The Infant and Toddler Curiosity Questionnaire: A validated caregiver-report measure of curiosity in children from 5 to 24 months

Elena C. Altmann[a*], Marina Bazhydai[a], Didar Karadağ[a], and Gert Westermann[a]

elena.c.altmann@gmail.com, m.bazhydai@lancaster.ac.uk, d.karadag@lancaster.ac.uk, g.westermann@lancaster.ac.uk

[a]Department of Psychology, Lancaster University

Author note

Correspondence concerning this article should be addressed to Elena Altmann, Lancaster University, Lancaster LA1 4YF, email: elena.c.altmann@gmail.com; ORCID ID: 0000-0002-8602-9845. Gert Westermann was supported by the Economic and Social Research Council (ESRC) International Centre for Language and Communicative Development (LuCiD) [ES/S007113/1 and ES/L008955/1]. Data, materials, and analysis scripts presented in this manuscript are available from the OSF. We have no conflicts of interest to disclose.

Research Highlights

- The first validated caregiver-report measure of curiosity capturing individual differences in trait curiosity in infants and toddlers.
- Twenty-three items were found to reliably capture a general factor of curiosity as well as three emergent subfactors.
- The measure had good test-retest reliability, and we found curiosity and temperament to be related yet distinct, indicating its validity as a trait measure.
- Results strongly support the scale's reliability and validity, showcasing the ITCQ as a powerful tool for developmental research.

Abstract

Humans are curious. Especially children are known for their drive to explore and learn, which is crucial for developing in and navigating through our complex world. Naturally, some children may be more curious than others, leading to differences in how they structure their own learning experiences, subsequently impacting their developmental trajectories. However, there is a gap in the research field for a reliable measure of such differences early in development. Across three studies, we present the development and assessment of the Infant and Toddler Curiosity Questionnaire (ITCQ), the first caregiver report measure to fill this gap. Items cover observable exploration behaviours in 5- to 24-month-olds to capture general tendencies of their desire to actively explore their immediate surroundings and are evaluated on a 7-point Likert-scale. Exploratory factor analyses and structural equation modelling on a sample of $N = 370$ UK caregivers led to the final selection of 23 items and provided evidence that the scale allows the reliable computation of an overall curiosity score, with three emergent subscales (*Sensory, Investigative*, and *Interactive*) explaining additional variance in the data. Furthermore, the scale had good test-retest reliability after 7 to 14 days ($N = 67$) and related to the child's temperament ($N = 75$; positively with surgency and effortful control, negatively with negative affect) offering evidence of its validity as a trait measure. Together, these results support the scale's reliability and validity, showcasing the ITCQ as a powerful tool for developmental research.

Keywords: infant curiosity, individual differences, psychometrics, exploration, early development, trait curiosity

The Infant and Toddler Curiosity Questionnaire: A validated caregiver-report measure of curiosity in children from 5 to 24 months

While the study of curiosity and its effects on learning has a long history in adults (Berlyne, 1960; Gruber et al., 2014; Kang et al., 2009; Rossing & Long, 1981), only in recent years has infant curiosity become a focus of research and has provided insights into how infants actively engage in their own learning. Studies have shown, for example, that infants prefer to engage with information of intermediate complexity (Kidd et al., 2012, 2014) and that they alternate between visual exploration and exploitation driven by their active learning experience (Altmann et al., 2024) in the pursuit of maximizing their learning progress (Poli et al., 2020; Twomey & Westermann, 2018), that they actively request information from adults through social orienting (Bazhydai et al., 2020) and pointing (Liszkowski et al., 2007), with learning benefits shown for actively requested information (Begus et al., 2014) but also from being in a state of curiosity more generally (Chen et al., 2022; Stahl & Feigenson, 2015). What is common to all of these studies is that they assume infants to be curious learners by definition and that they investigate the implications of such inherent trait curiosity on in-the-moment behaviours. Only a few studies, however, have considered how differences in infants' individual interests (Ackermann et al., 2020) and sensory seeking (Piccardi et al., 2020) affect their preference to engage with specific information. Overall, there has been no systematic investigation on individual differences in trait curiosity on infants' exploratory behaviour, learning, and later outcomes. To enable such research, it is necessary to have validated measures of infant curiosity.

This is different from research in adults where multiple scales exist to assess variation in trait curiosity (for reviews, see Grossnickle, 2016; Jirout & Klahr, 2012; Wagstaff et al., 2021), measuring general accounts of curiosity (e.g., Day, 1971; Kashdan et al., 2020; Litman & Jimerson, 2004; Litman & Spielberger, 2003; Naylor, 1981; Spielberger, 1979) but also more specific domains (e.g., social curiosity, Renner, 2006; work-related curiosity, Mussel et al., 2012) in the form of self-report questionnaires. These measures typically ask responders how commonly or intensely they experience a desire for knowledge and learning, thereby requiring meta-cognitive awareness (e.g., Goupil & Proust, 2023; Loewenstein, 1994). However, some questionnaires also conceptualised curiosity as the intrinsic motivation behind exploration in the pursuit of knowledge, leading to items focusing on more observable behaviours (e.g., Kashdan et al., 2009). Research using such trait measures has shown positive associations between curiosity and job performance as well as academic achievement (e.g., Grossnickle,

2016; Hardy et al., 2017; Kashdan & Yuen, 2007; Mussel, 2013; Reio & Wiswell, 2000; Reio & Callahan, 2004), highlighting its impact on life outcomes.

To investigate trait curiosity in children, some self-report measures designed for school-aged cohorts do exist (Byman, 2005; Maw & Maw, 1968; Olson, 1986; Penney & McCann, 1964) but their validity may be limited by the children's lack of motivation to self-reflect and reliably answer numerous repetitive items (Jirout & Klahr, 2012). An alternative to self-reports, especially relevant for younger children, are *other*-reports (primarily caregivers and teachers, e.g., Harty & Beall, 1984; Lee et al., 2023; Maw & Maw, 1970; Piotrowski et al., 2014). Other-reports enable measurement without age-restriction but do require another person to assess a latent (not directly observable) construct in the child, bringing its own challenges. Yet, extensive research from infancy onward has shown that such measures can generate reliable, valid, and longitudinally informative data by, for instance, having items address observable behaviours in which the latent construct manifests itself. Prominent examples include the Ages and Stages questionnaire (e.g., Klamer et al., 2005; Lepine et al., 2022; Richter & Janson, 2007; Salomonsson & Sleed, 2010), the MacArthur-Bates Communicative Development Inventory (CDI; Bornstein & Putnick, 2012; Can et al., 2013; Feldman et al., 2005; Fenson, 2002; Fenson et al., 1994, 2006; Marchman & Fernald, 2008), and temperament scales from infancy to early childhood (e.g., Putnam et al., 2008; Rothbart, 1986, 2011; Slagt et al., 2016; Wright & Jackson, 2022). However, there is a clear gap in the scientific literature regarding infant trait curiosity. Even though some emerging work has aimed to assess differences in early curiosity through caregiver reports (Lee et al., 2023; Piotrowski et al., 2014) the target group of infants and toddlers has thus far been neglected.

As mentioned, for other-reports it is important to create items based on observable behaviour in which the construct manifests. Regarding curiosity, this manifestation is commonly assumed to be active exploration and interaction with the environment (for review see Bazhydai et al., 2021; Jirout et al., 2024). Previous research across the first two years of life has found individual differences in exploration throughout various experimental paradigms, such as visual exploration (e.g., Colombo et al., 1991; Franchak et al., 2016; Piccardi et al., 2020; Wass & Smith, 2014), manual exploration (e.g., Fortner-wood & Henderson, 1997; Mandler et al., 1987; Muentener et al., 2018) and free play exploration (e.g., Bornstein et al., 2013; Slone et al., 2019; Smith & Yu, 2013), letting us plausibly assume that infants already differ in their trait curiosity. Some of these studies also found that exploration differences were predictive of variability in learning, later vocabulary, cognitive development, and academic achievement (e.g., Berg & Sternberg, 1985; Bornstein et al., 2013; Muentener et

al., 2018; Smith & Yu, 2013) highlighting its role in and importance across development. In fact, one study (Shah et al., 2018) used a subset of five caregiver-report items as an ad hoc measure of curiosity (e.g., “Likes to try new things.”, "Shows eagerness to learn new things.") and was able to find a positive relation with academic achievement in kindergarten. It is to be noted, however, that the other three included items rather capture skills *related* to curiosity, such as metacognitive communication (“Appropriately uses a variety of words to describe feelings.”), creativity (“Shows imagination in work and play.”) and temperament (“Easily adjusts to a new situation.”), but not curiosity directly. Thus, it is important to stress that measures used in psychological research need to be reliably constructed and validated with regards to the construct they aim to capture (e.g., Flake & Fried, 2020). Nevertheless, this finding does hint at the impact a systematically developed and validated caregiver report measure of curiosity could have. A gap in the literature remains regarding a measure which can capture individual differences in curiosity through active exploration tendencies in pre-verbal infants.

The current paper

We present the Infant and Toddler Curiosity Questionnaire (ITCQ) applicable for 5-24-month-old infants and toddlers as a new caregiver-report measure for capturing individual differences in trait curiosity. Considering the lack of consensus for a functional definition of curiosity especially early in development (Jirout & Klahr, 2012; Kidd & Hayden, 2015), we decided to base our approach on a folk psychology definition of infant curiosity, that is, *a keen desire or tendency to actively explore one’s immediate surroundings*. Here we report three studies on the ITCQ’s development and assessment, evidencing its reliability and validity in line with rigorous practices (e.g., Downing, 2003; Flake & Fried, 2020; Messick, 1995). Study 1 describes the questionnaire development, including its content and structural validity (sample size and general analytical approaches were pre-registered at https://aspredicted.org/19J_291). Study 2 supports the ITCQ’s test-retest reliability after 7-14 days, and study 3 explores the measure’s criterion validity via its relation to the well-established trait measure of temperament. Studies were given ethical approval by the University Faculty’s research ethics committee, and data as well as analysis scripts are available on the Open Science Framework (OSF).

Study 1: Questionnaire Development & Structural Validity

In this first study we describe the principles underlying the ITCQ's creation, a reduction of items to generate coherent and reliable responses, as well as offering evidence for its content and structural validity (Downing, 2003; Messick, 1995). Content validity concerns whether the items are representative and well formulated, and whether ambiguities were resolved, which was largely the focus of the item development and an initial pilot study. Structural validity, on the other hand, concerns the scale's dimensionality and internal consistency, which was the focus of the main analysis of this study. While we created the items with the purpose of measuring a single factor of general curiosity, it is recommended for new scales to explore and consider the best fitting emerging factor structure to explain additional variance in the data (e.g., McCoach et al., 2013). Furthermore, we aimed to provide evidence supporting our intention of one general factor (Artino Jr et al., 2014) to justify the computation of an overall mean score (Dunn & McCray, 2020).

