ANNEX B · RECEPTION AND KEY STAGE 1

# The evidence, the method, and the error we found in it.

**The instrument, the raw numbers on the scale they were actually collected on, the arithmetic mistake that put a wrong figure into everything we published, and the research we do and do not rely on.**

**Pocket Money Adventures CIC**
Company 16994988 · ICO ZC124930

This annex is built for scrutiny. It presents the raw method, the corrections and the limitations in full, including the instrument we have retired. If our numbers do not survive being checked by you, they were never worth having.

# Seven levels, and exactly where we sit on each

| What the evidence would need to show | Where we are | On what basis | How it gets answered, and when |
|---|---|---|---|
| **It can be delivered in a primary classroom** | SMALL FORMATIVE EVIDENCE | Story-led classroom delivery was completed in summer 2026, in Stoke-on-Trent and Staffordshire schools | Already answered for the Spark Session. **The seven-week sequence is answered by the first complete cycle** |
| **Children engage with it** | EARLY TEACHER-REPORTED EVIDENCE | Reported by class teachers in every setting, not observed by us | The weekly check-in now records participation every week rather than at the end |
| **They can perform the routine on cue** | FIDELITY CHECK, IN PLACE | Children perform Tap, Tap, Pause on cue at the end of the session. **This verifies teaching, not change** | Captured in every Spark Session from the next cohort |
| **They can tell a need from a want, and name save, spend and share** | BEING MEASURED NOW | Immediate outcomes, captured whole-class inside the session | **The 2026 to 2027 pilot.** Whole-class responses at the Spark and at week seven |
| **They pause before deciding, without the routine being named** | BEING MEASURED NOW | **This is the primary outcome.** Observed anchor scenes inside the session, not a teacher opinion | The same pilot. It is the outcome the whole design exists to test |
| **It persists to week seven** | BEING MEASURED NOW | Intermediate outcome. One teacher follow-up at one school so far, which is a signal and not a finding | The week seven recall check, across a full cohort rather than one class |
| **They use it in daily life without being asked** | EXPLORATORY | Teacher-reported signal only, and labelled as such wherever it appears | **Honestly, this may never be answerable at this age without a design we cannot yet fund.** We report it as a signal and refuse to convert it into a finding |
| **Sustained use of the decision routine over time** | IN THE THEORY, NOT MEASURED | An ultimate outcome in our Theory of Change, marked not measured so the distinction survives a screenshot | **A separately agreed longitudinal follow-up, with its own cohort, schools, funding and protocol. None of that is secured, so we promise no endline date** |
| **Lifelong financial behaviour and improved life chances** | A GOAL, NEVER AN OUTCOME | This is what the work is for. It is not something a seven-week programme can evidence, and we do not present it as one | It stays a goal. **Anyone claiming to have evidenced this from a primary programme is overclaiming** |

**Read the bottom three rows before you read any number we publish.** The line between what is being built and what is direction is drawn in our Theory of Change, not left to a reader to guess.

#### WHAT WE WILL TELL YOUR GOVERNORS

- Story-led classroom delivery was completed during summer 2026 and provides **small formative evidence** that the session, classroom discussion and short reinforcement can be delivered in practice
- **The Spark Session is the settled part.** The seven-week pathway has not yet completed a full end-to-end delivery cycle and its week-by-week outlines are being finalised, so we are among the first to run the complete sequence
- The 2026 to 2027 pilot outcome evaluation is what tests the immediate and intermediate outcomes
- A quality-mark submission is under assessment. Nothing has been awarded
- Teachers reported children using the routine without being asked to. That is an **exploratory signal**, not a measurement

#### WHAT WE WILL NOT TELL THEM

- Any effect size, or any comparison with an Education Endowment Foundation benchmark
- That anything is proven, evidence-based or relevant to an inspection judgement
- That the routine has become a habit, a reflex, or automatic
- That it changes real-world money behaviour, or prevents debt, scams or poor mental health later
- Any percentage of children, because our instruments measure a class
- **Any social return ratio.** We will not publish one before July 2027, and any figure quoted for us before then did not come from us

**Correction notice.** Earlier documents contained an incorrectly calculated percentage and a mislabelled domain. **Both have been withdrawn.** Annex B holds the raw instrument, the corrected calculations and the full method record.

