Education Endowment Foundation

30th October 2026

Develop an outcome measure for evaluations of programmes and approaches for children under three

Request for Proposals

Closing date for proposals: 26th November 2026

Background and context

In September 2026, the EEF and DfE launched a partnership to strengthen the early years evidence base in some areas, including some not already covered by the EEF's existing endowment funding such as effective programmes and approaches for children aged 0-2. However, delivering on this priority is constrained by a lack of good outcome measures for the under-3[1] age range. We are therefore commissioning a research team to develop an outcome measure for education trials covering children under 3, which we would expect to be used in impact evaluations that we commission for this age range as part of the partnership.

Prime areas of development for under 3s

The Early Years Foundation Stage (EYFS) statutory framework[2] identifies three 'prime' areas as foundational to all early learning and development:

- Communication and Language
- Personal, Social and Emotional Development (PSED)
- Physical Development

These are distinguished from the four 'specific' areas – literacy, mathematics, understanding the world, expressive arts and design – because they are areas through which all other learning and development is enabled. For children under 3, development is focused on these prime areas. This is given statutory weight through the mandatory progress check between age two and three[3], wherein practitioners must review each child's progress and provide parents and/or carers with a short, written summary of development specifically in the prime areas.

At the same time the EYFS framework characterises the seven areas of learning – and the prime areas in particular for this age group – as interconnected and interdependent, with development in one area influencing development in the others. For example, communication supports social-emotional regulation, motor skills enable exploration, and early relationships shape cognitive growth[4]. Educators are consequently encouraged to take a holistic view when assessing a child's development rather than assessing progress in each domain independently.

Given this, programmes or approaches for children in this age range will likely identify one or several of the prime areas as their outcomes of interest. An evaluation that builds evidence on such an intervention would benefit from capturing progress across all three domains to reflect the EYFS's holistic view of the child's interlinked skills and to understand whether and how a programme's effects on development differ across the three prime areas.

At the same time the EYFS framework characterises the seven areas of learning – and the prime areas in particular for this age group – as interconnected and interdependent, with development in one area influencing development in the others. For example, communication supports social-emotional regulation, motor skills enable exploration, and early relationships shape cognitive growth[4]. Educators are consequently encouraged to take a holistic view when assessing a child's development rather than assessing progress in each domain independently.

Given this, programmes or approaches for children in this age range will likely identify one or several of the prime areas as their outcomes of interest. An evaluation that builds evidence on such an intervention would benefit from capturing progress across all three domains to reflect the EYFS's holistic view of the child's

interlinked skills and to understand whether and how a programme's effects on development differ across the three prime areas.

Rapid, uneven and non-linear development for under 3s

Development in the early years and specifically from 0-3 is both rapid and highly individual – children develop and learn at different rates and not always in a linear pattern. Key early years guidance reflects this in how it defines learning trajectories in this age range. Specifically, Development Matters[5], DfE's non-statutory guidance, sets out children's development across three broad age bands – Birth to 3, 3- and 4-year-olds, and children in Reception. It recognises that learning of young children is not neat and orderly and that "babies and young children do not develop in a fixed way. Their development is like a spider's web with many strands, not a straight line."

The sector-developed non-statutory guidance, Birth to 5 Matters[6], takes this further by organising development into six overlapping 'Ranges' spanning birth to 5. The guidance suggests prioritising ranges over fixed ages as children may developmentally be at different stages and each child's needs should be met according to their development and not their age. It also notes that because development is so rapid during the first three years, children move through the early ranges quickly. This means that children of similar ages can still be developmentally quite far apart, so a single age-labelled group like a baby room will typically include children spanning several different Ranges at once.

Taken together, outcome measurement within the 0–3 age range would need to consider the rapid, non-linear and uneven pace of development. We expect the measure to include age-differentiated content sensitive to the pace of change within narrow developmental windows. It would also need a scoring or analytic approach capable of capturing uneven, non-linear patterns of change over time.

Cross-domain measures for under 3s in EEF's EYMD

In EEF's recently updated Early Years Measures Database (EYMD), of the 248 measures currently catalogued, 35 cover the 0-3 age range. Of these, only 18 cover three or more EYMD domains and 17 specifically cover all three EYFS prime areas (Communication and Language, PSED, Physical Development).