Questionnaire development

We developed an initial set of 34 statements capturing a wide range of behaviours infants can produce to interact with their physical and social environment as the manifestation of their curiosity. For this, we reviewed scientific and general sources (e.g., parent forums) regarding exploration behaviours children typically express throughout their first two years of life (looking, grabbing, mouthing, pointing, etc. typically directed at objects; Adolph & Hoch, 2019; Lockman, 2000) as well as everyday situations and locations in which they could be observed (e.g., at home or in new environments). Furthermore, we avoided focusing on curiosity about new people as this special domain of curiosity may be too heavily influenced by differences in temperament (e.g., shyness). Based on these sources, various items were proposed, and those deemed to capture the construct best by the research team were eventually selected. This process resulted in a list of 34 items which allowed for a diverse selection of behaviours while only requiring a manageable amount of time and effort from responders. These items covered behaviours such as interacting with objects, enjoyment of new discoveries, and interactions to gain information (for a full list of final items, see Table 1).

Due to the necessarily developmental perspective, some behaviours may not yet be observable in younger infants such as interacting socially (e.g., "When reading a picture book together, my child directs me (e.g., by pointing) towards what they want to know more about.")

whereas other items were expected to be equally applicable across ages (e.g., "When my child encounters an object, they typically seem interested in its properties (e.g., how it feels, tastes or sounds like, etc.)"). To constrain the variance in applicability of items, we decided to focus on an age range from five to 24 months. The minimum of five months was chosen based on a notable expansion in behaviours infants can produce from this age, whereas 24 months was chosen as the upper limit because from around the second birthday onwards verbal expressions of curiosity, such as question asking, become more prevalent.

Three of the items were reverse coded and described non-curious behaviour (e.g., "My child does not typically engage with (look at, point at, reach for, inspect) a lot of things in their environment."). While it is recommended to include such items to enhance data quality by making the reader slow down (Boley et al., 2021), they can also reduce the scale's overall reliability due to inattentive responses and lack of clarity (Rossiter, 2002; Salazar, 2015; Weijters & Baumgartner, 2012), often leading to their exclusion during the structural validity investigations. With this in mind, we included two additional items which were the mirrored versions of other positive items, solely meant to increase the responders' attention but not to be analysed as part of the final dataset (see Supplementary Materials S1.1). Even though they are not considered part of the final item list, they can be optionally included (e.g., for larger online studies with potentially lower response quality).

We chose a 7-point Likert-scale from 1 ("strongly disagree") to 7 ("strongly agree") as the response scale, with an option of "not applicable (NA)" if respondents could not think of any recent situation allowing them to rate a specific item or because their child had not yet been able to show the behaviour. Items were created in English initially targeting British caregivers and were repeatedly reviewed and improved based on the topical expertise of the authors, as well as through discussions with parents and native speakers to ensure their content validity. For further considerations regarding the response scale and online presentation, see Supplementary Materials (S2).

Piloting

A pilot sample (N = 22, age in months: M = 11.5, SD = 1.6, 41% female; £5 travel reimbursement given), collected from primary caregivers (all mothers) participating in an in-person study with typically developing 10- to 12-month-olds in the north-west of England, provided the first support for the questionnaire's construction. The exclusion of two items was suggested to improve the scale's homogeneity, without which the measure had very good

coefficients commonly used to indicate internal consistency (Coefficient alpha = .87, lambda-6 = .93). This offered first evidence that the scale was constructed sensibly enough to continue wider data collection (Nunnally & Bernstein, 1994).

At the end of the survey, caregivers were also invited to provide qualitative responses of additional behaviours capturing curiosity. These responses were found to reflect very similar behaviours and situations already covered in the questionnaire items (e.g., being interested in how things feel, trying to see what objects are on the table, etc.). They also supported our conceptualisation of curiosity being in line with how parents intuitively understand the construct; thus together, these findings evidenced the scale's content validity. In-person comments received from caregivers after completing the questionnaire offered additional insights as they mentioned that the questionnaire was clearly formulated and easy to complete. Some parents mentioned that the items let them easily differentiate between behavioural tendencies of their youngest child and their older siblings which they found fascinating.

Due to these overall promising preliminary results, we continued with wider data collection across the full age-range making no changes to the scale (i.e., also not removing the two ill-fitting items at this point in case their effect was due to the constrained sample). Based on this decision, the pilot data was deemed suitable to be included in the following main analyses.

Methods

Participants

A minimum sample size of *N*=360 was preregistered following a rule of thumb with 10 participants per item (e.g., Nunnally, 1978). A total of *N* = 370 responses were included in the final analyses (age range in months: 4.5-24.4, *M* = 13.5, *SD* = 5.2, see Figure 1; 51% female). Of these, *n* = 243 were recruited via social media, *n* = 72 attended in-lab visits, and *n* = 54 were contacted from the Babylab's database (which includes contact details for families willing to take part in infancy studies in the north-west of England) to directly complete the questionnaire alongside a temperament questionnaire (see study 3). Fifty-five additional responses were excluded due to being outside the preregistered age-range (*n* = 16), not being from the UK (*n* = 5), prematurity (*n* = 17), developmental concerns (*n* = 14), or poor data quality (*n* = 2, where all responses were either NA or the exact same response including reverse coded items). From the final sample, 97% indicated their child to be monolingual English, 82% of caregivers (predominantly mothers as per social media engagement, database recruitment and direct

communication) indicated to have achieved a degree in higher education (e.g., bachelor's degree and above), 50% of children were said to be the first born, 40% second born, and 10% were reported to have at least 2 older siblings (including stepsiblings). Participants recruited online via social media were invited to complete the survey without reimbursement whereas the sub-samples directly recruited from the Babylab's database received reimbursement as per university guidelines: in-lab visiting participants received £5 for their travel, and the sample which completed the longer version including the temperament scale received a £5 online gift voucher of their choice (via express.giftpay.com). The reported studies (1-3) were conducted according to guidelines laid down in the Declaration of Helsinki, with written informed consent obtained from a parent or guardian for each child before any assessment or data collection. All procedures received ethical approval under FST21068 at Lancaster University.

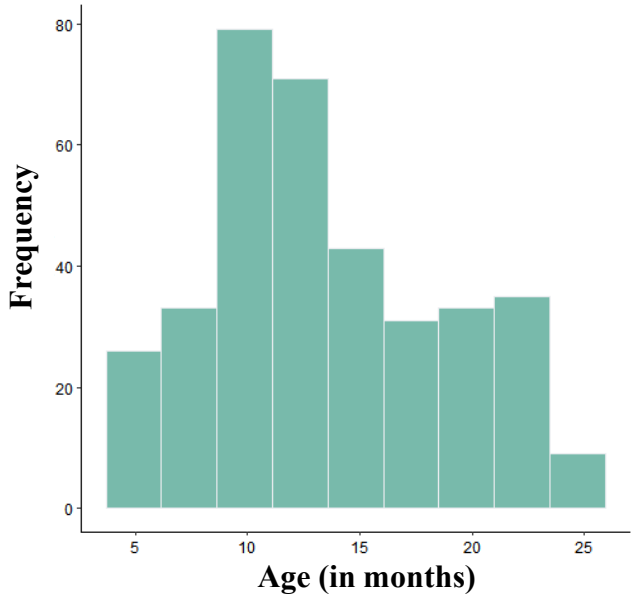

Figure 1. Age distribution of the full sample.

Materials

Aside from the piloted 34 items, the questionnaire also included two items directly addressing the construct: an item asking about the child's curiosity directly ("I would describe my child as curious") and one item about the child's curiosity in comparison to their peers (on an ad-hoc five-point scale from 1 ("a lot less curious") to 5 ("a lot more curious")). Additionally, respondents were asked to provide demographic information (prematurity, developmental concerns, country, languages spoken, birth order, and socio-economic status (SES) via their educational level; Singh et al., 2023), and could optionally contribute qualitative

responses regarding additional behaviours in which they saw their child's curiosity manifested (exact items in Supplementary Materials S1.3).

Procedure

Primary caregivers across the UK were invited to complete this online questionnaire on Qualtrics (Qualtrics, Provo, UT). Participating caregivers were provided with an information sheet before giving informed consent and received instructions to think about their child's typical behaviour to evaluate each statement. They indicated their child's age and sex after which the items were presented in a randomised order with four to five items per page. These were followed by the two items assessing direct curiosity, optional qualitative responses, and demographic questions. Lastly, caregivers were given the opportunity to sign up to receive an automatic email a week later for a re-test. It took most respondents under 15 minutes (M = 11.43, SD = 5.45) to complete the survey.

Analyses

All analyses were conducted in R (Version 4.1.2).

Item reduction and emerging sub-factors

The scale was designed to measure infants' and toddlers' trait curiosity as one general construct represented in the tendencies with which the infant explores their surroundings. Thus, the aim of the exploratory factor analysis was to reduce the item list to coherently capture this construct, as well as to better understand the additional variance in the data. This analysis required initial steps (following Pett et al., 2003) of identifying possibly ill-fitting items and assessing the scale's sampling adequacy and factorability. We then fitted a unidimensional structure eliminating items which did not sufficiently load onto the general curiosity factor (< .32). Subsequently, we explored how many sub-factors the items grouped into to further investigate the scale's dimensionality (Dunn & McCray, 2020), where the number of sufficient factors was indicated using a scree plot and a parallel analysis. For the two latter steps, we conducted Exploratory Factor Analyses (function "fa" from the *psych* package) using default minimum residual factoring (*minres*) and oblique rotation allowing for sub-factors to correlate. The *minres* factoring method is recommended for questionnaire data (Fabrigar et al., 1999) since it gives robust estimates even for skewed items, whereas correlation among emerging sub-factors was justified by our theoretical assumption of one underlying construct.

Structural validity

Having multiple items can lead to an emergent sub-factor structure if item topics and wordings result in correlated response behaviour. The current scale, however, was constructed to measure a general factor of trait curiosity due to lacking a strong theoretical basis for assuming multi-dimensionality of curiosity in early childhood. Here, we explored the scale's dimensionality to support the computation of an overall curiosity mean score (Dunn & McCray, 2020) by fitting the previously identified sub-factor structure using Structural Equation Modelling (SEM; function "sem" from the *lavaan* package; Rosseel, 2012). This allowed us to compare a unidimensional model to a correlational sub-factors model. Note that we initially also fitted a bi-factor model which, however, was considered less interpretable and was consequently removed from this report. Due to the expected occurrence of missing values (N/A responses for not yet or not recently observed behaviours), models were estimated using the robust full information maximum likelihood (estimator = "MLR" in lavaan). Model comparison included standard indices such as chi-square, comparative fit index (CFI), Root Mean Square Error of Approximation (RMSEA), and Bayesian Information Criterion (BIC).

Internal consistency

On the final set of items, commonly reported measures of internal consistency were computed, namely Coefficient alpha (also referred to as Cronbach's alpha, McNeish, 2018), Revelle's Omega total, and Guttman's lambda-6 (using "omega" from the *psych* package), as well as Revelle's coefficient Beta. Revelle's Beta estimates how much variation in the data can be attributed to some general underlying factor (Cooksey & Soutar, 2006), so that a general factor can be argued at beta values above .50, whereas values above .70 are recommended (John & Roedder, 1981; Revelle, 1979; Rossiter, 2002). Regarding the other three measures, values above .80 are considered good (but are said to be acceptable above .70 when not meant for diagnostic decisions; Pett et al., 2003) and the average Inter-item correlation is ideally between .20 and .40 (Piedmont, 2014). Additionally, we mentioned item-response-theory in the pre-registration, but this was omitted from this report due to limited informativeness above and beyond the here reported results.

