KINEDU · RESEARCH REPORT

Coupling and Differentiation in Early Development

A replication at population scale, and a psychometric characterisation of the Kinedu Assessment. Findings from 470 million milestone observations across 3.5 million children.

Date 29 July 2026 · revised 4 August 2026

Data Kinedu production records, January 2020 – December 2023

Instruments Kinedu Initial Assessment (structured, single administration) and post-activity milestone questions (longitudinal)

Revision, 4 August 2026. Two corrections since the 29 July edition. The analysed cohort is stated as 3.5 million children; the earlier “4.7 million” was the size of the dates-of-birth export rather than the children whose answers were analysed. And the C1 replication column now reads “consistent with the internal validation study”; it previously described that study as independent. No result, figure or conclusion changed.

Abstract

Stenhaug, Ram and Frank, analysing 21,861 children from Kinedu application data, reported that developmental domains are tightly coupled during the first year of life and separate thereafter. This study reproduces that finding on Kinedu’s full production records — approximately 160 times as many children — and characterises the psychometric properties of the underlying instrument.

Four principal results are reported. First, the expected ages assigned to every milestone match the empirically observed 75th percentile of acquisition with a median error of zero months, confirming at population scale a validation previously resting on 2,135 caregivers. Second, within-child coupling between developmental areas approximately halves across the first year — from 0.489 to 0.265 — and then settles approximately a quarter below its first-year level, remaining depressed through thirty-five months; the pattern replicates at two-, four- and six-month measurement intervals. Third, a single assessment provides real but age- and horizon-dependent information about a child’s subsequent developmental standing. In the main analysis, the association reached $r = 0.61$ (37% of variance); stratified analyses show weaker associations from early infancy across longer follow-up. The same assessment predicts rate of progress almost not at all ($r = 0.12$, 1.5%), establishing standing and movement as distinct quantities. Fourth, caregiver response style is measurable and accounts for approximately 11% of a child’s apparent standing, but for only about 5% of the observed correlation between developmental areas.

A methodological distinction proves decisive throughout: comparisons *between* children and comparisons *within* a child answer different questions and, on the question of differentiation, return opposite answers. Only the latter isolates development from stable individual differences.

Contents

1 · Claims and evidence

Each result is stated below with a certainty grade. Grades are assigned against four explicit criteria.

GRADE	SAMPLE	STATISTICAL SUPPORT	ROBUSTNESS	INDEPENDENT REPLICATION
A	>1M children	Effect confirmed by bootstrap interval excluding zero, or direct enumeration	Survives every specification tried	Reproduced on a second instrument, dataset or published study
B	>1M children	Same	Direction robust; *magnitude* depends on an analytic choice	One or the other, not both
C	Any	Effect statistically supported	Measurable only on a subset that may not represent the whole	Single design
D	Any	Not supported, or not testable with available data	—	—

All four columns must be satisfied for a grade. A claim meeting the A sample and statistical bar but failing robustness or replication is graded B. Grades describe the strength of the evidence, not the size of the effect: C7 is graded C despite a highly reliable measure, because the population it can be measured on is not representative.

REF	RESULT	CERT.
C1	**Expected ages match the empirically observed 75th percentile** Across eight candidate percentile thresholds, only the 75th produces an error centred on zero and symmetric between early and late. 57% of milestones land within one month.	A
C2	**The instrument discriminates properly from birth to four years** 93% of in-scope milestones reach the 75th percentile; the median acquisition curve runs from 26% to 90% of children answering affirmatively.	A
C3	**Developmental areas couple early, decouple through the first year, and remain decoupled** Within-child coupling roughly halves at the transition — 0.489 to 0.265 — then settles about a quarter below its first-year level and stays there through thirty-five months.	B
C4	**A single assessment carries useful, age- and horizon-dependent information about later developmental standing** In the main analysis, the assessment general factor predicts subsequent standing at r = 0.61, accounting for 37% of variance. Stratified analyses show weaker associations from early infancy across longer follow-up.	A
C5	**A child's starting point does not determine their rate of growth** The same factor that explains 37% of standing explains 1.5% of growth. Rate of progress is nearly independent of where a child began.	A
C6	**Physical skills arrive earliest and cluster tightest; social-emotional arrive last and vary most** Ordering of the four areas by typical age of acquisition is stable; the spacing between them depends on how right-censored curves are handled.	B
C7	**Roughly 11% of a child's apparent standing reflects how generously their caregiver answers** A caregiver-response measure built from items far above a child's level predicts apparent advancement at r ≈ 0.33, but accounts for only about 5% of the correlation between developmental areas.	C

Evidence supporting each grade

REF	SAMPLE	STATISTICAL SUPPORT	ROBUSTNESS	REPLICATION
C1	3,518,779 children · 280,595,197 answers	Median error ±0 mo; symmetric 24% early / 49% late	Eight thresholds tested; all others systematically biased	Consistent with the internal validation study (N=2,135)
C2	407 milestones · 48 skills	Only 1 of 407 fails to reach the 50th percentile	Stable under sample-size floors from 100 to 500 answers per cell	Curve shape consistent across all four developmental areas
C3	3,518,779 children (primary) · 1,758,085 (interval-controlled)	Peak-to-trough drop 0.224, 95% CI [0.205, 0.242], $p<0.0001$	Replicated at two-, four- and six-month measurement intervals	Reproduces Stenhaug, Ram & Frank (2021) in shape, timing and persistence
C4	1,109,370 children · assessment reliability 0.956	$r = 0.51–0.76$ across ten follow-up windows	Holds in every follow-up band from 2 to 48 months	Measured on an instrument independent of the one used at baseline
C5	1,109,370 children	$r = 0.12$ for growth against $r = 0.61$ for standing — 25× in variance	Partialling the factor out moves growth-coupling by –1%	Same decomposition holds in every age band tested
C6	407 milestones	Ordering invariant across specifications	Spacing varies with censoring treatment	Directionally consistent with the Stanford analysis
C7	370,767 children · style-measure reliability 0.898	$r = 0.32–0.35$ per area; coupling 0.728 → 0.694 when removed	Measurable only on the 11% of children receiving such items	Single design; not independently replicated

2 · Methods

Data

Two production tables were exported in full and parsed to columnar binary arrays. The first records milestone questions posed after an activity, one to three at a time, and is longitudinal: a child contributes observations repeatedly as they grow. The second records the Kinedu Initial Assessment, a structured instrument administered in a single sitting with a median of 32 items, and is cross-sectional: of 3,384,823 assessed children, 3,384,584 have exactly one administration. A third export supplied dates of birth (4,678,527 records containing only an identifier and a date; no identifying fields were requested or received).

Analysis covers records from January 2020 through December 2023, the period across which the instrument's response distribution is stable. Within that window, analysis is further restricted to the 48 validated skills carrying an established developmental-area mapping, and to milestones with an expected age of 48 months or less — the range over which the instrument was validated. This yields **280,595,197 observations across 407 milestones and 3,518,779 children** in the longitudinal instrument, and 189,419,544 observations across 3,300,095 children in the assessment.

Child age was taken from the stored value where present. Where derivation from date of birth and timestamp was required, the derivation was validated against the stored value on 48.2 million records: 100% coverage, median difference zero months, 92.5% within one month — the residual dispersion being the expected consequence of month-level timestamp granularity.

Scoring

Raw affirmative rates confound a child's development with the particular questions they were asked. Each observation is therefore expressed as a **residual**: the response (0 or 1) less the population affirmative rate for that specific milestone at that specific child age. A child's score in a developmental area is the mean residual across that area's items, which removes the effect of a changing question mix.

Correction for measurement reliability

Correlations between imperfectly measured quantities are attenuated toward zero, and where measurement precision varies systematically with age an artefactual trend results. Each area's items were therefore partitioned into two halves, alternating by item within area. Every area score was computed twice, once per half; the correlation between them, corrected by the Spearman–Brown formula, estimates that score's reliability. Reported correlations are disattenuated by the geometric mean of the two reliabilities involved.

This correction operates *against* the principal finding of C3 rather than in its favour: reliability rises with age (0.73 → 0.81), so the correction inflates later correlations more than earlier ones. Raw and corrected measures nonetheless yield the same 45–46% reduction.