# What was measured, by whom, and what it can carry

### WHAT IT IS

- **21 short items, in seven groups of three.** Each item scored 1 to 5
- So a group runs from 3 to 15, and the whole instrument from 21 to 105
- Completed by the **class teacher**, rating the class as a whole
- Administered three times: before the session, within 48 hours after, and again at six weeks
- Six groups ask about the children. **One asks the teacher about their own confidence**

### WHAT IT IS NOT

- **Not a test any child sat.** No pupil was assessed individually at any point
- Not blind. The same teacher rated before and after, and knew what had happened in between
- Not controlled. There is no comparison class and no comparison school
- Not independent. Nobody outside the classroom observed or scored anything
- Not a percentage of children, and it must never be printed beside a child count as if it were

**A separate three-question hands-up count exists**, taken by the teacher with the children in front of them. It is a rough classroom measure, it is affected by children copying each other, and at one school a different number of children were present for the two counts. We record it because it happened. We do not build anything on it.

## The cohort, in full

| School | Term | Classes | Year group | Outcome data |
|---|---|---|---|---|
| **School 1** | Summer 2026 | Two | Year 2 | Full: before, after, four weekly check-ins, six-week follow-up. **The historic figures in this annex** |
| **School 2** | Summer 2026 | One, mixed | Year 1 and Year 2 | Full, but on a different basis. **Cannot be compared with School 1 and is not pooled with it** |
| **School 3** | Summer 2026 | Three | Year 2 | Partial. A different rating scale was used and the standard instrument was not run |
| **School 4** | Summer 2026 | One, mixed | Year 1 and Year 2 | **None.** The follow-up was never returned. No report was written, because writing one would have meant inventing data |

Approximately 222 children across eight class groups in four schools received a Spark Session. Approximately 179 across six classes in three schools continued beyond it. **Earlier settings completed the activity agreed with them at the time, using differing combinations of Spark, reinforcement and follow-up. Completion of those earlier engagements does not mean that the exact School pathway now offered was delivered.** School 4 is outside the continuation count because the follow-up never came back. Schools are anonymised here; exact setting and class detail is retained in this annex and removed from the school-facing pack. **No Reception class has received the programme.**

# On the scale they were actually collected on

**HISTORIC RETIRED INSTRUMENT. NOT USED FOR CURRENT OUTCOME CLAIMS. These figures come from the 21-item questionnaire withdrawn in August 2026. They are printed so the correction can be checked, not as a result.**

**SOURCE, and the instrument this came from is now retired. One school, two Year 2 classes, around 73 children between them, rated as one cohort in a single return from the school's account. 21 short items each scored 1 to 5. Not a test any child sat, not 73 separate results, and not one result per class. Those classes were larger than the thirty-child unit we now sell. Cohorts from September 2026 use the four separate tools set out in our current pathway specification instead.**

**Financial knowledge** — **5 → 12 → 15** of 15

Need and want, save spend share, the routine itself

**Financial attitudes** — **7 → 11 → 13** of 15

Wanting to save, willingness to pause

**Decision-making** — **6 → 12 → 13** of 15

Using the routine, and giving a reason for a choice

**Financial vocabulary** — **6 → 12 → 12** of 15

Explaining money ideas in their own words

**Home and family** — **6 → 8 → 11** of 15

Teacher's expectation of family engagement

**Emotional wellbeing** — **8 → 12 → 14** of 15

Talking openly and calmly about money

Before · Immediately after, same day · Six-week teacher follow-up

**The six learning groups together: 38 of 90 before, 67 of 90 immediately after, 78 of 90 at the six-week follow-up.** On the teacher's own 1 to 5 scale that is a mean of **2.1, then 3.7, then 4.3** per item across those eighteen items. Across all twenty-one items, including the teacher-confidence group, it is 2.4, then 3.8, then 4.4. **Neither figure is quoted as a result for the current pathway.** Both are reproduced solely so the correction can be audited.