Exploration of demographic differences

We explored whether mean scores systematically differed across sex, age and SES (as approximated via the mother's education), to inform future applications of the scale, using multiple linear regression. Mean scores were computed on the final set of items, with unobserved items (N/A responses) not affecting individual scores.

First, however, structural equation models were tested in three steps regarding measurement invariance across sexes: the first step tests whether factor loadings are the same between groups (metric invariance), the second adds constraints on the intercepts (scalar invariance) and the third on residual variances (strict invariance). Invariance is achieved at each step if adding the respective constraints does not significantly worsen model fit. If it does, however, specific parameters need to be inspected and relaxed to achieve partial invariance, and their impact on observed scores considered and discussed (Vandenberg & Lance, 2000).

Lastly, we computed correlations between the overall mean scores and the two additional curiosity items as initial indications of construct validity, expecting positive relationships.

Results

Item reduction and emerging sub-factors

A schematic overview of item reduction process is presented in Figure 2. We first reverse coded the three negatively formulated items and computed a correlation matrix (using Spearman's rho) to investigate ill-fitting ones (e.g., Prett et al., 2003). This led to the exclusion of two items which correlated negatively with many of the rest (new foods & mouthing; Supplementary Materials S1.2). Then we ensured that the data was adequately sampled and factorable as indicated by a significant Bartlett's test of sphericity ($X^2(496) = 3474.50$, $p < .001$), a non-zero, positive matrix's determinant (.00006), and the Kaiser-Meyer-Olkin statistic above 0.70 for both, the overall sample (KMO = 0.89) as well as each individual item (lowest measure of sampling adequacy (MSA) = 0.82).

We then fitted a one-factor model on the remaining 32 items to further reduce the item list to only those loading onto a general factor. In this way, six additional items (Supplementary Materials S1.2) with loadings below the recommended .32 (Tabachnick & Fidell, 2007) were eliminated. The subsequent exploration of emerging factors only included the 26 remaining items that loaded strongly and positively onto this general factor. Drawing a line through the lower values in the scree plot (Cattell, 1966; Pett et al., 2003; Figure S1 in Supplemental Materials) suggested 3 or 4 factors. Similarly, the parallel analysis indicated four factors,

whereas the "elbow" of both graphs already occurred at factor 3. Accordingly, we fitted a 3-factor and a 4-factor model.

The 4-factor model had overall better fit values (TLI = .89, RMSEA = .048, BIC = -921.22) than the 3-factor model (TLI = .87, RMSEA = .052, BIC = -978.39). However, upon inspection of the emerging sub-factors, both models were highly similar except for an additional two-item factor in the 4-factor model. As it has been argued that a meaningful factor should consist of at least three items (Hair et al., 2010), we decided to continue with the 3-factor model. Table 1 shows the items' descriptives and subfactor-loadings in comparison to their loadings on the general factor. Overall, items were positively rated and only items on the third subfactor were within the ideal range of item difficulty. However, these items were also accumulating most NA responses indicating behaviours many infants had not recently or not yet expressed.

Table 1. Final questionnaire items (plus three excluded items), their descriptives (Mean, Standard deviation, Proportion NA, and Item difficulty) and Exploratory Factor loadings.

Item		M*(SD, Median)*	NA	Difficulty	General Factor	3-Factor Model		
						1	2	3
S1	When my child encounters an object, they typically seem interested in its properties (e.g., how it feels, tastes, or sounds like, etc.).	6.2 *(.81, 6)*	.00	.88	.34	**.61**		
S2	My child actively inspects a variety of objects, whether it be toys or ordinary household items.	6.6 *(.63, 7)*	.00	.94	.40	.50		
S3	My child usually inspects objects from all angles and sides.	5.6 *(1.15, 6)*	.01	.80	.44	.51		
S4	My child pokes at and probes objects to see how they feel.	6.0 *(1.01, 6)*	.02	.86	.43	.51		
S5	My child is interested in a wide variety of objects.	6.5 *(.70, 6)*	.00	.93	.41	.52		
S6	My child actively seeks out and enjoys new experiences.	5.8 *(1.05, 6)*	.05	.83	.53	.45		
S7	My child shows visible enjoyment (e.g., smiling, gurgling, babbling) when discovering something new.	6.3 *(.83, 6)*	.00	.90	.40	.44		
S8	My child enthusiastically explores new environments (e.g., a new house, the beach, etc.).	5.9 *(1.10, 6)*	.08	.85	.47	.43		

S9	My child likes to look around, scanning the environment for something new.	6.1 *(.94, 6)*	.00	.87	.39	.39		
Item		M*(SD, Median)*	NA	Difficulty	General Factor	3-Factor Model		
						1	2	3
S10	My child is interested in what other people next to them are doing. For example, when someone prepares food, my child closely observes their every move.	6.0 *(1.00, 6)*	.02	.86	.42	.37		
SR	My child does ***not*** typically engage with (look at, point at, reach for, inspect) a lot of things in their environment.	6.4 *(.88, 7)*	.01	.91	.43	.36		
Iv1	When my child looks into a container (e.g., a bag, kitchen drawer, etc.), they take out and inspect each of its contents.	6.1 *(1.10, 6)*	.06	.87	.57		**.70**	
Iv2	When something is hidden from my child (e.g., in closed boxes, rooms, cupboards etc.), they will actively try to uncover it.	6.0 *(1.21, 6)*	.09	.86	**.64**		.66	
Iv3	When I open my bag in front of my child, they will come and peek into it.	6.1 *(1.15, 6)*	.11	.87	.54		.64	
Iv4	My child often bangs objects to see what noise they make.	6.3 *(1.10, 7)*	.02	.89	.32		.40	
Iv5	If a toy has multiple functions, my child will typically discover and play with more than one of them.	5.6 *(1.28, 6)*	.05	.80	.62		.33	
Ia1	My child often leads me to/brings me things that they want to know more about.	5.2 *(1.56, 6)*	.21	.75	.55			**.77**
Ia2	When reading a picture book together, my child directs me (e.g., by pointing) towards what they want to know more about.	4.7 *(1.87, 5)*	.19	.68	.54			.71
Ia3	When we are in a new environment (e.g., the zoo, a shop, etc.), my child keeps pointing at all the things they find interesting.	5.1 *(1.87, 6)*	.18	.73	.53			.66
Ia4	When faced with a problem, my child will seek the help of others in order to solve it.	4.9 *(1.41, 5)*	.15	.70	.46			.61
Ia5	When faced with a problem (e.g., fitting a block into its respectively shaped hole), my child typically keeps trying to figure it out until they have solved it.	4.6 *(1.60, 5)*	.16	.66	.53			.38
Ia6	When someone shows my child how something works, they watch with continuous interest.	5.7 *(1.06, 6)*	.04	.81	.54			.36
Ia7	*When my child is confused by something, they look at me/another person for additional information.*	*5.7 (.98, 6)*	*.06*	*.82*	*.55*			

Ex1	*My child starts playing on their own, rather than waiting to be given something to play with.*	*6.1 (1.01, 6)*	*.02*	*.87*	*.32*
Ex2	*When my child plays with an assembly toy (e.g., building blocks, puzzle, a toy with detachable parts), they like to take it apart for further examination.*	*5.8 (1.25, 6)*	*.13*	*.83*	*.51*
Ex3	*When playing hide and seek, my child enjoys searching for the object or person that disappeared.*	*6.2 (.87, 6)*	*.16*	*.88*	*.42*

Notes. Items are ordered by subfactors and loadings on those subfactors (S = Sensory, Iv = Investigative, Ia = Interactive, Ex = Excluded). Item difficulty captures the likelihood of receiving the maximum score so that higher values indicate lower difficulty (ideally between .5 and .8 which is only given for the third subfactor). "NA" indicates the sample proportion of NA responses. Factor loadings smaller than .32 are not reported, the strongest loading per factor is marked in bold. Italicised items are those that loaded onto the general factor but did not load sufficiently onto any sub-factor.

The first sub-factor (11 items) could be labelled as *Sensory* as it includes items regarding more general manual and visual exploratory behaviours. The second subfactor (5 items) could be labelled as *Investigative* including items indicating a tendency to autonomously manipulate objects in their environment to gain information. The third subfactor (6 items) could be labelled as *Interactive* capturing to what degree the child uses and interacts with social partners to receive additional information. These labels aim to best describe each subfactor's contents especially considering the strongest loading items; but note item Ia5 as an exception.

Four items did not sufficiently load onto any of these subfactors. On inspection, three of these (Ex1, Ex2, Ex3) concerned play behaviour more so than exploration which may explain their distinctness. However, item Ia7 had one of the strongest loadings toward the general factor and also loaded onto the Interactive subfactor in the 4-factor model (at .41). As exploratory factor analysis is not meant to be a purely data-driven process, we decided to keep this item given its strong contribution to the general factor and its contextual fit with the Interactive subfactor (now 7 items) while excluding the other three. Thus, all subsequent analyses were conducted on the final set of 23 items.

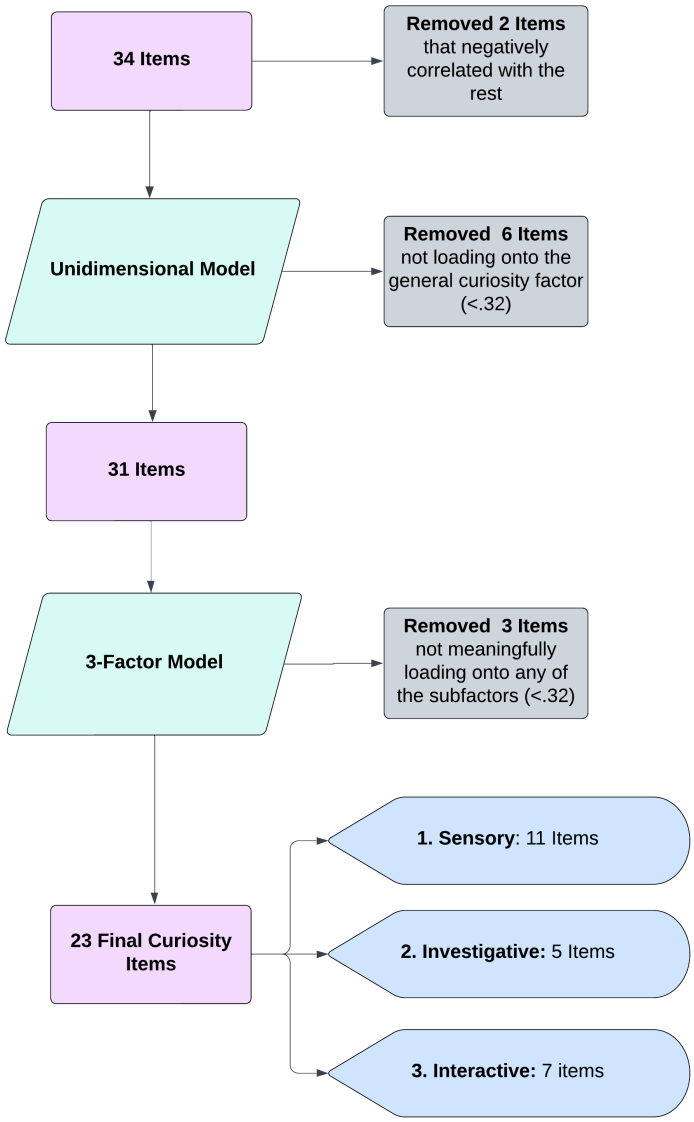

Figure 2. Schematic overview of the item reduction process.

