parent-completed screening for social emotional problems in young children: The effects of risk/disability status and gender on performance.* Infant Mental Health Journal, 2004. **25**(1): p. 62-73.

145. ASQ. *ASQ:SE Technical Qualities*. Available from: http://agesandstages.com/asq-products/asqse/asqse-technical-qualities/.

146. ASQ;SE. *Technical Report on ASQ-SE*. Available from: http://agesandstages.com/pdfs/asqse_technical_report.pdf.

147. ASQ. *ASQ:SE Product Details*. Available from: http://agesandstages.com/asq-products/asqse/asqse-product-details/.

148. Pearson. *Brief Infant Toddler Social Emotional Assessment (BITSEA)*. Available from: http://www.pearsonassessments.com/haiweb/cultures/en-us/productdetail.htm?pid=015-8007-352&Community=CA_Ed_AI_Early.

149. Briggs-Gowan, M.J. and A.S. Carter, *Social-emotional screening status in early childhood predicts elementary school outcomes.* Pediatrics, 2008. **121**(5): p. 957-62.

150. Pearson. *Infant Toddler Social Emotional Assessment (ITSEA)*. Available from: http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8007-387.

151. Carter, A.S., et al., *The Infant-Toddler Social and Emotional Assessment (ITSEA): factor structure, reliability, and validity.* Journal of Abnormal Child Psychology, 2003. **31**(5): p. 495-514.

152. ASEBA - Achenbach System of Empirically Based Assessment. *Preschool (Ages 1½-5) Assessments*. Available from: http://www.aseba.org/preschool.html.

153. CEBA - The California Evidence-Based Clearinghouse for Child Welfare. *Child Behavior Checklist for Ages 1.5-5 (CBCL/1.5-5)*. Screening and Assessment Tools for Child Welfare; Available from: http://www.cebc4cw.org/assessment-tool/child-behavior-checklist-for-ages-1-5-5-cbcl-1-5-5/.

154. Rescorla, L.A., *Assessment of young children using the Achenbach System of Empirically Based Assessment (ASEBA).* Mental retardation and developmental disabilities research reviews, 2005. **11**(3): p. 226-37.

155. Aston University - Life & Health Sciences. *Assessments for pre-school children - aged 1.5 to 5 years*. Available from: http://www1.aston.ac.uk/lhs/aseba/1-5/.

156. Pickstone, C., P. Hannon, and L. Fox, *Surveying and screening preschool language development in community-focused intervention programmes: a review of instruments.* Child Care Health Dev, 2002. **28**(3): p. 251-64.

157. Totsika, V. and K. Sylva, *The Home Observation for Measurement of the Environment Revisited.* Child and Adolescent Mental Health, 2004. **9**(1): p. 25-35.

158. ASU - School of Social and Family Dynamics - Family and Human Dynamics Research Institute. *HOME Inventory*. Available from: http://fhdri.clas.asu.edu/home/index.html.

159. Robins, D. *Official M-CHAT website*. Available from: http://www2.gsu.edu/~psydlr/DianaLRobins/Official_M-CHAT_Website.html.

160. *M-CHAT*. Available from: https://m-chat.org/index.php.

161. Robins, D.L., *Screening for autism spectrum disorders in primary care settings.* Autism, 2008. **12**(5): p. 537-556.

162. Eaves, L.C., H. Wingert, and H.H. Ho, *Screening for autism.* Autism, 2006. **10**(3): p. 229-242.

163. WPS Western Psychological Services. *Social Communication Questionnaire (SCQ)*. Available from: http://portal.wpspublish.com/portal/page?_pageid=53,70432&_dad=portal&_schema=PORTAL.

164. Hogrefe. *Social Communication Questionnaire (SCQ)*. Available from: http://www.hogrefe.co.uk/social-communication-questionnaire-scq.html.

165. Charman, T., et al., *Efficacy of three screening instruments in the identification of autistic-spectrum disorders.* The British Journal of Psychiatry, 2007. **191**(6): p. 554-559.

166. Chandler, S., et al., *Validation of the Social Communication Questionnaire in a Population Cohort of Children With Autism Spectrum Disorders.* Journal of the American Academy of Child & Adolescent Psychiatry, 2007. **46**(10): p. 1324-1332.

167. Corsello, C., et al., *Between a ROC and a hard place: decision making and making decisions about using the SCQ.* Journal of Child Psychology & Psychiatry & Allied Disciplines, 2007. **48**(9): p. 932-40.

168. Youth*in*mind. *SDQ - Information for researchers and professionals about the Strengths & Difficulties Questionnaires*. Available from: http://www.sdqinfo.com/.

169. Goodman, A., D.L. Lamping, and G.B. Ploubidis, *When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children.* Journal of Abnormal Child Psychology, 2010. **38**(8): p. 1179-91.

170. Goodman, R., et al., *Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample.* International Review of Psychiatry, 2003. **15**(1-2): p. 166-72.

171. Youthinmind. *Downloadable SDQs and related items*. Available from: http://www.sdqinfo.com/py/doc/b0.py.

172. Harris, F., *The first implementation of the Sure Start Language Measure*, 2002.

173. Harris, F., J. Law, and P. Roy, *The third implementation of the Sure Start Language Measure*, 2005.

174. Roy, P., H. Kersley, and J. Law, *The Sure Start Language Measure Standardisation Study. July 2004 - March 2005*, 2005.