A measure that is suitable for use in an evaluation would need to be sensitive to modest, programme-driven changes, have robust psychometric properties, be appropriate for use in England and be proportionate in administration time, cost, burden when used with children at scale. We reviewed the available measures against the requirements of an EEF trial (in Appendix 1) and concluded that none would be suitable for use in a trial in their current format. We summarise the strengths and limitations of the current measurement landscape for under-3s below:

Purpose

- All 17 measures were designed as diagnostic or screening tools rather than as research outcome measures. This does not rule them out as candidates for trial use, for example, ASQ-3 has been used as a secondary outcome in previous EEF trials with 2–4-year-olds (ParentChild+ and Maths Champions). However, their diagnostic purpose shapes how they are structured and scored which can limit their suitability for detecting programme effects (see Structure and scoring).

- Some measures were developed for a different population or purpose entirely and in turn are not fit for use or adaptation in the trial context - ERA assesses parent-child relational quality rather than child

developmental skill directly; PEDI is designed primarily for children with, or at risk of, disability; and VB-MAPP was developed and validated specifically for children with autism and language delays.

Constructs covered

- Coverage in the table in Appendix 1 does not confirm that a measure accurately covers the three EYFS prime areas. For example, ASQ-3 is described as covering five domains - Communication, Gross Motor, Fine Motor, Problem Solving, and Personal-Social – which have been mapped onto the three EYFS prime areas for the purpose of the table. However, ASQ-3's Personal-Social domain focuses on social and adaptive milestones (e.g. playing interactive games, imitating actions), which is not the same as self-regulation, a key construct in the EYFS PSED prime area.

Structure and scoring

- Binary thresholds and banded scoring formats often lack the granularity needed to detect a modest shift in outcomes between treatment and control groups. For example, when used as a secondary outcome in EEF's ParentChild+ trial, ASQ-3 (which uses banded scoring -yes/sometimes/not yet) showed ceiling effects and reduced score variation, particularly among higher-attaining children.

- Some measures offer a useful structural design for accounting for the rapid pace of development in this age range, where a single set of items is unlikely to be relevant or sensitive across the full range, for example, the items needed to capture a 6-month-old's development would differ substantially from those needed for a 30-month-old. ASQ-3, for example, uses 21 separate age-specific questionnaires spanning its full 0–5.5 years range, each with different, developmentally tailored item content for that narrow window.

Psychometric properties

- Across the psychometric properties of validity, construct validity[7] evidence is generally strong across the measures with 14 of the 17 rated moderate or strong. Criterion validity[8] evidence is considerably less consistent - only 11 of the 17 measures are rated moderate or strong[9].

- Across the psychometric properties of reliability, internal consistency[10] is well evidenced with 14 of the 17 measures rated moderate or strong. Test-retest[11] and inter-rater[12] reliability are less consistently evidenced - 10 of the 17 measures are rated moderate or strong on test-retest reliability, and 12 are rated moderate or strong on inter-rater reliability.

Relevance to England

- Very few existing cross-domain measures for under 3s were developed for use in England or are UK normed - of the 17 measures covering all three prime areas, only 2 were developed in England, and these are the same two that are UK-normed (Griffiths III and Schedule of Growing Skills). While UK population norming is not essential for a trial comparing treatment and control outcomes, the item content still needs to be culturally and linguistically appropriate for the target population as unfamiliar items could introduce measurement noise or floor effects unrelated to the programme's actual effect. A suitable measure would also need to have good construct validity in relation to the EYFS framework to be meaningful for the English context.

- Of the two UK-normed measures, standardisation sample evidence is either moderate or strong on both sample size and demographic representativeness. However, reporting of subgroup representativeness particularly for specific subgroups of interest for EEF trials (socio-economic

disadvantage, EAL and SEND status) is inconsistent. Schedule of Growing Skills' standardisation sample was directly compared against the general population across several relevant characteristics including ethnicity, English as an additional language, disability status, and parental qualifications and found to be broadly representative. By contrast, Griffiths III's standardisation sample is described only as representative of the UK and Ireland, with no reported breakdown by these characteristics.