Structural validity

We fitted a unidimensional and a correlational structural equation model to investigate the scale's dimensionality and better understand how the items are structured. The model of three correlating subfactors had a significantly better model fit than the unidimensional model ($\Delta F(3, 230) = 125.14$, $p < .001$) yet was not meeting the general criteria of acceptable fits. Thus, we considered computationally suggested modifications with the top three being correlated residuals between items S6 and S8, S10 and Ia6, as well as Ia2 and Ia3. With these modifications, the correlational model showed acceptable fit measures (Table 2). The chi-square test was significant in both models but this is to be expected at larger sample sizes (Schermelleh-Engel et al., 2003; Vandenberg, 2006). Expectedly, the subfactors were strongly

correlated (Sensory ~ Investigative: r = 0.64, Sensory ~ Interactive: r = .51, Investigative ~ Interactive: r = .75).

Table 2. Indices for the specified models using Structural Equation Modelling (SEM).

Model	Chi-Square (df, N)	CFI[a]	RMSEA[b]	AIC[c]	BIC[d]	adj. BIC[e]
Unidimensional	χ^2(230,370)=703.98***	.71	.075 [.069;.080]	22468.58	22738.61	22519.69
Correlational	χ^2(224,370)=388.63***	.90	.045 [.038;.051]	22098.38	22391.89	22153.95

Note: The correlational model includes the three modifications reported in structural validity section. *** p < .001; **a:** Comparative Fit Index (preferably ≥ .90); **b:** Root Mean Square Error of Approximation (preferably ≤ .060) and [95% Confidence Intervals]; **c:** Akaike Information Criterion; **d:** Bayesian Information Criterion; **e:** Sample-size adjusted BIC (all: smaller better)

Internal consistency

We computed common measures of internal consistency for both the complete scale and the emergent subfactors (Table 3). The overall scale was found to have high internal consistency, where a Revelle's beta of >.70 additionally supports our assumption of a general underlying factor. Furthermore, the separate subfactors also had good (>.70) indices supporting these data-driven options to explore additional variance in the sample. Consequently, we will refer to them as subscales.

Table 3. Measures of internal consistency for the full scale and the emergent subscales.

Scale	Number of Items	Coefficient α	Lambda-6	Revelle's omega total	Revelle's β
General	23	.87	.90	.89	.71
Sensory	11	.78	.78		
Investigative	5	.74	.71		
Social/Interactive	7	.81	.80		

Exploration of demographic differences

The mean curiosity scores were distributed around an average of M = 5.83 (SD = .59), evidencing that the scale captures variance in reported exploration tendencies (Figure 3). To inspect demographic differences in mean scores, we first tested measurement invariance

(metric, scalar, strict) across sex, that is, whether factor loadings, intercepts, and residual variances of each item in the correlational SEM model are the same for male and female infants. One factor loading (Sensory =~ S2) was relaxed to achieve partial metric invariance, beyond which also scalar and strict invariance were achieved (all model comparisons $p > .05$, indicating that additional constraints did not worsen the model fit). The differing item S2 concerns inspection of various types of objects as part of the Sensory subscale. However, we see no concern for the computation and comparison of mean scores across sexes because of the following three reasons: 1) This item has limited impact as it is one of 11 items in that subscale, 2) it did not generate significantly different scores between sexes ($W = 16812$, $p = .812$), and 3) because scalar invariance was achieved. Overall, this test supports the ITCQ's applicability across sexes.

Multiple linear regressions showed that age ($b = .04$, $p < .001$) but neither sex ($b_m = .06$, $p = .305$) nor SES ($b = -.02$, $p = .453$) predicted curiosity scores ($F(3, 324) = 18.05$, $p < .001$). As we did not expect age to be a significant predictor, we exploratorily investigated these relations for each of the subscales and found the same pattern, while revealing that the age effect was mainly driven by the Interactive subscale scores ($b = .10$, $p < .001$, $F(3, 321) = 38.02$, $p < .001$), somewhat smaller for the Investigative subscale scores ($b = .07$, $p < .001$, $F(3, 322) = 20.93$, $p < .001$) and non-significant for Sensory subscale scores ($b = .01$, $p = .092$, $F(3, 324) = 2.03$, $p = .120$). Additionally, we found that the age effect on mean scores disappeared from the age of 13 months ($b = .02$, $p = .087$). These subscale patterns are in line with those for item difficulties and proportion of NA responses (Table 1), the latter of which also significantly decreased with age ($b = -.31$, $p < .001$) but stabilised around the age of 14 months ($b = -.03$, $p = .066$).

Lastly, the overall mean scores and the two curiosity items were positively correlated (direct item: $r_s = .36$, $p < .001$; comparative item: $r_s = .21$, $p < .001$), offering a first indication of construct validity. It is to be noted, however, that these single items are not validated measures themselves and therefore cannot reliably capture differences in curiosity (e.g., median value for the direct item was 7 out of 7). Other potential reasons for these lower correlations are that directly asking about curiosity is more prone to inflict response bias or that parents have a broader meaning of curiosity in mind when responding to these single-item questions. Thus, these results may rather be considered as *initial* checks regarding the construct.

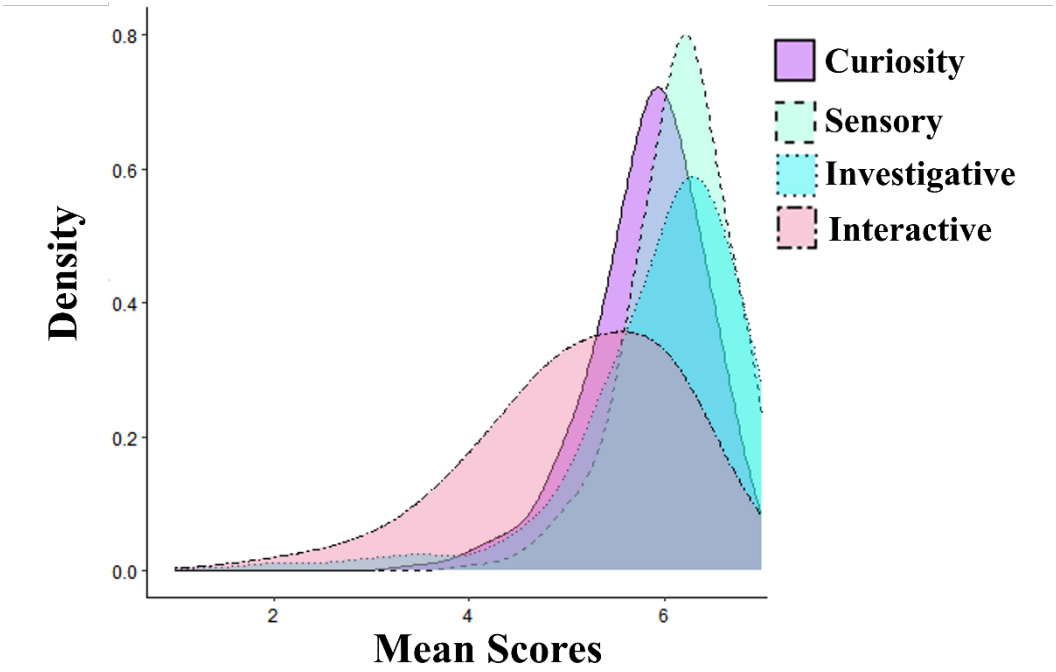

Figure 3. Mean score density plots for overall curiosity as well as the subscale scores.

Interim Discussion

In this first study, we described the development of the ITCQ from creating an initial set of items capturing various behaviours with which infants and toddlers can explore and interact with their environment to the final set of 23 items comprising this internally consistent and structurally valid measure of trait curiosity. An initial pilot study provided evidence for the scale's content validity. The subsequent main analysis included 370 responses across an age range from 5 to 24 months and offered sufficient evidence to justify the computation of an overall mean score, along with three emergent subscales which could explain additional variance in the data. These subscales seemed to capture different types of curiosity manifesting in broader, sensory exploration, more focused, investigative exploration, and interacting with social others to gain information. While high internal consistency for the full scale further supported the computation of an overall curiosity score, good indicators for the subscales made these a valid option for exploratory analyses.

We then confirmed the scale's applicability across sexes and found that mean scores were only predicted by age but neither sex nor SES (but note that SES was rather homogenous in this sample). The age-effect could be due to the scale items covering various exploratory behaviours which develop across the first two years of life and are thus bound to increase with

time, making parental observations more robust and caregivers more confident in their reported agreement. The differential correlations between age and the subscales (Sensory < Investigative < Interactive), in line with patterns of NA responses and item difficulty, as well as the disappearance of this effect from 13 months of age onward lend support to this notion. Therefore, this effect should be controlled for in cross-sectional studies including younger infants, for which the Sensory subscale would be most meaningful.

It is to be noted that a large number of responses concerned infants in the range of 10-12 months of age, which may have biased the item selection. This age group is somewhat in the middle of the full age-range and part of the group in which age-effects were observed, so that in fact, the selection would have been biased towards items that are applicable from earlier in development (making up around half of the scale as represented by the Sensory subscale). Nevertheless, longitudinal studies are needed to explore the developmental trajectory of item scores, for which we would expect to find rank stability (more curious children stay more curious) to support the scale's temporal consistency.

Study 2: Test-retest Reliability

Another important aspect of a scale's validation is test-retest reliability which indicates the clarity of the items via the responses' temporal stability (Crocker & Algina, 1986). If items are well constructed to capture observable behaviour that reflects the child's general tendencies, the responses should be consistent with each other. Here, the retest timeframe was set to 7 to 14 days so that participants were unlikely to remember their previous responses and the child would not have experienced a leap in behavioural development.

Methods

Participants

As mentioned previously, participants who provided consent and email address at the end of the survey were automatically contacted through Qualtrics one week after their initial response. From the participants included in the main analysis in Study 1, we collected $N = 67$ test-retest responses completed within 7-14 days ($M = 7.61$, $SD = 1.19$) of the first measurement. Three additional responses were excluded due to longer timeframes (18, 46, and 144 days, respectively). Babies of the final 67 responders were typically developing and representative of the full sample (age in months at first timepoint: $M = 12.7$, $SD = 5.1$, range: 5.1 – 24.2; 58% female). Caregivers provided consent to proceed to the questionnaire items

and completed this second measurement without any additional reward or compensation. Responses were matched via anonymous identification numbers, imbedded in the automatic emails.

Materials

The test-retest version of the full questionnaire included the original 36 items as well as the two curiosity questions. However, we conducted all analyses using only the final 23 items based on the results from Study 1.

Analyses

Test-retest reliability was investigated in two ways: how consistently participants responded to each item (using the function "testRetest" from the *psych* package; (Revelle, 2023), as well as the Intraclass Correlation Coefficient (ICC) of mean scores (using the function "icc" from the *irr* package; Gamer et al., 2019). The first analysis implemented the data's multi-level structure to provide reliability indices for items and participants over time and indicated variance for each of these components and their interactions. Furthermore, we specified the ICC of mean scores as a two-way mixed effect model with absolute agreement and single unit analysis as suggested by the literature (e.g., Koo & Li, 2016). Historically, ICC scores have been considered as *poor* at values smaller than .5, as *moderate* between .5 and .75, as *good* between .75 and .9, and as *excellent* above .9.