**The one we report separately, and why.** One of the seven groups, items 13 to 15, asks the teacher how confident *they* feel teaching this age group. It went 12 to 12 of 15 on the day and 14 of 15 at the follow-up. That is a real and useful result about the teacher. It is not a result about the children, so it does not sit with the six above and it never counts towards a headline. Our earlier documents published it as "impulse and spending behaviour", which it is not.

**What these numbers can carry.** Higher in all six learning groups immediately after the session. Six weeks later, higher again in five and level in one. Whole-class teacher judgement, one cohort, no comparison group, no independent observer, the same teacher before and after and therefore not blind. Both the before and after ratings were submitted on the same day, eleven days after the session, so they are retrospective rather than contemporaneous. It tells you a session landed in a classroom. It does not tell you what any individual child learned, and it cannot tell you what any child does with money.

# An arithmetic error, and what it cost us

**HISTORIC RETIRED INSTRUMENT. NOT USED FOR CURRENT OUTCOME CLAIMS.**

## What we got wrong, and the correction

Our earlier documents carried a headline: **59.5% before, 94.0% immediately after**. That figure is withdrawn. It is wrong, and the reason is arithmetic rather than interpretation, so it is worth setting out plainly.

**The instrument scores each item from 1 to 5.** Three items to a group, so a group runs from 3 to 15 and the whole instrument from 21 to 105. Every published percentage divided by 84, as though the scale ran 0 to 12 with a floor of zero. It does not. A teacher who rates every item at the lowest point still scores 21, not nothing.

Two consequences followed, and both were in print.

| What was published | What is true |
|---|---|
| **59.5% to 94.0%** | 50 of 105 to 79 of 105. Or, better, a mean teacher rating per item of 2.4 to 3.8 out of 5 |
| **A before figure compared against a six-week follow-up, at a second school** | **Two different denominators on the same instrument.** The before figure was computed out of 84 and the follow-up out of 105, so part of that apparent rise was the change of denominator rather than the children. **That comparison is withdrawn and the figures are not reprinted here**, because a corrected version of a number nobody should have published is still a number nobody should publish |
| **"Impulse / spending behaviour", already at 100% before delivery** | Items 13 to 15 are the instrument's teacher-confidence block. The first one reads "How confident are you in teaching financial literacy to this age group?" It was a teacher's self-rating published as children's spending behaviour |
| **"Scores rose across all seven areas"** | Six rose. The seventh, the teacher-confidence group, was unchanged on the day |

**What this means for a report about your classroom.** Where a single figure is useful, we give the mean teacher rating per item out of 5. Where a group is quoted, it is quoted out of 15, which is what the teacher's three answers can actually add up to. No percentage of children appears anywhere, because we have never measured a child.

# The timing, and what changed as a result

## The timing, which we also have to tell you

The before rating and the after rating were both submitted to the online form **on the same day, about two minutes apart, eleven days after the session**. The report of the time called that an administrative date. If a paper record was made on the day itself, we do not hold it. So this is a teacher's retrospective before-and-after rating rather than a contemporaneous one, and we describe it that way from now on. It is a weaker piece of evidence than we previously presented, and knowing that is worth more than the number was.

**What changed as a result.** No percentage of any kind appears on a cover, a certificate, a price sheet, a sponsor document or a funding application. **The instrument itself is retired** for cohorts from September 2026, and replaced by the four separate tools in our current pathway specification. The two schools that hold reports built on the old arithmetic have been sent a correction notice and a corrected replacement report, and the superseded versions are marked as superseded in our records.

**The general lesson, stated carefully.** It is not that a percentage is always wrong when a scale has a non-zero minimum. It is that **the earlier calculation divided by an incorrect denominator and did not account for the instrument's minimum**. Any transformed score we publish in future will state its formula, its scale range and how to read it. Raw counts and mean item ratings remain our preferred format, because they cannot be misread as a proportion of children.