Administration

- The measures covering this age range use a mix of parent/practitioner report (6/17), trained-observer assessment (6/17), and both combined (5/17). Parent or practitioner reports draw on observation of the child across many everyday contexts and over time, which may better reflect typical behaviour, but are subject to reporting bias. For example, in a trial context, practitioners would usually be aware of treatment status and those in the treatment group may have received training or support as part of the programme. This may lead them to rate children differently from educators in the control group. Trained-observer assessments offer more standardised, consistent scoring, but reflect the child's behaviour within a single assessment session which may not represent their typical functioning, particularly for very young children who may respond differently to an unfamiliar assessor. This is compounded by additional challenges specific to this age range - many children have limited receptive and expressive language, and in turn assessors must rely on observation, caregiver report, or structured play rather than verbal instruction. Children's engagement is also closely tied to caregiver presence and emotional security, and factors such as fatigue, hunger, or unfamiliar environments can substantially affect results[13]. Trained-observer measures like Griffiths III and Mullen also tend to require the highest qualification and cost.

- Some of the measures with the strongest sensitivity and psychometric properties also carry the highest qualification, time, and cost burden (e.g. Bayley-4, Griffiths III), making them less feasible to deploy across a trial-scale sample.

Taken together, our review supports commissioning either the adaptation of existing measure(s) or the development of a new cross-domain measure for under-3s that is fit for use in a trial context. As evidenced by the limited number of suitable measures, developing a robust measure for children this young is challenging. We therefore anticipate that by the end of this project, we will have developed an interim measure that is suitable for our first trial with this age group, but that further work may be needed to continue refining and validating the measure.

Considerations for measure development

We set out below some key considerations we expect the appointed research team to address in developing this new cross-domain measure for under-3s. This list is not intended to be exhaustive, and the research team's own expertise should inform the final design and approach. The research team is also advised to consult the EEF-commissioned user guide, Conducting Assessments with Infants and Toddlers (birth to three years).

Key considerations

- The measure should adequately cover the three EYFS prime areas - Communication and Language, PSED, and Physical Development, reflecting the child's interlinked skills and capturing how a programme's effects on development may differ across the three prime areas. The typical progression

across each prime area, as set out in guidance such as Development Matters and Birth to 5 Matters, may offer useful information for identifying the specific skills and content the measure should capture.

- The measure should be suitable for use in an impact evaluation, producing scores that can be compared over time and between treatment and control groups. Sub-scales of the overall scale will also need to be validated in its own right.

- The measure should be sensitive to modest, programme-driven change, capturing sufficient variation in outcomes across the full range of ability.

- The item content of the measure should reflect the rapid, uneven and non-linear development from 0-3. For example, this could be through age-differentiated item sets. Scores from different age-appropriate item sets would need to be comparable on the same underlying scale for children moving between age bands over the course of a trial. Additionally, because development at this age is not linear, the measure's scoring approach should be designed so that more sophisticated methods for analysing change over time can be used where needed. We are open to discussing the most feasible initial age range for the measure, given the time frame available and aims of the first use case for the measure (EEF funded under-2s trial).

- The measure should be designed and validated for a general early years population and should be culturally and linguistically appropriate for children in England. The research team should also consider what standard of UK-relevant standardisation evidence (sample size and demographic representativeness) the new measure should meet, given its intended use across trials in England more widely.

- We would like the research team to explore how the measure performs for children eligible for Early Years Pupil Premium (EYPP). We are also interested in the research team's views on the feasibility of exploring how the measure performs for children with EAL and SEND status.

- The measure should be validated on construct validity, criterion validity, and reliability (internal consistency, test-retest, and inter-rater), with evidence generated and reported for all properties. As a minimum, we would expect each property to meet the 'moderate' threshold from the psychometric appraisal criteria used in the EEF-commissioned EYMD[14]. Ideally the research team would assess all of these properties, but we are open to the team's views on where this may not be feasible.

- The research team should weigh the trade-offs between parent/practitioner report and trained-observer assessment considering factors such as accuracy, standardisation, cost, administrative burden and appropriate flexibility for this age group to determine what would be most suitable for trial purposes. Combining both approaches may also be worth considering, to establish a more accurate picture of a child's progress and development, provided the research team can determine an effective way to do so.

- The measure's administration time, cost, and assessor qualification requirements should be proportionate to deployment across a large trial sample at repeated timepoints.

The primary aim of this project is to develop an outcome measure for use in impact evaluations. All design decisions should prioritise this purpose. As a secondary consideration, and to the extent feasible within the project's timeline and budget, we would be interested in the research team considering whether the measure could also be used by educators for formative assessment, to aid their judgements at the statutory check or for informing decisions about targeted support. This is because it increases the return on investment if the tool

could be used widely in practice as well as research, could strengthen assessment practice in early years and over time research efficiency.