Results

We found good internal consistency at both timepoints (T1: Coefficient **α** = .87, **λ-6** = .94; T2: **α** = .88, **λ-6** = .96), indicating that the items correlated with each other to a similar extent. Furthermore, item scores were correlated across measurements at $r = .86$ ($p < .001$). The mean within-subject test-retest reliability of response patterns over items and time was good (*rqq* = .79) as was the reliability of all ratings across items and times (*RkF* = .97) (Revelle, 2023; Shrout & Lane, 2012). Multilevel components of variance further showed that most of the variance in scores could be attributed to the items (44%), participants (13%), and the interaction between items and participants (23%). Little to no variance, however, was attributed to time effects (time: 0%; participant*time interaction: 0.1%; items*time interaction: 0%). This suggests that participants responded to items with sufficient temporal stability to support the

scale's test-retest reliability. A good ICC of mean scores seconded this finding (ICC(A, 1) = .82; $F(66,47.9) = 11.3$, $p < .001$; 95%CI=[0.72; 0.89]).

Interim Discussion

Both measures supported the scale's temporal stability indicating that the items were well constructed to allow for reliable responses. It is to be noted, however, that this sample was self-selected and might represent highly motivated responders. Future research should aim to recruit a more representative sample and explore different timespans between measurements for a more in-depth investigation.

Study 3: Criterion Validity

Another source of validity evidence is the measure's relationship to other variables (Downing, 2003). This includes correlating them with scores of other existing measures with well-known characteristics. Therefore, we decided to compare the new scale to facets of temperament as we would expect them to be related yet distinct. While the temperament scales mostly capture how the child typically reacts to certain situations, the ITCQ mostly captures infant-initiated exploratory behaviours. As behavioural expressions may well be affected by how the child reacts to certain situations, we expected the temperamental facets to differentially correlate with the curiosity scores. Furthermore, it had previously been stated that an established relation between a curiosity measure and temperament would support the notion of it capturing curiosity as a trait (Piotrowski et al., 2014).

Temperament is viewed as an early equivalent to adult personality traits, and its measures (Infant Behavior Questionnaire or IBQ; Early Child Behavior Questionnaire or ECBQ) have been shown to be reliable, valid, and informative both in personality related research but also for predicting behavioural outcomes (e.g., Putnam et al., 2008; Rothbart, 1986, 2011), making them appropriate measures for exploring the ITCQ's criterion validity. In adults, positive accounts of curiosity similar to our conceptualisation (e.g., Interest-Type Epistemic Curiosity, Litman, 2008; CEI Exploration Subscale, Kashdan et al., 2009) have been linked to higher Extraversion and Conscientiousness as well as lower Neuroticism (e.g., Hunter et al., 2016; Kashdan et al., 2018). Consequently, similar relations between the ITCQ and the equivalent temperament subscales would support the measure's validity and extend these findings into infancy.

Methods

Participants

From the participants included in the main analysis in Study 1, N = 75 caregivers (children's age in months: M = 14.1, SD = 4.5, range: 6.5-24.2; 50.7% female) additionally completed the temperament survey, two thirds of which indicated having a degree in higher education. Most of these respondents were recruited directly from the Babylab's database, and thus, received £5 as reimbursement in the form of an online gift voucher of their choice (n = 55). The rest (n = 20) completed the temperament survey without additional rewards after their in-person study visit for which they had received £5 travel reimbursement and a book for the child. Participants provided written consent prior to answering any questions.

Materials

While the full temperament measure consists of around 200 items across multiple facets of temperament, the "very short form" versions (IBQ-vsf and ECBQ-vsf) each consists of 36 items evaluated on a 7-point Frequency-scale from 1 ("Never") to 7 ("Always") and an option of "NA - not applicable", which have been validated to capture three broader dimensions: Surgency, Negative Affect, and Effortful Control (e.g., Putnam et al., 2010, 2014). Surgency items capture facets such as Approach, High Intensity Pleasure, Activity Level, and Perceptual Sensitivity, making this factor comparable to the personality dimension of *Extraversion*. Negative Affect items capture levels of Sadness, Distress to Limitations, and Fear, making this factor comparable to the personality dimension of *Neuroticism*. Lastly, Effortful Control items capture Duration of Orienting, and levels of Low Intensity Pleasure, Cuddliness, and Soothability, making this factor comparable to the personality dimension of *Conscientiousness*.

Participants first completed the full ITCQ, followed by either the very short form of the IBQ or ECBQ depending on the child's age: IBQ if the child was between 5 and 12 months old and the ECBQ for ages 13 months and over.

Hypotheses

We expected Surgency to positively correlate with curiosity, as a more extraverted child may exhibit more exploratory behaviours across contexts. Second, we expected Negative Affect to negatively correlate with curiosity, as a more fearful and distressed child may exhibit fewer exploratory behaviours across contexts. Lastly, we did not have a clear prediction on how Effortful Control may correlate with curiosity but could hypothesise a positive relation

with longer exploratory engagement in line with links found in adults (e.g., Hunter et al., 2016; Kashdan et al., 2018).

Analyses

We computed mean scores for all scales (exploratorily also for the curiosity subscales) and conducted Spearman correlations between the temperament and curiosity scores. We treated scores from the IBQ and ECBQ equally, as items form into the same three dimensions and because of their assessed longitudinal stability (Putnam et al., 2008; Rothbart, 1986).

Results

Correlations between the facets of temperament and curiosity are shown in Table 4. We found significant, positive correlations of moderate effect sizes between both surgency as well as effortful control and the mean curiosity score. Exploratory correlations revealed these to be strongest for the Sensory subscale. Additionally, we found a negative correlation between curiosity and negative affect. This relation seemed to be mainly driven by lower scores on the Interactive subscale so that young children reported to be more fearful and distressed were especially unlikely to interact with social others in the pursuit of information.

Table 4.

Spearman correlation matrix between curiosity and temperament mean scores.

	Overall Curiosity	Sensory	Investigative	Interactive
Surgency	**0.39*****	**0.47*****	**0.31****	0.16
Negative Affect	-0.27*	-0.24*	-0.14	**-0.3****
Effortful Control	0.25*	**0.3****	0.17	0.08

Note. $p < .001$ '***', $p < .01$ '**', $p < .05$ '*' with strongest correlations (≥ .3) in bold.

Interim Discussion

We investigated how the ITCQ related to other early traits measures, more specifically facets of temperament, to obtain evidence of its criterion validity. We found significant correlations of moderate effect size between all three temperament dimensions and overall curiosity in line with previous personality links found in adults (e.g., Hunter et al., 2016; Kashdan et al., 2018), whereas the subscales offered additional insights. The negative correlation between Curiosity and Negative Affect, that is being more fearful and distressed, is

furthermore in accordance with previous adult research that showed anxiety to be negatively associated with epistemic curiosity (Collins et al., 2004; Kashdan & Roberts, 2004; Litman & Jimerson, 2004; Litman & Spielberger, 2003; Naylor, 1981). Additionally, we observed the strongest negative correlation with the Interactive subscale which is consistent with the idea that Neuroticism may specifically inhibit social interactions and respective exploratory behaviours (Green & Campbell, 2000). Together, these results provide evidence that curiosity and facets of temperament are related but still capture unique characteristics of the child's personality comparable to links found in adult personality research.

General Discussion

Recognising the need to measure individual differences in trait curiosity in infants and toddlers, we developed the Infant and Toddler Curiosity Questionnaire (ITCQ) as the first caregiver report measure to assess trait curiosity in this targeted age group, with items capturing observable exploration behaviours specific to infants and toddlers between 5 and 24 months of age. Across three studies we reported evidence for the scale's reliability and validity, suggesting that the ITCQ could become a powerful tool for developmental research.

The first study focused on the initial questionnaire development leading to a final set of 23 items, selected based on internal consistency and exploratory factor analyses. Three methodologically emergent subscales captured additional co-variance among the items and developmental exploration skills: sensory curiosity, investigative curiosity, and interactive curiosity as in gaining new information by interacting with social others. The well-fitting correlational model using structural equation modelling and the strong correlations between subfactors offered sufficient support for the computation of an overall mean score. As the full scale but also each of the subfactors had good measures of internal consistency, we considered these subfactors as curiosity subscales. Furthermore, the scale seemed to work the same for male and female infants, but scores did increase with age until around 13 months. Together, this work offers multiple avenues to disentangle effects of trait curiosity but also of more specific types of curiosity manifestations. The second study then showed that the final scale had good test-retest reliability after 7 to 14 days.

Lastly, study three indicated criterion validity as the ITCQ scores were significantly related to the well-established trait measure of temperament (Putnam et al., 2014). Here, we found positive correlations between Curiosity and Surgency, which is considered a precursor of Extraversion, as well as Curiosity and Effortful Control, a precursor of Conscientiousness.

In contrast, Curiosity negatively correlated with Negative Affect, a precursor of Neuroticism. Together, these findings are in line with theoretical considerations as well as previous personality research in adults (e.g., Collins et al., 2004; Hunter et al., 2016; Kashdan et al., 2018; Kashdan & Roberts, 2004; Litman & Jimerson, 2004; Litman & Spielberger, 2003; Naylor, 1981), providing crucial evidence for construct and criterion validity.

Limitations & future research

Questionnaire development is a strenuous process for which there is no gold-standard as evidenced by numerous open discussions. Using the best practices as a guide, we created and assessed the ITCQ to be a reliable and valid measure (Downing, 2003; Pett et al., 2003). Yet, future studies are needed to collect multiple independent samples replicating these findings in more diverse socio-economic populations and across different cultures. Even though the recruitment for the questionnaire study was largely conducted online via social media, and thus had the potential to reach a more representative population, the final sample ultimately was affected by a self-selection bias of mainly highly educated, white responders. Consequently, the ITCQ's generalisability requires further investigation of potential socio-cultural differences. For example, the encouragement of curiosity might be a privilege of those who have the time and resources, as well as the environmental safety to allow it. Studies in Tanzanian and Chinese parents, for example, have indicated that despite cultural norms emphasizing obedience over independence and curiosity, parental education and financial security were strong predictors of more positive and encouraging attitudes towards these constructs (Chuang & Su, 2009; Jukes et al., 2021). Furthermore, it is possible that some items are not applicable across cultures in which children's interactions with social others to seek help and information are less pronounced (Little et al., 2016). However, this research avenue has already gained traction with planned investigations into cultural differences regarding the applicability of the ITCQ in non-western cultures such as Japan, China, and India.

Another limitation concerns the necessity for longitudinal data to establish temporal stability of the trait measure, its developmental trajectories, and its convergent validity to other measures of early curiosity (Lee et al., 2023), problem solving (Hoicka et al., 2023) or observation-based curiosity scores (Fortner-Wood & Henderson, 1997). Nevertheless, our reported studies here suggest that the ITCQ is a promising measure for application in psychological research to potentially explain variance in observed exploration behaviours (e.g., Mandler et al., 1987; Piccardi et al., 2020; Slone et al., 2019; Smith & Yu, 2013) as well as developmental trajectories (e.g., (Berg & Sternberg, 1985; Bornstein et al., 2013; Muentener et al., 2018; Shah et al., 2018). In fact, preliminary reports of our measure have already gained

international interest so that a German, Dutch, and Italian version of the ITCQ are currently undergoing validation, and a child version for 2–5-year-olds has also been developed (Altmann et al., 2023).