## What a reader should take from this

**ONE**

The programme still did the same thing in the same classroom. What changed is what we are entitled to say about it, which was always the weaker part.

**TWO**

Our best-looking number is gone and nothing replaced it. If a competitor shows you a bigger figure, ask which scale it was collected on and who did the rating.

**THREE**

The two schools holding reports built on the old arithmetic have been sent a correction notice and a corrected replacement, with the date and recipient logged and the superseded version marked as superseded.

## How this was found

Not by us, in the first instance. An outside reader looked at the seven published domains and asked why one of them was already at its ceiling before the session. Chasing that question back to the raw form is what exposed the mislabelled group, and chasing the mislabelled group back to the scoring is what exposed the denominator.

**The general point, which matters more than this particular error.** A number that nobody can trace back to a raw record is not evidence, however carefully it is presented. Every figure in this pack now has a path from the form a teacher filled in to the sentence you are reading. If you want to walk that path with us, ask, and we will show you the working rather than the summary.

# Every item, so you can check the labels yourself

**HISTORIC RETIRED INSTRUMENT. NOT USED FOR CURRENT OUTCOME CLAIMS.**

The instrument below was used for cohorts up to July 2026 and is **retired**. It is printed in full because the mislabelling in section three above cannot be checked without it, and because a provider that asks you to trust a domain name should show you the question behind it.

**21 items, seven groups of three, each scored 1 to 5** by the class teacher for the class as a whole. Group range 3 to 15. Instrument range 21 to 105.

| # | Group, as the app names it | Published as | Status |
|---|---|---|---|
| **1–3** | Financial Knowledge | Need vs want recognition | **RETIRED.** Replaced by class task Q1 and Q2 |
| **4–6** | Financial Attitudes | Tap, Tap, Pause awareness | **RETIRED.** Attitude items removed; not directly observable |
| **7–9** | Decision-Making | Save / spend / share | **RETIRED.** Replaced by class task Q2 and Q4 |
| **10–12** | Financial Vocabulary | Value of money | **RETIRED.** Replaced by class task Q3, which asks for the reason |
| **13–15** | **Teacher Capability.** Item 13 reads "How confident are you in teaching financial literacy to this age group?" | **Impulse / spending behaviour** | **THE MISLABELLING.** A teacher's self-rating published as children's spending behaviour. Moved to the implementation tool where it belongs |
| **16–18** | Home & Family. Item 16 asks how many families the teacher **expects** to engage | Home / parental engagement | **RETIRED.** An expectation, not an observation. Nothing replaces it, because we cannot measure it |
| **19–21** | Emotional Wellbeing | Financial confidence / agency | **RETIRED.** Moved to the observation record, labelled teacher-reported |

**Two of the seven groups were never about the children at all.** Items 13 to 15 asked the teacher about themselves. Items 16 to 18 asked the teacher what they expected families to do. Both were being added into a total presented as a learning result. Separating them is the single biggest reason the instrument was retired rather than repaired.

The replacement is set out in our frozen pathway specification: a direct class task, an implementation and usability check, and a labelled observation record, reported in three separate sections and never combined.

# What is captured, when, by whom, and what it can carry

**CURRENT INSTRUMENTS. In use for the 2026 to 2027 pilot outcome evaluation. Printed in full so they can be inspected before they are trusted, which is the opposite of how the retired one was handled.**

Measurement sits **inside** delivery. There is no separate visit, no researcher in the room, and **nothing that identifies a child leaves the classroom.** Four things are recorded, and they are kept apart on purpose.