Approach

We are open to different approaches to meet the key objectives of the project, including either adapting existing measure(s) or developing a new one, provided the choice is well justified. If the research team proposes adapting existing measures, we would expect them to be responsible for ensuring they have the necessary permissions to do the adaptation and for the subsequent use of the measure. Any pre-existing intellectual property will remain with its existing owner, but any new or adapted materials developed in whole or in part through EEF funding will be subject to Crown copyright. We anticipate the work will involve defining the constructs to be measured, developing an initial item bank, conducting expert review and cognitive testing, field testing and psychometric calibration, and analysis.

Key outputs

We anticipate the outputs from this work to include:

Output 1: A published protocol pre-specifying the planned approach to meet the objectives of the project, including plans to develop an initial item bank, conducting expert review and cognitive testing, field testing and psychometric calibration, and analysis.

Output 2: A finalised measure ready for use in trials with under 3s and accompanying practical administration and scoring guidance.

Output 3: A technical report documenting the measure's development and full psychometric evidence base.

Budget

We have currently allocated a budget of up to £250,000 for this project. Where research teams expect to require additional funds to deliver the outputs above, this should be clearly specified in the proposal. Proposals will be assessed on their value for money. We expect the project to fall outside the scope for VAT.

We require applicants to provide a detailed budget for the project using the budget template provided (not included in word count).

Project Timeline

This work will begin in December 2026. Baseline testing for our first under-2s trial is expected to begin in September 2028, so a finalised measure, with administration and scoring guidance, must be ready by late May/early June 2028 at the latest to train administrators and deliver the anticipated baseline testing in October, there will be no room for delay. An earlier completion date would be preferable.

We have not proposed specific phases and timeline for this work and welcome research teams' proposal for how the project should be phased and sequenced. Proposals should include a detailed timeline covering the proposed activities, including construct definition, item bank development, expert review and cognitive testing, field testing and psychometric calibration, and analysis. In developing the phases and timeline, research teams should consider the following:

- We anticipate the measure will need to be administered at least twice: an initial field test to inform revision of items, followed by a final round of testing to generate psychometric evidence.
- We would welcome research teams' views on whether the timing of administration within the academic year matters, for example because of settling-in periods for new children in the autumn term.
- Where test-retest reliability is proposed, research teams should set out the proposed interval between administrations, given the pace of development in this age range.
- Where possible, the measure should be administered and scored during field testing by people similar to those who would administer it in a trial, rather than only by the measure developers.

In parallel with this project, EEF will select the delivery organisation and appoint an evaluator for the first under-2s trial. We expect the delivery organisation to be selected by April 2027 and the evaluator to be appointed by November 2027. The evaluator will need enough information about the measure to budget for and plan its administration, including recruiting and training test administrators with appropriate qualifications well ahead of baseline testing. The research team will therefore be expected to work closely with EEF during evaluator commissioning, and with EEF and the appointed evaluator during the trial set-up period. This will include producing an interim summary of the measure's design and administration requirements in time for evaluator commissioning in July 2027 and supporting the evaluator to train assessors ahead of baseline testing. Where the research team uses trained administrators for field testing, we would also be interested in whether they could support the trial evaluation.

Commissioning process and Timeline

Who should apply?

We encourage applications from research teams with measurement and psychometric expertise, including experience of developing or adapting, and validating outcome measures in the early years. We also expect the team to bring experience of assessing children in early years settings in England and/or expertise in child development from birth to 3. Applications may come from teams or from consortia that combine complementary expertise; in the latter case, applicants should demonstrate their ability to collaborate effectively and ensure the quality and efficiency of the project. This call for proposals is open – researchers do not need to be part of EEF's panel of evaluators. Please note that we accept applications only from legally constituted organisations, not from individuals.

What does the application process look like?

Interested research teams are invited to submit an in-depth proposal of approximately **4,000 words**, excluding references and CVs (indicative length, the focus should be on quality rather than quantity). The proposal should:

- Involve a **strong team** with the expertise set out above, including both measurement/psychometric and early years expertise. Please include full CVs of proposed research team members, highlighting relevant expertise, and summarise the team's experience in the proposal itself.
- Include a **detailed description of the key objectives** of the project.

- Provide a **detailed description of the proposed approach** to meeting the key objectives of the project, including a justification for developing a new measure or adapting an existing one and how the approach addresses the key considerations set out above.

- Provide a **clear plan for the design and delivery of the key outputs** of the project, including setting out what psychometric evidence can realistically be generated within the project timelines.