Conclusion

In this paper, we present the development of a newly constructed caregiver report questionnaire (ITCQ) and showcase that it effectively captures early exploration tendencies as a manifestation of individual differences in infants' and toddlers' trait curiosity. Importantly, the ITCQ fills an important gap in the scientific landscape of infancy research. Across three studies we demonstrated evidence for the measure's reliability and validity following rigorous practice to ensure that future applications of the ITCQ will offer new and powerful insights into early human development.

Acknowledgements

We thank all families who participated and made this study possible. We would also like to thank the following Babylabs for their support in our online recruitment efforts: Cambridge Babylab, Birkbeck Babylab & Toddlerlab, Birmingham Babylab, University of Essex Babylab, Liverpool Language Lab, Oxford University Babylab, Oxford Brooks Babylab, Plymouth Babylab, UoM Child Study Centre. Gert Westermann was supported by the Economic and Social Research Council (ESRC) International Centre for Language and Communicative Development (LuCiD) [ES/S007113/1 and ES/L008955/1]. The authors have no conflicts of interest to disclose.

Author contributions

Elena Altmann: Conceptualisation; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Visualisation; Writing - original draft; and Writing - review & editing. **Marina Bazhydai:** Conceptualisation; Funding acquisition; Methodology; Supervision; Writing - review & editing. **Didar Karadağ:** Conceptualisation; Methodology; Writing - review & editing. **Gert Westermann:** Conceptualisation; Funding acquisition; Methodology; Supervision; Writing - review & editing.

References

Ackermann, L., Hepach, R., & Mani, N. (2020). Children learn words easier when they are interested in the category to which the word belongs. *Developmental Science*, *23*(3). https://doi.org/10.1111/desc.12915

Adolph, K. E., & Hoch, J. E. (2019). Motor development: Embodied, embedded, enculturated, and enabling. *Annual Review of Psychology*, *70*, 141–164.

Altmann, E. C., Bazhydai, M., & Westermann, G. (2025). Curious Choices: Infants' moment-to-moment information sampling is driven by their exploration history. *Cognition*, *254*, 105976. https://doi.org/10.1016/j.cognition.2024.105976

Altmann, E. C., de Boer, E., Eiteljoerge, S., Bazhydai, M., Hunnius, S., Bounia-Mastrogianni, P., Meyer, M., Karadağ, D., Braithwaite, E. K., & Westermann, G. (2023, August 23). *The Development of the Early Child Curiosity Questionnaire (ECCQ)*. [Poster Presentation] 8th Lancaster Conference on Infant & Early Child Development, Lancaster, UK.

Artino Jr, A. R., La Rochelle, J. S., Dezee, K. J., & Gehlbach, H. (2014). Developing questionnaires for educational research: AMEE Guide No. 87. *Medical Teacher*, *36*(6), 463–474.

Bazhydai, M., Twomey, K., & Westermann, G. (2021). Curiosity and Exploration. In J. B. Benson (Ed.), *Encyclopedia of Infant and Early Childhood Development* (2nd ed; Issue 2nd ed). Elsevier. https://eprints.lancs.ac.uk/id/eprint/136590/

Bazhydai, M., Westermann, G., & Parise, E. (2020). “I don’t know but I know who to ask”: 12-month-olds actively seek information from knowledgeable adults. *Developmental Science*, *23*(5), e12938. https://doi.org/10.1111/desc.12938

778 Begus, K., Gliga, T., & Southgate, V. (2014). Infants Learn What They Want to Learn:
779 Responding to Infant Pointing Leads to Superior Learning. *PLoS ONE*, *9*(10), e108817.
780 https://doi.org/10.1371/journal.pone.0108817
781 Berg, C. A., & Sternberg, R. J. (1985). Response to novelty: Continuity versus discontinuity in
782 the developmental course of intelligence. *Advances in Child Development and*
783 *Behavior*, *19*, 1–47.
784 Berlyne, D. E. (1960). *Conflict, arousal, and curiosity* (pp. xii, 350). McGraw-Hill Book
785 Company. https://doi.org/10.1037/11164-000
786 Boley, B. B., Jordan, E., & Woosnam, K. M. (2021). Reversed polarity items in tourism scales:
787 Best practice or dimensional pitfall? *Current Issues in Tourism*, *24*(4), 466–478.
788 Bornstein, M. H., Hahn, C.-S., & Suwalsky, J. T. D. (2013). Physically Developed and
789 Exploratory Young Infants Contribute to Their Own Long-Term Academic
790 Achievement. *Psychological Science*, *24*(10), 1906–1917.
791 https://doi.org/10.1177/0956797613479974
792 Bornstein, M. H., & Putnick, D. L. (2012). Stability of Language in Childhood: A Multi-Age,
793 -Domain, -Measure, and -Source Study. *Developmental Psychology*, *48*(2), 477–491.
794 https://doi.org/10.1037/a0025889
795 Byman, R. (2005). Curiosity and sensation seeking: A conceptual and empirical examination.
796 *Personality and Individual Differences*, *38*(6), 1365–1379.
797 https://doi.org/10.1016/j.paid.2004.09.004
798 Can, D. D., Ginsburg-Block, M., Golinkoff, R. M., & Hirsh-Pasek, K. (2013). A long-term
799 predictive validity study: Can the CDI Short Form be used to predict language and early
800 literacy skills four years later? *Journal of Child Language*, *40*(4), 821–835.
801 https://doi.org/10.1017/S030500091200030X

802 Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral*
803 *Research*, *1*(2), 245–276. https://doi.org/10.1207/s15327906mbr0102_10
804 Chen, X., Twomey, K. E., & Westermann, G. (2022). Curiosity enhances incidental object
805 encoding in 8-month-old infants. *Journal of Experimental Child Psychology*, *223*,
806 105508. https://doi.org/10.1016/j.jecp.2022.105508
807 Chuang, S. S., & Su, Y. (2009). Do we see eye to eye? Chinese mothers' and fathers' parenting
808 beliefs and values for toddlers in Canada and China. *Journal of Family Psychology*,
809 *23*(3), 331–341. https://doi.org/10.1037/a0016015
810 Collins, R. P., Litman, J. A., & Spielberger, C. D. (2004). The measurement of perceptual
811 curiosity. *Personality and Individual Differences*, *36*(5), 1127–1141.
812 https://doi.org/10.1016/S0191-8869(03)00205-8
813 Colombo, J., Mitchell, D. W., Coldren, J. T., & Freeseman, L. J. (1991). Individual Differences
814 in Infant Visual Attention: Are Short Lookers Faster Processors or Feature Processors?
815 *Child Development*, *62*(6), 1247–1257. https://doi.org/10.2307/1130804
816 Cooksey, R. W., & Soutar, G. N. (2006). Coefficient beta and hierarchical item clustering: An
817 analytical procedure for establishing and displaying the dimensionality and
818 homogeneity of summated scales. *Organizational Research Methods*, *9*(1), 78–98.
819 Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory.* Holt,
820 Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887 ($44.75).
821 https://eric.ed.gov/?id=ed312281
822 Day, H. I. (1971). *Intrinsic Motivation: A New Direction in Education.*
823 Downing, S. M. (2003). Validity: On the meaningful interpretation of assessment data. *Medical*
824 *Education*, *37*(9), 830–837.

825 Dunn, K. J., & McCray, G. (2020). The Place of the Bifactor Model in Confirmatory Factor
826 Analysis Investigations Into Construct Dimensionality in Language Testing. *Frontiers*
827 *in Psychology*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyg.2020.01357
828 Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use
829 of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3),
830 272.
831 Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H.
832 E., & Paradise, J. L. (2005). Concurrent and Predictive Validity of Parent Reports of
833 Child Language at Ages 2 and 3 Years. *Child Development*, *76*(4), 856–868.
834 https://doi.org/10.1111/j.1467-8624.2005.00882.x
835 Fenson, L. (2002). *MacArthur Communicative Development Inventories: User's guide and*
836 *technical manual*. Paul H. Brookes.
837 Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., Tomasello, M.,
838 Mervis, C. B., & Stiles, J. (1994). Variability in Early Communicative Development.
839 *Monographs of the Society for Research in Child Development*, *59*(5), i–185.
840 https://doi.org/10.2307/1166093
841 Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2006).
842 *MacArthur-Bates Communicative Development Inventories, Second Edition*.
843 https://doi.org/10.1037/t11538-000
844 Flake, J. K., & Fried, E. I. (2020). Measurement Schmeasurement: Questionable Measurement
845 Practices and How to Avoid Them. *Advances in Methods and Practices in*
846 *Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393
847 Fortner-wood, C., & Henderson, B. B. (1997). Individual Differences in Two-Year-Olds'
848 Curiosity in the Assessment Setting and in the Grocery Store. *The Journal of Genetic*
849 *Psychology*, *158*(4), 495–497. https://doi.org/10.1080/00221329709596686

Franchak, J. M., Heeger, D. J., Hasson, U., & Adolph, K. E. (2016). Free Viewing Gaze Behavior in Infants and Adults. *Infancy*, *21*(3), 262–287. https://doi.org/10.1111/infa.12119

Gamer, M., Lemon, J., & Singh, I. F. P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement* (Version 0.84.1) [Computer software]. https://cran.r-project.org/web/packages/irr/index.html

Goupil, L., & Proust, J. (2023). Curiosity as a metacognitive feeling. *Cognition*, *231*, 105325. https://doi.org/10.1016/j.cognition.2022.105325

Green, J. D., & Campbell, W. K. (2000). Attachment and Exploration in Adults: Chronic and Contextual Accessibility. *Personality and Social Psychology Bulletin*, *26*(4), 452–461. https://doi.org/10.1177/0146167200266004

Grossnickle, E. M. (2016). Disentangling Curiosity: Dimensionality, Definitions, and Distinctions from Interest in Educational Contexts. *Educational Psychology Review*, *28*(1), 23–60. https://doi.org/10.1007/s10648-014-9294-y

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of Curiosity Modulate Hippocampus-Dependent Learning via the Dopaminergic Circuit. *Neuron*, *84*(2), 486–496. https://doi.org/10.1016/j.neuron.2014.08.060

Hair, J. F., Black, W. C., & Babin, B. J. (2010). *Multivariate Data Analysis: A Global Perspective*. Pearson Education.