1 · WHOLE-CLASS CAPTURE

### Inside the session, and at week seven

- The same short questions at the start and end of the Spark, giving your class its own before and after
- **Observed anchor scenes** for the primary outcome: does a child pause before deciding, **without the routine being named in the question**
- The recall check at weeks six to seven

**Whole-class responses only.** No child is named, marked or ranked, and nothing goes on a child's record.

2 · FIDELITY AND DOSAGE

### Did it actually run

- Can the children perform the routine on cue at the end of the session. **This verifies teaching, not change**
- Which weekly sessions ran, and which did not
- How long each actually took, against the five minutes of preparation we specify
- Participation, adaptations, and any wording you had to explain

This is where our two-hour figure stops being a design estimate.

3 · EXPLORATORY SIGNALS

### Labelled, never counted

Vocabulary you heard, any use in class without the routine being named, and anything a family mentioned. **Family Mission returns are not a measurement and are not reported as one**, because a returned slip represents only the families who chose and were able to return it. None of this is converted into a figure or merged with the capture above.

4 · THE TEACHER OUTCOME

### Held distinct

One short rating before and after, recorded against the class: do you enjoy teaching about money, and do you feel more confident delivering it inclusively. **It is a real outcome and it is about you, not the children**, so it is never added into a learning total.

**Why the separation matters.** Our previous instrument put children's learning, the teacher's confidence and expected family engagement into one twenty-one item questionnaire and produced a single number. It could not tell those things apart. **It has been retired rather than repaired**, and Annex B prints it item by item so you can see why.

## Administration

| Field | Specification |
|---|---|
| **When** | **Baseline before delivery, a capture at the start and end of the Spark Session, weekly check-ins, and a recall check at weeks six to seven.** Free continued-learning packs then carry the routine across the rest of the term |
| **Where** | **Inside delivery.** No separate data-collection visit, no researcher in the room, and no information that identifies a child leaves the classroom |
| **Who administers** | The facilitator for the in-session capture and the anchor scenes; the class teacher for the weekly check-ins and the recall check. **Neither is blind to the delivery stage, and we say so rather than implying independence** |
| **Unit** | **The class.** Whole-class responses and observed scenes. No child is named, marked, ranked or given a record with us |
| **Response methods** | Standing and sitting, hands, cards, pointing, partner talk, or an equivalent method agreed with the teacher. **A child who takes part in any way is counted as responding.** A child who does not is counted as not responding, never as wrong |
| **Attendance** | Recorded every time. **Counts are never presented as though the same children answered on each occasion**, and where attendance differs materially the report says so beside the figures rather than in a footnote |
| **The primary outcome** | Whether a child pauses before deciding, without the routine being named, in an observed anchor scene. **An observed scene, not an adult's opinion of a child** |
| **The fidelity check** | Can the children perform the routine on cue at the end of the session. **This verifies teaching, not change**, and is reported as an output rather than an outcome |
| **Family Mission returns** | **A signal from the families who return them, never a rate of family use.** Never converted into a percentage of families |

**What these instruments cannot do.** They are not blind, not controlled and not independent. Every count is a count of a classroom moment with adults present. **They can show whether a class can do something at three points in time. They cannot show that the programme caused it**, and nothing in the design pretends otherwise.

# What we think changes, in what order, and which parts are measured

Every programme should be able to say what it thinks it changes, in what order, and which parts of that it has actually measured. Ours is one page, dated 4 August 2026, and it uses a deliberate typographic rule: **upright text is measured in the 2026 to 2027 pilot; anything marked not measured is in the theory and is not being claimed.** The rule exists so the distinction survives a screenshot.