- Provide an approach to **research ethics, safeguarding, equity, diversity and inclusion (EDI) and data protection** when working with children, parents/carers and early years settings.

- Provide a **clear project management plan** for completing the project in achievable timelines, including:

  - An assessment of risks to the completion of the project and how they will be mitigated.

  - A timeline which includes key touchpoints at which the team will brief EEF on emerging themes/findings and the key outputs.

- Represent **good value for money**. Proposals must be clearly costed using the provided budget template and instructions, with costs aligned to the proposed plan and key costing assumptions set out.

Proposals will be assessed against the following criteria, with their respective weighting:

1. **Capability and relevant experience of core project team (35%)**

   a. Evidence of the research team having measurement and psychometrics expertise, including experience developing or adapting and validating outcome measures in the early years.

   b. Evidence of the research team having experience assessing children in the early years settings in England and/or expertise in child development from birth to 3. Consortia will need to demonstrate that the team can work well together to quality and efficiency of the project.

2. **Methodology and approach (40%)**

   a. The key objectives of the project are well defined and are appropriate, feasible and well justified.

   b. The proposed approach is well defined, and it is clear how the proposed approach will achieve the key objectives of the project. The approach is appropriate, feasible and well justified.

   c. The proposed plan for the design and delivery of the key outputs of the project is well defined. The plan is appropriate, feasible and well justified.

3. **Approach to research governance (10%)**

   a. The proposal identifies key risks to the project delivery and proposes appropriate strategies to mitigate these risks.

   b. The proposal describes the research team's approach to research ethics, to equity, diversity and inclusion (EDI), and to data protection (including but not limited to GDPR compliance, specifying legal bases for processing personal and special category data).

4. **Value for money (15%)**

    a. The proposal is clearly costed using the provided budget template and instructions. Costs are in line with the proposed plan for the design and delivery of the key outputs and relevant assumptions on costing model are provided.

    b. There is evidence that the proposal demonstrates value for money. The proposed budget demonstrates careful management of available resources, efficiency, and effectiveness to achieve desired project outputs.

The scoring scale (Table 3) will be used to assess the proposals against the criteria detailed above.

Table 3 – Scoring Scale	
0	Totally fails to meet the requirement - information not available
1	Meets some of the requirements with limited supporting information
2	Meets some of the requirements with reasonable explanation
3	Fully meets the requirements with detailed explanation and evidence
4	Exceeds the requirements with extensive explanation and evidence

What does the commissioning timeline look like?

Table 4 – Commissioning Timeline	
Activity	**Date**
EEF advertises Invitations to Tender (ITT) on website	30th October 2026
Deadline for proposals	26th November 2026 **(by 5pm)**
EEF appoints research team	2nd December 2026

How to apply?

Interested teams should submit their **proposals and budget template** via this link by **5pm on 26th November 2026**. If you have any questions, please contact Flavy Sharma (flavy.sharma@eefoundation.org.uk).

Appendix 1: Cross-domain measures for under 3s in EEF's EYMD

The table below covers shortlisted measures that cover all three EYFS prime areas for children aged 0-3. All measures in this table passed the following filters:

1. Covers all three EYFS prime areas (Communication & Language, PSED, Physical Development) – measures may cover additional EYMD EYMD domains beyond the three (e.g. Literacy, Mathematics).
2. Age range covers ages 0–3.