Hardy, J. H., Ness, A. M., & Mecca, J. (2017). Outside the box: Epistemic curiosity as a predictor of creative problem solving and creative performance. *Personality and Individual Differences*, *104*, 230–237. https://doi.org/10.1016/j.paid.2016.08.004

Harty, H., & Beall, D. (1984). Toward the development of a children's science curiosity measure. *Journal of Research in Science Teaching*, *21*(4), 425–436. https://doi.org/10.1002/tea.3660210410

875 Hoicka, E., Powell, S., Rose, S. E., Reindl, E., & Tennie, C. (2023). The Early Independent
876 Problem Solving Survey (EIPSS): Its psychometric properties in children aged 12–47
877 months. *Cognitive Development*, *68*, 101366.
878 https://doi.org/10.1016/j.cogdev.2023.101366
879 Hunter, J. A., Abraham, E. H., Hunter, A. G., Goldberg, L. C., & Eastwood, J. D. (2016).
880 Personality and boredom proneness in the prediction of creativity and curiosity.
881 *Thinking Skills and Creativity*, *22*, 48–57. https://doi.org/10.1016/j.tsc.2016.08.002
882 Jirout, J. J., Evans, N. S., & Son, L. K. (2024). Curiosity in children across ages and contexts.
883 *Nature Reviews Psychology*, 1–14. https://doi.org/10.1038/s44159-024-00346-5
884 Jirout, J. J., & Klahr, D. (2012). Children's scientific curiosity: In search of an operational
885 definition of an elusive concept. *Developmental Review*, *32*(2), 125–160.
886 https://doi.org/10.1016/j.dr.2012.04.002
887 John, G., & Roedder, D. L. (1981). Reliability assessment: Coefficients alpha and beta. *The*
888 *Changing Marketing Environment: New Theories and Applications*, 354–357.
889 Jukes, M. C. H., Mgonda, N. L., Tibenda, J. J., Gabrieli, P., Jeremiah, G., Betts, K. L.,
890 Williams, J., & Bub, K. L. (2021). Building an assessment of community-defined
891 social-emotional competencies from the ground up in Tanzania. *Child Development*,
892 *92*(6), e1095–e1109. https://doi.org/10.1111/cdev.13673
893 Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., &
894 Camerer, C. F. (2009). The Wick in the Candle of Learning: Epistemic Curiosity
895 Activates Reward Circuitry and Enhances Memory. *Psychological Science*, *20*(8), 963–
896 973. https://doi.org/10.1111/j.1467-9280.2009.02402.x
897 Kashdan, T. B., Disabato, D. J., Goodman, F. R., & McKnight, P. E. (2020). The Five-
898 Dimensional Curiosity Scale Revised (5DCR): Briefer subscales while separating overt

and covert social curiosity. *Personality and Individual Differences*, *157*, 109836. https://doi.org/10.1016/j.paid.2020.109836

Kashdan, T. B., Gallagher, M. W., Silvia, P. J., Winterstein, B. P., Breen, W. E., Terhar, D., & Steger, M. F. (2009). The curiosity and exploration inventory-II: Development, factor structure, and psychometrics. *Journal of Research in Personality*, *43*(6), 987–998. https://doi.org/10.1016/j.jrp.2009.04.011

Kashdan, T. B., & Roberts, J. E. (2004). Trait and State Curiosity in the Genesis of Intimacy: Differentiation From Related Constructs. *Journal of Social and Clinical Psychology*, *23*(6), 792–816. https://doi.org/10.1521/jscp.23.6.792.54800

Kashdan, T. B., Stiksma, M. C., Disabato, D. J., McKnight, P. E., Bekier, J., Kaji, J., & Lazarus, R. (2018). The five-dimensional curiosity scale: Capturing the bandwidth of curiosity and identifying four unique subgroups of curious people. *Journal of Research in Personality*, *73*, 130–149. https://doi.org/10.1016/j.jrp.2017.11.011

Kashdan, T. B., & Yuen, M. (2007). Whether highly curious students thrive academically depends on perceptions about the school learning environment: A study of Hong Kong adolescents. *Motivation and Emotion*, *31*(4), 260–270. https://doi.org/10.1007/s11031-007-9074-9

Kidd, C., & Hayden, B. Y. (2015). The Psychology and Neuroscience of Curiosity. *Neuron*, *88*(3), 449–460. https://doi.org/10.1016/j.neuron.2015.09.010

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple Nor Too Complex. *PLoS ONE*, *7*(5), e36399. https://doi.org/10.1371/journal.pone.0036399

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks Effect in Infant Auditory Attention. *Child Development*, n/a-n/a. https://doi.org/10.1111/cdev.12263

923 Klamer, A., Lando, A., Pinborg, A., & Greisen, G. (2005). Ages and Stages Questionnaire used
924 to measure cognitive deficit in children born extremely preterm. *Acta Paediatrica*,
925 *94*(9), 1327–1329. https://doi.org/10.1111/j.1651-2227.2005.tb02095.x
926 Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation
927 Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–
928 163. https://doi.org/10.1016/j.jcm.2016.02.012
929 Lee, N., Lazaro, V., Wang, J. J., Şen, H. H., & Lucca, K. (2023). Exploring individual
930 differences in infants' looking preferences for impossible events: The Early
931 Multidimensional Curiosity Scale. *Frontiers in Psychology*, *13*.
932 https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1015649
933 Lépine, J., Gagnon, K., Prud'homme, J., Vinay, M. C., Doussau, A., Fourdain, S., Provost, S.,
934 Belval, V., Bernard, C., Gallagher, A., Poirier, N., & Simard, M.-N. (2022). Utility of
935 the Ages and Stages Questionnaires 3rd Edition for Developmental Screening in
936 Children with Surgically Repaired Congenital Heart Disease. *Developmental*
937 *Neurorehabilitation*, *25*(2), 125–132. https://doi.org/10.1080/17518423.2021.1960918
938 Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Pointing out new news, old news, and
939 absent referents at 12 months of age. *Developmental Science*, *10*(2), F1–F7.
940 https://doi.org/10.1111/j.1467-7687.2006.00552.x
941 Litman, J. A. (2008). Interest and deprivation factors of epistemic curiosity. *Personality and*
942 *Individual Differences*, *44*(7), 1585–1595. https://doi.org/10.1016/j.paid.2008.01.014
943 Litman, J. A., & Jimerson, T. L. (2004). The Measurement of Curiosity As a Feeling of
944 Deprivation. *Journal of Personality Assessment*, *82*(2), 147–157.
945 https://doi.org/10.1207/s15327752jpa8202_3

946 Litman, J. A., & Spielberger, C. D. (2003). Measuring Epistemic Curiosity and Its Diversive
947 and Specific Components. *Journal of Personality Assessment*, *80*(1), 75–86.
948 https://doi.org/10.1207/S15327752JPA8001_16
949 Little, E. E., Carver, L. J., & Legare, C. H. (2016). Cultural Variation in Triadic Infant–
950 Caregiver Object Exploration. *Child Development*, *87*(4), 1130–1145.
951 https://doi.org/10.1111/cdev.12513
952 Lockman, J. J. (2000). A perception–action perspective on tool use development. *Child*
953 *Development*, *71*(1), 137–144.
954 Loewenstein, G. (n.d.). *The Psychology of Curiosity: A Review and Reinterpretation*. 24.
955 Mandler, J. M., Fivush, R., & Reznick, J. S. (1987). The development of contextual categories.
956 *Cognitive Development*, *2*(4), 339–354. https://doi.org/10.1016/S0885-
957 2014(87)80012-6
958 Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge
959 in infancy predict cognitive and language outcomes in later childhood. *Developmental*
960 *Science*, *11*(3), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x
961 Maw, W. H., & Maw, E. W. (1968). Self-appraisal of Curiosity. *Journal of Educational*
962 *Research*, *61*(10), 462–465.
963 https://www.proquest.com/docview/1290479073/citation/8DA93BAA53CC409FPQ/
964 1
965 Maw, W. H., & Maw, E. W. (1970). Self-Concepts of High- and Low-Curiosity Boys. *Child*
966 *Development*, *41*(1), 123–129. https://doi.org/10.2307/1127394
967 McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). Evidence Based on the Internal
968 Structure of the Instrument: Factor Analysis. In D. B. McCoach, R. K. Gable, & J. P.
969 Madura (Eds.), *Instrument Development in the Affective Domain: School and*

Corporate Applications (pp. 109–161). Springer. https://doi.org/10.1007/978-1-4614-7135-6_4

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741.

Muentener, P., Herrig, E., & Schulz, L. (2018). The efficiency of infants' exploratory play is related to longer-term cognitive development. *Frontiers in Psychology*, *9*, 635.

Mussel, P. (2013). Intellect: A theoretical framework for personality traits related to intellectual achievements. *Journal of Personality and Social Psychology*, *104*(5), 885–906. https://doi.org/10.1037/a0031918

Mussel, P., Spengler, M., Litman, J. A., & Schuler, H. (2012). Development and validation of the German Work-Related Curiosity Scale. *European Journal of Psychological Assessment*, *28*(2), 109–117. https://doi.org/10.1027/1015-5759/a000098

Naylor, F. D. (1981). A State-Trait Curiosity Inventory. *Australian Psychologist*, *16*(2), 172–183. https://doi.org/10.1080/00050068108255893

Nunnally, J. C. (1978). An overview of psychological measurement. *Clinical Diagnosis of Mental Disorders: A Handbook*, 97–146.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric Theory New York. *NY: McGraw-Hill*.

Olson, E. (1986). *Measurement of curiosity in junior high school students*.

Penney, R. K., & McCann, B. (1964). The Children's Reactive Curiosity Scale. *Psychological Reports*, *15*(1), 323–334. https://doi.org/10.2466/pr0.1964.15.1.323

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. sage.

995 Piccardi, E. S., Johnson, M. H., & Gliga, T. (2020). Explaining individual differences in infant
996 visual sensory seeking. *Infancy*, *25*(5), 677–698. https://doi.org/10.1111/infa.12356
997 Piedmont, R. L. (2014). Inter-item Correlations. In A. C. Michalos (Ed.), *Encyclopedia of*
998 *Quality of Life and Well-Being Research* (pp. 3303–3304). Springer Netherlands.
999 https://doi.org/10.1007/978-94-007-0753-5_1493
1000 Piotrowski, J. T., Litman, J. A., & Valkenburg, P. (2014). Measuring Epistemic Curiosity in
1001 Young Children: Brief Report. *Infant and Child Development*, *23*(5), 542–553.
1002 https://doi.org/10.1002/icd.1847
1003 Poli, F., Serino, G., Mars, R. B., & Hunnius, S. (2020). Infants tailor their attention to maximize
1004 learning. *Science Advances*, *6*(39), eabb5053. https://doi.org/10.1126/sciadv.abb5053
1005 Putnam, S. P., Helbig, A. L., Gartstein, M. A., Rothbart, M. K., & Leerkes, E. (2014).
1006 Development and Assessment of Short and Very Short Forms of the Infant Behavior
1007 Questionnaire–Revised. *Journal of Personality Assessment*, *96*(4), 445–458.
1008 https://doi.org/10.1080/00223891.2013.841171
1009 Putnam, S. P., Jacobs, J. F., Gartstein, M. A., & Rothbart, M. K. (2010). *Development and*
1010 *Assessment of Short and Very Short Forms of the Early Childhood Behavior*
1011 *Questionnaire*. [Poster] Annual International Conference on Infant Studies, Baltimore,
1012 MD.
1013 Putnam, S. P., Rothbart, M. K., & Gartstein, M. A. (2008). Homotypic and heterotypic
1014 continuity of fine-grained temperament during infancy, toddlerhood, and early
1015 childhood. *Infant and Child Development*, *17*(4), 387–405.
1016 https://doi.org/10.1002/icd.582
1017 Reio Jr., T. G., & Callahan, J. L. (2004). Affect, curiosity, and socialization-related learning:
1018 A path analysis of antecedents to job performance. *Journal of Business and Psychology*,
1019 *19*, 3–22.