Thresholds and inference

A milestone is taken to reach a percentile threshold at the first age where its affirmative rate attains it *and* the following three ages remain within six percentage points, which prevents a single noisy cell from registering as acquisition. Age cells containing fewer than 200 observations are discarded; milestones retaining fewer than six cells are excluded. Confidence intervals derive from bootstrap resampling of children (400–500 replicates), never of observations, since observations within a child are not independent.

C1 · Expected ages

Certainty **A** **Very high** 280,595,197 answers · 407 milestones

Each milestone carries an expected age, assigned at the 75th percentile of acquisition following American Academy of Pediatrics practice. That assignment derives from an internal validation study of 2,135 caregivers. It can be tested directly: for every milestone, the age at which the observed affirmative rate first crosses a given percentile is compared against the stored value.

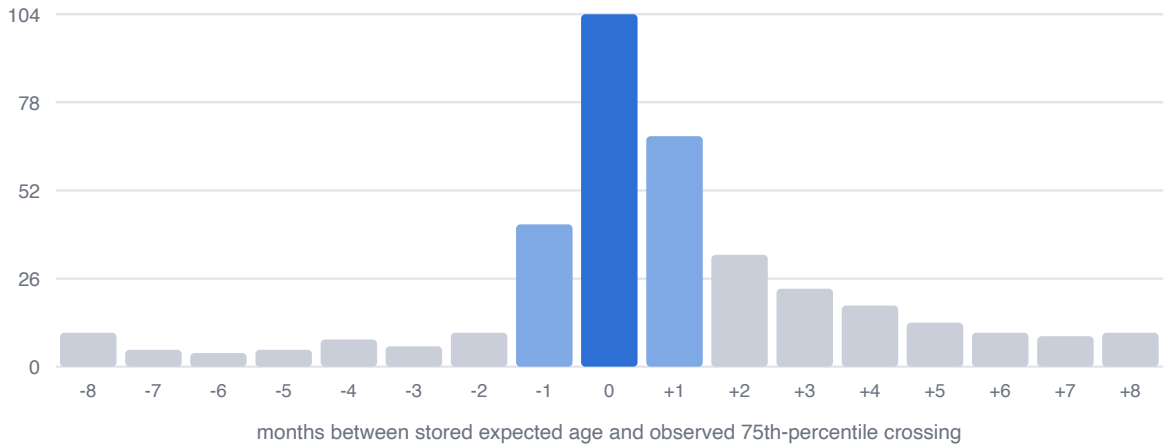

Figure 1. Difference between each milestone's stored expected age and the age at which 75% of children are observed to have acquired it. Centred on zero and symmetric. End bars pool everything beyond ±8 months.

THRESHOLD	MILESTONES REACHING IT	MEDIAN ERROR	WITHIN ±1 MO	EARLY / LATE
50th percentile	406	−2 mo	42%	89% / 1%
70th percentile	607	−1 mo	43%	57% / 23%
75th percentile	**378**	**±0 mo**	**57%**	**24% / 49%**
80th percentile	465	±0 mo	39%	33% / 49%
90th percentile	251	+2 mo	35%	2% / 91%

Eight thresholds were tested between the 50th and 95th percentiles. The 75th is the only one at which the error is both centred on zero and symmetric between early and late; below it the stored age reads systematically late, above it systematically early. **57% of in-scope milestones land within a single month of the empirically observed value**, on a sample roughly 1,600 times the size of the original validation study.

C2 · Instrument behaviour

Certainty **A** **Very high** 407 milestones · 48 skills

Within scope the instrument behaves as an acquisition measure should. Of 407 milestones with usable curves, **93% reach the 75th percentile** and only one fails to reach the 50th. The median curve runs from **25.9%** of children answering affirmatively at the earliest well-sampled age to **89.5%** at the latest: a low floor, a steep climb, a plateau.

Figure 2. Four representative acquisition curves. Dashed line marks the 75th percentile.

A residual 3.6% of curves begin above 80%. These are almost entirely newborn reflexes — rooting, sucking, the Babinski response — which are present at birth rather than acquired, and for which a high floor is the correct behaviour rather than a defect.

Sequence and dispersion across developmental areas

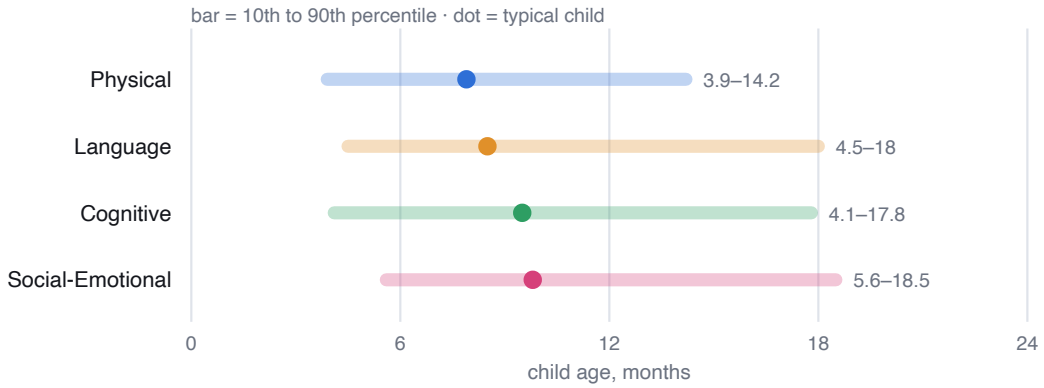

Figure 3. Age at which the 10th, 50th and 90th percentile of children reach the typical milestone in each area, averaged over milestones crossing all three thresholds within the observed range.

AREA	MILESTONES	10TH PCT	TYPICAL	90TH PCT	DISPERSION
Physical	136	3.9	7.9	14.2	10.3 mo
Language	28	4.5	8.5	18	13.5 mo
Cognitive	48	4.1	9.5	17.8	13.7 mo
Social-Emotional	39	5.6	9.8	18.5	12.9 mo

Physical skills arrive earliest and cluster most tightly; social-emotional skills arrive last and vary most between children. The ordering is invariant across specifications. The spacing between areas is not, and depends on the treatment of milestones that never reach the 90th percentile within the observed age range — which is why C6 is graded B.

> **Note on dispersion estimates.** Percentile spreads are averaged only over milestones crossing all three thresholds. Averaging over whichever milestones happen to reach each threshold separately produces incoherent output, since different milestones reach different thresholds — under that treatment one area returns a 90th percentile earlier than its own median.

C3 · Coupling and differentiation

Certainty **B** **High** 3,518,779 children · 279,505,209 answers

Stenhaug, Ram and Frank report that developmental domains behave as one tightly coupled system in the first year of life and separate thereafter. Reproducing this requires distinguishing two questions that are easily conflated: whether children who are ahead in one area are ahead in others, and whether an individual child's areas advance together over time. These have different answers.

Between children

Scoring each child in each area and correlating across children yields coupling near 0.6 that *rises* with age — 0.587 to 0.660 on the structured assessment, 0.580 to 0.665 on post-activity answers. Both instruments agree. This measure captures everything stable about a child, and stable differences do not decouple; C4–C5 quantify this directly.

Within a child

Following the same child over time and correlating their *changes* removes every stable characteristic by construction. Coupling declines steeply through the first year, reaching its minimum at the 10–14 month transition.