Measures covering all three EYFS prime areas for children aged 0–3														
Measure	**Domains**	**Age**	**Respondent**	**UK Norming (Y/N)**	**Construct validity rating**	**Criterion validity rating**	**Internal consistency rating**	**Test-retest reliability**	**Inter-rater reliability**	**Administration time**	**Prior qualification required**	**Cost**	**Tool type**	**Additional notes on suitability for EEF trials**
Ages & Stages Questionnaires (ASQ-3)	3 – C&L, PSED, Physical development	0–5.5yrs	Parents/caregivers	N	✓✓	✓	Not reported	✓✓✓	✓	10-15 min		$295 starter kit [Available to purchase online, no qualification check]	Developmental screening	Banded scoring (Yes/sometimes/not yet) built for referral flagging, so likely not sensitive enough to detect continuous, incremental programme driven change; has been used in previous EEF trials like ParentChild+ and Maths Champions - ceiling effects observed when using this measure in the ParentChild+ trial which confirms reduced sensitivity, especially for higher-attaining children and reduced score variation.
Battelle Developmental Inventory (BDI-3 Screening Test)	3 – C&L, PSED, Physical development	0–7yrs	Child (observation), Parent/caregiver (interview)	N	N/A	N/A	✓✓✓	✓✓✓	✓✓✓	10–30 min	1-day training	$1,297.76 for complete kit [Available to purchase online but requires self-attested qualification level]	Developmental screening	Items are scored on a 3-point scale (0/1/2), but the Screening Test specifically condenses this into a Pass/Refer classification based on a standard-deviation cut-off - the final output is binary, not continuous. Additionally, there is no evidence base to assess whether its reliably measuring the intended construct.
Bayley Scales of Infant and	3 – C&L; Literacy; Mathem	0–3.5yrs	Child (observation),	N	✓✓✓	✓✓	✓✓✓	✓✓✓	✓✓	30-70 min	Level B qualified assessor —	$1,191.60–$1,499.60	Clinical diagnostic	Strong psychometric properties and continuous, norm-referenced scoring (0-2 item

Toddler Development 4 (BSID)	atics, PSED, Physical Development		Caregiver (interview)								formal, supervised training specific to assessing children (via a relevant professional degree, e.g. psychology, occupational therapy, speech-language pathology)	[Available to purchase online but requires self-attested qualification level]		scoring system – 0=not present, 1=emerging, 2=mastery) unlike banded screening tools; But requires a Level B-qualified assessor and 30-70-min administration and is costly - not scalable for 1:1 use in trials.
Brigance Diagnostic Inventory of Early Development 3	3 – C&L, PSED, Physical development	0–8yrs	Child	N	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓	30-60 min		Not published online	Developmental screening / diagnostic	Diagnostic/screening tool built to flag developmental delay. Unlike banded or binary scored screening tools, it produces continuous standard scores, percentiles, and age equivalents from a norm-referenced item bank, preserving finer-grained variation between children.
Developmental Assessment of Young Children (DAYC-2)	3 - C&L, PSED, Physical development	0-5yrs	Child	N	✓✓✓	Not reported	✓✓✓	✓✓✓	✓✓✓	10-20 min/domain	Level C qualified assessor - a doctorate in psychology, education, or a closely related field	Kits $441; forms from $54 [Available to purchase online but requires self-attested qualification level]	Developmental screening / diagnostic	Item-level scoring is binary pass/fail (0/1), but these are summed and converted into continuous standard scores, percentile ranks, and age equivalents at the domain level.
Developmental Profile 4 (DP-4)	3 – C&L, PSED, Physical development	0–21yrs	Parent/Teacher	N	✓✓	✓✓✓	✓✓✓	✓✓	✓✓✓	20-40 min	Level B qualified assessor	£477 starter kit [Available to purchase online but requires self-attested	Developmental screening	Item-level responses are simple yes/no, but these convert into continuous, norm-referenced standard scores, percentiles, and age equivalents. Uses a single fixed item bank (190 items) with age-based start/stop rules and norm-referenced scoring across a birth-21yr span, rather

											qualification level]		than separate age-specific forms.
Infant-Toddler Developmental Assessment-2 (IDA-2)	3 – C&L, PSED and Physical development	0-3yrs	Trained observer, parent	N	Not reported	Not reported	✓✓✓	Not reported	✓✓✓	Not reported		Kit $737.00	Developmental screening	Does not produce a score or percentile at all – built instead around checklists of developmental items to build a picture of the child's current functioning and inform a care/service plan. Additionally, critical evidence of its psychometric properties is unreported.
Parents' Evaluation of Developmental Status–Developmental Milestones	3 – C&L, PSED and Physical development	0-7yrs	Parent or child	N	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓	~10-15 min		~$339–$369 starter kit	Developmental screening	Binary milestone checklist format (asks whether a child either has or has not reached a milestone).
Griffiths Mental Development Scale III	3 – C&L, PSED and Physical development	0-6yrs	Trained observer	Y UK Standardisation sample size: ✓✓ UK Standardisation sample characteristics: ✓✓✓	✓✓	✓✓✓	Not reported	✓✓✓	✓✓✓	60 min	Psychologist/ paediatrician/ allied health professional	Kit £1,690 excl. VAT [Only holders of an ARICD Certificate of Registration as User can purchase and use this measure]	Clinical developmental screening	Produces continuous developmental quotients, scaled scores, and percentiles per subscale; Is reported as can be used of monitoring development over time and measuring impact of interventions. But it requires a psychologist/paediatrician/allied health professional and 60-min administration - not scalable for 1:1 use in trials.
Merrill-Palmer-Revised (M-P-R)	3 – C&L, PSED and Physical development	0-6.5yrs	Child; parent/examiner for rating scales	N	✓✓✓	✓✓✓	✓✓✓	Not reported	Not reported	~45 min	Level C qualified assessor - a doctorate in psychology, education, or	Kit £1,195.00 [Available to purchase online but requires self-	Clinical diagnostic	Provides normed standard scores, percentiles, age equivalents. Also separately provides an additional scoring output – 'growth scores' which is built to track small,