| Stage | What it says |
| --- | --- |
| **Activities** | A Spark Session of 60 to 75 minutes inside a slot of up to 90, supported by the Ava book series, Sidekick Cards, Pulse micro-activities, Family Mission parent packs and Mini-Mission cards |
| **Outputs** | Reach, delivery, **a fidelity check** that children can perform the routine on cue, dosage logged weekly, Family Missions distributed, and delivery quality. **The fidelity check verifies teaching, not change** |
| **Immediate outcomes** | **Primary: children pause before deciding, without the routine being named in the question, in a money choice.** Children can tell a need from a want. They can name save, spend and share and give an example of each. They can say what they would wait and save for. They feel more confident talking about money. **Teachers enjoy teaching it and feel more confident delivering it inclusively, held distinct as a teacher outcome** |
| **Intermediate outcomes** | Persistence to week seven. Use in daily life without being asked, **exploratory and teacher-reported**. Improved financial knowledge, **derived from the three knowledge outcomes and not separately measured**. Families talk about money at home, **returner-only and exploratory** |
| ***Ultimate outcomes*** | ***Not measured.*** *Children have strengthened financial decision-making, and sustained use of the decision routine over time.* ***Repeated classroom practice may support familiarity, recall and later application. Whether it produces sustained or real-world behaviour change is not established*** |
| ***Goals*** | ***Not measured. A goal, never an outcome.*** *Children and young people have positive lifelong financial behaviour and improved life chances, avoid future debt, and experience less financial inequality* |

**Measurement sits inside delivery.** Everything measured is captured at one of three points inside the programme: whole-class responses and observed anchor scenes in the Spark Session, the week seven check-in, and one slip on a Family Missions, plus a short teacher rating and a one-off school context form. **No separate data collection visit, and no information that identifies a child leaves the classroom.**

**How to read it.** The pathway is logical, not chronological. Children learn at different paces, so an outcome may be reached at any point rather than in the order the page happens to list them.

# What this is ultimately for, and the limit we hold

## Why this age, and what it is ultimately for

Schools and funders reasonably ask what the point of all this is in twenty years. Our answer has a shape and a limit, and the limit is the part most providers leave out.

### THE FOUR THINGS THIS IS AIMED AT

- **Problem debt in adulthood.** Clients reaching debt advice typically carry unsecured problem debt in the order of £15,000 to £17,000, on StepChange's published figures
- **Scam victimisation.** UK Finance reported authorised push payment fraud losses of £450.7m in 2024. A pause before a purchase is a behavioural precursor, nothing more
- **Money-related mental health.** The Money and Mental Health Policy Institute finds people with long-term mental health problems are **nine times** more likely to have struggled financially
- **Lost productivity and life chances.** Quantified in Resolution Foundation, Institute for Fiscal Studies and Financial Conduct Authority work on chronic financial stress

### AND THE LIMIT, STATED FIRST NOT LAST

**We do not publish a social return ratio.** We will not publish one until longitudinal data exists and the study, evaluator, funding and protocol are secured, which on our current plan is **not before July 2027**. Any multiplier quoted for us before that date is not authorised by us and did not come from us.

**We do not claim to prevent any of the four harms above.** The claim is directional and it is a contribution claim: money habits begin forming in early childhood, our age band sits inside that window, and that is a reason to work here rather than evidence that the work succeeded.

We cite no international cohort evidence, because none of it was carried out on this programme, this age group or in a British primary classroom., and **we do not assume it transfers cleanly to a British primary school.** Closing that gap is exactly what the longitudinal arm is designed to do.

**The honest version of the vision.** A seven-week programme cannot evidence a changed adult. What it can do is teach a routine at the age the literature says the foundations form, measure whether children can use it, and be truthful about the distance between those two things. **The distance is the work.**

**Where we stand with quality marks, in full, because there are three.**

**Financial Education Quality Mark, Young Enterprise.** A full assessor pack was submitted in August 2026: the book, the Intervention Manual, the teacher materials, the thirty-card Sidekick library, the seven-week roadmap, the logic model and four evaluation documents. **It is under assessment. Nothing has been awarded.**

**PSHE Association and the Association for Citizenship Teaching.** A separate resource, four teacher-led Key Stage 1 lesson packs of about 40 to 45 minutes each, went to both. **Those submissions expressly exclude the Spark Session and the seven-week programme**, because a facilitated session is a service rather than a teaching resource.

**Until an award is made and its scope confirmed in writing, we do not describe the programme, the organisation or the seven weeks as quality marked, accredited or approved.** If a mark is awarded we will say which body gave it and exactly which resource it covers.