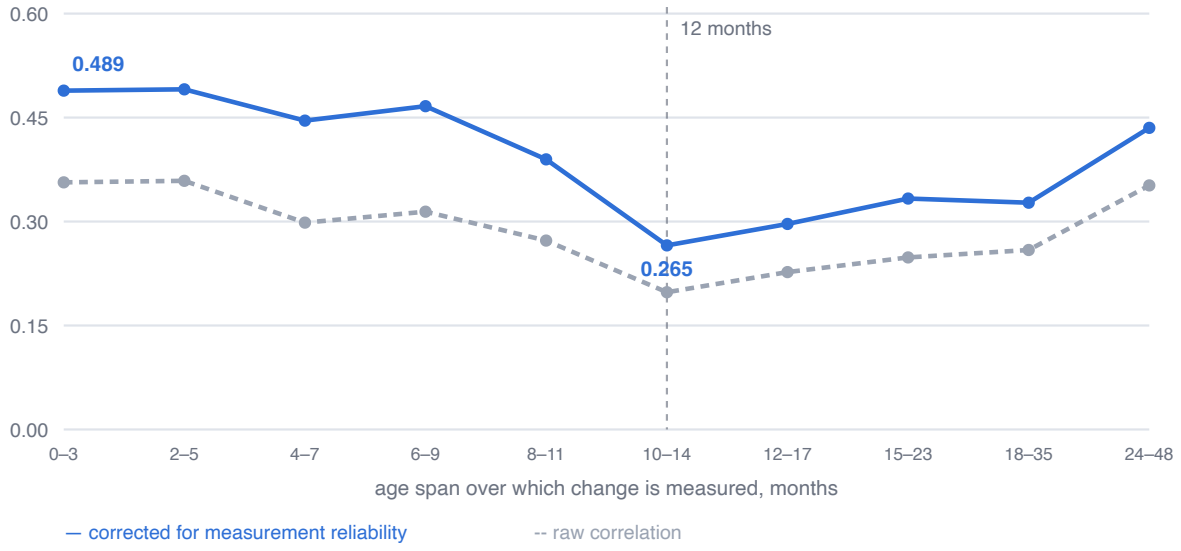

Figure 4. Within-child coupling between developmental areas across consecutive age bands, full population. Raw and reliability-corrected. Coupling approximately halves between the first measured transition and the 10–14 month minimum.

MEASURE	FIRST TRANSITION (0–3 MO)	MINIMUM (10–14 MO)	REDUCTION
Raw correlation	0.357	0.198	45%
Corrected for reliability	**0.489**	**0.265**	**46%**

The drop from the first measured transition to the minimum is **0.224**, 95% CI [0.205, 0.242], $p < 0.0001$ across 500 bootstrap resamples of children. Raw and corrected measures give the same 45–46% reduction, so the result does not depend on the reliability correction.

Persistence, with the measurement interval held constant

Consecutive age bands widen with age, and a longer interval mechanically raises coupling because it contains more total development. To establish what happens *after* the first year, the interval must be held fixed. Three were tested.

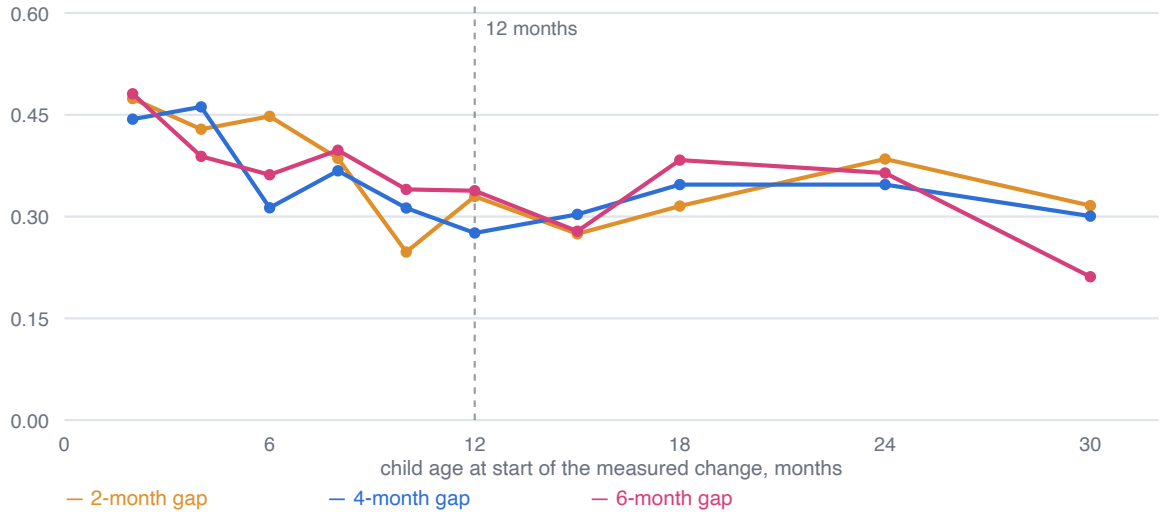

Figure 5. Within-child coupling at three fixed measurement intervals. Each line holds the interval constant and varies only the age at which change is measured.

INTERVAL	BEFORE 10 MONTHS	FROM 12 MONTHS ON	REDUCTION
2 months	0.434	0.324	25%
4 months	0.396	0.315	21%
6 months	0.407	0.315	23%

All three converge on **0.315–0.324 after twelve months** — three near-independent measurements returning the same value. On the four-month interval the sustained drop is 0.082, 95% CI [0.052, 0.114], $p < 0.0001$.

Two features merit note. The decline begins **by six to ten months**, earlier than the twelve-month figure usually cited. And it does not reverse at any subsequent age tested, through thirty-five months.

> **What separation does and does not mean.** Coupling roughly halves at the transition and settles about a quarter below its first-year level. It does not approach zero. The four areas soften from moving as effectively one process to moving as four loosely related ones — they remain the development of a single child, and continue to share a substantial common component at every age measured.

> **Why B rather than A.** The test is genuinely longitudinal and uses one instrument at both ends, as the design requires. That instrument is post-activity questions asked one to three at a time — thinner than the

> structured assessment, which cannot be used because it is administered once per child. Repeat structured assessments would settle it.

C4–C5 · Standing and movement

Certainty **A** **Very high** 1,109,370 children · baseline reliability 0.956

Crossing the two instruments answers a question neither can answer alone. Each child's assessment yields a general factor — a single number summarising how far along they were across all four areas at baseline. The amount of information that number provides about later standing depends on the child's age at assessment and on the follow-up horizon.

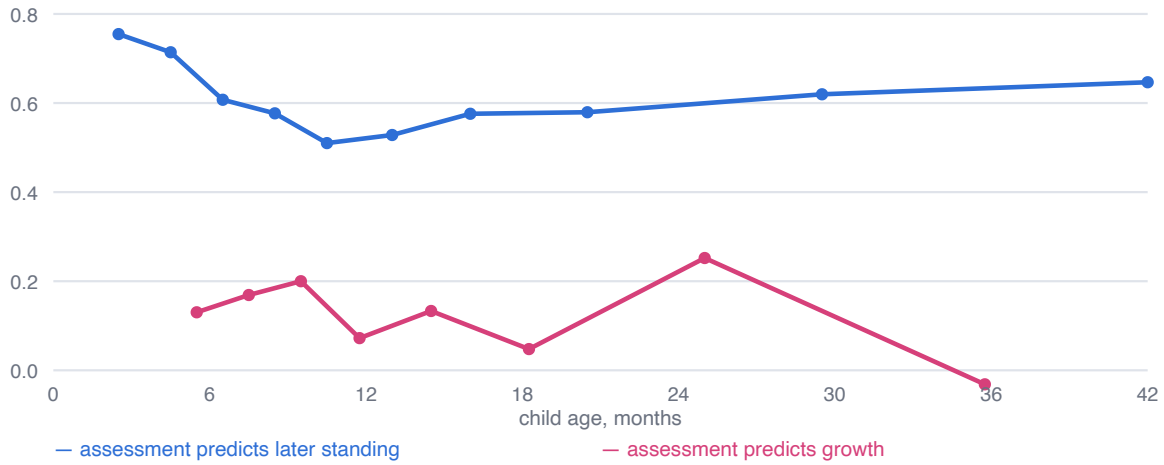

Figure 6. What a child's assessment general factor predicts. Upper line: their later standing in each developmental area. Lower line: how much they change. Both corrected for measurement reliability.

THE ASSESSMENT GENERAL FACTOR PREDICTS…	CORRELATION	VARIANCE EXPLAINED
A child's **later standing** in each area	0.61 (range 0.51–0.76)	**37%**
How much a child **changes**	0.12 (range −0.03–0.25)	**1.5%**
Ratio		**25×**

> **Interpretation across ages and horizons.** The r = 0.61 summary is not a universal forecast. In stratified follow-up analysis, a child assessed at 3–4 months shows r = 0.42 at a three-month horizon and r = 0.23 across two years; a child assessed at 24–25 months shows r = 0.61 at a three-month horizon. The amount of information a baseline assessment provides depends materially on when it is taken and how far ahead it is used.