											a closely related field	attested qualification level]		incremental changes in the same child over time.
Mullen Scales of Early Learning	4 – C&L, Mathematics, PSED, Physical Development	0–5.5yrs	Child	Y (partial)	✓✓✓	✓✓✓	✓	✓	✓	15 min (1yr) – 60 min (5yrs)	Degree in psychology/education + training	£1,326.36 excl. VAT [Available to purchase online but requires self-attested qualification level]	Clinical diagnostic	Produces continuous T-scores (mean 50, SD 10) per domain. Can detect fine grained differences between children.
Parent-Child Early Relational Assessment (ERA)	3 – C&L, PSED, Physical Development	0–5yrs	Trained observer and parent	N	✓✓✓	✓✓	✓✓✓	Not reported	✓✓✓	5-min interaction + 45–60 min coding	40hrs group training	Not specified	Clinical/relational assessment	This assesses parent-child relational quality (not child skill development directly) – not suitable.
Pediatric Evaluation of Disability Inventory (PEDI)	3 – C&L, PSED, Physical Development	0.5–7yrs	Parent/caregiver	N	✓✓✓	✓✓✓	✓✓✓	✓✓✓	Not reported	31-60 min		£170.94 (manual) + £65.10 (25 scoring sheets) excl. VAT [Available to purchase online but requires self-attested qualification level]	Clinical diagnostic	The measure is designed primarily for use in clinical and rehabilitative contexts, assessing functional and adaptive skills, often for children with, or at risk of, disability, rather than as a general-population developmental progress measure. Produces continuous scaled scores (0–100 or T-scores) explicitly designed to support monitoring of change over time for intervention planning.
Preschool Evaluation Scale (PES-2)	3 – C&L, PSED, Physical development	0–6yrs	Parent/guardian (Home); educational personnel (School)	N	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓	20 min		Kit $335; forms from $49 [Available to purchase online but requires self-attested qualification level]	Developmental screening	Converts raw scores to continuous standard scores and percentiles via a norm-referenced item bank.

Schedule of Growing Skills	3 – C&L, PSED, Physical development	0-5yrs	Child (observation)	Y UK Standardisation sample size: ✓✓ UK Standardisation sample characteristics: ✓✓	✓	N/A	✓✓✓	N/A	N/A	~20-30 min		£100	Developmental screening	Unverifiable scoring format. The evidence of its psychometric properties is weak or unreported.
The Ounce Scale	3 – C&L, PSED, Physical development	0–42 months	Parent, caregiver, professional	N	✓✓	✓✓✓	✓✓	Not reported	✓✓	Not reported		Forms from $34.40	Portfolio/ ongoing observational assessment	Structured as ongoing portfolio assessment, synthesised into a binary rating ('developing as expected' / 'needs development') at each of 8 defined age-level points - practitioners and parents gradually build up a 'portfolio' of observations, work samples, and anecdotal notes about the child as they naturally occur during everyday routines and interactions, which is then synthesised into the rating rather than derived from a single administration event. The binary format is likely to constrain sensitivity to modest, incremental change and the ongoing nature of the assessment would make it difficult to administer in a trial.
VB-MAPP	6 – C&L, Early Learning Environment, Literacy, Mathem	0–4yrs	Trained observer	N	✓✓✓	Not reported	✓✓✓	Not reported	✓✓	Not reported	Behaviour analysts	[Available to purchase online, no qualification check]	Criterion-referenced / ABA assessment	Criterion-referenced against a fixed checklist of curriculum steps, developed for Applied Behavioural Analysis (ABA) contexts (developed and validated specifically for children with autism and

	atics, PSED, Physical Develop ment													language delays) - used by practitioner to track a child's progress over time.