# And, in every case, the limit of it

Pocket Money ADVENTURES

### Why four to seven

A review commissioned by the Money Advice Service, now the Money and Pensions Service, concluded that the cognitive foundations of later financial behaviour, including self-control, working memory and the ability to apply simple rules, are typically in place by around age seven.

**It is a review of existing developmental research, not a study of any programme. The popular shorthand that money habits are fixed by seven overstates what it found, and we do not use it.**

Whitebread, D. and Bingham, S. (2013) *Habit Formation and Learning in Young Children*. University of Cambridge, for the Money Advice Service.

### Why a pause

Inhibitory control, the ability to stop and think before acting, is one of the core executive functions. Diamond's 2013 review finds that some sustained practice regimes improve executive function in children.

**That is a general educational principle about creating time to think before acting. We have not established that Tap, Tap, Pause improves executive function or inhibitory control, and we should not be read as saying so.**

Diamond, A. (2013) Executive functions. *Annual Review of Psychology* 64, 135–168.

### Why the money in a child's world is harder to see

A 2024 meta-analysis in the *Journal of Retailing* pooled 392 effect sizes from 71 papers, covering 11,257 participants, 338,513 transactions and 17 countries. It found a small but reliable tendency to spend more when paying by card or phone than with cash, stronger for purchases that signal status and absent when donating or tipping.

**It is a finding about adults, and the effect is small. We use it to explain why an exchange a child can watch has largely disappeared, not as evidence about children.**

Schomburgk, L., Belli, A. and Hoffmann, A.O.I. (2024) Less cash, more splash? A meta-analysis on the cashless effect. *Journal of Retailing* 100(3), 382–403.

### Why now

Following the Curriculum and Assessment Review, the government has said citizenship will become statutory at key stages 1 and 2, which is school Years 1 to 6, with financial education inside it. The final national curriculum is expected in spring 2027, for first teaching from September 2028.

**The detailed programme of study has not been published. Nobody yet knows what primary financial education will be required to contain, us included, and any provider telling you their scheme already meets it is guessing.**

Department for Education, government response to the Curriculum and Assessment Review, 2025.

### On timelines, since we used to quote one

We used to print that habits take about 66 days. The study behind that number followed 96 adults who each chose one daily eating, drinking or activity behaviour. Estimated time to near-maximum automaticity varied enormously, from 18 days to 254.

**Adults, self-chosen, daily, in one setting. It sets no timeline for a child, for a weekly classroom activity, or for this programme, and we have stopped citing it as though it did.**

Lally, P., van Jaarsveld, C.H.M., Potts, H.W.W. and Wardle, J. (2010) How are habits formed. *European Journal of Social Psychology* 40(6), 998–1009.

**What we do not cite, on purpose.** No effect size, no Education Endowment Foundation comparison, and no claim that any of the work above was carried out on our programme, our age group, or in a British primary classroom. **Two of the four are adult behavioural studies. The other two are broader developmental reviews, not evaluations of anything we do.** None of them is evidence that Tap, Tap, Pause works. They explain why the design is plausible.

# Where every piece of evidence in our documents comes from

Pocket Money ADVENTURES

An anonymised school can still be identified from a year group, a date, a cohort size and a bit of local context. So this ledger records where each piece of evidence came from, whether we have permission to use it, and where it may appear.

| Evidence | Source | Permission | Where it may be used |
|---|---|---|---|
| **Class ratings, 21 items** | Teacher submission, dated form record | HELD Under the school agreement | Anywhere, anonymised, with the limitations attached |
| **Hands-up counts** | Teacher tally in class | HELD | As a rough classroom measure only, never as a headline |
| **Teacher observations in weekly check-ins** | Teacher's own written submission | HELD Anonymous by design | Anonymised, attributed by role, described as reported not observed by us |
| **Two parent-relayed examples that reached us with a child's name attached** | Relayed to us by the class teacher | NOT NEEDED | **Anonymised permanently.** No name, no year group, no school. Not used in sponsor or sales material. **A permission would not change this, because we do not name children** |
| **Named teacher quotations** | Signed permission register | HELD for four named people | Only where the register records permission. **We do not name teachers in marketing** |
| **School names** | — | NOT USED | **No school is named in any outward document.** No reach numbers in marketing |
| **Children's names** | — | NEVER | **Never in any learning, reporting or sponsor context. A safeguarding, incident or complaints record may identify a child where that is necessary** |
| **Photographs of children** | — | NOT TAKEN | None taken unless a school has gathered permissions and agreed it in writing first |