Partial correlation confirms the separation from the other direction: removing the assessment factor reduces between-child coupling in **levels** by 32% on average and by as much as 57% at the youngest ages, while moving coupling in **changes** by −1%.

There exists, therefore, a single stable characteristic of each child — already measurable in the first months of life — that shapes standing across all four areas simultaneously and durably, and that has almost no bearing on the rate at which the child subsequently progresses. The between-child and within-child results of C3 are consequently not in tension: they are orthogonal. Differentiation is a property of development; the strong between-child correlation is a property of starting conditions.

The substantive reading of C5. Rate of progress is very nearly independent of starting position. Children who begin behind do not advance more slowly, and the 98.5% of variance in growth left unexplained by starting point must be attributable to other factors — environment, caregiving, and whatever else acts on a child after birth. Existing gaps therefore persist without widening, and the determinants of how fast a child progresses lie almost entirely outside where that child began.

Not established. The composition of the stable factor. Family income, schooling and material resources are plausible contributors and none was measured; the available fields are an identifier and a date of birth. Its existence and magnitude are established, its constituents are not.

C7 · Caregiver response style

Certainty C **Moderate** 370,767 children

A caregiver inclined to answer affirmatively would raise all four of a child's area scores together, producing the between-child pattern described in C3 without any developmental basis. This was tested rather than assumed.

Approximately 10% of assessment items concern milestones expected twelve or more months after the child's current age; caregivers answer affirmatively to 15.3% of them. A child genuinely two months advanced should not be performing behaviours expected a year hence, so a child's score on those items approximates a measure of response style. Area scores were computed on a disjoint set of near-level items, so any association between the two is substantive rather than mechanical.

QUANTITY	VALUE
Reliability of the response-style measure	0.898
Correlation with apparent advancement, per area	0.32 – 0.35
Share of apparent standing attributable to it	≈ 11%
Between-area coupling before removal	0.728
Between-area coupling after removal	0.694
Share of coupling attributable to response style	**≈ 5%**

Response style is real, stable and reliably measured; roughly one part in nine of a child's apparent standing reflects it. It accounts for only about 5% of the correlation between developmental areas, however. **The remainder is substantive** — children advanced in one area are genuinely advanced in the others.

> **Why C.** Only 11% of children receive enough far-above-level items to be measured. If the assessment advances to harder items following an affirmative answer, that subset is pre-selected for generous responding, compressing the range and understating the effect. The 5% figure is a lower bound.

8 · Discussion

The central methodological result of this study is that between-child and within-child analyses of the same data return opposite answers on the question of differentiation, and that the discrepancy is not an artefact requiring explanation but a substantive finding in its own right.

Comparing children with one another, coupling between developmental areas is high at every age and *increases* across early childhood. Comparing each child with their own earlier state, coupling halves across the first year and remains depressed thereafter. C4–C5 identify the source of the divergence: a stable characteristic present from the first months of life accounts for 37% of a child's standing in every area simultaneously, and for 1.5% of their rate of progress. Between-child comparison is dominated by that stable characteristic; within-child comparison eliminates it by construction. Differentiation is therefore a property of the developmental process, while the strong cross-sectional correlation among areas is a property of the conditions from which children begin.

The replication itself is directional rather than numerical. The reduction observed here is milder than the published 0.75 → 0.15, which is expected: the longitudinal instrument available is thinner than a structured assessment, the scoring method differs, and the statistical model differs. The shape of the decline, its timing, and its persistence all reproduce.

Two features extend the original finding. The decline begins by six to ten months, earlier than the twelve-month transition commonly cited; and it does not reverse at any age subsequently tested. Coupling does not approach zero — the four areas soften from behaving as effectively one process to behaving as four loosely related ones, retaining a substantial common component throughout.

The independence of growth from starting position (C5) carries an implication beyond the immediate question. If 98.5% of the variance in how rapidly a child progresses is unrelated to where that child began, then the determinants of developmental pace lie almost entirely in what happens after birth. Existing differences between children persist without widening, and the rate at which any individual child advances remains open.

9 · Limitations

1. **The structured assessment is administered once per child.** Of 3,384,823 assessed children, 3,384,584 have exactly one administration. No test of change can therefore employ the structured instrument at both endpoints, which is the sole reason C3 is graded B.
2. **The longitudinal instrument is thin.** Post-activity questions arrive one to three at a time and depend on which activity the caregiver selected, so which areas are measured is not under experimental control. Residualising against the population rate for each milestone and age removes the resulting composition bias from score levels but cannot make the measurement as rich as a full assessment.
3. **The population is self-selected.** These are families who chose to install a child-development application and continued to respond. Absolute ages of acquisition may not generalise. Comparisons *within* this population, on which every result here rests, are unaffected.
4. **Attrition is severe and non-random.** Observation volume falls from 20.6% of the total in the first three months of life to 0.3% beyond 54 months. Estimates at later ages rest on families who persisted.
5. **No socioeconomic, geographic or clinical covariates were available.** The stable factor of C4 is measured but not characterised.
6. **Age is resolved to the month.** Effects operating on a finer timescale would not be detected.
7. **Seven skills were initially unmapped** to a developmental area owing to nomenclature mismatches between the skill table and the reference map, silently excluding 77,978 observations. This was identified and corrected; the mapping covers 98.6% of in-scope observation volume, with the remaining 1.4% carried as unassigned rather than imputed.

10 · References

Stenhaug, B. A., Ram, N., & Frank, M. C. *The Structure of Developmental Variation in Early Childhood.* Stanford University, Language and Cognition Lab. Preprint: osf.io/preprints/psyarxiv/95erq

Kinedu. *Initial Assessment: scale construction and psychometric analysis.* Internal validation study, N = 2,135 caregivers of children aged 0–55 months; Cronbach's α = .996; ages of acquisition assigned at the 75th percentile following American Academy of Pediatrics practice.

Appendix A · Acquisition curves

Every milestone and skill in scope, plotted as the share of children answering affirmatively against child age. All curves use the 2020–2023 window and the same inclusion rules as the main analysis: age cells of at least 200 observations, at least six surviving cells per milestone.

A.1 · By skill, with dispersion

Each panel shows one skill. The line is the mean across that skill's milestones at each age; the band is ±1 standard deviation across those milestones — a direct reading of how uniformly the skill's components arrive. A narrow band means the milestones within that skill are acquired at close to the same age; a wide band means the skill is a loose grouping of components that arrive at quite different times.

Panels are ordered by the age at which the skill's mean curve reaches 50%. The dashed vertical marks that age. Skills are computed on a balanced panel: only ages at which at least 60% of the skill's milestones are observed contribute, which prevents the mean from shifting simply because the set of milestones being asked changes with age.

Figure A1. Skill-level acquisition curves with ±1 SD, part 1 of 2 — the 24 earliest-acquired skills.