1020 Reio Jr., T. G., & Wiswell, A. (2000). Field investigation of the relationship among adult
1021 curiosity, workplace learning, and job performance. *Human Resource Development*
1022 *Quarterly*, *11*(1), 5–30. https://doi.org/10.1002/1532-1096(200021)11:1<5::AID-
1023 HRDQ2>3.0.CO;2-A
1024 Renner, B. (2006). Curiosity About People: The Development of a Social Curiosity Measure
1025 in Adults. *Journal of Personality Assessment*, *87*(3), 305–316.
1026 https://doi.org/10.1207/s15327752jpa8703_11
1027 Revelle, W. (1979). Hierarchical cluster analysis and the internal structure of tests. *Multivariate*
1028 *Behavioral Research*, *14*(1), 57–74.
1029 Revelle, W. (2023). *psych: Procedures for Psychological, Psychometric, and Personality*
1030 *Research* (Version 2.3.9) [Computer software]. https://cran.r-
1031 project.org/web/packages/psych/index.html
1032 Richter, J., & Janson, H. (2007). A validation study of the Norwegian version of the Ages and
1033 Stages Questionnaires. *Acta Paediatrica*, *96*(5), 748–752.
1034 https://doi.org/10.1111/j.1651-2227.2007.00246.x
1035 Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of*
1036 *Statistical Software*, *48*, 1–36.
1037 Rossing, B. E., & Long, H. B. (1981). Contributions of Curiosity and Relevance To Adult
1038 Learning Motivation. *Adult Education*, *32*(1), 25–36.
1039 https://doi.org/10.1177/074171368103200102
1040 Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing.
1041 *International Journal of Research in Marketing*, *19*(4), 305–335.
1042 Rothbart, M. K. (1986). Longitudinal observation of infant temperament. *Developmental*
1043 *Psychology*, *22*(3), 356–365. https://doi.org/10.1037/0012-1649.22.3.356

Rothbart, M. K. (2011). *Becoming who we are: Temperament and personality in development*. Guilford Press.

Salazar, M. S. (2015). The dilemma of combining positive and negative items in scales. *Psicothema*, *27*(2), 192–199.

Salomonsson, B., & Sleed, M. (2010). The Ages & Stages Questionnaire: Social–Emotional: A validation study of a mother-report questionnaire on a clinical mother–infant sample. *Infant Mental Health Journal*, *31*(4), 412–431. https://doi.org/10.1002/imhj.20263

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). *Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures*. *8*(2).

Shah, P. E., Weeks, H. M., Richards, B., & Kaciroti, N. (2018). Early childhood curiosity and kindergarten reading and math academic achievement. *Pediatric Research*, *84*(3), Article 3. https://doi.org/10.1038/s41390-018-0039-3

Shrout, P. E., & Lane, S. P. (2012). Psychometrics. In *Handbook of research methods for studying daily life* (pp. 302–320). The Guilford Press.

Singh, L., Barokova, M. D., Baumgartner, H. A., Lopera-Perez, D. C., Omane, P. O., Sheskin, M., Yuen, F. L., Wu, Y., Alcock, K. J., & Altmann, E. C. (2023). A unified approach to demographic data collection for research with young children across diverse cultures. *Developmental Psychology*.

Slagt, M., Dubas, J. S., Deković, M., & van Aken, M. A. G. (2016). Differences in sensitivity to parenting depending on child temperament: A meta-analysis. *Psychological Bulletin*, *142*(10), 1068–1110. https://doi.org/10.1037/bul0000061.supp

Slone, L. K., Smith, L. B., & Yu, C. (2019). Self-generated variability in object images predicts vocabulary growth. *Developmental Science*, *22*(6), e12816. https://doi.org/10.1111/desc.12816

1069 Smith, L. B., & Yu, C. (2013). Visual Attention Is Not Enough: Individual Differences in
1070 Statistical Word-Referent Learning in Infants. *Language Learning and Development*,
1071 *9*(1), 25–49. https://doi.org/10.1080/15475441.2012.707104
1072 Spielberger, C. D. (1979). *State-trait personality inventory*.
1073 Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and
1074 exploration. *Science*, *348*(6230), 91–94. https://doi.org/10.1126/science.aaa3799
1075 Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics. Boston, MA: Pearso*
1076 *n Education*. Inc.
1077 Twomey, K. E., & Westermann, G. (2018). Curiosity-based learning in infants: A
1078 neurocomputational approach. *Developmental Science*, *21*(4), e12629.
1079 https://doi.org/10.1111/desc.12629
1080 Vandenberg, R. J. (2006). Introduction: Statistical and Methodological Myths and Urban
1081 Legends: Where, Pray Tell, Did They Get This Idea? *Organizational Research*
1082 *Methods*, *9*(2), 194–201. https://doi.org/10.1177/1094428105285506
1083 Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement
1084 Invariance Literature: Suggestions, Practices, and Recommendations for
1085 Organizational Research. *Organizational Research Methods*, *3*(1), 4–70.
1086 https://doi.org/10.1177/109442810031002
1087 Wagstaff, M. F., Flores, G. L., Ahmed, R., & Villanueva, S. (2021). Measures of curiosity: A
1088 literature review. *Human Resource Development Quarterly*, *32*(3), 363–389.
1089 https://doi.org/10.1002/hrdq.21417
1090 Wass, S. V., & Smith, T. J. (2014). Individual Differences in Infant Oculomotor Behavior
1091 During the Viewing of Complex Naturalistic Scenes. *Infancy*, *19*(4), 352–384.
1092 https://doi.org/10.1111/infa.12049

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research*, *49*(5), 737–747.

Wright, A. J., & Jackson, J. J. (2022). Childhood temperament and adulthood personality differentially predict life outcomes. *Scientific Reports*, *12*(1), Article 1. https://doi.org/10.1038/s41598-022-14666-0

Supplementary Materials

The r-script and raw data to reproduce the analyses, as well as a paper-pen-version for the questionnaire have been made available on the OSF.

1. Additional Items

1.1 Mirrored Items for attention

Ia6-M. When someone shows my child how something works, they are usually ***not*** very interested. *(Mirror of Final Item Ia6, Table 1; Can be additionally included in the middle of the item list, perhaps for online research with concerns over data quality, but must then be excluded from aggregate scores.)*

Ex1-M. My child usually waits to be given a toy to play with, rather than start playing by themselves. *(Mirror of excluded item Ex1, Table 1)*

1.2 Excluded Items (original numbering)

Excluded due to negatively correlating with the rest:

Item.5 When my child encounters an object, they are likely to put it in their mouth for further inspection (e.g., to see what it feels or tastes like).

Item.21 My child is usually happy to try new foods they haven't eaten before.

Excluded due to not loading onto the general factor:

Item.1 When I hold or move a toy or object in front of my child, they follow it with their eyes.

Item.4R When my child is introduced to something new, they are often not very interested.

Item.9 My child is constantly reaching for objects to explore.

Item.14 Once my child was able to crawl, they used this new skill to explore their environment on their own terms.

Item.29 R My child does not seem to care when we go somewhere new, they still prefer to engage with familiar objects they brought from home (e.g., their pacifier or favourite toy).

Item.36 My child is usually interested in new people.

Excluded due to not loading onto any subfactor:

Item.15/Ex1 My child starts playing on their own, rather than waiting to be given something to play with.

Item.18/Ex2 When my child plays with an assembly toy (e.g., building blocks, puzzle, a toy with detachable parts), they like to take it apart for further examination.

Item.25/Ex3 When playing hide and seek, my child enjoys searching for the object or person that disappeared.

1.3 Open Ended Questions

In reference to their response to the comparative item:

Please provide a short example that illustrates the option you just indicated. *(If more or less curious)*

Please provide a short example of how your child has recently (or indicate an approximate age) explored their environment in a way which was not reflected in the statements you rated. *(If equally curious)*

2. Additional considerations regarding the response scale

As the response scale we decided on a 7-point Likert scale from 1 ("strongly disagree") to 7 ("strongly agree"). We considered using a frequency scale from *never* to *always* which is used in several infant and early childhood questionnaires (e.g., temperament scales IBQ and ECBQ). While both agreement (that is, Likert) and frequency scales can be used to generate aggregate scores, it was found in a systematic review that agreement scales lead to better fit and response quality and that frequency scales can be problematic in their interpretation (Brown, 2004). Thus, we decided on the format of agreement.

As the questionnaire was administered online using the secure software Qualtrics (Qualtrics, Provo, UT), we also had to consider in which way the response options would be presented. Next to the more conventional "radio-button" responses (one for each scale-point), a "slider" was discussed where caregivers could indicate their level of agreement anywhere between 0 and 100. While reviews suggest that such sliders can be more engaging with comparable data quality, they are in fact more time-intensive and can lead to frustration (Sikkel et al., 2014) and higher drop-out rates (Couper et al., 2006; Cook et al., 2001). Furthermore, they seem to add cognitive complexity which would not be recommended for wider representation of the caregiver population (e.g., Funke et al., 2014; Stanley & Jenkins, 2007). Based on these considerations, we decided to implement a conventional 7-point Likert scale with a button for each response option.

3. Number of exploratory factors

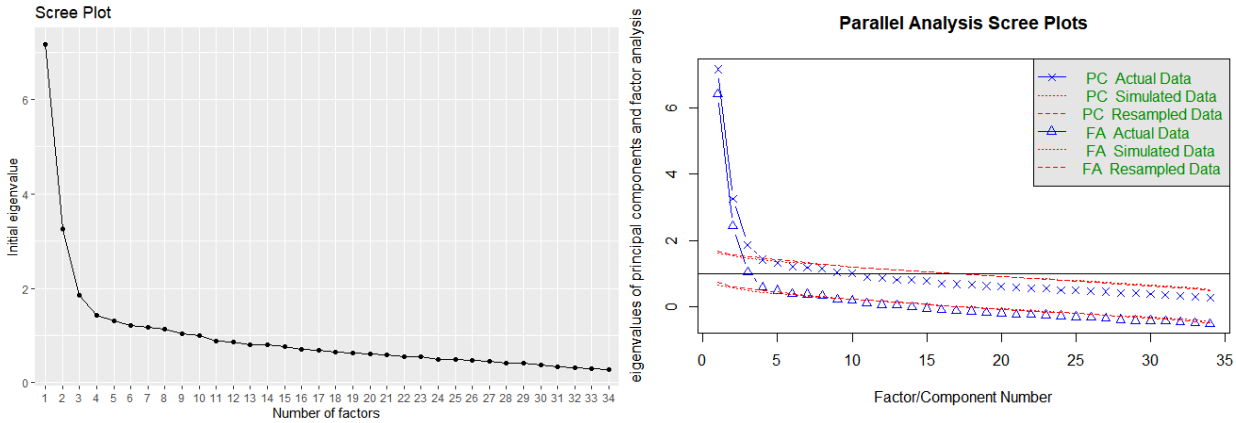

Figure S1. Scree plot & parallel analysis to determine the adequate number of factors suggesting 3 or 4 factors as indicated by their explanatory eigenvalues and the "elbow" in the graphs.