**One thing on the record.** Two examples reached us with a child's name attached, in a teacher's own message. Those names were stripped before anything was written and they will never appear. We asked the school on 15 June 2026 whether we could use the examples without surnames and received no reply, so they are used in anonymised form in reports only, never in sponsor or sales material. **A written permission would not unlock a name.** Nothing does.

## References


1. Whitebread, D. and Bingham, S. (2013) *Habit Formation and Learning in Young Children*. University of Cambridge, commissioned by the Money Advice Service, now the Money and Pensions Service.
2. Diamond, A. (2013) Executive functions. *Annual Review of Psychology* 64, 135–168.
3. Schomburgk, L., Belli, A. and Hoffmann, A.O.I. (2024) Less cash, more splash? A meta-analysis on the cashless effect. *Journal of Retailing* 100(3), 382–403. 392 effect sizes, 71 papers, 11,257 participants, 338,513 transactions, 17 countries.
4. Lally, P., van Jaarsveld, C.H.M., Potts, H.W.W. and Wardle, J. (2010) How are habits formed: modelling habit formation in the real world. *European Journal of Social Psychology* 40(6), 998–1009.
5. Department for Education, government response to the Curriculum and Assessment Review, 2025. Citizenship becomes statutory at key stages 1 and 2, with financial education inside it. Final national curriculum expected spring 2027, first teaching from September 2028.

**We publish no effect size of our own, make no Education Endowment Foundation comparison, and cite no study we have not read.** The effect-size count in reference 3 belongs to that meta-analysis and describes its own scope. Where a source is about adults, we say so. Where it does not apply to us, we say that too. Every claim in every document we publish traces to a line in this annex or to a dated record we hold.

## How to check us

### ASK FOR THE RAW FORM

The twenty-one items, in the wording the teacher saw. It is the only way to tell whether a group label describes what the question actually asked. Ours did not, in one case, and that is how it was found.

### ASK FOR THE SCALE

Every percentage has a denominator and somebody chose it. Ask what the lowest possible score is. If the answer is not zero, a percentage of the maximum is misleading and should not be used.

### ASK WHO RATED IT

A teacher rating her own class after teaching it is not blind and is not independent. That does not make the rating worthless. It makes it one teacher's judgement, which is what we say it is.

Those three questions work on any provider, and they are the ones we would rather a school asked us than took our word for.

**Before this annex is shared widely with funders, sponsors or academic contacts, it needs one signed external methods note** covering the retired-instrument correction, the direct class task, the counting rules, the missing-data treatment, the evidence ladder and the wording of the claims. **That review has not yet happened, and until it does this annex is our own account of our own method.**

## What would move us up the ladder, and what it would take

| The next rung | What would have to happen |
|---|---|
| **Children can explain their reasoning** | Question 3 of the class task run across a full cohort, with one administration by an adult who did not teach the lessons |
| **They still hold it after a delay** | The week 7 checkpoint run as specified, then a further check a term later. **Nobody has funded a longer follow-up, which is why it has never happened** |
| **They apply it to a new scenario** | Question 4 run at every administration rather than only at the end, across more than one school |
| **Anything above that** | A comparison group, an independent observer and a controlled design over years. **We are not close, and we will say so for as long as it is true** |

**If a sponsor renews, the most useful thing that money buys next is a longer follow-up in the same class, not a second classroom.** We can already show that a session lands. What nobody has shown, including us, is whether anything is still there two terms later.

hello@pocketmoneyadventures.co.uk · company 16994988 · ICO ZC124930
Annex B