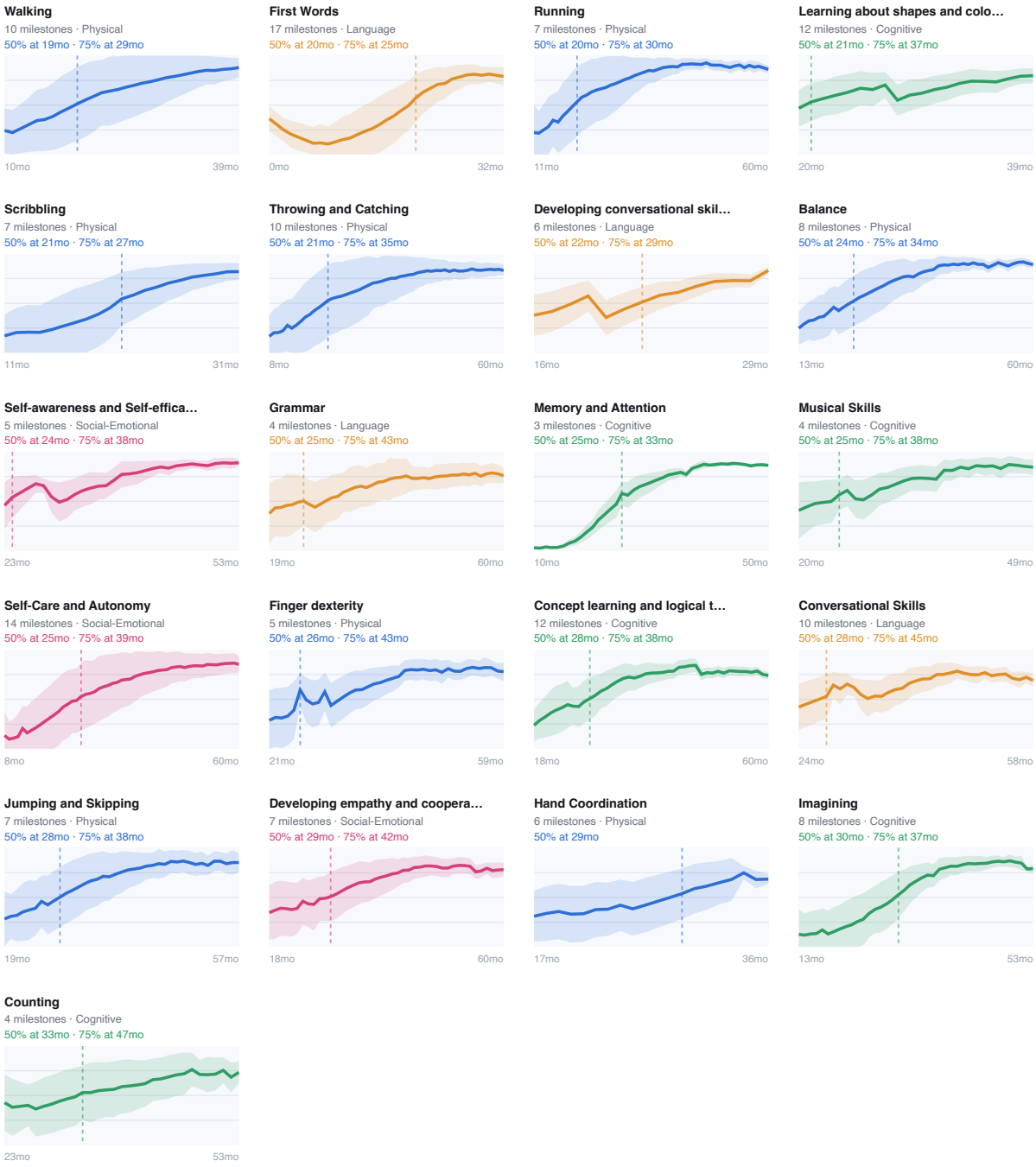

Figure A2. Skill-level acquisition curves with ±1 SD, part 2 of 2.

A.2 · Dispersion within each skill

The standard deviation at the 50% crossing, tabulated. Low values identify skills whose milestones behave as a coherent unit; high values identify skills that group together components arriving at materially different ages.

SKILL	AREA	MILESTONES	50% AT	75% AT	SD AT 50%
Newborn Reflexes and Posture	Physical	10	0 mo	0 mo	5.9
Memory and Attention	Cognitive	3	25 mo	33 mo	12.0
Developing conversational skills	Language	6	22 mo	29 mo	15.6
Musical Skills	Cognitive	4	25 mo	38 mo	15.8
Finger dexterity	Physical	5	26 mo	43 mo	16.6
First Steps	Physical	7	12 mo	13 mo	17.4
Early Phonetics	Language	10	18 mo	28 mo	17.7
Learning about shapes and colors	Cognitive	12	21 mo	37 mo	19.4
Self-awareness and Self-efficacy	Social-Emotional	5	24 mo	38 mo	20.9
Concept learning and logical thinking	Cognitive	12	28 mo	38 mo	21.3
Conversational Skills	Language	10	28 mo	45 mo	22.3
Imagining	Cognitive	8	30 mo	37 mo	23.5
Head Control	Physical	6	0 mo	3 mo	23.7
First Words	Language	17	20 mo	25 mo	24.2
Grammar	Language	4	25 mo	43 mo	24.5
Developing empathy and cooperation	Social-Emotional	7	29 mo	42 mo	24.9
Balance	Physical	8	24 mo	34 mo	25.1
Hand Coordination	Physical	6	29 mo	—	25.4
Developing musical skills	Cognitive	3	12 mo	15 mo	25.6
Babbling	Language	9	5 mo	8 mo	26.4
Scribbling	Physical	7	21 mo	27 mo	27.0
Counting	Cognitive	4	33 mo	47 mo	27.4
Standing Up	Physical	10	10 mo	12 mo	28.8
Rolling Over and Sitting Up	Physical	9	5 mo	7 mo	29.8
Jumping and Skipping	Physical	7	28 mo	38 mo	30.3
Imitating	Cognitive	12	13 mo	18 mo	30.7

Developing finger dexterity	Physical	10	9 mo	11 mo	31.2
Communicating through gestures	Language	5	8 mo	10 mo	31.8
Kicking	Physical	5	17 mo	25 mo	32.8
Crawling	Physical	15	7 mo	10 mo	34.0
Self-Care and Autonomy	Social-Emotional	14	25 mo	39 mo	34.4
Early movements and coordination	Physical	14	3 mo	5 mo	34.5
Basic Vocabulary	Language	17	16 mo	25 mo	35.4
Running	Physical	7	20 mo	30 mo	35.4
Early knowledge and curiosity	Cognitive	10	13 mo	17 mo	35.9
Throwing and Catching	Physical	10	21 mo	35 mo	36.2
Object Exploration	Cognitive	12	3 mo	5 mo	36.3
Walking	Physical	10	19 mo	29 mo	37.1
Secure Attachment	Social-Emotional	11	4 mo	7 mo	37.9
Developing Independence	Social-Emotional	7	11 mo	15 mo	38.0
First interactions	Social-Emotional	8	3 mo	8 mo	38.8
Developing hand coordination	Physical	17	11 mo	16 mo	39.1
Developing the senses	Physical	12	2 mo	7 mo	41.4
Developing memory and attention	Cognitive	14	9 mo	16 mo	43.6
Developing self-awareness	Social-Emotional	4	8 mo	19 mo	44.0

> **Reading the extremes.** Newborn Reflexes and Posture shows a standard deviation of 5.9 — its ten milestones are present at birth and behave almost identically. Developing self-awareness (44.0) and Developing memory and attention (43.6) sit at the other end: their milestones span such different ages that the skill's mean curve describes no individual milestone well. Skills in the upper part of this table are candidates for splitting.

A.3 · Why there is no area-level acquisition curve

The same aggregation applied at the level of developmental area produces standard deviations of 38 to 43 percentage points, and two of the four areas never reach 75% at any age. This is not a defect in the calculation: a developmental area is not a thing acquired at a point in time. Physical spans 175 milestones across four years, from newborn reflexes to jumping. Averaging their affirmative rates yields a curve that describes no child and no milestone.

This limitation is specific to **acquisition curves** — to pooling raw affirmative rates. It does not extend to residualised measures. Once each observation is expressed relative to the population rate for its own milestone at its own age, the age structure that makes pooling invalid has already been removed, and area-level aggregates are well defined. Every area-level result in this report — the coupling analysis of C3, the standing-and-movement decomposition of C4–C5 — is computed exactly that way and behaves properly, with per-area reliabilities of 0.82 to 0.95.

The distinction is worth stating plainly because it determines which views are meaningful at which level of the hierarchy. **Acquisition curves belong to milestones and skills**, where the components share an age window and pooling is legitimate. **Residualised scores and their change over time belong to areas and to the child overall**, where the age structure has been normalised away. The area-level view corresponding to a curve is not a curve at all: it is the sequence and dispersion of acquisition *ages* given in C2.

A.4 · Every milestone

All 407 in-scope milestones, grouped by developmental area and then by skill, ordered within skill by expected age. The vertical line in each panel marks the stored expected age; the dashed horizontal marks 75%. Each caption gives the expected age, the observed age range, and the affirmative rate at the first and last observed age.

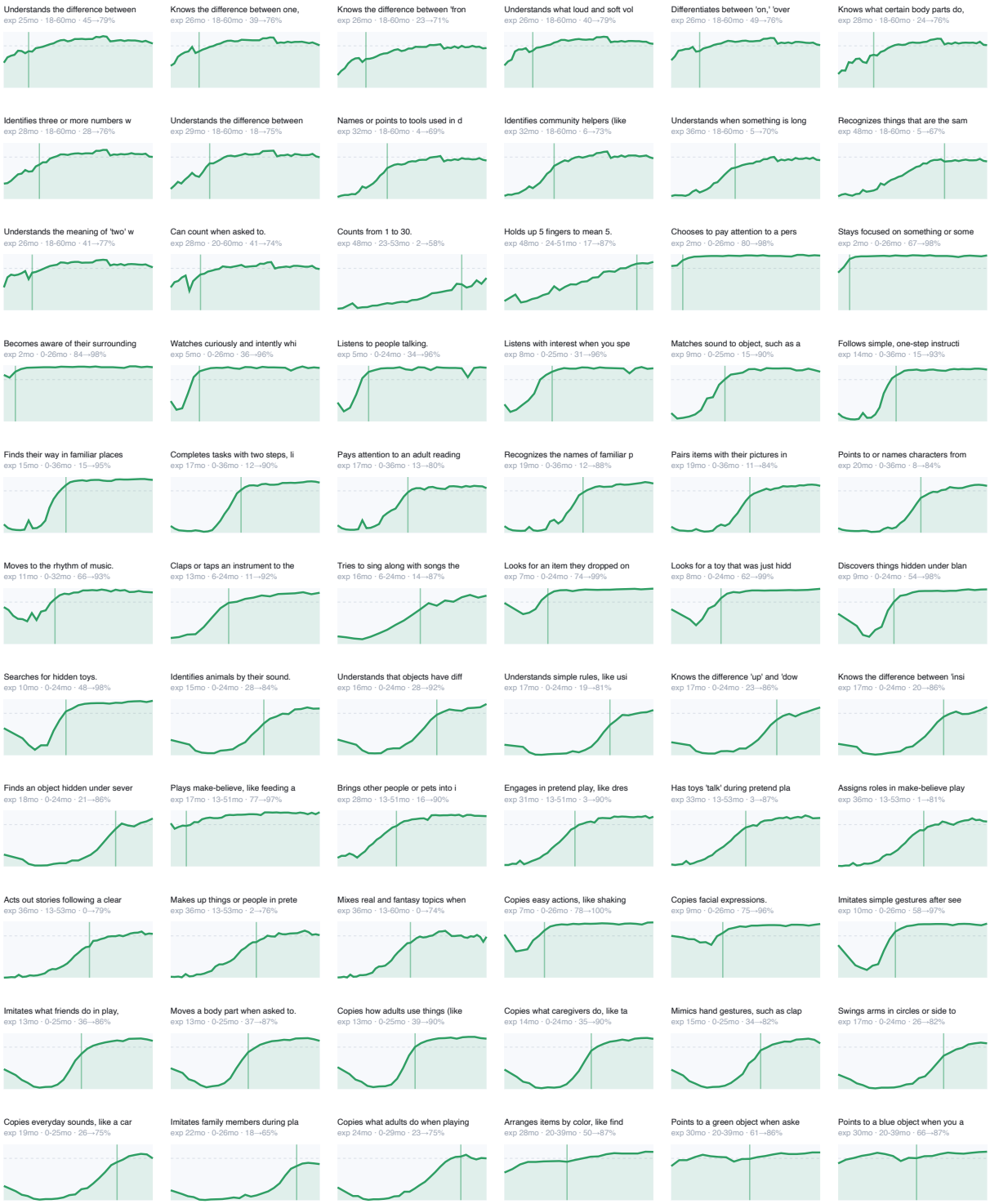

Figure A3. Milestone acquisition curves, part 1 of 7.

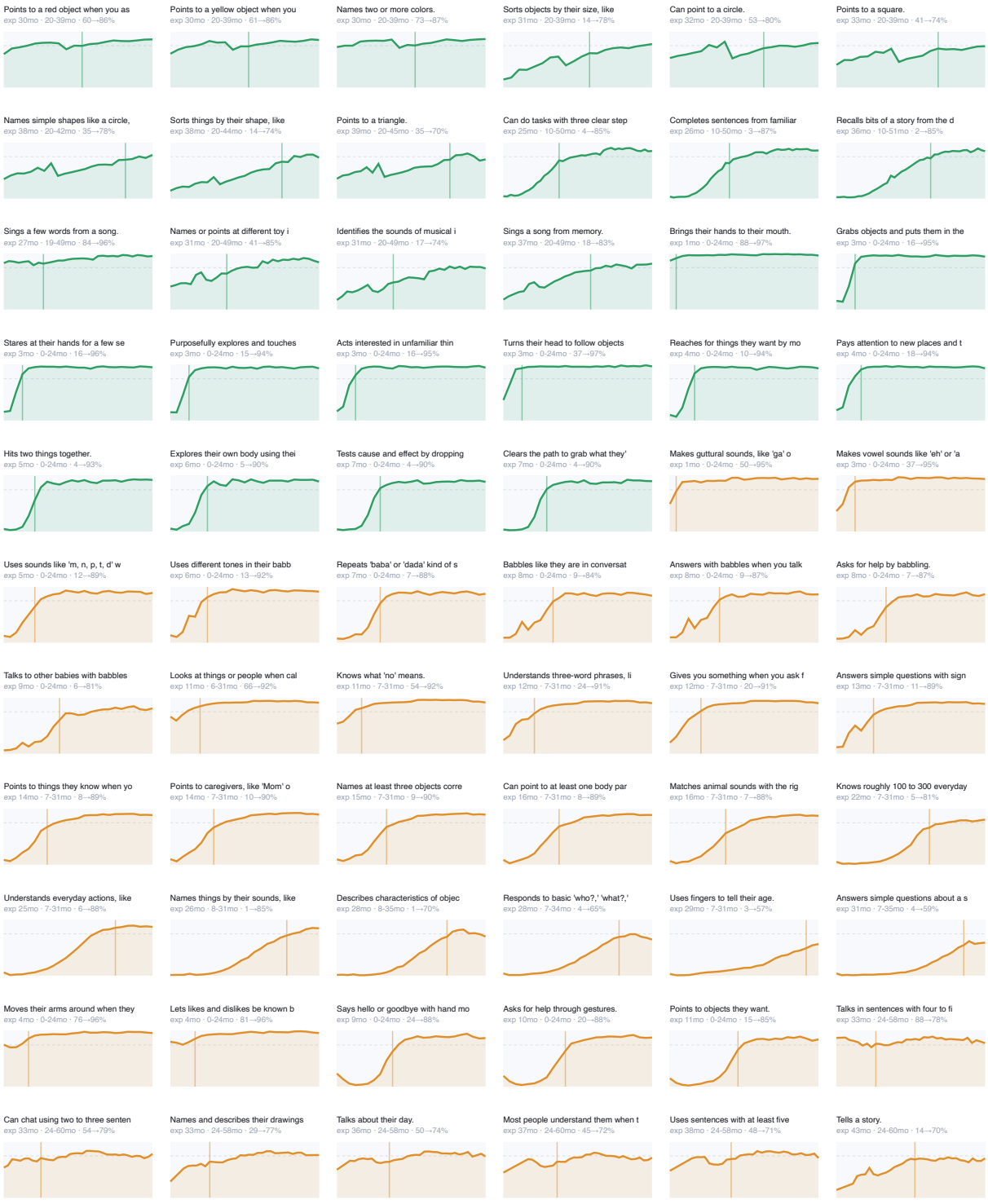

Figure A4. Milestone acquisition curves, part 2 of 7.

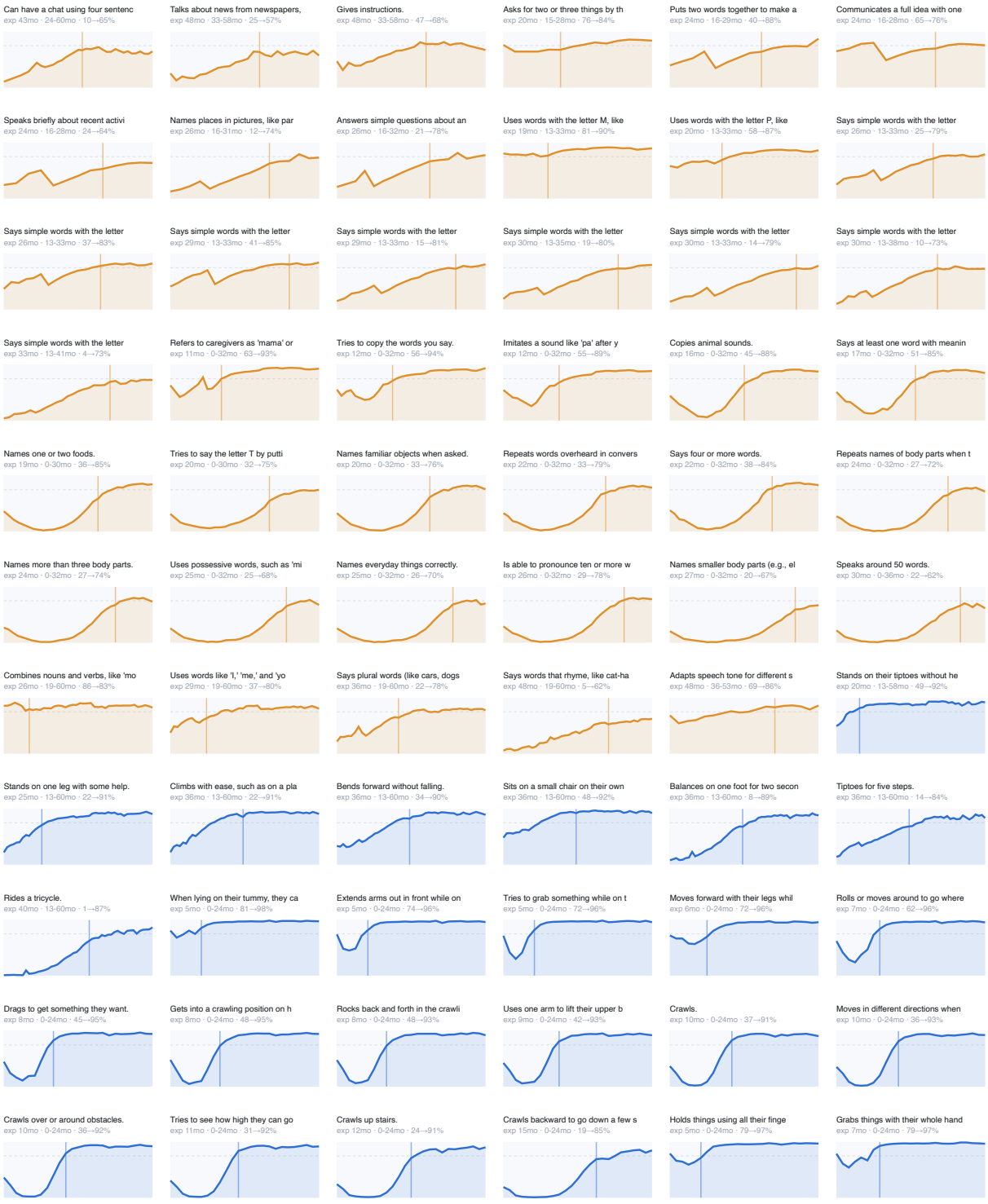

Figure A5. Milestone acquisition curves, part 3 of 7.

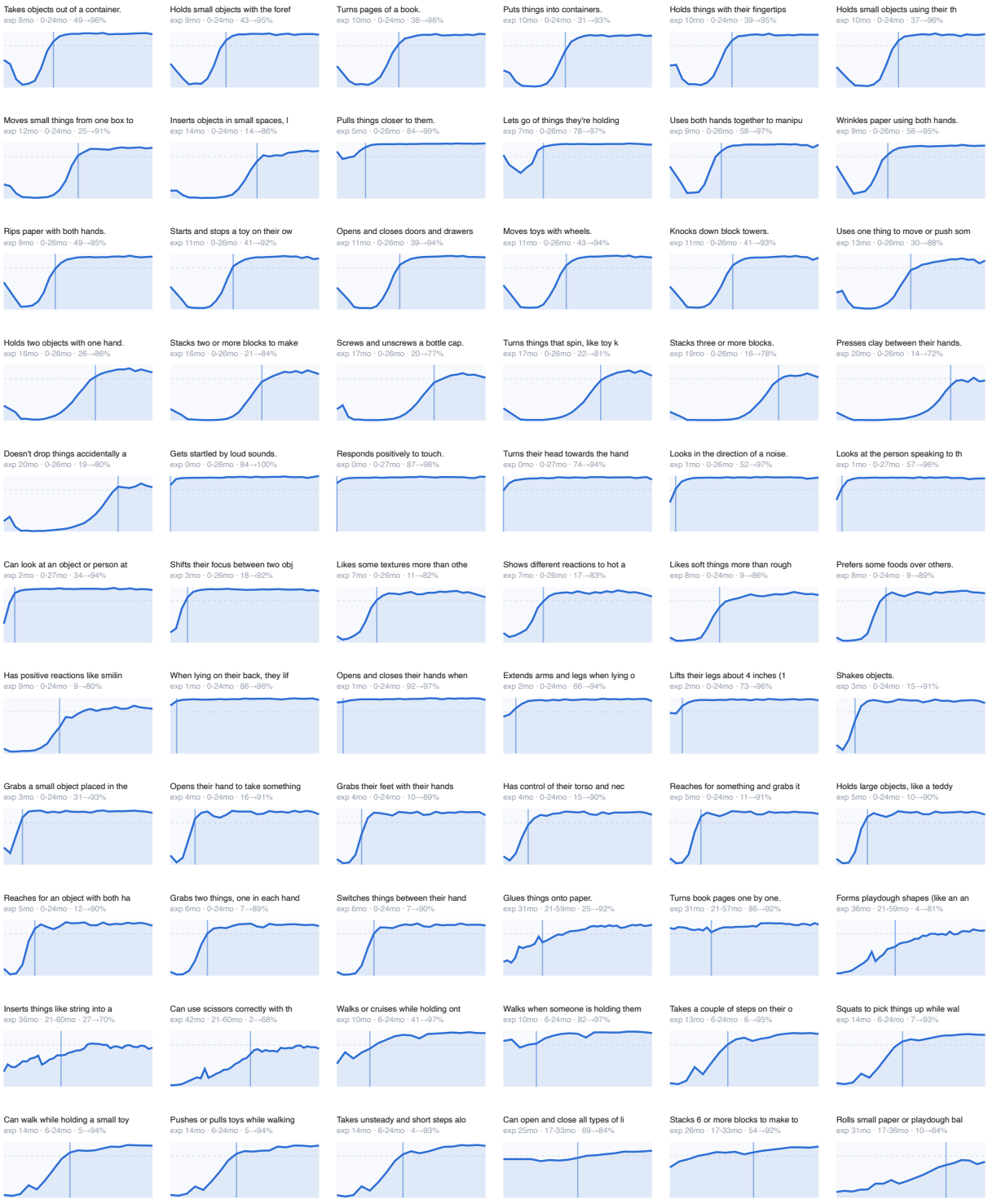

Figure A6. Milestone acquisition curves, part 4 of 7.

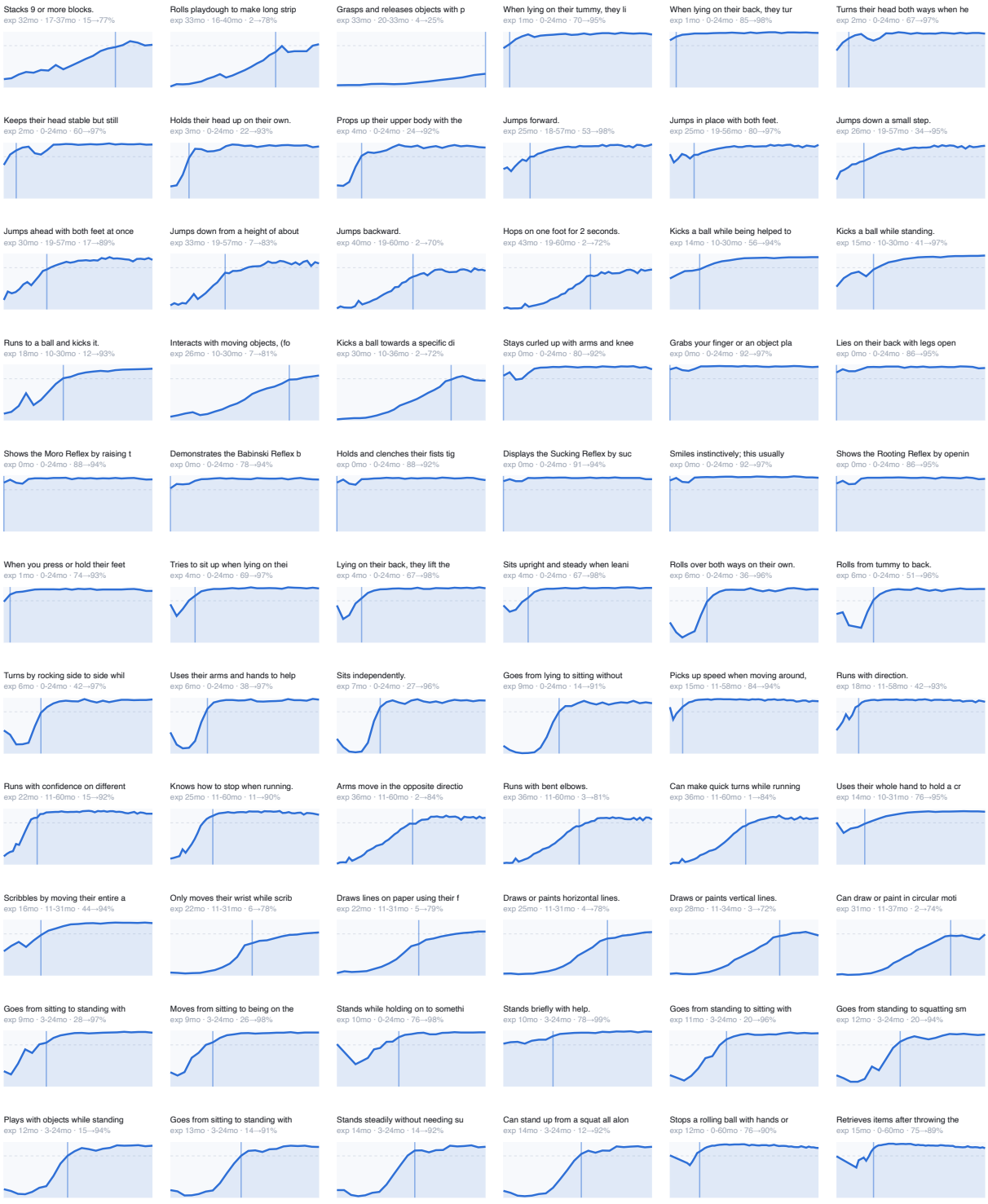

Figure A7. Milestone acquisition curves, part 5 of 7.

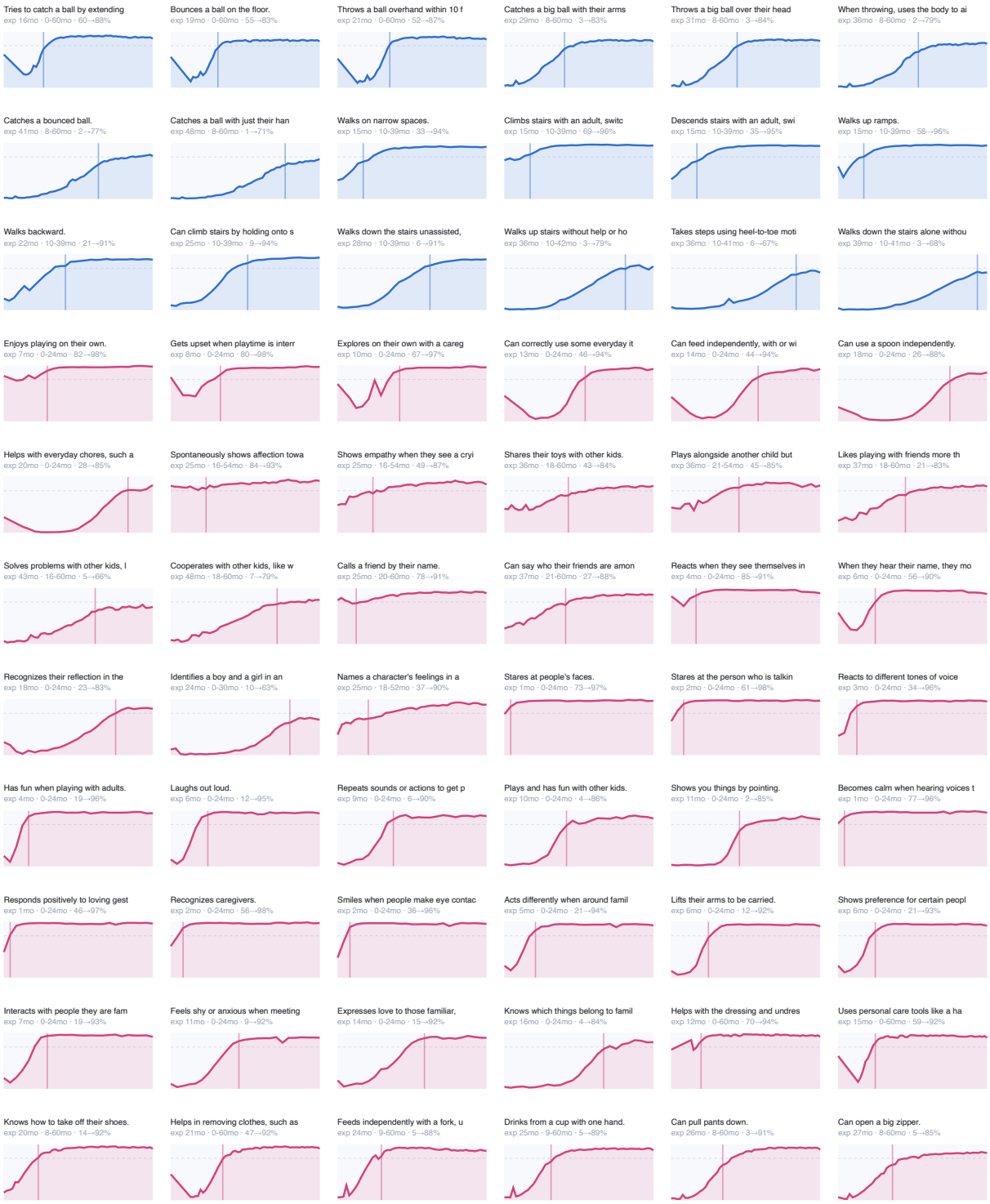

Figure A8. Milestone acquisition curves, part 6 of 7.

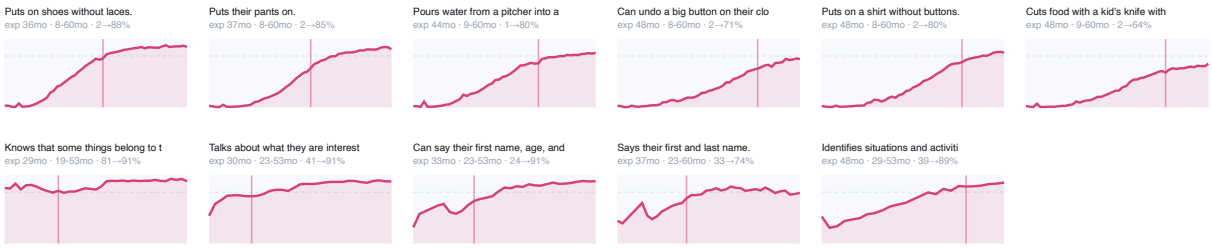

Figure A9. Milestone acquisition curves, part 7 of 7.

Analysis window January 2020 – December 2023. The Kinedu Assessment is a developmental surveillance instrument and is not diagnostic.